HAL open science

Feedback-aided complexity reductions in ML and Lattice decoding

Arun Kumar Singh, Petros Elia

▶ To cite this version:

Arun Kumar Singh, Petros Elia. Feedback-aided complexity reductions in ML and Lattice decoding. IEEE International Symposium on Information Theory (ISIT'12), Jul 2012, United States. pp.5. hal-00707823

HAL Id: hal-00707823
https://hal.science/hal-00707823

Submitted on 13 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Feedback-aided complexity reductions in ML and Lattice decoding

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

***Abstract*—The work analyzes the computational-complexity savings that a single bit of feedback can provide in the computationally intense setting of non-ergodic MIMO communications. Specifically we derive upper bounds on the *feedback-aided complexity exponent* required for the broad families of ML-based and lattice based decoders to achieve the optimal diversity-multiplexing behavior. The bounds reveal a complexity that is reduced from being exponential in the number of codeword bits, to being at most exponential in the rate. Finally the derived savings are met by practically constructed ARQ schemes, as well as simple lattice designs, decoders, and computation-halting policies.**

I. Introduction

The current work is a continuation of studies on the rate-reliability-complexity limits of non-ergodic MIMO communications[1]. In this setting, computational complexity and rate-reliability performance are highly intertwined, in the sense that limitations to computational resources (commonly measured by floating point operations - flops), bring about substantial degradation in the system performance. In the high rate setting of interest, the lion's share of computational costs is due to decoding algorithms, on which we here focus, specifically considering the broad family of ML-based and regularized (MMSE-preprocessed) lattice decoding algorithms.

a) Error and complexity exponents: In terms of reliability, the diversity multiplexing tradeoff (DMT, cf. [1]) has been extensively used to quantify the relationship between the rate, denoted as R, and the probability of error P_{err}. In the high SNR regime (SNR will be henceforth denoted as ρ), this relationship was described in [1] using the high SNR measures of multiplexing gain $r := R/\log\rho$ and diversity gain $d(r) := -\lim_{\rho\to\infty}\log P_{\text{err}}/\log\rho$. As a result the same work revealed, for the case of no feedback, the optimal DMT in the form of the maximum possible diversity gain $d^*(r)$ for a given r.

The work in [2]–[4] provided a similar treatment for complexity. Specifically for $N_{\max}$ denoting the amount of computational reserves, in flops per duration of one codeword, that the transceiver is endowed with[2], the work in [2], [3] introduced the complexity exponent to take the form

$$c(r) := \lim_{\rho\to\infty}\frac{\log N_{\max}}{\log\rho}, \tag{1}$$

where the value of the above exponent was derived as a function of the desired r and $d(r)$. In the specific setting of quasi-static MIMO, without feedback and in the absence of lattice reduction techniques[3], the above work revealed that to achieve the optimal DMT $d^*(r)$, the complexity exponent is upper bounded by a piecewise linear function which, at integer r takes the form

$$c(r) = r(n_{\mathrm{T}} - r). \tag{2}$$

This bound was shown to be tight for a broad range of practical settings, and it also revealed a complexity that scales exponentially with the number of codeword bits.

This very considerable complexity brought to the fore the need for methods that manage to achieve the same near-optimal performance, but do so with much reduced computational resources.

b) Feedback gains: In terms of feedback, the work in [5] utilized the DMT machinery to analyze the reliability gains of feedback, and to specifically show that an L-round ARQ scheme can provide for a much increased feedback-aided DMT which[4] was shown to take the form $d^*(r/L)$.

Motivated by the considerable magnitude of the complexity exponent in (2), we here seek to understand the role of feedback in reducing complexity, rather than in improving reliability. For this we seek to quantify the feedback-aided complexity exponent required to achieve the original $d^*(r)$ in the presence of a modified version of the above mentioned L-round MIMO ARQ. Specifically we will derive upper bounds on the minimum[5] complexity exponent required by ML and regularized (MMSE-preprocessed) lattice based sphere decoders (SD) to achieve the optimal DMT $d^*(r)$. We will focus on the family of minimum delay ARQ schemes (to be described later on). The derivations focus on ML-based decoding, but given the equivalence of ML and regularized lattice based decoding shown in [3], these same results extend automatically to the regularized lattice decoding case. We note that the validity of the presented bounds depends on the existence of actual schemes that meet them. These schemes will be here provided, together with the associated lattice designs, decoders, as well as halting and ordering polices.

[1] By non-ergodic MIMO we refer to the setting where there is considerable channel state information at the receiver (CSIR), and very little if any channel state information at the transmitter (little or no CSIT).

[2] In the sense that after $N_{\max}$ flops the transceiver must simply terminate potentially prematurely and before completion of its task.

[3] We here note that while lattice reduction (LR) indeed allows here for near-optimal behavior at very manageable complexity, it is the case that there exist scenarios for which these same LR methods cannot be readily applied. Such problematic cases include the ubiquitous scenario where outer binary codes are employed and decoded using soft information. It is for this exact reason that we focus on the complexity analysis of non LR-aided schemes which remains of strong interest for many pertinent communication scenarios.

[4] This held for the setting of quasi-static fading and no power adaptation - which is the setting of interest here.

[5] By minimum we refer to a minimization over all lattice code designs (which must vary accordingly depending on the setting), all policies of computational halting, and all policies on decoding ordering. A decoding ordering policy describes the order in which the transmitted information symbols are decoded by sphere decoding algorithm.

The analysis and the constructed feedback schemes tell us how to properly utilize a single bit of feedback to alleviate the adverse effects of computational constraints, as those seen in the derived rate-reliability-complexity tradeoffs of [4].

Before proceeding to a brief description of the MIMO ARQ signaling, we quickly note that we here employ an ARQ variant which reduces the L-round scheme to a two-round scheme with uneven but fixed durations, and we do so by disregarding all but the first and last rounds. Such a scheme requires just one bit of feedback. We will however, for clarity of exposition, maintain use of the notation of the better known L-round scheme.

A. MIMO-ARQ signaling

We here present the general $n_T \times n_R$ MIMO-ARQ signaling setting, and focus on the details which are necessary for our exposition. For further understanding of the MIMO-ARQ channel, the reader is referred to [5] as well as [6].

Under ARQ signaling, each message is associated to a unique block $[\mathbf{X}_C^1\ \mathbf{X}_C^2 \cdots \mathbf{X}_C^L]$ of signaling matrices, where each $\mathbf{X}_C^i \in \mathbb{C}^{n_T \times T},\ i = 1, \cdots, L$, corresponds to the $n_{\mathrm{T}} \times T$ matrix of signals sent during the ith round. The accumulated code matrix at the end of round ℓ, $\ell = 1, \cdots, L$, takes the form $\mathbf{X}_C^{ARQ,\ell} = [\mathbf{X}_C^1\ \mathbf{X}_C^2 \cdots \mathbf{X}_C^\ell] \in \mathbb{C}^{n_T \times \ell T}$. We note that the signals $\mathbf{X}_C^{ARQ,L}$ are drawn from a lattice design that ensures unique decodability at every round[6].

In the quasi-static case of interest, the received signal accumulated at the end of the ℓ-th round takes the form

$$\mathbf{Y}_C^\ell = \theta \mathbf{H}_C \mathbf{X}_C^{ARQ,\ell} + \mathbf{W}_C^\ell, \quad \ell = 1, \cdots, L, \tag{3}$$

where $\mathbf{H}_C \in \mathbb{C}^{n_R \times n_T}$, where the scaling factor θ is chosen such that $\mathbb{E}(\|\theta \mathbf{X}_C^i\|^2) \leq \rho T,\ 1 \leq i \leq \ell$.

We proceed with quantifying the complexity reductions due to ARQ feedback.

II. Complexity reduction using ARQ feedback

We here seek to analyze the complexity reductions due to MIMO ARQ feedback. Specifically for $d^*(r)$ denoting the optimal DMT of the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ MIMO channel in the absence of feedback, we here seek to describe the feedback-aided complexity exponent required to meet the same $d^*(r)$ with the assistance now of an L-round ARQ scheme. As stated before, our analysis focuses on the setting of $L \leq n_{\mathrm{T}}$ and of *minimum-delay ARQ schemes*, corresponding to $T = 1$. The derived exponent is to be compared with the exponent in (2) (cf. [2, Theorem 6] and [3, Corollary 1b]) corresponding to no feedback. The following holds for the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ ($n_{\mathrm{R}} \geq n_{\mathrm{T}}$), i.i.d. regular fading[7] MIMO channel.

Theorem 1: Let $c(r)$ be the minimum complexity exponent required to achieve $d^*(r)$, minimized over all lattice designs, all ARQ schemes with $L \leq n_{\mathrm{T}}$ rounds of ARQ, all halting policies and all decoding order policies. Then $c(r) \leq \overline{c}_{red}(r)$ where

$$\overline{c}_{red}(r) \triangleq \frac{1}{n_T}\left[r(n_T - \lfloor r \rfloor - 1) + (n_T \lfloor r \rfloor - r(n_T - 1))^+\right],$$

which is a piecewise linear function that, for integer r, takes the form

$$\overline{c}_{red}(r) = \frac{1}{n_T} r(n_T - r), \text{ for } r = 0, 1, \cdots, n_T.$$

The proof of the above theorem will be presented in Appendix A, together with the proofs for the upcoming Propositions 1 and 2, and it will include the derivation of the upper bound, and the constructive achievement of this bound which is presented in Propositions 1, 2. The constructive part of the proof is based on designing ARQ schemes and implementations (lattice designs and halting policies) that meet the bound. We proceed with these propositions where we identify cases for which the above complexity bound suffices to achieve $d^*(r)$ with the help of feedback.

An important aspect in ARQ schemes is knowing when to decode and when not to decode across the different rounds. Towards this we have the following definition.

Definition 1 (Aggressive intermediate halting policies): We define *aggressive intermediate halting policies* to be the family of policies that halt decoding in the first round whenever the minimum singular value of the channel scales as $\rho^{-\epsilon}$ for some $\epsilon > 0$, which do not decode in the second to the L-1 round, and which decode at the last round iff a) they have not decoded in the first round and b) the channel is not in outage with respect to the effective rate of ARQ scheme.

Given such aggressive halting policies, the L round scheme reduces to a two round scheme where the second round comprises of $(L-1)T$ channel uses. As noted before, for notational uniformity with earlier works in [5], [6], we will continue to use the notation of the L-round schemes but again clarify that only one bit of ARQ feedback is needed.

Furthermore we will henceforth use the term *ARQ-compatible, minimum delay, NVD, rate-1 lattice designs* to refer to the family of $n_{\mathrm{T}} \times n_{\mathrm{T}}$ lattice designs $\mathbf{X}_C^{ARQ,L}$ with total number of transmitted integers $\kappa = 2n_{\mathrm{T}}$, with non-vanishing determinant (NVD)[8] for $r \leq 1$, and with all the information appearing in all rounds.

Proposition 1: A minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$, irrespective of the ARQ-compatible, minimum delay, NVD, rate-1 lattice design, for any aggressive intermediate halting policy, and any sphere decoding order policy.

The following describes a very simple MIMO ARQ coding implementation that achieves $d^*(r)$ with $c(r) \leq \overline{c}_{red}(r)$. The proof of this proposition will appear later on, and is crucial in the achievability part of the proof of Theorem 1.

Proposition 2: The minimum delay ARQ scheme with $L = n_{\mathrm{T}}$ rounds, implemented with any aggressive intermediate halting policy, any sphere decoding order policy, and a rate-1

[6]Loosely speaking, unique decodability means that, for any $\ell = 1, \cdots, L$, the corresponding $\mathbf{X}_C^{ARQ,\ell}$ carries all bits of information.

[7]The i.i.d. regular fading statistics satisfy the general set of conditions as described in [7], where a) the near-zero behavior of the fading coefficients h is bounded in probability as $c_1|h|^t \leq p(h) \leq c_2|h|^t$ for some positive and finite c_1, c_2 and t, where b) the tail behavior of h is bounded in probability as $p(h) \leq c_2 e^{-b|h|^\beta}$ for some positive and finite c_2, b and β, and where c) $p(h)$ is upper bounded by a constant K.

[8]A code has a non-vanishing determinant if, without power normalization, there is a lower bound on the minimum determinant that does not depend on the constellation size. The determinant of any non-normalized difference matrix is lower bounded by a constant independent of ρ (see [8]).

lattice design $\mathbf{X}_C^{ARQ,L}$ drawn from the center of perfect codes (cf. [8], [9])[9], achieves $d^*(r)$ with $c(r) \leq \bar{c}_{red}(r)$.

Theorem 1 has quantified the computational reserves that are sufficient to achieve DMT optimality. These computational reserves can be seen to be smaller than those required to achieve the same optimal DMT $d^*(r)$ without feedback. For example, given any known minimum-delay DMT optimal design which remains fixed for all r, in the absence of feedback, the exponent needed to achieve $d^*(r)$ is that in (2) (cf. [2, Theorem 6] and [3, Corollary 1b]) and takes the form

$$c(r) = r(n_T - r), \tag{5}$$

(for integer $r = 0, 1, \cdots, n_T$), whereas as we have just seen, for $L = n_\mathrm{T}, T = 1$ this exponent reduces to a much smaller

$$c(r) \leq \frac{1}{n_T} r(n_T - r).$$

We proceed with a few examples.

Example 1 (Corresponding to Theorem 1 and Proposition 2): For the general $n_\mathrm{T} \times n_\mathrm{R}$ setting with $n_\mathrm{R} \geq n_\mathrm{T}$, and for $r = n_\mathrm{T}/2$, the computational resources required to achieve the optimal $d^*(r)$ with existing DMT optimal (minimum delay) non-feedback schemes (cf. [2, Theorem 6]), scales as[10]

$$N_{\max} \doteq \rho^{n_\mathrm{T}^2/4} \doteq 2^{Rn_\mathrm{T}/2},$$

whereas the feedback aided complexity required by the feedback scheme in Proposition 2 scales as

$$N_{\max} \doteq \rho^{n_\mathrm{T}/4} \doteq 2^{R/2}.$$

Generally, given a rate that scales linearly with n_T, in the absence of feedback the complexity exponent of achieving $d^*(r)$ scales with n_T^2, whereas the feedback aided complexity exponent scales with n_T.

Example 2: Figure 1 considers the case of $n_\mathrm{T} = 4 \leq n_\mathrm{R}$ and Rayleigh fading, and compares the above complexity upper bound in the presence of feedback ($L = 4, T = 1$), to the equivalent complexity exponent in (5) of achieving the same optimal DMT $d^*(r)$ without ARQ feedback (Perfect codes and natural, fixed decoding ordering (cf. [2])). The feedback-aided complexity exponent reveals an exponential reduction by a factor of $n_\mathrm{T} = 4$.

[9]For general lattice designs derived from cyclic division algebra (CDA) (cf. [8], [9]), $\mathbb{F}$ and $\mathbb{L}$ are number fields, with $\mathbb{L}$ a finite, cyclic Galois extension of $\mathbb{F}$ of degree n. Let σ denote a generator of the Galois group $\mathrm{Gal}(\mathbb{L}/\mathbb{F})$. Let z be an indeterminate satisfying $lz = z\sigma(l)$, $\forall\, l \in \mathbb{L}$ and $z^n = \gamma$ for some non-norm element $\gamma \in \mathbb{F}^*$. Then the set of all elements of the form $\sum_{i=0}^{n-1} z^i l_i$ forms a CDA $D(\mathbb{L}/\mathbb{F}, \sigma, \gamma)$ with center $\mathbb{F}$ and maximal subfield $\mathbb{L}$. The mentioned codes are limited in the center of the division algebra and take the simple form

$$\mathbf{X}_C^{ARQ,L} = \begin{bmatrix} f_0 & \gamma f_{n_T-1} & \cdots & \gamma f_1 \\ f_1 & f_0 & \cdots & \gamma f_2 \\ \vdots & \vdots & \ddots & \vdots \\ f_{n_T-1} & f_{n_T-2} & \cdots & f_0 \end{bmatrix} \in \mathbb{C}^{n_T \times n_T}, \tag{4}$$

where f_i belong to the QAM constellation.

[10]We use $\doteq$ to denote the *exponential equality*, i.e., we write $f(\rho) \doteq \rho^B$ to denote $\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log \rho} = B$, and $\dot{\leq}, \dot{<}$, and $\dot{\geq}, \dot{>}$ are defined similarly.

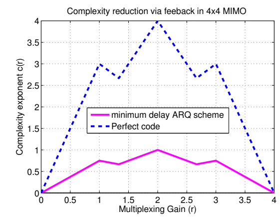


Fig. 1. Complexity reduction with minimum delay ARQ schemes.

A. Feedback reduction for asymmetric channels: $n_\mathrm{R} \leq n_\mathrm{T}$

We now consider the case of $n_\mathrm{R} \leq n_\mathrm{T}$, and specifically the case where $n_\mathrm{R} | n_\mathrm{T}$ (i.e., n_T is an integer multiple of n_R), to observe again how simple implementations offer substantial reductions in complexity. In terms of statistics, the results hold for any i.i.d. regular fading distribution.

Theorem 2: In the MIMO ARQ channel with $n_\mathrm{R} | n_\mathrm{T}$, the minimum complexity exponent $c(r)$ required to achieve $d^*(r)$, minimized over all lattice designs, all halting policies, and all minimum delay ARQ schemes with $L \leq n_\mathrm{T}$ rounds of ARQ, is bounded as $c(r) \leq \bar{c}_{red}(r)$ where

$$\bar{c}_{red}(r) \triangleq \frac{1}{n_R} \left[r(n_R - \lfloor r \rfloor - 1) + (n_R \lfloor r \rfloor - r(n_R - 1))^+ \right],$$

which is a piecewise linear function that, for integer r, takes the form

$$\bar{c}_{red}(r) = \frac{1}{n_\mathrm{R}} r(n_\mathrm{R} - r), \text{ for } r = 0, 1, \cdots, n_\mathrm{R}.$$

Applying as the constructive part of the proof of the above theorem, the following describes a very simple MIMO ARQ block-diagonal repetition coding implementation that achieves $d^*(r)$ with a much reduced $c(r) \leq \bar{c}_{red}(r)$.

Proposition 3: A minimum delay ARQ scheme with $L = n_\mathrm{T}, T = 1$, implemented with any aggressive intermediate halting policy, any sphere decoding order policy, and a rate-$\frac{n_\mathrm{R}}{n_\mathrm{T}}$ block-diagonal repetition lattice design $\mathbf{X}_C^{ARQ,L}$ where the (rate-1) block component code is drawn from the center of $n_\mathrm{R} \times n_\mathrm{R}$ perfect codes, achieves $d^*(r)$ with $c(r) \leq \bar{c}_{red}(r)$ from Theorem 2.

The proof of Theorem 2 and Proposition 3 will be presented in Appendix B. Of interest is the special MISO-ARQ case of $n_\mathrm{R} = 1$, where the above described scheme will allow for a zero complexity exponent, and for a complexity that scales as a subpolynomial function of ρ and as a subexponential function of the number of codeword bits and of the rate.

Corollary 2a: Over the $n_\mathrm{T} \times 1$ MISO channel, the minimum delay ARQ scheme with $L = n_\mathrm{T}$ rounds, implemented with a rate-$\frac{1}{n_\mathrm{T}}$ repetition QAM design $\mathbf{X}_C^{ARQ,L}$, achieves $d^*(r)$ with $c(r) = 0$.

This corollary follows directly from Theorem 2.

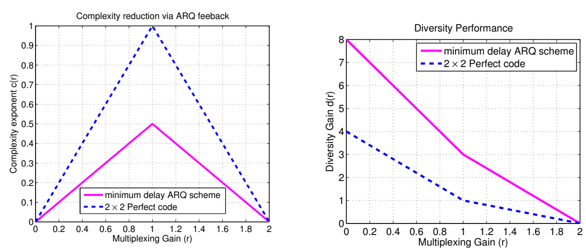
We proceed with a few examples.

Example 3 (Corresponding to Theorem 2 and Proposition 3): For the 4×2 MIMO channel with $L = 4, T = 1$, applying a lattice design of the form

$$\mathbf{X}_C^{ARQ,L} = \begin{bmatrix} f_0 & \gamma f_1 & 0 & 0 \\ f_1 & f_0 & 0 & 0 \\ 0 & 0 & f_0 & \gamma f_1 \\ 0 & 0 & f_1 & f_0 \end{bmatrix} \in \mathbb{C}^{4\times 4},$$

where $f_0, f_1 \sim QAM$, together with an aggressive intermediate halting policy for the first round decoder, and with any sphere decoding ordering policy, can achieve the optimal $d^*(r)$ of the 4×2 channel, and can do so with computational resources of $N_{\max} \doteq \rho^{\bar{c}_{red}(r)}$ flops, which for integer r translates to $N_{\max} \doteq \rho^{\frac{1}{n_R} r(n_R - r)} = \rho^{\frac{r}{2}(2-r)}$.

Example 4 (Corresponding to Theorem 2 and Proposition 3): Figure 2 compares two schemes: the 2×2 MIMO channel (minimum delay, DMT optimal lattice design), and the 4×2 minimum delay MIMO-ARQ channel with $L = n_T = 4$, 1 bit of feedback, and the implementation of Proposition 3. We see a considerably reduced complexity of the feedback aided scheme (Fig. 2(a), lower line) which, at the same time, achieves a much higher DMT performance (Fig. 2(b), upper line) than its non-feedback counterpart.



(a) Complexity exponent (b) DMT

Fig. 2. Complexity reduction $n_R|n_T$ i.i.d. Rayleigh channel with ARQ feedback.

APPENDIX A
PROOF OF THEOREM 1

The proof follows from the footsteps of the [2, proof of Theorem 2]. Due to space limitations we restrict this exposition to the proof steps that are necessary to understand the complexity exponent for the novel ARQ schemes discussed in this paper. For further understanding of encoders and decoders considered here, the reader is referred to [2], [4]. We begin by establishing necessary conditions for L-round ARQ scheme to achieve $d^*(r)$ over an $n_T \times n_R$ $(n_R \geq n_T)$ MIMO.

Condition 1: To achieve maximum diversity gain of $n_R.n_T$ the total number of channel uses $LT \geq n_T$. For minimum delay $(T = 1)$ L-round ARQ schemes with $L \leq n_T$, it then follows that $L = n_T$.

Condition 2: To achieve maximum multiplexing gain of n_T the total number of integers transmitted $\kappa \geq 2n_T T = 2n_T$.

It can be seen from [2] that the complexity of sphere decoder increases with κ. Thus, for minimum delay L-round ARQ schemes with $L \leq n_T$, the tightest upper bound on the complexity exponent is established by considering $T = 1$, $L = n_T$ and $\kappa = 2n_T$. Setting $T = 1$ and $L = n_T$ implies use of at least rate-1 lattice designs for L-round ARQ scheme for achieving DMT performance of $d^*(r)$ for $0 \leq r \leq n_T$. We will later show that encoding-decoding policy described in Proposition 1 and Proposition 2 achieve DMT performance of $d^*(r)$ for $T = 1$, $L = n_T$ and $\kappa = 2n_T$, which in turn implies that $\lim_{\rho\to\infty} r = r_1$, where r_1 is the multiplexing gain for the first round of ARQ.

Having established the necessary parameters we proceed to prove the claim of Theorem 1. Following the footsteps of the [2, proof of Theorem 2] we can show that in the presence of aggressive halting policy and SD with search radius $\xi > \sqrt{d^*(\frac{r}{L})\log\rho}$, an upper bound on the complexity exponent for first round decoder can be obtained as the solution to a constrained maximization problem according to

$$\bar{c}_1(r) \triangleq \max_{\substack{\{\mu_1 < \epsilon, \\ \mu_1 \geq \cdots \geq \mu_{n_T} \geq 0\}}} \sum_{j=1}^{n_T} \min\left(\frac{r}{n_T} - (1-\mu_j), \frac{r}{n_T}\right)^+,$$

where $\mu_j \triangleq -\frac{\log \sigma_j(\mathbf{H}_C^H \mathbf{H}_C)}{\log\rho}$, $j = 1, \cdots, n_T$ with $\mu_1 \geq \cdots \geq \mu_{n_T}$, where σ_j denotes j-th singular value of $\mathbf{H}_C^H\mathbf{H}_C$ and where we have made use of the fact that $\lim_{\rho\to\infty} r = r_1$. In the limit $\epsilon \to 0$ this upper bound simplifies to

$$\bar{c}_1(r) = 0.$$

To establish the L-th round complexity exponent, we proceed with the L-th round system model given by

$$\mathbf{Y}_C^L = \theta \mathbf{H}_C \mathbf{X}_C^{ARQ,L} + \mathbf{W}_C^L,$$

where for rate-1 lattice designs we have $\theta^2 = \rho^{1-r_L}$, where $r_L = \frac{r}{L}$ denotes multiplexing gain for L-th round of ARQ. The vectorized real valued representation of L-th round system model takes the form

$$\mathbf{y}^L = \theta \mathbf{H}^L \mathbf{x}^L + \mathbf{w}^L, \quad \text{where} \tag{6}$$

$$\mathbf{H}^L = \mathbf{I}_L \otimes \mathbf{H}_R, \text{ with } \mathbf{H}_R = \begin{bmatrix} Re\{\mathbf{H}_C\} & -Im\{\mathbf{H}_C\} \\ Im\{\mathbf{H}_C\} & Re\{\mathbf{H}_C\} \end{bmatrix},$$

$\mathbf{x}^L = (\mathbf{x}_1^{\mathrm{T}}, \cdots, \mathbf{x}_L^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^{2n_T L}$ with $\mathbf{x}_t = [Re\{\mathbf{X}_{t,C}^{ARQ,L}\}^{\mathrm{T}}, Im\{\mathbf{X}_{t,C}^{ARQ,L}\}^{\mathrm{T}}]^{\mathrm{T}}$ for $t = 1, \cdots, L$, where $\mathbf{X}_{t,C}^{ARQ,L}$ is t-th column of $\mathbf{X}_C^{ARQ,L}$, $\mathbf{w}^L$ and $\mathbf{y}^L$ can be defined similarly. The vectorized codeword $\mathbf{x}_L$ takes the form (cf. [3])

$$\mathbf{x}^L = \mathbf{G}\mathbf{s}, \qquad \mathbf{s} \in \mathbb{S}_r^\kappa \triangleq \mathbb{Z}^\kappa \cap \rho^{\frac{r_L}{2}}\mathcal{R}, \tag{7}$$

where $\mathbf{G} \in \mathbb{R}^{2Ln_T\times\kappa}$ is the lattice generator matrix, where $\kappa = 2n_T$ and where $\mathcal{R} \subset \mathbb{R}^\kappa$ is a natural bijection of the code shaping region 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 $\mathbb{R}^\kappa$, although this could be relaxed. Combining (6) and (7) yields the equivalent system model

$$\mathbf{y}^L = \mathbf{M}^L\mathbf{s} + \mathbf{w}^L, \tag{8a}$$

$$\text{where} \quad \mathbf{M}^L \triangleq \rho^{\frac{1}{2}-\frac{r_L}{2}} \mathbf{H}^L \mathbf{G} \in \mathbb{R}^{2n_R L\times\kappa}. \tag{8b}$$

Let $\mathbf{G} = [\mathbf{\Gamma}_1^T \ \mathbf{\Gamma}_2^T \cdots \mathbf{\Gamma}_L^T]^T$, where $\mathbf{\Gamma}_i \in \mathbb{C}^{2n_T\times 2n_T}$, for $i = 1, \cdots, L$.

Then the equivalent code-channel matrix ($\boldsymbol{M}^L$) takes the form

$$\mathbf{M}^L = \rho^{\frac{1}{2}-\frac{r_L}{2}} \begin{bmatrix} \mathbf{H}_R & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{H}_R \end{bmatrix} \begin{bmatrix} \mathbf{\Gamma}_1 \\ \vdots \\ \mathbf{\Gamma}_L \end{bmatrix}, \tag{9}$$

$$= \rho^{\frac{1}{2}-\frac{r_L}{2}} [\mathbf{\Gamma}_1^T \mathbf{H}_R^T \ \cdots \ \mathbf{\Gamma}_L^T \mathbf{H}_R^T]^T. \tag{10}$$

In order to compute the singular values of $\boldsymbol{M}^L$ we note that

$$
\begin{aligned}
(\mathbf{M}^L)^H(\mathbf{M}^L) =&\rho^{1-r_L}(\mathbf{\Gamma}_1^H\mathbf{H}_R^H\mathbf{H}_R\mathbf{\Gamma}_1+\cdots+\mathbf{\Gamma}_L^H\mathbf{H}_C^H\mathbf{H}_C\mathbf{\Gamma}_L),\\
\succeq&\rho^{1-r_L}\mathbf{\Gamma}_1^H\mathbf{H}_R^H\mathbf{H}_R\mathbf{\Gamma}_1, \qquad (11)
\end{aligned}
$$

where $\mathbf{A}\succeq\mathbf{B}$ denotes that $\mathbf{A}-\mathbf{B}$ is positive-semidefinite. Without loss of generality we can assume that $\mathbf{\Gamma}_1$ is full-rank. It then follows that the singular values of $\mathbf{M}^L$ can be lower bounded as

$$
\begin{aligned}
\sigma_i((\mathbf{M}^L)^H(\mathbf{M}^L)) \geq&\rho^{1-r_L}\sigma_i(\mathbf{\Gamma}_1^H\mathbf{H}_R^H\mathbf{H}_R\mathbf{\Gamma}_1),\\
\dot{\geq}&\rho^{1-r_L-\mu_{\iota_2(i)}}, \qquad (12)
\end{aligned}
$$

where we have made use of the fact that $\sigma_{min}(\mathbf{\Gamma}_1)\doteq\rho^0$ and where $\iota_2(i)\triangleq\left\lceil\frac{i}{2}\right\rceil$.

Using (12) and following the footsteps of the [2, proof of Theorem 2], the upper bound on the complexity exponent for the L-th round decoding of minimum delay L-round ARQ schemes achieving $d^*(r)$ can be obtained as the solution to a constrained maximization problem according to

$$
\begin{aligned}
\bar{c}_{red}(r)\triangleq&\max_{\boldsymbol{\mu}}\sum_{i=1}^{n_T}\min\left(\frac{r}{n_T}-(1-\mu_i),\frac{r}{n_T}\right)^+ \qquad (13a)\\
&\text{s.t.}\quad I(\boldsymbol{\mu})\leq d^*(r), \qquad (13b)\\
&\qquad \mu_1\geq\cdots\geq\mu_{n_T}\geq 0, \qquad (13c)
\end{aligned}
$$

where we have made use of the fact that $L=n_T$ and $r_L=\frac{r}{L}$. The solution to this optimization problem takes the form

$$
\bar{c}_{red}(r)=\frac{1}{n_T}\left[r(n_T-\lfloor r\rfloor-1)+(n_T\lfloor r\rfloor-r(n_T-1))^+\right],
$$

which for integer multiplexing gain values simplifies to

$$
\bar{c}_{red}(r)=\frac{1}{n_T}r(n_T-r),\ \text{ for } r=0,1,\cdots,n_T.
$$

For the proof to be complete we must now prove that the aforementioned family of ARQ schemes, halting policies and lattice designs can indeed achieve the desired DMT $d^*(r)$. For this purpose we recall the following lemma from [10]:

Lemma 1: For an i.i.d. regular fading channel, a minimum delay ARQ scheme with $L=n_{\mathrm{T}}$ rounds achieves $d^*(r)$ for all ARQ-compatible, minimum delay, NVD, rate-1 lattice designs, all intermediate aggressive halting policies and a sphere decoder with search radius $\xi>\sqrt{d^*(\frac{r}{L})\log\rho}$.

In the presence of Lemma 1 and upper bound $\bar{c}_{red}(r)$, it is direct to see that a minimum delay ARQ scheme with $L=n_{\mathrm{T}}$ rounds achieves $d^*(r)$ with $c(r)\leq\bar{c}_{red}(r)$, irrespective of the ARQ-compatible, minimum delay, NVD, rate-1 lattice design, for any aggressive intermediate halting policy, and any decoding order policy. This proves Theorem 1, Proposition 1 and Proposition 2. □

APPENDIX B
SKETCH PROOF OF THEOREM 2 AND PROPOSITION 3

In this section we present the sketch of the proof for Theorem 2 and Proposition 3. Let lattice design $\theta\mathbf{X}_C^{ARQ,L}$ from Proposition 3 takes the form

$$
\mathbf{X}_C^{ARQ,L}=\begin{bmatrix}\mathbf{X}&\cdots&\mathbf{0}\\ \vdots&\ddots&\vdots\\ \mathbf{0}&\cdots&\mathbf{X}\end{bmatrix}\in\mathbb{C}^{n_T\times n_T}, \qquad (14)
$$

where for rate-$\frac{n_R}{n_T}$ lattice designs we have $\theta^2=\rho^{1-\frac{n_T r_L}{n_R}}$, where $r_L=\frac{r_1}{L}$ denotes multiplexing gain for L-th round of ARQ, and where block component code $\mathbf{X}\in\mathbb{C}^{n_R\times n_R}$. The L-th round received signal is given by

$$
\mathbf{Y}_C^L=\theta\mathbf{H}_C\mathbf{X}_C^{ARQ,L}+\mathbf{W}_C^L. \qquad (15)
$$

Let

$$
\mathbf{H}_C=\begin{bmatrix}\mathbf{H}_1&\cdots&\mathbf{H}_{\frac{n_T}{n_R}}\end{bmatrix}\in\mathbb{C}^{n_R\times n_T},
$$

where $\mathbf{H}_i\in\mathbb{C}^{n_R\times n_R}$, for $i=1,\cdots,\frac{n_T}{n_R}$. After substituting for $\mathbf{H}_C$ and $\mathbf{X}_C^{ARQ,L}$ in (15) we get that

$$
\begin{aligned}
\mathbf{Y}_C^L=&\theta\begin{bmatrix}\mathbf{H}_1&\cdots&\mathbf{H}_{\frac{n_T}{n_R}}\end{bmatrix}\begin{bmatrix}\mathbf{X}&\cdots&\mathbf{0}\\ \vdots&\ddots&\vdots\\ \mathbf{0}&\cdots&\mathbf{X}\end{bmatrix}+\mathbf{W}_C^L,\\
=&\theta\tilde{\mathbf{H}}_C\mathbf{X}+\mathbf{W}_C^L,
\end{aligned}
$$

where $\tilde{\mathbf{H}}_C=[\mathbf{H}_1^T\ \cdots\ \mathbf{H}_{\frac{n_T}{n_R}}^T]^T\in\mathbb{C}^{n_T\times n_R}$.

We observe that the lattice design in (14) converts the system into an equivalent channel $\tilde{\mathbf{H}}_C$ with inverted channel dimensions. It follows that for this equivalent channel $\tilde{\mathbf{H}}_C$ the system parameters are given by $n_T'=n_R$, $n_R'=n_T$, $T'=1$, $L'=n_T'$ and $r_L'=\frac{n_T}{n_R}r_L$. Since for the new system we have $n_R'\geq n_T'$, Proposition 3 and Theorem 2 can now be proved by following the footsteps of the proof of Theorem 1. This completes the proof of Theorem 2 and Proposition 3. □

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] J. Jaldén and P. Elia, "Sphere decoding complexity exponent for decoding full rate codes over the quasi-static mimo channel," Feb. 2012, to appear in *IEEE Trans. Inform. Theory*, available on arXiv:1102.1265 [cs.IT].

[3] A. Singh, P. Elia, and J. Jaldén, "Achieving a vanishing performance-gap to exact lattice decoding at a subexponential complexity," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3692 – 3707, Jun. 2012.

[4] ——, " Complexity analysis for ML-based sphere decoder achieving a vanishing performance-gap to brute force ML decoding," in *Proc. Int. Zurich Seminar on Communications (IZS)*, Mar. 2012, pp. 127–130.

[5] H. El Gamal, G. Caire, and M. O. Damen, "The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3601–3621, Aug. 2006.

[6] S. A. Pawar, K. Raj Kumar, P. Elia, P. Vijay Kumar, and B. A. Sethuraman, "Space-time codes achieving the DMD tradeoff of the MIMO-ARQ channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, Jul. 2009.

[7] L. Zhao, W. Mo, Y. Ma, and Z. Wang, "Diversity and multiplexing tradeoff in general fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1547–1557, Apr. 2007.

[8] J.-C. Belfiore and G. Rekaya, "Quaternionic lattices for space-time coding," in *Proc. IEEE Information Theory Workshop (ITW)*, Paris, France, Mar. 2003.

[9] B. A. Sethuraman, B. Sundar Rajan, and V. Shashidhar, "Full-diversity, high-rate, space-time block codes from division algebras," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2596–2616, Oct. 2003.

[10] A. Singh, " Rate-Reliability-Complexity limits in ML and lattice based decoding for MIMO, multiuser and cooperative communications," Ph.D. dissertation, EURECOM/Telecom ParisTech: France, 2012.