Fundamental Limits of Infinite Constellations in MIMO Fading Channels

Yair Yona
Dept. of EE-Systems
Tel-Aviv University
Tel-Aviv 69978, Israel
Email: yairyo@eng.tau.ac.il

Meir Feder
Dept. of EE-Systems
Tel-Aviv University
Tel-Aviv 69978, Israel
Email: meir@eng.tau.ac.il

***Abstract*—The fundamental and natural connection between the infinite constellation (IC) dimension and the best diversity order it can achieve is investigated in this paper. In the first part of this work we develop an upper bound on the diversity order of IC's for any dimension and any number of transmit and receive antennas. By choosing the right dimensions, we prove in the second part of this work that IC's in general and lattices in particular can achieve the optimal diversity-multiplexing tradeoff of finite constellations. This work gives a framework for designing lattices for multiple-antenna channels using lattice decoding.**

I. Introduction

The use of multiple antennas in wireless communication has certain inherent advantages. On one hand, using multiple antennas in fading channels allows to increase the transmitted signal reliability, i.e. diversity. For instance, diversity can be attained by transmitting the same information on different paths between transmitting-receiving antenna pairs with i.i.d Rayleigh fading distribution. The number of independent paths used is the diversity order of the transmitted scheme. On the other hand, the use of multiple antennas increases the number of degrees of freedom available by the channel. In [1],[2] the ergodic channel capacity was obtained for multiple-input multiple-output (MIMO) systems with M transmit and N receive antennas, where the paths have i.i.d Rayleigh fading distribution. It was shown that for large signal to noise ratios (SNR), the capacity behaves as $C(SNR) \approx \min(M,N)\log(SNR)$. The multiplexing gain is the number of degrees of freedom utilized by the transmitted scheme.

For the quasi-static Rayleigh flat-fading channel, Zheng and Tse [3] characterized the dependence between the diversity order and the multiplexing gain, by deriving the optimal tradeoff between diversity and multiplexing, i.e. for each multiplexing gain found the maximal diversity order. They showed that the optimal diversity-multiplexing tradeoff (DMT) can be attained by ensemble of i.i.d Gaussian codes, given that the block length is greater or equal to $N+M-1$. For this case, the tradeoff curve takes the form of the piecewise linear function that connects the points $(N-l)(M-l)$, $l=0,1,\dots,\min(M,N)$.

The material in this paper was presented in part at the IEEE International Symposium in Information Theory (ISIT) 2011.

This research was supported by THE ISRAEL SCIENCE FOUNDATION grant No. 634/09.

Space time codes are coding schemes designed for MIMO systems. There has been an extensive work in this field [4],[5] [6] and references therein. Some of these works present schemes that maximize the diversity order, others maximize the multiplexing gain, and there are also works aimed at achieving the optimal DMT. In [7], El Gamal et al presented lattice space time (LAST) codes. These space time codes are subsets of an infinite lattice, where the lattice dimensionality equals to the number of degrees of freedom available by the channel, i.e. $\min(M,N)$. By using an ensemble of nested lattices, common randomness, generalized minimum Euclidean lattice decoding and modulo lattice operation (that in a certain sense takes into account the finite code book), they showed that LAST codes can achieve the optimal DMT.

The authors in [7] also derived a lower bound on the diversity order, for the case $N\geq M$, for LAST codes shaped into a sphere with regular lattice decoding, i.e. decoding over the infinite lattice without taking into consideration the finite codebook. For sufficiently large block length they showed that $d(r)\geq(N-M+1)(M-r)$ where r is the multiplexing gain and the lattice dimension is M. Taherzadeh and Khandani showed in [8] that this is also an upper bound on the diversity order of any LAST code shaped into a sphere and decoded with regular lattice decoding. These results show that LAST codes together with regular lattice decoding are suboptimal compared to the optimal DMT of power constrained constellations.

Infinite constellations (IC's) are structures in the Euclidean space that have no power constraint. In [9], Poltyrev analyzed the performance of IC's over the additive white Gaussian noise (AWGN) channel. In this work we first extend the definitions of diversity order and multiplexing gain to the case where there is no power constraint. We also introduce a new term: the average number of dimensions per channel use, which is essentially the IC dimension divided by the number of channel uses. Then we extend the methods used in [9] in order to derive an upper bound on the diversity of any IC with certain average number of dimensions per channel use, as a function of the multiplexing gain. It turns out that for a given number of dimensions per channel use, the diversity is a straight line as a function of the multiplexing gain that depends on the IC number of transmit and receive antennas. This analysis holds for any M and N, and also for lattices and regular lattice decoding. We also find the average number of dimensions per channel use for which the upper bounds coincide with the optimal DMT of finite constellations. Finally, we show that for the aforementioned average number of dimensions per channel use, together with sufficient amount of channel uses, there exist sequences of lattices that attain different segments of the optimal DMT with regular lattice decoding, i.e. for each point in the DMT of [3] there exists a lattice sequence of certain dimension that achieves it with regular lattice decoding.

This work gives a framework for designing lattices for multiple-antenna channels using regular lattice decoding. It also shows the fundamental and natural connection between the IC dimension and its optimal diversity order. For instance, it is shown that for the case $M = N = 2$, the maximal diversity order of 4 can be achieved (with regular lattice decoding) by a lattice that has at most $\frac{4}{3}$ average number of dimensions per channel use. On the other hand the Alamouti scheme [10], that also has maximal diversity order of 4, utilizes only a single dimension per channel use in this set up. Hence, there is still a room for improvement of $\frac{1}{3}$ dimensions per channel use, while attaining full diversity.

The outline of the paper is as follows. In section II basic definitions for the fading channel and IC's are given. Section III presents a lower bound on the average decoding error probability of IC for any channel realization, and an upper bound on the diversity order. An upper bound on the error probability for each channel realization, a transmission scheme that attains the best diversity order and some averaging arguments regarding the achievable diversity order of IC's, are all presented on section IV.

II. Basic Definitions

We refer to the countable set $S = \{s_1, s_2, \dots\}$ in $\mathbb{C}^n$ as infinite constellation (IC). Let $\mathrm{cube}_l(a) \subset \mathbb{C}^n$ be a (probably rotated) l-complex dimensional cube ($l \leq n$) with edge of length a centered around zero. An IC S_l is l-complex dimensional if there exists rotated l-complex dimensional cube $\mathrm{cube}_l(a)$ such that $S_l \subset \lim_{a\to\infty} \mathrm{cube}_l(a)$ and l is minimal. $M(S_l, a) = |S_l \bigcap \mathrm{cube}_l(a)|$ is the number of points of the IC S_l inside $\mathrm{cube}_l(a)$. In [9], the n-complex dimensional IC density for the AWGN channel was defined as the upper limit (the limit supremum) of the ratio $\gamma_{\mathrm{G}} = \limsup_{a\to\infty} \frac{M(S,a)}{a^{2n}}$ and the volume to noise ratio (VNR) was given as $\mu_{\mathrm{G}} = \frac{\gamma_{\mathrm{G}}^{-\frac{1}{n}}}{2\pi e\sigma^2}$.

The Voronoi region of a point $x \in S_l$, denoted as $V(x)$, is the set of points in $\lim_{a\to\infty} \mathrm{cube}_l(a)$ closer to x than to any other point in the IC. The effective radius of the point $x \in S_l$, denoted as $r_{\mathrm{eff}}(x)$, is the radius of the l-complex dimensional ball that has the same volume as the Voronoi region, i.e. $r_{\mathrm{eff}}(x)$ satisfies

$$|V(x)| = \frac{\pi^l r_{\mathrm{eff}}^{2\cdot l}(x)}{\Gamma(l+1)}. \quad (1)$$

We consider a quasi static flat-fading channel with M transmit and N receive antennas. We assume for this MIMO channel perfect channel knowledge at the receiver and no channel knowledge at the transmitter. The channel model is as follows:

$$\underline{y}_t = H \cdot \underline{x}_t + \rho^{-\frac{1}{2}} \underline{n}_t \qquad t = 1, \dots, T \quad (2)$$

where $\underline{x} = \{\underline{x}_1, \dots, \underline{x}_T\} \in S_l \subset \mathbb{C}^{MT}$ belongs to the infinite constellation with density $\gamma_{tr} = \limsup_{a\to\infty} \frac{M(S_l,a)}{a^{2\cdot l}}$ (where $a^{2\cdot l}$ is the volume of $cube_l(a)$), $\underline{n}_t \sim CN(\underline{0}, \frac{2}{2\pi e} I_N)$ where CN denotes complex-normal, I_N is the N-dimensional unit matrix, and $\underline{y}_t \in \mathbb{C}^N$. H is the fading matrix with N rows and M columns where $h_{i,j} \sim CN(0,1)$, $1 \leq i \leq N$, $1 \leq j \leq M$, and $\rho^{-\frac{1}{2}}$ is a scalar that multiplies each element of $\underline{n}_t$, where ρ plays the role of average SNR in the receive antenna for power constrained constellations that satisfy $\frac{1}{T} E\{\|\underline{x}\|^2\} \leq \frac{2M}{2\pi e}$.

By defining H_{ex} as an $NT \times MT$ block diagonal matrix, where each block on the diagonal equals H, $\underline{n}_{\mathrm{ex}} = \rho^{-\frac{1}{2}} \cdot \{\underline{n}_1, \dots, \underline{n}_T\} \in \mathbb{C}^{NT}$ and $\underline{y}_{\mathrm{ex}} \in \mathbb{C}^{NT}$ we can rewrite the channel model in (2) as

$$\underline{y}_{\mathrm{ex}} = H_{\mathrm{ex}} \cdot \underline{x} + \underline{n}_{\mathrm{ex}}. \quad (3)$$

In the sequel we use L to denote $min(M, N)$. We define as $\sqrt{\lambda_i}$, $1 \leq i \leq L$ the real valued, non-negative singular values of H. We assume $\sqrt{\lambda_L} \geq \cdots \geq \sqrt{\lambda_1} > 0$. Our analysis is done for large values of ρ (large VNR at the transmitter). We state that $f(\rho) \dot{\geq} g(\rho)$ when $\lim_{\rho\to\infty} -\frac{f(\rho)}{\ln(\rho)} \leq -\frac{g(\rho)}{\ln(\rho)}$, and also define $\dot{\leq}$, $\doteq$ in a similar manner by substituting $\leq$ with $\geq$, $=$ respectively.

We now turn to the IC definitions in the transmitter. We define the average number of dimensions per channel use as the IC dimension divided by the number of channel uses. We denote the average number of dimensions per channel use by K. Let us consider a KT-complex dimensional sequence of IC's $S_{KT}(\rho)$, where $K \leq L$, and T is the number of channel uses. First we define $\gamma_{tr} = \rho^{rT}$ as the density of $S_{KT}(\rho)$ in the transmitter. The IC multiplexing gain is defined as

$$MG(r) = \lim_{\rho\to\infty} \frac{1}{T} \log_\rho(\gamma_{\mathrm{tr}} + 1) = \lim_{\rho\to\infty} \frac{1}{T} \log_\rho(\rho^{rT} + 1). \quad (4)$$

Note that $MG(r) = max(0, r)$. For $0 \leq r \leq K$, $r = MG(r)$ has the meaning of multiplexing gain. Roughly speaking, $\gamma_{tr} = \rho^{rT}$ gives us the number of points of $S_{KT}(\rho)$ within the KT-complex dimensional region $cube_{KT}(1)$. In order to get the multiplexing gain, we normalizing the exponent of the number of points within $cube_{KT}(1)$, rT, by the number of channel uses - T. Note that the IC multiplexing gain, r, can be directly translated to finite constellation multiplexing gain r by considering the IC points within a shaping region. The VNR in the transmitter is

$$\mu_{\mathrm{tr}} = \frac{\gamma_{\mathrm{tr}}^{-\frac{1}{KT}}}{2\pi e\sigma^2} = \rho^{1-\frac{r}{K}} \quad (5)$$

where $\sigma^2 = \frac{\rho^{-1}}{2\pi e}$ is each dimension noise variance. Now we can understand the role of the multiplexing gain for IC's. The AWGN variance decreases as ρ^{-1}, where the IC density increases as ρ^{rT}. When $r = 0$ we get constant IC density as a function of ρ, where the noise variance decreases, i.e. we get the best error exponent. In this case the number of words within $cube_{KT}(1)$ remains constant as a function of ρ. On the other hand, when $r = K$, we get VNR $\mu_{tr} = 1$, and from [9] we know that it inflicts average error probability that is bounded away from zero. In this case, the increase in the number of IC words within $cube_{KT}(1)$ is at maximal rate.

Now we turn to the IC definitions in the receiver. First we define the set $H_{ex} \cdot cube_{KT}(a)$ as the multiplication of each point in $cube_{KT}(a)$ with the matrix H_{ex}. In a similar manner $S'_{KT} = H_{ex} \cdot S_{KT}$. The set $H_{ex} \cdot cube_{KT}(a)$ is almost surely KT-complex dimensional (where $K \leq L$) and in this case $M(S_{KT}, a) = |S_{KT} \bigcap \mathrm{cube}_{KT}(a)| = |S'_{KT} \bigcap (H_{ex} \cdot \mathrm{cube}_{KT}(a))|$. We define the receiver density as

$$\gamma_{\mathrm{rc}} = \limsup_{a \to \infty} \frac{M(S_{KT}, a)}{Vol(H_{ex}.\mathrm{cube}_{KT}(a))}$$

i.e., the upper limit of the ratio of the number of IC words in $H_{ex}.\mathrm{cube}_{KT}(a)$, and the volume of $H_{ex}.\mathrm{cube}_{KT}(a)$. The volume of the set $H_{ex} \cdot cube_{KT}(a)$ is smaller than $a^{2KT} \cdot \lambda_L^T \dots \lambda_{L-B+1}^T . \lambda_{L-B}^{\beta T}$, assuming $K = B + \beta$ where $B \in \mathbb{N}$ and $0 < \beta \leq 1$, i.e. the volume is smaller than the multiplication of the $B+1$ strongest singular values, raised to the power of the maximal amount of channel uses each can take place in. Hence we get

$$\gamma_{\mathrm{rc}} \geq \rho^{rT} \lambda_L^{-T} \dots \lambda_{L-B+1}^{-T} \cdot \lambda_{L-B}^{-\beta T} \qquad (6)$$

and the receiver VNR is

$$\mu_{\mathrm{rc}} \leq \rho^{1-\frac{r}{K}} \cdot \lambda_L^{\frac{1}{K}} \dots \lambda_{L-B+1}^{\frac{1}{K}} \cdot \lambda_{L-B}^{\frac{\beta}{K}}. \qquad (7)$$

Note that for $N \geq M$ and $K = M$ we get $\gamma_{\mathrm{rc}} = \rho^{rT} \cdot \prod_{i=1}^{M} \lambda_i^{-T}$ and $\mu_{\mathrm{rc}} = \rho^{1-\frac{r}{M}} \cdot \prod_{i=1}^{M} \lambda_i^{\frac{1}{M}}$. The average decoding error probability over the IC points of $S_{KT}(\rho)$, for a certain channel realization H, is defined as

$$\overline{Pe}(H, \rho) = \limsup_{a \to \infty} \frac{\sum_{\underline{x}' \in S'_{KT} \bigcap (H_{ex} \cdot \mathrm{cube}_{KT}(a))} Pe(\underline{x}', H, \rho)}{M(S_{KT}, a)} \qquad (8)$$

where $Pe(\underline{x}', H, \rho)$ is the error probability of $\underline{x}'$. The average decoding error probability of $S_{KT}(\rho)$ over all channel realizations is $\overline{Pe}(\rho) = E_H\{\overline{Pe}(H, \rho)\}$. Hence the $diversity\ order$ equals

$$d = -\lim_{\rho \to \infty} \log_\rho(\overline{Pe}(\rho)) \qquad (9)$$

III. Upper Bound on the Diversity Order

In this section we derive an upper bound on the diversity order of any IC with average number of dimensions per channel use K and any value of T, M and N. We begin by deriving a lower bound on the average decoding error probability of $S_{KT}(\rho)$ for each channel realization. As in [3] and [7], we also define $\lambda_i = \rho^{-\alpha_i}$, $1 \leq i \leq L$. For very large ρ, the Wishart distribution is of the form $\rho^{-\sum_{i=1}^{L}(|N-M|+2i-1)\alpha_i}$ and we can assume $0 \leq \alpha_L \leq \dots \leq \alpha_1$. By assigning in (6), (7) respectively, we can write

$$\gamma_{\mathrm{rc}} \geq \rho^{T(r + \sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B})}$$

and

$$\mu_{\mathrm{rc}} \leq \rho^{1 - \frac{1}{K}(r + \sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B})}.$$

Theorem 1. *For any KT-complex dimensional IC $S_{KT}(\rho)$ with transmitter density $\gamma_{\mathrm{tr}} = \rho^{rT}$ and channel realization $\underline{\alpha} = (\alpha_1, \dots, \alpha_L)$, we have the following lower bound on the average decoding error probability for $0 \leq r \leq K$*

$$\overline{Pe}(H, \rho) > \frac{C(KT)}{4} e^{-\mu_{\mathrm{rc}} \cdot A(KT) + (KT-1)\ln(\mu_{\mathrm{rc}})}$$

where $A(KT) = e \cdot \Gamma(KT+1)^{\frac{1}{KT}}$ and $C(KT) = \frac{e^{KT-\frac{3}{2}} \Gamma(KT+1)^{\frac{KT-1}{KT}}}{2 \cdot \Gamma(KT)}$.

Proof: We divide the proof into 2 parts. In the first part we prove the result for lattices, that constitute a symmetric structure for which the Voronoi regions of different lattice points are identical. In the second part we prove the result for IC's with receiver density γ_{rc} and no restrictions other than that on the IC structure. As the second part proof is somewhat more involved, we present it on appendix A.

We begin by proving the result for lattices. Lattices constitute a discrete subgroup of the Euclidean space, with the ordinary vector addition operation. Consider a KT-complex dimensional lattice, $S'_{KT}(\rho)$, in the receiver with density γ_{rc}. The lattice points have identical Voronoi regions up to a translate. Hence, the volume of each Voronoi region equals

$$|V(x)| = \frac{1}{\gamma_{rc}} \quad \forall x \in S'_{KT}(\rho).$$

According to the definition of the effective radius in (1), we get that $r_{\mathrm{eff}}(x) = r_{\mathrm{eff}}(\gamma_{rc})$, $\forall x \in S'_{KT}(\rho)$. Note that in lattices the maximum-likelihood (ML) decoding error probability is identical for all lattice points, i.e. the average and maximal error probabilities are identical. It has been proven in [9], [11] that the error probability of any lattice point in the receiver fulfils

$$Pe_{S'_{KT}} > Pr(\|\underline{\tilde{n}}_{\mathrm{ex}}\| \geq r_{\mathrm{eff}}(\gamma_{rc}))$$

where $Pe_{S'_{KT}}$ is the ML decoding error probability of any lattice point, and $\underline{\tilde{n}}_{\mathrm{ex}}$ is the effective noise in the KT-complex dimensional hyperplane where $S'_{KT}(\rho)$ resides. We find an explicit expression to the lower bound

$$\Pr\left(\|\underline{\tilde{n}}_{\mathrm{ex}}\| \geq r_{\mathrm{eff}}(\gamma_{\mathrm{rc}})\right) > \Pr\left(\|\underline{\tilde{n}}_{\mathrm{ex}}\| \geq r_{\mathrm{eff}}(\frac{\gamma_{\mathrm{rc}}}{2})\right) >$$
$$\int_{r_{\mathrm{eff}}^2}^{r_{\mathrm{eff}}^2+\sigma^2} \frac{r^{KT-1} e^{-\frac{r}{2\sigma^2}}}{\sigma^{2KT} 2^{KT} \Gamma(KT)} dr \geq \frac{r_{\mathrm{eff}}^{2KT-2} e^{-\frac{r_{\mathrm{eff}}^2}{2\sigma^2}}}{\sigma^{2KT-2} 2^{KT} \Gamma(KT) \sqrt{e}}. \qquad (10)$$

By assigning $r_{\mathrm{eff}}^2 = (\frac{2 \cdot \Gamma(KT+1)}{\gamma_{rc} \pi^{KT}})^{\frac{1}{KT}}$ we get

$$Pe_{S'_{KT}} > C(KT) \cdot e^{-\frac{\gamma_{\mathrm{rc}}^{-\frac{1}{KT}}}{2\pi e \sigma^2} A(KT) + (KT-1)\ln(\frac{\gamma_{\mathrm{rc}}^{-\frac{1}{KT}}}{2\pi e \sigma^2})}$$

and by assigning $\mu_{rc} = \frac{\gamma_{\mathrm{rc}}^{-\frac{1}{KT}}}{2\pi e \sigma^2}$ we get

$$Pe_{S'_{KT}} > \frac{C(KT)}{4} \cdot e^{-\mu_{rc} A(KT) + (KT-1)\ln(\mu_{rc})}. \qquad (11)$$

Note that in (10) we lower bounded the error probability with $r_{\text{eff}}(\frac{\gamma_{rc}}{2})$ instead of $r_{\text{eff}}(\gamma_{rc})$, and also in (11) we multiplied by $\frac{1}{4}$, in order to be consistent with the general lower bound for IC's. For lattices we have $\overline{Pe}(H,\rho) = Pe_{S'_{KT}}$. Hence this concludes the proof. We give the proof for general IC's in appendix A. ■

Next, we would like to use this lower bound to average over the channel realizations and get an upper bound on the diversity order.

Theorem 2. *The diversity order of any KT-complex dimensional sequence of IC's $S_{KT}(\rho)$, with K average number of dimensions per channel use, is upper bounded by*

$$d_{KT}(r) \leq d^*_K(r) = M \cdot N(1 - \frac{r}{K})$$

for $0 < K \leq \frac{M \cdot N}{N+M-1}$, *and*

$$d_{KT}(r) \leq d^*_K(r) = (M-l)(N-l)\frac{K}{K-l}(1 - \frac{r}{K})$$

for $\frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1 < K \leq \frac{(M-l)(N-l)}{N+M-1-2 \cdot l} + l$ *and* $l = 1, \ldots, L-1$. *In all of these cases* $0 \leq r \leq K$.

Proof: For any IC with VNR μ_{rc}, assigning $\mu'_{rc} > \mu_{rc}$ in the lower bound from Theorem 1 also gives a lower bound on the error probability

$$\overline{Pe}(H,\rho) > \frac{C(KT)}{4} e^{-\mu'_{\text{rc}} \cdot A(KT) + (KT-1)\ln(\mu'_{\text{rc}})}.$$

It results from the fact that inflating the IC into an IC with VNR μ'_{rc} must decrease the error probability, where $\frac{C(KT)}{4} e^{-\mu'_{\text{rc}} \cdot A(KT) + (KT-1)\ln(\mu'_{\text{rc}})}$ is a lower bound on the error probability of any IC with VNR μ'_{rc}. Hence, for the case $\mu_{rc} \leq 1$ we can lower bound the error probability by assigning 1 in the lower bound and get $\frac{C(KT)}{4} e^{-A(KT)}$, i.e. for $\mu_{rc} \leq 1$ the average decoding error probability is bounded away from 0 for any value of ρ. We can give the event $\mu_{rc} \leq 1$ the interpretation of outage event.

We would like to set a lower bound for the error probability for each channel realization $\underline{\alpha}$, which we denote by $P_e^{LB}(\rho, \underline{\alpha})$. We know that $\mu_{\text{rc}} \leq \rho^{1 - \frac{1}{K}(r + \sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B})}$. For the case $\sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B} < K - r$, we take

$$P_e^{LB}(\rho, \underline{\alpha}) = \frac{C(KT)}{4} e^{-L(\rho,\underline{\alpha}) \cdot A(KT) + (KT-1)\ln(L(\rho,\underline{\alpha}))}$$

where $L(\rho, \underline{\alpha}) = \rho^{1 - \frac{1}{K}(r + \sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B})} > 1$. For the case $\sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B} \geq K - r$ we get that $\mu_{rc} \leq 1$, and we take

$$P_e^{LB}(\rho, \underline{\alpha}) = \frac{C(KT)}{4} e^{-A(KT)}.$$

In order to find an upper bound on the diversity order, we would like to average $P_e^{LB}(\rho, \underline{\alpha})$ over the channel realizations. In our analysis we consider large values of ρ, and so we calculate

$$\overline{P_e}(\rho) \dot{>} \int_{\underline{\alpha} \geq 0} P_e^{LB}(\rho, \underline{\alpha}) \cdot \rho^{-\sum_{i=1}^{L}(|N-M|+2i-1)\alpha_i} d\underline{\alpha} \quad (12)$$

where $\underline{\alpha} \geq 0$ signifies the fact that $\alpha_1 \geq \cdots \geq \alpha_L \geq 0$. By defining $\mathcal{A} = \{\underline{\alpha} | \sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B} < K - r; \underline{\alpha} \geq 0\}$ and $\overline{\mathcal{A}} = \{\underline{\alpha} | \sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B} \geq K - r; \underline{\alpha} \geq 0\}$ we can split (12) into 2 terms

$$\overline{P_e}(\rho) \dot{>} \int_{\underline{\alpha} \in \mathcal{A}} P_e^{LB}(\rho, \underline{\alpha}) \cdot \rho^{-\sum_{i=1}^{L}(|N-M|+2i-1)\alpha_i} d\underline{\alpha} + \int_{\underline{\alpha} \in \overline{\mathcal{A}}} P_e^{LB}(\rho, \underline{\alpha}) \cdot \rho^{-\sum_{i=1}^{L}(|N-M|+2i-1)\alpha_i} d\underline{\alpha}. \quad (13)$$

Hence, we get

$$\overline{P_e}(\rho) \dot{>} \int_{\underline{\alpha} \in \overline{\mathcal{A}}} P_e^{LB}(\rho, \underline{\alpha}) \cdot \rho^{-\sum_{i=1}^{L}(|N-M|+2i-1)\alpha_i} d\underline{\alpha}. \quad (14)$$

In a similar manner to [3], [7], for very large ρ, we approximate the average by finding the most dominant exponential term in the integral, i.e. We would like to find the minimal value of

$$\lim_{\rho \to \infty} -\log_\rho (P_e^{LB}(\rho, \underline{\alpha}) \cdot \rho^{-\sum_{i=1}^{L}(|N-M|+2i-1)\alpha_i})$$

for the case $\underline{\alpha} \in \overline{\mathcal{A}}$. For $\underline{\alpha} \in \overline{\mathcal{A}}$, we get that $P_e^{LB}(\rho, \underline{\alpha})$ is bounded away from 0 for any value of ρ. Hence, in order to find the most dominant error event we would like to find $\min_{\underline{\alpha}} \sum_{i=1}^{L}(|N-M| + 2i - 1)\alpha_i$ given that $\underline{\alpha} \in \overline{\mathcal{A}}$. The minimal value is achieved for the case where $\sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B} = K - r$ and $\underline{\alpha} \geq 0$. Hence, for any $K \leq L$ we state that

$$d_{KT}(r) \leq \min_{\underline{\alpha}} \sum_{i=1}^{L}(|N-M| + 2i - 1)\alpha_i, \qquad 0 \leq r \leq K \quad (15)$$

where $\sum_{i=0}^{B-1} \alpha_{L-i} + \beta \alpha_{L-B} = K - r$ and $\alpha_1 \geq \cdots \geq \alpha_L \geq 0$. Basically this optimization problem is a linear programming problem whose solution is as follows. For $0 < K \leq \frac{M \cdot N}{N+M-1}$ the optimization problem solution is $\alpha_i = 1 - \frac{r}{K}$, $i = 1, \ldots, L$. For $\frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1 < K \leq \frac{(M-l)(N-l)}{N+M-1-2 \cdot l} + l$ and $l = 1, \ldots, L-1$ the optimization problem solution is $\alpha_L = \cdots = \alpha_{L-l+1} = 0$ and $\alpha_{L-l} = \cdots = \alpha_1 = \frac{K-r}{K-l}$. By assigning the optimization problem solution, we get the desired upper bound. The optimization problem is solved on appendix B. ■

From Theorem 2 we get an upper bound on the diversity order by assuming transmission of the KT complex dimensions over the $B+1$ strongest singular values. This assumption is equivalent to assuming *beamforming* which may improve the coding gain, but does not increase the diversity order. This assumption allowes us to derive a ower nound on the average decoding error probability. However, we still get maximal diversity order of MN in this case.

Let us consider as an illustrative example the case of $M = N = 2$. In this case, for $0 < K \leq \frac{4}{3}$ we get $d^*_K(r) = 4(1 - \frac{r}{K})$. For $\frac{4}{3} < K \leq 2$ we get $d^*_k(r) = \frac{K}{K-1}(1 - \frac{r}{K})$. In both cases $0 \leq r \leq K$. For this set up we have 2 singular values and so $\alpha_1 \geq \alpha_2 \geq 0$. The optimization problem is of the form $\min_{\underline{\alpha} \geq 0} \alpha_1 + 3\alpha_2$, where for $0 < K \leq 1$ the constraint is $\beta \alpha_2 = K - r$, and for $1 < K \leq 2$ the constraint is $\alpha_2 + \beta \alpha_1 = K - r$. For the case $0 < K < \frac{4}{3}$ the optimization problem solution is $\alpha_1 = \alpha_2 = 1 - \frac{r}{K}$, i.e. in this case the most dominant error event occurs when both singular values are very small. For the case $K = \frac{4}{3}$ the constraint is of the form $\alpha_2 + \frac{\alpha_1}{3} = \frac{4}{3} - r$, and the optimization problem solution is achieved for both $\alpha_1 = \alpha_2 = 1 - \frac{3r}{4}$ and $\alpha_2 = 0,\ \alpha_1 = 4 - 3r$. For the case $\frac{4}{3} < K \leq 2$ the optimization problem solution is achieved for $\alpha_2 = 0,\ \alpha_1 = \frac{K-r}{K-1}$, i.e. one strong singular value and another very weak singular value.

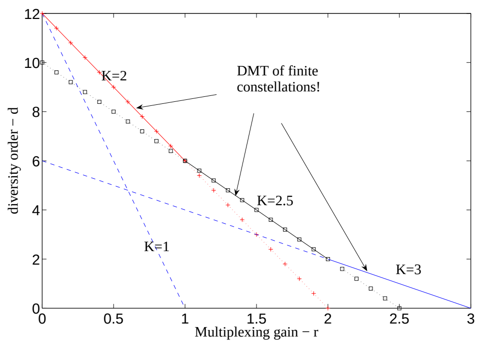

Fig. 1. The diversity order as a linear function of the multiplexing gain r for $M = 4$, $N = 3$ and $K = 1,\ 2,\ 2.5$ and 3.

Corollary 1. *For* $0 < K \leq \frac{M \cdot N}{N+M-1}$ *we get* $d^*_K(0) = MN$. *For* $\frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1 < K \leq \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$, $l = 1, \ldots, L-1$ *we get* $d^*_K(l) = (M-l)(N-l)$.

Proof: The proof is straight forward from $d^*_K(r)$ properties. ■

From Corollary 1 we get that the range of K can be divided into segments, where for each segment we have a set of straight lines, that are all equal at a certain integer point. Note that at these points, we get the same values as the finite constellations optimal DMT.

Corollary 2. *In the range* $l \leq r \leq l+1$, *the maximal possible diversity order is achieved at dimension* $K_l = \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$ *and gives*

$$d^*_{K_l}(r) = (M-l)(N-l)\frac{K_l}{K_l - l}(1 - \frac{r}{K_l})$$
$$= (M-l)(N-l) - (r-l)(N+M-2\cdot l-1)$$

where $l = 0, \ldots, L-1$.

Proof: The proof is straight forward from $d^*_K(r)$ properties. ■

From Corollary 2 we can see that $d^*_{K_l}(l) = (M-l)(N-l)$ and $d^*_{K_l}(l+1) = (M-l-1)(N-l-1)$. We also know that $d^*_{K_l}(r)$ is a straight line. Also, the finite constellations optimal DMT consists of a straight line in the range $l \leq r \leq l+1$, that equals $(N-l)(M-l)$ when $r = l$ and $(M-l-1)(N-l-1)$ when $r = l+1$. Hence, in the range $l \leq r \leq l+1$ for $K_l = \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$, we get an upper bound that equals to the optimal DMT of finite constellations presented in [3]. As for each $l = 0, \ldots, L-1$, we have such K_l, taking

$$\max_{0 \leq K \leq L} d^*_K(r) \quad 0 \leq r \leq L$$

gives us the optimal DMT of finite constellations.

Figure 1 illustrates the properties of $d^*_K(r)$ presented in Corrolaries 1, 2. We take the example of $M = 4$, $N = 3$. For $0 \leq K \leq 2$ we get upper bounds that have diversity order 12 for $r = 0$. We can see that in the range $0 \leq r \leq 1$, the upper bound of $K = 2$ is maximal and equals to the finite constellations optimal DMT. In the range $2 < K \leq 2.5$ we can see that the upper bounds have the same diversity order 6 at $r = 1$. In the range $1 \leq r \leq 2$, the upper bound of $K = 2.5$ is maximal and equals to the finite constellations optimal DMT in this range. For $2.5 < K \leq 3$, the upper bounds equal to 2 at $r = 2$. In the range $2 < r \leq 3$, the upper bound of $K = 3$ is maximal and again equals to the finite constellations optimal DMT in this range.

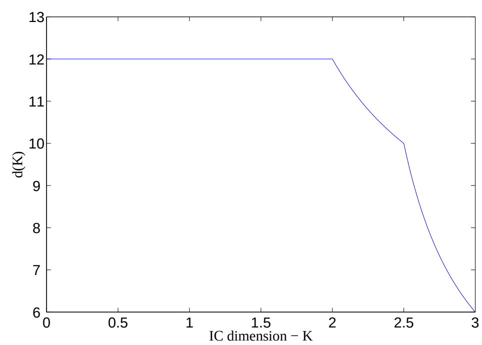

Fig. 2. $d^*_K(0)$ as a function of the IC dimensions per channel use K, for $M = 4$, $N = 3$.

On figure 2 we present the maximal diversity order that can be attained for different average number of dimensions per channel use, i.e. the upper bound on the diversity order for $r = 0$, $d^*_K(0)$, where $0 \leq K \leq L$. We present as an example the case where $M = 4$, $N = 3$. For this case, in the range $0 \leq K \leq 2$ we get $d^*_K(0) = 12$. It coincides with the result presented on figure 1, where we showed that in this range the straight lines have the same value for $r = 0$. Hence, for IC's, one can use up to 2 average number of dimensions per channel use without compromising on the diversity order. Starting from $K \geq 2$, the tradeoff starts to kick-in and the maximal diversity order starts to reduce as we increase the average number of dimensions per channel use. Also note that for $K = 3$ the diversity order is 6 when $r = 0$.

IV. Attaining the Best Diversity Order

In this section we show that the upper bound derived in section III is achievable by a sequence of IC's in general and lattices in particular. First we present a transmission scheme for any M, N, $K_l = \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$ and $T_l = N + M - 1 - 2\cdot l$ where $l = 0, \ldots, L-1$, and as previously defined $L = \min(M, N)$. Then we introduce the effective channel of the transmission scheme. Afterwards we extend the methods presented in [9], in order to derive an upper bound on the average decoding error probability of ensemble of IC's, for each channel realization. By averaging the upper bound over the channel realizations, we find the achievable DMT of IC's at these dimensions and show it coincides with the optimal DMT of finite constellations. Finally, we discuss peak to average properties of the transmission scheme, and show that there exists a single sequence of IC's that attains the optimal DMT of finite constellations.

A. The Transmission Scheme

The transmission matrix G_l, $l = 0, \ldots, L-1$, has M rows that represent the transmission antennas $1, \ldots, M$, and $T_l = N+M-1-2\cdot l$ columns that represent the number of channel uses.

We begin by describing the transmission matrix structure in general for any M and N.

1) For $N \geq M$ and $K_{M-1} = \frac{M(N-M+1)}{N-M+1} = M$: the matrix G_{M-1} has $N-M+1$ columns (channel uses). On each channel use, transmit different M symbols on antennas $1, \ldots, M$.
2) For $M > N$ and $K_{N-1} = \frac{N(M-N+1)}{M-N+1} = N$: the matrix G_{N-1} has $M-N+1$ columns. On the first column transmit symbols $x_1, \ldots, x_N$ on antennas $1, \ldots, N$ and on the $M-N+1$ column transmit symbols $x_{N(M-N)+1}, \ldots, x_{N(M-N+1)}$ on antennas $M-N+1, \ldots, M$.
3) For K_l, $l = 0, \ldots, L-2$: the matrix G_l has $M+N-1-2\cdot l$ columns. We add to G_{l+1}, the transmission scheme of K_{l+1}, two columns in order to get G_l. In the first added column transmit $l+1$ symbols on antennas $1, \ldots, l+1$. On the second added column transmit different $l+1$ symbols on antennas $M-l, \ldots, M$.

Example: $M = 4$, $N = 3$. In this case the transmission scheme for $K = 3$, 2.5 and 2 (G_2, G_1 and G_0 respectively) is as follows:

$$\underbrace{\underbrace{\underbrace{\begin{pmatrix} x_1 & 0 \\ x_2 & x_4 \\ x_3 & x_5 \\ 0 & x_6 \end{pmatrix.}_{K_2=\frac{6}{2}}\begin{matrix} x_7 & 0 \\ x_8 & 0 \\ 0 & x_9 \\ 0 & x_{10} \end{matrix}}_{K_1=\frac{10}{4}}\begin{matrix} x_{11} & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & x_{12} \end{matrix}\Bigg)}_{K_0=\frac{12}{6}}. \tag{16}$$

B. The Effective Channel

Next we define the effective channel matrix induced by the transmission scheme. In accordance with the channel model from (2), the multiplication $H \cdot G_l$ yields a matrix with N rows and T_l columns, where each column equals to $H \cdot \underline{x}_t$, $t = 1 \ldots T_l$, as in (2). We are interested in transmitting $K_l T_l$-complex dimensional IC with $K_l T_l$ complex symbols. Hence, in the proposed transmission scheme, G_l has exactly $K_l T_l$ non-zero complex entries that represent the $K_l T_l$-complex dimensional IC within $\mathbb{C}^{MT_l}$. For each column of G_l, denoted by $\underline{g}_i$, $i = 1 \ldots T_l$, we define the effective channel that $\underline{g}_i$ sees as $\widehat{H}_i$. It consists of the columns of H that correspond to the non-zero entries of $\underline{g}_i$, i.e. $H \cdot \underline{g}_i = \widehat{H}_i \cdot \widehat{\underline{g}}_i$, where $\widehat{\underline{g}}_i$ equals the non-zero entries of $\underline{g}_i$. As an example assume without loss of generality that the first l_i entries of $\underline{g}_i$ are not zero. In this case $\widehat{H}_i$ is an Nxl_i matrix equals to the first l_i columns of H. In accordance with (3), $H_{\text{eff}}^{(l)}$ is an NT_lx$K_l T_l$ block diagonal matrix consisting of T_l blocks. Each block corresponds to the multiplication of H with different column of G_l, i.e. $\widehat{H}_i$ is the $i'th$ block of $H_{\text{eff}}^{(l)}$. Note that in the effective matrix $NT_l \geq K_l T_l$.

We would like to elaborate on the structure of $H_{\text{eff}}^{(l)}$ blocks. For this reason we denote the columns of H as $\underline{h}_i$, $i = 1, \ldots, M$.

1) The case where $N \geq M$. For this case the transmission scheme has $N+M-1-2\cdot l$ columns. The first $N-M+1$ columns of G_l, $\underline{g}_1, \ldots, \underline{g}_{N-M+1}$, contain $M \cdot (N-M+1)$ different complex symbols, i.e. there are no zero entries in these columns. Hence, in this case the first $N-M+1$ blocks of $H_{\text{eff}}^{(l)}$ are
$$\widehat{H}_i = H \qquad i = 1, \cdots, N-M+1. \tag{17}$$
After the first $N-M+1$ columns we have $M-1-l$ pairs of columns. For each pair we have
$$\widehat{H}_{N-M+2k} = \{\underline{h}_1, \ldots, \underline{h}_{M-k}\} \tag{18}$$
and
$$\widehat{H}_{N-M+2k+1} = \{\underline{h}_{k+1}, \ldots, \underline{h}_M\} \tag{19}$$
where $k = 1, \ldots, M-1-l$.
2) The case where $M > N$. Again the transmission scheme has $N+M-1-2\cdot l$ columns. By the definition of the first $M-N+1$ columns of G_l, we get that
$$\widehat{H}_i = \{\underline{h}_i, \ldots, \underline{h}_{N+i-1}\} \qquad i = 1, \cdots, M-N+1. \tag{20}$$
We have additional $N-1-l$ pairs of columns in G_l. For each of these pairs we get
$$\widehat{H}_{M-N+2k} = \{\underline{h}_1, \ldots, \underline{h}_{N-k}\} \tag{21}$$
and
$$\widehat{H}_{M-N+2k+1} = \{\underline{h}_{M-N+k+1}, \ldots, \underline{h}_M\} \tag{22}$$
where $k = 1, \ldots, N-1-l$.

Example: consider $M = 4$, $N = 3$ as presented in (16). In this case $l = 0, 1, 2$ and we have $K_2 = 3$, $K_1 = 2.5$ and $K_0 = 2$ respectively.

1) $K_2 = 3$: $H_{\text{eff}}^{(2)}$ is generated from the multiplication of the 3x4 matrix H with the first 2 columns of the transmission matrix. In this case $H_{\text{eff}}^{(2)}$ is a 6x6 block diagonal matrix, consisting of 2 blocks. Each block is a 3x3 matrix. We get that $\widehat{H}_1 = \{\underline{h}_1, \underline{h}_2, \underline{h}_3\}$ and $\widehat{H}_2 = \{\underline{h}_2, \underline{h}_3, \underline{h}_4\}$.
2) $K_1 = \frac{10}{4}$: $H_{\text{eff}}^{(1)}$ is a 12x10 block diagonal matrix consisting of 4 blocks. The first 2 blocks are identical

to the blocks of $H_{\text{eff}}^{(2)}$. The additional 2 blocks (multiplication with columns 3-4) are 3x2 matrices. We get that $\widehat{H}_3=\{\underline{h}_1,\underline{h}_2\}$ and $\widehat{H}_4=\{\underline{h}_3,\underline{h}_4\}$.

3) $K_0=2$: $H_{\text{eff}}^{(0)}$ consists of 6 blocks. In this case the last 2 blocks are 3x1 vectors. We get that $\widehat{H}_5=\underline{h}_1$ and $\widehat{H}_6=\underline{h}_4$.

We present $H_{\text{eff}}^{(0)}$ of our example in equation (23). Note that $\underline{h}_i\in\mathbb{C}^3$ for $1\le i\le 4$, and $\underline{0}$ is a $3x1$ vector.

From the sequential structure of $H_{\text{eff}}^{(l)}$ blocks (17)-(19), (20)-(22) it is easy to see that when two columns of H occur in a certain block of $H_{\text{eff}}^{(l)}$, the columns of H between them must also occur in the same block, i.e. if $\underline{h}_1$, $\underline{h}_5$ occur in a certain block, then $\underline{h}_2,\underline{h}_3,\underline{h}_4$ also occur in the same block. Next we prove a property of the transmission scheme G_l, that relates to the number of occurrences of the columns of H in the blocks of $H_{\text{eff}}^{(l)}$. For each set of columns in H, we give an upper bound on the amount of its appearances in different blocks.

Lemma 1. *Consider the transmission scheme G_l, $l=0,\dots L-1$. In case $0\le i-j<L$, the columns $\underline{h}_j,\dots,\underline{h}_i$ may occur together in $N-i+j$ blocks of $H_{\text{eff}}^{(l)}$ at most. In case $i-j\ge L$ they can not occur together in any block of $H_{\text{eff}}^{(l)}$.*

Proof: See appendix C. ■

C. Upper Bound on The Error Probability

Next we would like to derive an upper bound on the average decoding error probability of ensemble of K_lT_l-complex dimensional IC, for each channel realization. We define $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|=\rho^{-\sum_{i=1}^{K_lT_l}\eta_i}$, where $\rho^{-\frac{\eta_i}{2}}$ is the $i'th$ singular value of $H_{\text{eff}}^{(l)}$, $1\le i\le K_lT_l$. We also define $\underline{\eta}=(\eta_1,\dots,\eta_{K_lT_l})^T$. Note that $NT_l\ge K_lT_l$.

Theorem 3. *There exists a sequence of K_lT_l-complex dimensional IC's, with channel realization $H_{\text{eff}}^{(l)}$ and receiver VNR $\mu_{rc}=\rho^{1-\frac{r}{K_l}-\frac{\sum_{i=1}^{K_lT_l}\eta_i}{K_lT_l}}$, that has average decoding error probability*

$$\begin{aligned}\overline{Pe}(H_{\text{eff}}^{(l)},\rho)=\overline{Pe}(\underline{\eta},\rho)&\le D(K_lT_l)\rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i}\\&=D(K_lT_l)\rho^{-T_l(K_l-r)}\cdot|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|^{-1}\end{aligned}$$

where $D(K_lT_l)$ is a constant independent of ρ, and $\eta_i\ge 0$ for every $1\le i\le K_lT_l$.

Proof: We base our proof on the techniques developed by Poltyrev [9] for the AWGN channel. However, the channel considered here is colored. In spite of that, we show that what affects the average decoding error probability is the singular values product, which is encapsulated by the receiver VNR, μ_{rc}. This observation enables us to facilitate this colored channel analysis.

Based on [9] we have the following upper bound on the maximum-likelihood (ML) decoding error probability of each K_lT_l-complex dimensional IC point $\underline{x}'\in S_{K_lT_l}$

$$Pe(\underline{x}')\le Pr(\|\underline{\tilde{n}}_{\text{ex}}\|\ge R)+\sum_{\underline{l}\in Ball(\underline{x}',2R)\bigcap S_{K_lT_l},\underline{l}\neq\underline{x}'}Pr(\|\underline{l}-\underline{x}'-\underline{\tilde{n}}_{ex}\|<\|\underline{\tilde{n}}_{ex}\|)\quad(24)$$

where $Ball(\underline{x}',2R)$ is a K_lT_l-complex dimensional ball of radius $2R$ centered around $\underline{x}'$, and $\underline{\tilde{n}}_{\text{ex}}$ is the effective noise in the K_lT_l-complex dimensional hyperplane where the IC's resides. Note that the second term in (24) represents the pairwise error probability to points within $Ball(\underline{x}',2R)$, i.e. the decision region is at distance R at most.

Next we upper bound the average decoding error probability of an ensemble of constellations drawn uniformly within $cube_{K_lT_l}(b)$. Each code-book contains $\lfloor\gamma_{\text{tr}}b^{2K_lT_l}\rfloor$ points, where each point is drawn uniformly within $cube_{K_lT_l}(b)$. In the receiver, the random ensemble is uniformly distributed within $\{H_{\text{eff}}^{(l)}\cdot\text{cube}_{K_lT_l}(b)\}$. Let us consider a certain point, $\underline{x}'\in\{H_{\text{eff}}^{(l)}\cdot\text{cube}_{K_lT_l}(b)\}$, from the random ensemble in the receiver. We denote the ring around $\underline{x}'$ by $Ring(\underline{x}',i\Delta)=Ball(\underline{x}',i\Delta)\setminus Ball(\underline{x}',(i-1)\Delta)$. The average number of points within $Ring(\underline{x}',i\Delta)$ of the random ensemble is

$$\begin{aligned}&Av(\underline{x}',i\Delta)=\gamma_{\text{rc}}|H_{\text{eff}}^{(l)}\cdot\text{cube}_{K_lT_l}(b)\bigcap Ring(\underline{x}',i\Delta)|\\&\le\gamma_{\text{rc}}|Ring(\underline{x}',i\Delta)|\le\frac{\gamma_{\text{rc}}\pi^{K_lT_l}2K_lT_l}{\Gamma(K_lT_l+1)}(i\Delta)^{2K_lT_l-1}\Delta\quad(25)\end{aligned}$$

where $\gamma_{\text{rc}}=\rho^{rT_l+\sum_{i=1}^{K_lT_l}\eta_i}$. By using the upper bounds on the error probability (24), and the average number of points within the rings (25), we get for a certain channel realization the following upper bound on the average decoding error probability of the finite constellations ensemble, at point $\underline{x}'$

$$\begin{aligned}&\overline{P_e^{FC}}(\underline{x}',\rho,\underline{\eta})\le Pr(\|\underline{\tilde{n}}_{\text{ex}}\|\ge R)+\\&\gamma_{\text{rc}}Q(K_lT_l)\sum_{i=1}^{\lceil\frac{2R}{\Delta}\rceil}Pr(\tilde{n}_{ex,1}>\frac{(i-1)\Delta}{2})\cdot(i\Delta)^{2K_lT_l-1}\Delta\end{aligned}\quad(26)$$

where $Q(K_lT_l)=\frac{\pi^{K_lT_l}2K_lT_l}{\Gamma(K_lT_l+1)}$, and $\tilde{n}_{ex,1}$ is the first component of $\underline{\tilde{n}}_{ex}$ (the pairwise error probability has scalar decision region). By taking $\Delta\to 0$ we get

$$\begin{aligned}&\overline{P_e^{FC}}(\underline{x}',\rho,\underline{\eta})\le Pr(\|\underline{\tilde{n}}_{\text{ex}}\|\ge R)+\\&\gamma_{\text{rc}}Q(K_lT_l)\int_0^{2R}Pr(\tilde{n}_{ex,1}>\frac{x}{2})x^{2K_lT_l-1}dx.\end{aligned}\quad(27)$$

Note that this upper bound applies for any value of $R\ge 0$ and b, and does not depend on $\underline{x}'$, i.e. $\overline{P_e^{FC}}(\underline{x}',\rho,\underline{\eta})=\overline{P_e^{FC}}(\rho,\underline{\eta})$.

Now we divide the channel realization into 2 subsets: $\mathcal{A}=\{\underline{\eta}\mid\sum_{i=1}^{K_lT_l}\eta_i\le T_l(K_l-r),\eta_i\ge 0\}$, where $\underline{\eta}=(\eta_1,\dots,\eta_{K_lT_l})$ and $\overline{\mathcal{A}}=\{\underline{\eta}\mid\sum_{i=1}^{K_lT_l}\eta_i>T_l(K_l-r),\eta_i\ge 0\}$. For each set we upper bound the error probability. We begin with the case $\underline{\eta}\in\mathcal{A}$. For this case we upper bound the terms in (27) and find an upper bound on the error probability as a function of the receiver VNR, $\mu_{rc}=\rho^{1-\frac{r}{K_l}-\frac{\sum_{i=1}^{K_lT_l}\eta_i}{K_lT_l}}$.

$$H_{\text{eff}}^{(0)} = \begin{pmatrix} \underline{h}_1 & \underline{h}_2 & \underline{h}_3 & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{h}_2 & \underline{h}_3 & \underline{h}_4 & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{h}_1 & \underline{h}_2 & \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{h}_3 & \underline{h}_4 & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{h}_1 & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{0} & \underline{h}_4 \end{pmatrix} \tag{23}$$

We begin by upper bounding the integral of the second term in (27). Note that

$$Pr(\tilde{n}_{\text{ex},1} \geq \frac{x}{2}) \leq e^{-\frac{x^2}{8\sigma^2}}.$$

Hence, the integral of the second term in (27) can be upper bounded by

$$\sigma^{2K_lT_l}\Gamma(K_lT_l)2^{3K_lT_l-2}\int_0^{2R}\frac{e^{-\frac{x^2}{8\sigma^2}}x^{2K_lT_l-1}}{\sigma^{2K_lT_l}\Gamma(K_lT_l)2^{3K_lT_l-2}}dx$$

where $\int_0^{2R}\frac{e^{-\frac{x^2}{8\sigma^2}}x^{2K_lT_l-1}}{\sigma^{2K_lT_l}\Gamma(K_lT_l)2^{3K_lT_l-2}}dx = Pr(\|\tilde{n}_{\text{ex}}\| \leq 2R) \leq 1$. As a result we get the following upper bound

$$\int_0^{2R} Pr(\tilde{n}_{ex,1} > \frac{x}{2})x^{2K_lT_l-1}dx \leq \sigma^{2K_lT_l}\Gamma(K_lT_l)2^{3K_lT_l-2}. \tag{28}$$

By assigning this upper bound in the second term of (27) we get

$$\begin{aligned} &\gamma_{\text{rc}}Q(K_lT_l)\int_0^{2R} Pr(\tilde{n}_{ex,1} > \frac{x}{2})x^{2K_lT_l-1}dx \\ &\leq \frac{\gamma_{\text{rc}}\sqrt{\pi}^{2K_lT_l}2K_lT_l\sigma^{2K_lT_l}\Gamma(K_lT_l)2^{3K_lT_l-2}}{\Gamma(K_lT_l+1)} \\ &= \rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i}\cdot\frac{4^{K_lT_l}}{2e^{K_lT_l}}. \end{aligned} \tag{29}$$

In the next step we upper bound $Pr(\|\tilde{n}_{\text{ex}}\| \geq R)$, the first term in (27). We choose

$$R^2 = R_{\text{eff}}^2 = \frac{2K_lT_l}{2\pi e}\gamma_{rc}^{-\frac{1}{K_lT_l}} = \frac{2K_lT_l}{2\pi e}\rho^{-\frac{r}{K_l}-\sum_{i=1}^{K_lT_l}\frac{\eta_i}{K_lT_l}}.$$

For $\underline{\eta} \in \mathcal{A}$ we get that

$$\frac{R_{\text{eff}}^2}{2K_lT_l\cdot\sigma^2} = \rho^{1-\frac{r}{K_l}-\sum_{i=1}^{K_lT_l}\frac{\eta_i}{K_lT_l}} \geq 1.$$

By using the upper bounds from [9], we know that for the case $\frac{R_{\text{eff}}^2}{2K_lT_l\cdot\sigma^2} \geq 1$, $Pr(\|\tilde{n}_{\text{ex}}\| \geq R_{\text{eff}}) \leq e^{-\frac{R_{\text{eff}}^2}{2\sigma^2}}(\frac{R_{\text{eff}}^2 e}{2K_lT_l\sigma^2})^{K_lT_l}$. Hence we get

$$\begin{aligned} &Pr(\|\tilde{n}_{\text{ex}}\| \geq R_{\text{eff}}) \leq \\ &e^{-K_lT_l\rho^{1-\frac{r}{K_l}-\sum_{i=1}^{K_lT_l}\frac{\eta_i}{K_lT_l}}}\cdot\rho^{T_l(K_l-r)-\sum_{i=1}^{K_lT_l}\eta_i}\cdot e^{K_lT_l}. \end{aligned} \tag{30}$$

The fact that $\underline{\eta} \in \mathcal{A}$ has 2 significant consequences: the VNR is greater or equal to 1, and as ρ increases the maximal VNR in the set also increases. For very large VNR in the receiver, the upper bound of the first term, (30), is negligible compared to the upper bound on the second term, (29). On the other hand, the set of rather small VNR values is fixed for increasing ρ (the VNR is grater or equal to 1). Hence there must exist a coefficient $D^{'}(K_lT_l)$ that gives us

$$\overline{P_e^{FC}}(\rho,\underline{\eta}) \leq D^{'}(K_lT_l)\rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i} \tag{31}$$

for any ρ and $\underline{\eta} \in \mathcal{A}$, where $\overline{P_e^{FC}}(\rho,\underline{\eta})$ is the average decoding error probability of the ensemble of constellations, for a certain channel realizations.

Note that we could also take $R \geq R_{\text{eff}}$, as the upper bound in (29) does not depend on R and the upper bound in (30) would only decrease in this case. It results from the fact that we are interested in the exponential behavior of the error probability, and we consider fixed VNR (as a function of ρ) as an outage event. This allows us to take cruder bounds than [9] on (29), that do not depend on R.

For the case $\underline{\eta} \in \overline{\mathcal{A}}$, we get

$$\rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i} \geq 1.$$

Hence, we can upper bound the error probability for $\underline{\eta} \in \overline{\mathcal{A}}$ by 1. We can also upper bound the error probability for this case by the upper bound from equation (31), as long as we state that $D^{'}(K_lT) \geq 1$. Hence, the upper bound from (31) applies for $\eta_i \geq 0$, $1 \leq i \leq K_lT_l$.

Up until now we upper bounded the average decoding error probability of ensemble of finite constellations. Now we extend those finite constellations into an ensemble of IC's with density γ_{tr}, and show that the upper bound on the average decoding error probability does not change. Let us consider a certain finite constellation, $C_0(\rho,b) \subset cube_{K_lT_l}(b)$, from the random ensemble. We extend it into IC

$$IC(\rho,K_lT_l) = C_0(\rho,b) + (b+b^{'})\cdot\mathbb{Z}^{2K_lT_l} \tag{32}$$

where without loss of generality we assumed that $cube_{K_lT_l}(b) \in \mathbb{C}^{K_lT_l}$. In the receiver we have

$$IC(\rho,K_lT_l,H_{\text{eff}}^{(l)}) = H_{\text{eff}}^{(l)}\cdot C_0(\rho,b) + (b+b^{'})H_{\text{eff}}^{(l)}\cdot\mathbb{Z}^{2K_lT_l}. \tag{33}$$

By extending each finite constellation in the ensemble to IC according to the method presented in (32), we get new ensemble of IC's. We would like to set b and $b^{'}$ to be large enough such that the IC's ensemble average decoding error probability has the same upper bound as in (31), and density that equals γ_{rc} up to a coefficient. First we would like to set a value for $b^{'}$. Increasing $b^{'}$ decreases the error probability inflicted by the codewords outside the set $\{H_{\text{eff}}^{(l)}\cdot C_0(\rho,b)\}$. Without loss of generality, we upper bound the error probability of the words $x \in \{H_{\text{eff}}^{(l)}\cdot C_0(\rho,b)\} \subset IC(\rho,K_lT_l,H_{\text{eff}}^{(l)})$, denoted by

$P_e^{IC}(H_{\text{eff}}^{(l)}\cdot C_0)$. Due to the tiling symmetry, $P_e^{IC}(H_{\text{eff}}^{(l)}\cdot C_0)$ is also the average decoding error probability of the entire IC. We begin with $\underline{\eta}\in\mathcal{A}$. For this case, we upper bound the IC error probability in the following manner

$$P_e^{IC}(H_{\text{eff}}^{(l)}\cdot C_0)\leq P_e^{FC}(H_{\text{eff}}^{(l)}\cdot C_0)+Pe\big(H_{\text{eff}}^{(l)}\cdot(IC\setminus C_0)\big)$$

where $P_e^{FC}(H_{\text{eff}}^{(l)}\cdot C_0)$ is the error probability of the finite constellation $\{H_{\text{eff}}^{(l)}\cdot C_0\}$, and $Pe\big(H_{\text{eff}}^{(l)}\cdot(IC\setminus C_0)\big)$ is the average decoding error probability to points in the set $\{H_{\text{eff}}^{(l)}\cdot(IC\setminus C_0)\}$. For the case $\underline{\eta}\in\mathcal{A}$, we know that $0\leq\eta_i\leq T_l(K_l-r)$. Hence, the constriction caused by the channel in each dimension can not be smaller than $\rho^{-\frac{T_l}{2}(K_l-r)}$. As a result, for any $x_1\in\{H_{\text{eff}}^{(l)}\cdot C_0\}$ and $x_2\in\{H_{\text{eff}}^{(l)}\cdot(IC\setminus C_0)\}$ we get $\|\underline{x}_1-\underline{x}_2\|\geq 2b^{'}\cdot\rho^{-\frac{T_l}{2}(K_l-r)}$. By choosing $b^{'}=\sqrt{\frac{K_lT_l}{\pi e}}\rho^{\frac{T_l}{2}(K_l-r)+\epsilon}$, we get for $\underline{\eta}\in\mathcal{A}$ that $\|\underline{x}_1-\underline{x}_2\|\geq 2\sqrt{\frac{K_lT_l}{\pi e}}\rho^{\epsilon}$. Hence we get

$$Pe\big(H_{\text{eff}}^{(l)}\cdot(IC\setminus C_0)\big)\leq Pr(\|\underline{\tilde{n}}_{\text{ex}}\|\geq\sqrt{\frac{K_lT_l}{\pi e}}\rho^{\epsilon}).$$

For $\rho\geq 1$ we get according to the bounds in [9] that

$$Pr(\|\underline{\tilde{n}}_{\text{ex}}\|\geq\sqrt{\frac{K_lT_l}{\pi e}}\rho^{\epsilon}))\leq e^{-K_lT_l\rho^{1+\epsilon}}\rho^{K_lT_l(1+\epsilon)}e^{K_lT_l}.$$

As a result, there exists a coefficient $D^{''}(K_lT_l)$ such that

$$Pe\big(H_{\text{eff}}^{(l)}\cdot(IC\setminus C_0)\big)\leq D^{''}(K_lT_l)\rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i}$$

for $\underline{\eta}\in\mathcal{A}$ and $\rho\geq 1$. This bound applies for any IC in the ensemble. From (31) we can state that $\overline{P_e^{FC}}(\rho,\underline{\eta})=E_{C_0}\big(P_e^{FC}(H_{\text{eff}}^{(l)}\cdot C_0)\big)\leq D^{'}(K_lT_l)\rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i}$. Hence, we get that

$$\overline{Pe}(\rho,\underline{\eta})\leq D(K_lT_l)\rho^{-T_l(K_l-r)+\sum_{i=1}^{K_lT_l}\eta_i}\tag{34}$$

where $\overline{Pe}(\rho,\underline{\eta})=E_{C_0}\big(P_e^{IC}(H_{\text{eff}}^{(l)}\cdot C_0)\big)$ is the average decoding error probability of the ensemble of IC's defined in (33), and $D=2\max(D^{'},D^{''})>1$.

Next, we set the value of b to be large enough such that each IC density from the ensemble in (33), $\gamma_{rc}^{'}$, equals γ_{rc} up to a factor of 2. By choosing $b=b^{'}\cdot\rho^{\epsilon}$ we get

$$\gamma_{rc}^{'}=\gamma_{rc}(\frac{b}{b+b^{'}})^{2K_lT}=\gamma_{rc}\frac{1}{1+\rho^{-\epsilon}}.$$

For each value $\rho\geq 1$, we get $\frac{1}{2}\gamma_{rc}\leq\gamma_{rc}^{'}\leq\gamma_{rc}$. As a result we have

$$\mu_{rc}\leq\mu_{rc}^{'}=\frac{(\gamma_{rc}^{'})^{-\frac{1}{K_lT}}}{2\pi e\sigma^2}\leq 2\mu_{rc}.$$

Note that in our proof we referred to matrix of dimension $NT_lxK_lT_l$. However these results apply for any full rank matrix with number of rows which is greater or equal to the number of columns. ■

By averaging arguments we know that there exists a sequence of IC's that satisfies these requirements.

D. Achieving the Optimal DMT

In this subsection we calculate the DMT of the proposed transmission scheme. We upper bound the determinant of the effective channel inverse, $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|^{-1}$, based on the effective channel properties presented in subsection IV-B. In Theorem 3 we showed that the upper bound on the error probability depends on this determinant. Hence, the upper bound on the determinant gives us a new upper bound on the average decoding error probability. We average the new upper bound over all channel realizations and get the transmission scheme DMT.

The channel matrix H consists of $N\cdot M$ i.i.d entries, where each entry has distribution $h_{i,j}\sim CN(0,1)$. Without loss of generality we consider the case where the columns of H are drawn sequentially from left to right, i.e. $\underline{h}_1$ is drawn first, then $\underline{h}_2$ is drawn et cetera. Column $\underline{h}_j$ is an N-dimensional vector. Given $\underline{h}_{\min(1,j-N+1)},\ldots,\underline{h}_{j-1}$, we can write

$$\underline{h}_j=\Theta(\underline{h}_{\min(1,j-N+1)},\ldots,\underline{h}_{j-1})\cdot\underline{\widetilde{h}}_j$$

where $\Theta(\cdot)$ is an NxN unitary matrix. $\Theta(\cdot)$ is chosen such that:

1) The first entry of $\underline{\widetilde{h}}_j$, $\widetilde{h}_{1,j}$, is in the direction of $\underline{h}_{j-1}$.
2) The second entry, $\widetilde{h}_{2,j}$, is in the direction orthogonal to $\underline{h}_{j-1}$, in the hyperplane spanned by $\{\underline{h}_{j-1},\underline{h}_{j-2}\}$.
3) The $\min(j,N)-1$ entry, $\widetilde{h}_{\min(j,N)-1,j}$, is in the direction orthogonal to the hyperplane spanned by $\{\underline{h}_{\max(2,j-N+2)},\ldots,\underline{h}_{j-1}\}$ inside the hyperplane spanned by $\{\underline{h}_{\max(1,j-N+1)},\ldots,\underline{h}_{j-1}\}$.
4) The rest of the $N-\min(j,N)+1$ entries are in directions orthogonal to the hyperplane $\{\underline{h}_{\max(1,j-N+1)},\ldots,\underline{h}_{j-1}\}$.

Note that $\widetilde{h}_{i,j}$, $1\leq i\leq N$, $1\leq j\leq M$ are i.i.d random variables with distribution $CN(0,1)$. Let us denote by $\underline{h}_{j\perp j-1,\ldots,j-k}$ the component of $\underline{h}_j$ which resides in the $N-k$ subspace which is perpendicular to the space spanned by $\{\underline{h}_{j-1},\ldots,\underline{h}_{j-k}\}$. In this case we get

$$\|\underline{h}_{j\perp j-1,\ldots,j-k}\|^2=\sum_{i=k+1}^{N}|\widetilde{h}_{i,j}|^2\quad 1\leq k\leq\min(j,N)-1.\tag{35}$$

If we assign $|\widetilde{h}_{i,j}|^2=\rho^{-\xi_{i,j}}$, we get that the probability density function (PDF) of $\xi_{i,f}$ is

$$f(\xi_{i,j})=C\cdot\log\rho\cdot\rho^{-\xi_{i,j}}\cdot e^{-\rho^{-\xi_{i,j}}}\tag{36}$$

where C is a normalization factor. In our analysis we assume very large values for ρ. Hence we can neglect events where $\xi_{i,j}<0$ since in this case the PDF (36) decreases exponentially as a function of ρ. For very large ρ, $\xi_{i,j}\geq 0$, $1\leq i\leq N$ and $1\leq j\leq M$ the PDF takes the following form

$$f(\xi_{i,j})\propto\rho^{-\xi_{i,j}}\qquad\xi_{i,j}\geq 0.\tag{37}$$

In this case by assigning in (35) the vector $\underline{\xi}_j=(\xi_{1,j},\ldots,\xi_{N,j})^T$, that has PDF which is proportional to

$\rho^{-\sum_{i=1}^{N}\xi_{i,j}}$, we get that

$$\|\underline{h}_{j\perp j-1,\dots,j-k}\|^2 \doteq \rho^{-\min_{s\in\{k+1,\dots,N\}}\xi_{s,j}} = \rho^{-a(k,\underline{\xi}_j)} \quad (38)$$

where $1 \leq k \leq \min(j,L)-1$ and $a(k,\underline{\xi}_j) = \min_{s\in\{k+1,\dots,N\}}\xi_{s,j}$. In addition

$$\|\underline{h}_j\|^2 \doteq \rho^{-\min_{s\in\{1,\dots,N\}}\xi_{s,j}} = \rho^{-a(0,\underline{\xi}_j)}. \quad (39)$$

Note that

$$a(\min(j,L)-1,\underline{\xi}_j) \geq \cdots \geq a(0,\underline{\xi}_j) \geq 0. \quad (40)$$

Next we wish to quantify the contribution of a certain column in the channel matrix, $\underline{h}_j$, to the determinant $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|$. $H_{\text{eff}}^{(l)}$ is a block diagonal matrix. Hence the determinant of $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|$ can be expressed as

$$|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}| = \prod_{i=1}^{T_l}|\widehat{H}_i^\dagger\widehat{H}_i|. \quad (41)$$

Assume $\widehat{H}_i = (\widehat{\underline{h}}_1,\dots,\widehat{\underline{h}}_m)$, i.e. $\widehat{H}_i$ has m columns. In this case we can state that the determinant

$$|\widehat{H}_i^\dagger\widehat{H}_i| = \|\widehat{\underline{h}}_1\|^2\|\widehat{\underline{h}}_{2\perp 1}\|^2\cdots\|\widehat{\underline{h}}_{m\perp m-1,\dots,1}\|^2.$$

Note that $\widehat{H}_i$ also has more rows than columns. The columns of $\widehat{H}_i$ are subset of the columns of the channel matrix H. Hence we are interested in the blocks where $\underline{h}_j$ occurs. We know that the contribution of $\underline{h}_j$ to those determinants can be quantified by taking into account the columns to its left in each block. We consider two cases:

- The case $N \geq M$. In this case we can see from (17)-(19) that $\underline{h}_j$ may occur with $\{\underline{h}_1,\dots,\underline{h}_{j-1}\}$ to its left in different blocks.
- The case $M > N$. In this case we can see from (20)-(22) that $\underline{h}_j$ may occur only with $\{\underline{h}_{\max(1,j-N+1)},\dots,\underline{h}_{j-1}\}$ to its left in different blocks.

Based on (38) and (39) we can quantify the contribution of $\underline{h}_j$ to $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|$ by

$$\|\underline{h}_j\|^{2b_j(0)}\prod_{k=1}^{\min(j,L)-1}\|\underline{h}_{j\perp j-1,\dots,j-k}\|^{2b_j(k)} \doteq \rho^{-\sum_{k=0}^{\min(j,L)-1}b_j(k)a(k,\underline{\xi}_j)} \quad (42)$$

where $b_j(k)$ is the number of occurrences of $\underline{h}_j$ in $H_{\text{eff}}^{(l)}$ blocks, with only $\{\underline{h}_{j-1},\dots,\underline{h}_{j-k}\}$ to its left. $b_j(0)$ is the number of occurrences of $\underline{h}_j$ with no columns to its left. Note that from the definition of the transmission scheme we get that for $l=0$, $b_j(k) > 0$ for $1 \leq k \leq \min(j,L)-1$.

In the following theorem we calculate the DMT of the proposed transmission scheme.

Theorem 4. *There exists a sequence of K_lT_l-complex dimensional IC's with transmitter density $\gamma_{tr} = \rho^{rT_l}$ and T_l channel uses that has diversity order*

$$d_{K_lT_l}(r) \geq (M-l)(N-l) - (r-l)(N+M-2\cdot l-1)$$

where $0 \leq r \leq K_l$ and $l = 0,\dots,L-1$.

Proof: The proof outline is as follows. The upper bound on the error probability from Theorem 3 depends on $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|^{-1}$. We upper bound this determinant value and average over different realizations of $H_{\text{eff}}^{(l)}$ in order to find G_l diversity order. We begin by lower bounding $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|$. Based on the sequential structure of G_l, we lower bound the contribution of a certain column of H, $\underline{h}_j$, $1 \leq j \leq M$ to the determinant. This gives us a new upper bound on the error probability for each channel realization. We average the new upper bound on the error probability, by averaging over $\widetilde{\underline{h}}_1,\dots,\widetilde{\underline{h}}_M$. From this averaging we get the required diversity order.

Specifically, we first lower bound the contribution of $\underline{h}_j$ to the determinant (42), by upper bounding $\sum_{k=0}^{\min(j,L)-1}b_j(k)a(k,\underline{\xi}_j)$. Based on Lemma 1, and the fact that when two columns of H occur together in a block of $H_{\text{eff}}^{(l)}$, all the columns of H between them must also occur in the same block, we get

$$\sum_{s=k}^{\min(j,L)-1}b_j(s) \leq N-k \qquad 0 \leq k \leq \min(j,L)-1. \quad (43)$$

where $\sum_{s=k}^{\min(j,L)-1}b_j(s)$ is the number of occurrences of $\{\underline{h}_j,\dots,\underline{h}_{j-k}\}$ in $H_{\text{eff}}^{(l)}$ blocks. Hence, we can state that $\sum_{s=0}^{\min(j,L)-1}b_j(s) \leq N$, by assigning $k=0$ in (43). Also note that for $l=0$, the sum $\sum_{s=0}^{\min(j,L)-1}b_j(s)a(s,\underline{\xi}_j)$ is larger than for any other $1 \leq l \leq L-1$. From the inequalities in (40), and the fact that for $l=0$ we get $b_j(k) > 0$ for any $1 \leq k \leq \min(j,L)-1$, we can state that

$$\sum_{s=0}^{\min(j,L)-1}b_j(s)a(s,\underline{\xi}_j) \leq \sum_{s=0}^{\min(j,L)-2}a(s,\underline{\xi}_j) + (N-\min(j,L)+1)a(\min(j,L)-1,\underline{\xi}_j) = c(j). \quad (44)$$

Using (42) and (44) we can state that for a vector $\underline{\xi}_j$, that has PDF $\rho^{-\sum_{i=1}^{N}\xi_{i,j}}$, we can lower bound the contribution of $\underline{h}_j$ to $|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}|$ by

$$\|\underline{h}_j\|^{2b_j(0)}\prod_{k=1}^{\min(j,L)-1}\|\underline{h}_{j\perp j-1,\dots,j-k}\|^{2b_j(k)} \geq \rho^{-c(j)}. \quad (45)$$

By taking into account the contribution of each column $\underline{h}_j$ to the determinant we get that

$$|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}| = \prod_{j=1}^{M}\|\underline{h}_j\|^{2b_j(0)}\prod_{k=1}^{\min(j,L)-1}\|\underline{h}_{j\perp j-1,\dots,j-k}\|^{2b_j(k)}. \quad (46)$$

By considering the set of vectors $\underline{\xi}_1,\dots,\underline{\xi}_M$, that have PDF $\rho^{-\sum_{j=1}^{M}\sum_{i=1}^{N}\xi_{i,j}}$, and by using the lower bound from (45) we get

$$|H_{\text{eff}}^{(l)\dagger}H_{\text{eff}}^{(l)}| \geq \rho^{-\sum_{j=1}^{M}c(j)} \quad (47)$$

The upper bound on the error probability presented in Theorem 3 is proportional to

$$\rho^{-T_l(K_l-r)} \cdot |H_{\text{eff}}^{(l)\dagger} H_{\text{eff}}^{(l)}|^{-1} = \rho^{-T_l(K_l-r)+\sum_{i=1}^{K_l T} \eta_i} \tag{48}$$

for $\eta_i \geq 0$ and $1 \leq i \leq K_l T_l$, where $\rho^{-\frac{\eta_i}{2}}$ are the singular values of $H_{\text{eff}}^{(l)}$. Hence, in order to use the upper bound from Theorem 3 in our analysis, we need to show that by taking $\xi_{i,j} \geq 0$, $1 \leq i \leq N$, $1 \leq j \leq M$ we also get that $\eta_i \geq 0$, $1 \leq i \leq K_l T_l$. Note that the entries of $H_{\text{eff}}^{(l)}$ are elements of the channel matrix H. Also, all $H's$ columns must appear in $H_{\text{eff}}^{(l)}$. Hence, from trace considerations we get

$$\frac{\rho^{-\min_{i,j}(\xi_{i,j})}}{K_l T_l} \leq \rho^{-\min_s(\eta_s)} \leq N \cdot K_l T_l^2 \rho^{-\min_{i,j}(\xi_{i,j})}.$$

As a result we get that $\min_{i,j}(\xi_{i,j}) \geq 0$ if and only if $\min_s(\eta_s) \geq 0$, and so $\eta_s \geq 0$ for every $1 \leq s \leq K_l T_l$. As the upper bound on the error probability in (48) applies for $\eta_i \geq 0$, $1 \leq i \leq K_l T_l$, this upper bound also applies whenever $\xi_{i,j} \geq 0$, $1 \leq i \leq N$ and $1 \leq j \leq M$. In equation (47) we found a lower bound on the determinant. We use this lower bound to upper bound the determinant of the matrix inverse $|H_{\text{eff}}^{(l)\dagger} H_{\text{eff}}^{(l)}|^{-1}$

$$|H_{\text{eff}}^{(l)\dagger} H_{\text{eff}}^{(l)}|^{-1} \leq \rho^{\sum_{j=1}^{M} c(j)}. \tag{49}$$

and as a consequence we can upper bound the error probability.

We can express the average decoding error probability over the ensemble of IC's for large ρ as follows

$$\overline{Pe}(\rho) = \int_H Pe(\rho, H) f(H) dH \doteq \int_{\xi_{\underline{i},\underline{j}} \geq 0} Pe(\rho, \xi_{\underline{i},\underline{j}}) f(\xi_{\underline{i},\underline{j}}) d\xi_{\underline{i},\underline{j}} \tag{50}$$

where $Pe(\rho, H) = Pe(\rho, \xi_{\underline{i},\underline{j}})$ is the ensemble average decoding error probability per channel realization, and $\xi_{\underline{i},\underline{j}} \geq 0$ means $\xi_{i,j} \geq 0$ for $1 \leq i \leq N$ and $1 \leq j \leq M$. We divide the integration range into 2 sets: $\mathcal{A} = \{\xi_{\underline{i},\underline{j}} \mid \sum_{i=1}^{N}\sum_{j=1}^{M} \xi_{i,j} \leq T_l(K_l - r); \xi_{\underline{i},\underline{j}} \geq 0\}$ and $\overline{\mathcal{A}} = \{\xi_{\underline{i},\underline{j}} \mid \sum_{i=1}^{N}\sum_{j=1}^{M} \xi_{i,j} > T_l(K_l - r); \xi_{\underline{i},\underline{j}} \geq 0\}$. Hence, we can write the average decoding error probability as follows

$$\overline{Pe}(\rho) \doteq \int_{\xi_{\underline{i},\underline{j}} \in \mathcal{A}} Pe(\rho, \xi_{\underline{i},\underline{j}}) f(\xi_{\underline{i},\underline{j}}) d\xi_{\underline{i},\underline{j}} + \int_{\xi_{\underline{i},\underline{j}} \in \overline{\mathcal{A}}} Pe(\rho, \xi_{\underline{i},\underline{j}}) f(\xi_{\underline{i},\underline{j}}) d\xi_{\underline{i},\underline{j}}. \tag{51}$$

We begin by upper bounding the first term of the error probability in (51). Based on Theorem 3, the average decoding error probability per channel realization is upper bounded by $Pe(\rho, H) \leq \rho^{-T_l(K_l-r)+\sum_{i=1}^{K_l T_l} \eta_i}$. Using the upper bound on the determinant (49) and the fact that $|H_{\text{eff}}^{(l)\dagger} H_{\text{eff}}^{(l)}|^{-1} = \rho^{\sum_{i=1}^{K_l T_l} \eta_i}$, we get that the first term of the error probability (51) is upper bounded by

$$\int_{\xi_{\underline{i},\underline{j}} \in \mathcal{A}} \rho^{-T_l(K_l-r)+\sum_{j=1}^{M}(c(j)-\sum_{i=1}^{N}\xi_{i,j})} d\xi_{\underline{i},\underline{j}}. \tag{52}$$

Now we prove a Lemma that shows that the exponent of the integrand in the upper bound from (52) is negative for $\xi_{\underline{i},\underline{j}} \geq 0$.

Lemma 2. *consider $\xi_{i,j} \geq 0$ for $1 \leq i \leq N$ and $1 \leq j \leq M$. The sum*

$$c(j) - \sum_{i=1}^{N} \xi_{i,j} \leq 0$$

for every $1 \leq j \leq M$.

Proof: See appendix D. ■

In a similar manner to [3], [7], for very large ρ and finite integration range, we can approximate the integral by finding the most dominant exponential term in (52). Based on Lemma 2 we know that the exponent of the integrand is always negative. Hence, we can approximate the upper bound by finding

$$\min_{\xi_{\underline{i},\underline{j}} \in \mathcal{A}} T_l(K_l - r) + \sum_{j=1}^{M}(\sum_{i=1}^{N}\xi_{i,j} - c(j)).$$

As $\sum_{i=1}^{N}\xi_{i,j} - c(j) \geq 0$ the minimum is achieved when $\sum_{i=1}^{N}\xi_{i,j} - c(j) = 0$ for $1 \leq j \leq M$. This can be achieved for instance by taking $\xi_{i,j} = 0$ for $1 \leq i \leq N$, $1 \leq j \leq M$. In this case we get that the diversity order equals $T_l(K_l - r)$ which is the best diversity order possible for IC's of complex dimension $K_l T_l$.

Next we upper bound the second term of the error probability from (51). For $\xi_{\underline{i},\underline{j}} \in \overline{\mathcal{A}}$ we upper bound the average decoding error probability per channel realization by 1. In this case we get

$$\int_{\xi_{\underline{i},\underline{j}} \in \overline{\mathcal{A}}} \rho^{-\sum_{j=1}^{M}\sum_{i=1}^{N}\xi_{i,j}} d\xi_{\underline{i},\underline{j}}.$$

Again we approximate this integral by calculating the most dominant exponential term, i.e. $\min_{\xi_{\underline{i},\underline{j}} \in \overline{\mathcal{A}}} \sum_{i=1}^{N}\sum_{j=1}^{M}\xi_{i,j}$. The minimal value for this case is also $T_l(K_l - r)$. Hence, we get diversity order $T_l(K_l - r)$ also for the second term in (51). As a result we can state that for both terms in (51) we get the same diversity order, and the transmission scheme diversity order is upper bounded by $T_l(K_l - r)$. By assigning the values of T_l and K_l we get the theorem upper bound. This concludes the proof. ■

The diversity order attained on Theorem 4 for K_l, T_l coincides with the optimal DMT of finite constellations in the range $l \leq r \leq l+1$. Hence, by considering $0 \leq l \leq L-1$, we can attain the optimal DMT with L sequences of IC's.

We present as an illustrative example the case of $M = N = 2$. Let us consider the case where $l = 0$. In this case $K_0 = \frac{4}{3}$, and $T_0 = 3$, i.e. we transmit 4-complex dimensional IC. The transmission scheme diversity order in this case is $4 - 3r$, $0 \leq r \leq \frac{4}{3}$. In this case the effective channel matrix, $H_{\text{eff}}^{(0)}$, consists of 3 blocks: $\widehat{H}_1 = (\underline{h}_1, \underline{h}_2)$, $\widehat{H}_2 = \underline{h}_1$ and $\widehat{H}_3 = \underline{h}_2$. According to our definitions we get that

$$|\widehat{H}_1^\dagger \widehat{H}_1| = \|\underline{h}_1\|^2 \cdot \|\underline{h}_{2\perp 1}\|^2 = \rho^{-\min(\xi_{1,1},\xi_{2,1})} \cdot \rho^{-\xi_{2,2}}$$

and also $\|\underline{h}_1\|^2 = \rho^{-\min(\xi_{1,1},\xi_{2,1})}$, $\|\underline{h}_2\|^2 = \rho^{-\min(\xi_{1,2},\xi_{2,2})}$.

In accordance with (51) we divide the integral into 2 terms. In the first term we solve the optimization problem

$$\min_{\xi_{\underline{i},\underline{j}}\in\mathcal{A}}(4-3r)-\left(\xi_{2,2}+2\cdot\min\left(\xi_{1,1},\xi_{2,1}\right)\right.$$
$$\left.+\min\left(\xi_{1,2},\xi_{2,2}\right)\right)+\sum_{i=1}^{2}\sum_{j=1}^{2}\xi_{i,j}.$$

One solution to this problem is $\xi_{i,j}=0$ for $1\leq i\leq 2$, $1\leq j\leq 2$. In this case we get an exponential term that equals $4-3r$. For the second integral we solve the optimization problem

$$\min_{\xi_{\underline{i},\underline{j}}\in\overline{\mathcal{A}}}\sum_{i=1}^{2}\sum_{j=1}^{2}\xi_{i,j}.$$

In this case the optimization problem solution is $\sum_{i=1}^{2}\sum_{j=1}^{2}\xi_{i,j}=4-3r$. Hence, all together we get diversity order that equals $4-3r$, that coincides with the optimal DMT of finite constellations in the range $0\leq r\leq 1$.

In the next theorem we prove the existence of a sequence of lattices that has the same lower bound as in Theorem 4.

Theorem 5. *There exists a sequence of $2K_lT_l$-real dimensional lattices with transmitter density $\gamma_{tr}=\rho^{rT_l}$ and T_l channel uses that has diversity order*

$$d_{K_lT_l}(r)\geq(M-l)(N-l)-(r-l)(N+M-2\cdot l-1)$$

where $0\leq r\leq K_l$ and $l=0,\ldots,L-1$.

Proof: See appendix E ■

Note that we consider a $2K_lT_l$-real dimensional lattice, as in the transmission scheme G_l we spread the first K_lT_l dimensions of the lattice on the real part of G_l's non-zero entries, and the other K_lT_l dimensions of the lattice on the imaginary part of G_l's non-zero entries. By doing that we don't necessarily transmit a K_lT_l-complex dimensional lattice. Considering the $2K_lT_l$-real dimensional lattice enables us to use the *Minkowski-Hlawaka-Siegel* Theorem [9],[12], and prove Theorem 5.

E. Power Peak to Average Ratio

For practical reasons, such as power peak to average ratio, one may prefer to have a transmission scheme that spreads the transmitted power equally over time and space. The transmitting matrix G_l contains exactly K_lT_l non-zero entries, where the rest of the entries are zero. In order to spread the power more equally over time and space we use the following unitary operations

$$U_LG_lU_R.$$

U_L is an MxM unitary matrix that spreads each column of G_l, i.e. spreads over space. U_R is a T_lxT_l unitary matrix that spreads each raw of G_l, i.e. spreads over time. As the distribution of H and $H\cdot U_L$ are identical, multiplying U_L with G_l gives exactly the same performance. Based on the notations from (2) we can state that

$$G_l\cdot U_R=\left(\underline{x}_1,\ldots,\underline{x}_{T_l}\right)$$

where $\left(\underline{x}_1,\ldots,\underline{x}_{T_l}\right)$ are the channel inputs. In the receiver we can state that the received signals are $\left(\underline{y}_1,\ldots,\underline{y}_{T_l}\right)$. By multiplying with $U_R^{\dagger}$ we get

$$\left(\underline{y}_1,\ldots,\underline{y}_{T_l}\right)\cdot U_R^{\dagger}=G_l+\left(\underline{n}_1,\ldots,\underline{n}_{T_l}\right)U_R^{\dagger}.$$

The distribution of $\left(\underline{n}_1,\ldots,\underline{n}_{T_l}\right)$ is identical to the distribution of $\left(\underline{n}_1,\ldots,\underline{n}_{T_l}\right)U_R^{\dagger}$. Hence, multiplying G_l with U_R gives also exactly the same performance. For instance, in order to achieve full diversity and spread the power more uniformly, we take G_0 and duplicate its structure s times to create the transmission scheme $G_0^{(s)}$. In this case the transmission matrix $G_0^{(s)}$ consists of sK_0T_0 complex non-zero entries, i.e we transmit an sK_0T_0 complex dimensional IC within the sMT_0 complex space. $G_0^{(s)}$ is an MxsT_0 dimensional matrix, that has exactly the same diversity order as G_0 (it duplicates the structure of G_0 s times). Each row of $G_0^{(s)}$ has exactly sN non-zero entries. We define $U_R^{(s)}$ as sT_0xsT_0 unitary matrix. For large enough s, the multiplication $G_0^{(s)}\cdot U_R^{(s)}$ spreads the power more uniformly over space and time, and still achieves full diversity.

F. Averaging Arguments

In this subsection we show that there exist L sequences of lattices that attain the optimal DMT, where each sequence out of the L sequences attains different segment on the optimal DMT curve. In addition we show that there exists a single IC that attains the optimal DMT by diluting its points and adapting its dimensionality.

As a consequence of Theorem 3 and Theorem 4 we can state the following

Corollary 3. *Consider a sequence of KT-complex dimensional IC's $S_{KT}(\rho)$ with density $\gamma_{tr}=1$, that attains diversity order d. This sequence of IC's also attains diversity order $d(1-\frac{r}{K})$ when the sequence density is scaled to $\gamma_{tr}=\rho^{rT}$.*

Proof: As $\gamma_{tr}=1$ for every ρ, $S_{KT}(\rho)$ has multiplexing gain $r=0$. We denote the error probability of $S_{KT}(\rho)$ by $Pe_S(\rho,0)$, where 0 represents the multiplexing gain. Assume that the error probability of $S_{KT}(\rho)$ equals

$$Pe_S(\rho,0)=A^{'}(\rho)\rho^{-d}$$

where $-\lim_{\rho\to\infty}\log_{\rho}Pe_S(\rho,0)=d$, i.e. $S_{KT}(\rho)$ has diversity order d. By scaling the sequence of IC's such that

$$\overline{S}_{KT}(\rho)=S_{KT}(\rho)\cdot\rho^{-\frac{r}{2K}}\qquad 0\leq r\leq K,$$

i.e., scaling $S_{KT}(\rho)$ by a factor of $\rho^{-\frac{r}{2K}}$, we get that $\overline{S}_{KT}(\rho)$ has density $\gamma_{tr}=\rho^{rT}$, multiplexing gain r and error probability

$$Pe_{\overline{S}}(\rho,r)=Pe_S(\rho^{1-\frac{r}{K}},0)=A^{'}(\rho^{1-\frac{r}{K}})\rho^{-d(1-\frac{r}{K})}.$$

As a result we get $-\lim_{\rho\to\infty}\log_{\rho}Pe_{\overline{S}}(\rho,r)=d(1-\frac{r}{K})$, i.e. $\overline{S}_{KT}(\rho)$ has diversity order $d(1-\frac{r}{K})$. ■

Corollary 4. *The optimal DMT is attained by exactly L sequences of $2K_lT_l$-real dimensional lattices, $l=0,\ldots,L-1$,* *where each sequence attains different segment of the optimal DMT.*

Proof: From Theorem 5 we know that there exists a $2K_lT_l$-real dimensional sequence of lattices with density $\gamma_{tr} = 1$ that attains diversity $(M-l)(N-l)+l(N+M-2\cdot l-1)$. Hence, based on Corollary 3 we can state that by scaling this $2K_lT_l$-real dimensional sequence of lattices into a sequence of lattices with density $\gamma_{tr} = \rho^{rT_l}$ we get diversity order $(M-l)(N-l)-(r-l)(N+M-2\cdot l-1)$, i.e. the sequence of lattices attains the optimal DMT line in the range $l \leq r \leq l+1$. The optimal DMT is the maximal value between L lines, for each $0 \leq r \leq L$. Hence, there exist L sequences of lattices that attain the optimal DMT. ∎

Next, we show that there exists a single sequence of IC's that attains the optimal DMT. The optimal DMT consists of L segments of straight lines. Each segment is attained by reducing the IC's dimensionality to the correct dimension, and diluting their points to get the desired density. Note that in Theorem 4 we showed that for each multiplexing gain, r, there exists a sequence of IC's that attains the optimal DMT. On the other hand, on Corollary 5 we show that a single sequence of IC's attains the optimal DMT for any r, by adapting its dimensionality and diluting its points. Also note that $K_0T_0 > K_1T_1 > \cdots > K_{L-1}T_{L-1}$.

Corollary 5. *There exists a single sequence of K_0T_0-complex dimensional IC's, that attains the L segments of the optimal DMT:*

$$(M-l)(N-l)-(r-l)(N+M-2\cdot l-1) \quad 0 \leq r \leq K_l$$

where $l = 0, \cdots, L-1$. The $l'th$ segment is attained by reducing the IC's complex dimensionality to K_lT_l, and by diluting their points to get density $\gamma_{tr} = \rho^{T_l r}$.

Proof: See Appendix F. ∎

V. Conclusion

In this work we introduced the fundamental limits of IC's/lattices in MIMO fading channels. We believe that this work can set a framework for designing lattices for MIMO channels using lattice decoding.

Acknowledgment

The authors wish to thank Joseph J. Boutros for interesting discussions regarding this work.

References

[1] I. Telatar, "Capacity of multi-antenna gaussian channels," *Europ. Trans. Telecommu*, vol. 10, pp. 585 –595, Nov. 1999.

[2] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, vol. 1, no. 2, pp. 41 –59, 1996.

[3] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. on Inf. Theory*, vol. 49, no. 5, pp. 1073 – 1096, 2003.

[4] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. on Inf. Theory*, vol. 44, no. 2, pp. 744 –765, mar. 1998.

[5] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. on Inf. Theory*, vol. 45, no. 5, pp. 1456 –1467, jul. 1999.

[6] P. Elia, K. Kumar, S. Pawar, P. Kumar, and H.-F. Lu, "Explicit space time codes achieving the diversity multiplexing gain tradeoff," *IEEE Trans. on Inf. Theory*, vol. 52, no. 9, pp. 3869 –3884, sep. 2006.

[7] H. El Gamal, G. Caire, and M. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of mimo channels," *IEEE Trans. on Inf. Theory*, vol. 50, no. 6, pp. 968 –985, 2004.

[8] M. Taherzadeh and A. K. Khandani, "On the limitations of the naive lattice decoding," *IEEE Trans. on Inf. Theory*, vol. 56, no. 10, pp. 4820 –4826, oct. 2010.

[9] G. Poltyrev, "On Coding Without Restrictions for the AWGN Channel," *IEEE Trans. on Inf. Theory*, vol. 40, no. 2, pp. 409–417, 1994.

[10] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451 –1458, oct 1998.

[11] V. Tarokh, A. Vardy, and K. Zeger, "Universal bound on the performance of lattice codes," *Information Theory, IEEE Transactions on*, vol. 45, no. 2, pp. 670 –681, mar 1999.

[12] P. Gruber and C. Lekkerkerker, "Geometry of Numbers," *Nort-Holland Mathematical Library*.

Appendix A
Proof of Theorem 1

We prove the result for any IC with density γ_{rc}. The proof outline is as follows. We prove the theorem by contradiction. First, for a given IC with receiver density γ_{rc}, we assume average decoding error probability that equals to the lower bound we wish to prove. Then, we derive a "regular" IC from the given IC with the same density γ_{rc} and the same average decoding error probability. Regularizing the IC allows us to find a lower bound on the IC maximal error probability that depends on its density. We expurgate half of the codewords with largest error probability and get another regular IC with density $\frac{\gamma_{rc}}{2}$. Based on the average decoding error probability, we upper bound the expurgated IC maximal error probability, and based on its density we lower bound the same maximal error probability, and get contradiction.

Let us consider a KT-complex dimensional IC in the receiver, $S'_{KT}(\rho)$, with receiver density γ_{rc} and average decoding error probability

$$\overline{Pe}(H,\rho) = (1-\epsilon^*)\frac{\overline{C}(KT)}{4}e^{-\mu_{\mathrm{rc}}\cdot\overline{A}(KT)+(KT-1)\ln(\mu_{\mathrm{rc}})} \tag{53}$$

where $\overline{A}(KT) = (\frac{1}{(1-\epsilon_1)(1-\epsilon_2)})^{\frac{1}{KT}}e \cdot \Gamma(KT+1)^{\frac{1}{KT}}$, $\overline{C}(KT) = (\frac{1}{(1-\epsilon_1)(1-\epsilon_2)})^{\frac{KT-1}{KT}}\frac{e^{KT-\frac{3}{2}}\Gamma(KT+1)^{\frac{KT-1}{KT}}}{2\cdot\Gamma(KT)}$ and $0 < \epsilon_1, \epsilon_2 < 1$.

Next we construct a regularized IC, $S''_{KT}(\rho)$, from $S'_{KT}(\rho)$, that has bounded and finite volume Voronoi regions, i.e. there exists a finite radius r such that $V(x) \subset Ball(x,r)$, $\forall x \in S''_{KT}(\rho)$, where $Ball(x,r)$ is a KT-complex dimensional ball centered around x. We construct $S''_{KT}(\rho)$ in the following manner. Let us define $C_0(\rho,H) = \{S'_{KT}(\rho)\bigcap(H_{ex}\cdot cube_{KT}(b))\}$, i.e. a finite constellation derived from $S'_{KT}(\rho)$. We turn this finite constellation into an IC by tiling $C_0(\rho,H)$ in the following manner

$$S''_{KT}(\rho) = C_0(\rho,H) + (b+b')\tilde{H}_{ex}\mathbb{Z}^{2KT} \tag{54}$$

where for simplicity we assumed that $cube_{KT}(b) \subset \mathbb{C}^{KT}$, i.e. contained within the first KT complex dimensions. Correspondingly, under this assumption, $\tilde{H}_{ex}$ equals the first KT complex columns of H_{ex}. In this case, the tiling of $C_0(\rho, H)$ is done according to the complex integer combinations of $\tilde{H}_{ex}$ columns. In general, $cube_{KT}(b)$ may be a rotated cube within $\mathbb{C}^{MT}$. In this case the tiling is done according to some KT complex linearly independent vectors, consisting of linear combinations of H_{ex} columns. An alternative way to construct $S^{''}_{KT}(\rho)$ is by considering the transmitter IC $S_{KT}(\rho)$. In this case we can construct another IC in the transmitter

$$\overline{S}_{KT}(\rho) = \{S_{KT}(\rho) \bigcap cube_{KT}(b)\} + (b + b^{'})\mathbb{Z}^{2KT} \quad (55)$$

where without loss of generality we assumed again that $cube_{KT}(b) \in \mathbb{C}^{KT}$. In this case $S^{''}_{KT}(\rho) = \{H_{ex} \cdot \overline{S}_{KT}(\rho)\}$.

Next we would like to set b and $b^{'}$ to be large enough such that $S^{''}_{KT}(\rho)$ has average decoding error probability smaller or equal to $\frac{\overline{C}(KT)}{2} e^{-\mu_{\rm rc} \cdot \overline{A}(KT) + (KT-1)\ln(\mu_{\rm rc})}$ and density larger or equal to γ_{rc}. Due to the symmetry that results from the tiling (54), it is sufficient to upper bound the average decoding error probability of the words $x \in C_0(\rho, H) \subset S^{''}_{KT}(\rho)$ denoted by $Pe_{S^{''}_{KT}}(C_0)$ in order to upper bound the entire IC $S^{''}_{KT}(\rho)$ average decoding error probability. Hence $Pe_{S^{''}_{KT}}(C_0)$ is also the average decoding error probability for the IC $S^{''}_{KT}(\rho)$. We can upper bound the error probability in the following manner

$$Pe_{S^{''}_{KT}}(C_0) \leq Pe(C_0) + Pe(S^{''}_{KT} \setminus C_0) \quad (56)$$

where $Pe(C_0)$ is the average decoding error probability of the finite constellation $C_0(\rho, H)$ and $Pe(S^{''}_{KT} \setminus C_0)$ is the average decoding error probability to points in the set $\{S^{''}_{KT} \setminus C_0(\rho, h)\}$, i.e. the error probability inflicted by the replicated codewords outside the set $C_0(\rho, H)$.

We begin by upper bounding $Pe(S^{''}_{KT} \setminus C_0)$ by choosing $b^{'}$ to be large enough. By the tiling at the transmitter (55) and the fact that we have finite complex dimension KT, for a certain channel realization H_{ex} we get that there exists $\delta(H_{ex})$ such that any pair of points $x_1 \in C_0(\rho, H)$, $x_2 \in \{S^{''}_{KT} \setminus C_0(\rho, h)\}$ fulfils $\|\underline{x}_1 - \underline{x}_2\| \geq 2b^{'} \cdot \delta(H_{ex})$. The term $\delta(H_{ex})$ is a factor that defines the minimal distance between these 2 sets for a given channel realization. Note that also for the case $M > N$, there must exist such $\delta(H_{ex})$, as we assumed that $S^{''}_{KT}(\rho)$ is KT-complex dimensional IC, i.e. the projected IC $S^{''}_{KT}(\rho) = H_{ex}\overline{S}_{KT}(\rho)$ is also KT-complex dimensional. Hence, we get that

$$Pe(S^{''}_{KT} \setminus C_0) \leq Pr(\|\underline{\tilde{n}}_{\rm ex}\| \geq b^{'}\delta(H_{ex}))$$

where $\underline{\tilde{n}}_{\rm ex}$ is the effective noise in the KT-complex dimensional hyperplane where $S^{''}_{KT}(\rho)$ resides. By using the upper bounds from [9], we get that for $\frac{(b^{'}\delta(H_{ex}))^2}{2KT} > \sigma^2$

$$Pr(\|\underline{\tilde{n}}_{\rm ex}\| \geq b^{'}\delta(H_{ex})) \leq e^{-\frac{(b^{'}\delta(H_{ex}))^2}{2\sigma^2}} (\frac{(b^{'}\delta(H_{ex}))^2 e}{2KT\sigma^2})^{KT}.$$

Hence, for $b^{'}$ large enough we get that

$$Pe(S^{''}_{KT} \setminus C_0) \leq (1 - \epsilon^*) \frac{\overline{C}(KT)}{4} e^{-\mu_{\rm rc} \cdot \overline{A}(KT) + (KT-1)\ln(\mu_{\rm rc})}.$$

Now we would like to upper bound the error probability, $Pe(C_0)$, of the finite constellation $C_0(\rho, H)$. According to the definition of the average decoding error probability in (8), the definition of $C_0(\rho, H)$ and the assumption in (53), we get that

$$P_e(C_0) \leq \frac{(1 - \epsilon^*)(1 + \epsilon(b))}{4} \overline{C}(KT) e^{-\mu_{\rm rc} \cdot \overline{A}(KT)} \cdot \mu_{\rm rc}^{(KT-1)}$$

where $lim_{b \to \infty} \epsilon(b) = 0$. It results from the fact that in (8) we take the limit supremum, and so for b large enough the average decoding error probability of the IC must be upper bounded by the aforementioned term. Also, for any b the average decoding error probability of the finite constellation $C_0(\rho, H)$ is smaller or equal to the error probability, defined in (8), of decoding over the entire IC. Based on the upper bound from (56) we get the following upper bound on the error probability of $S^{''}_{KT}(\rho)$

$$Pe_{S^{''}_{KT}}(C_0) \leq \frac{(1-\epsilon^*)(1+\epsilon(b))}{2} \overline{C}(KT) e^{-\mu_{\rm rc} \cdot \overline{A}(KT)} \cdot \mu_{\rm rc}^{(KT-1)}. \quad (57)$$

According to the definition of γ_{rc} and due to the fact that we are taking limit supremum: for any $0 < \epsilon_1 < 1$ there exists b large enough such that

$$\frac{|C_0(\rho, H)|}{vol\big(H_{ex} \cdot cube_{KT}(b)\big)} \geq (1 - \epsilon_1)\gamma_{rc}. \quad (58)$$

where $|C_0(\rho, H)|$ is the number of words in $C_0(\rho, H)$. In fact there exists large enough b that fulfils both (57) and (58).

In (54) we tiled by $b + b^{'}$. If we had tiled $C_0(\rho, H)$ only by b, then for large enough b we would have got IC with density larger or equal to $(1 - \epsilon_1)\gamma_{rc}$. However , as we tile by $b + b^{'}$, we get for b large enough that $S^{''}_{KT}(\rho)$ has density greater or equal to $\frac{1-\epsilon_1}{1+\frac{b^{'}}{b}}\gamma_{rc}$. Hence, for any $0 < \epsilon_2 < 1$ there exists b large enough such that

$$\gamma^{''}_{rc} \geq (1 - \epsilon_1)(1 - \epsilon_2)\gamma_{rc}. \quad (59)$$

where $\gamma^{''}_{rc}$ is the density of $S^{''}_{KT}(\rho)$. Again, there also must exist large enough b that fulfils (57) and (59) simultaneously. Hence, for large enough b we can derive from $S^{'}_{KT}(\rho)$ an IC $S^{''}_{KT}(\rho)$ with density $\gamma^{''}_{rc} \geq (1 - \epsilon_1)(1 - \epsilon_2)\gamma_{rc}$ and average decoding error probability smaller or equal to $\frac{(1-\epsilon^*)(1+\epsilon(b))}{2} \overline{C}(KT) e^{-\mu_{\rm rc} \cdot \overline{A}(KT) + (KT-1)\ln(\mu_{\rm rc})}$.

By averaging arguments we know that expurgating the worst half of the codewords in $S^{''}_{KT}(\rho)$, yields an IC $S^{'''}_{KT}(\rho)$ with density

$$\gamma^{'''}_{rc} \geq (1 - \epsilon_1)(1 - \epsilon_2)\frac{\gamma_{rc}}{2} = \overline{\gamma_{rc}} \quad (60)$$

and maximal decoding error probability

$$sup_{x \in S^{'''}_{KT}} Pe_{S^{'''}_{KT}}(x) \leq (1 - \epsilon^*)(1 + \epsilon(b))\overline{C}(KT) e^{-\mu_{\rm rc} \cdot \overline{A}(KT)} \mu_{\rm rc}^{KT-1} \quad (61)$$

where $Pe_{S^{'''}_{KT}}(x)$ is the error probability of $x \in S^{'''}_{KT}(\rho)$.

From the construction method of $S^{''}_{KT}(\rho)$, defined in (54), it can be easily shown that tiling $C_0(\rho, H)$ yields bounded and finite volume Voronoi regions, i.e. there exists a finite radius r such that $V(x) \subset Ball(x, r)$, $\forall x \in S^{''}_{KT}(\rho)$. Due to the symmetry that results from $S^{''}_{KT}(\rho)$ construction (54), it also applies for $S_{KT}^{'''}(\rho)$. Hence, there must exist a point $x_0 \in S_{KT}^{'''}(\rho)$ that satisfies $|V(x_0)| \le \frac{1}{\gamma_{rc}^{'''}} \le \frac{1}{\overline{\gamma_{rc}}}$. According to the definition of the effective radius in (1), we get that $r_{\text{eff}}(x_0) \le r_{\text{eff}}(\overline{\gamma_{rc}})$. Hence, we get

$$sup_{x\in S_{KT}^{'''}} Pe_{S_{KT}^{'''}}(x) \ge Pe_{S_{KT}^{'''}}(x_0) > Pr\big(\|\underline{\tilde{n}}_{\text{ex}}\| \ge r_{\text{eff}}(x_0)\big) \ge Pr\big(\|\underline{\tilde{n}}_{\text{ex}}\| \ge r_{\text{eff}}(\overline{\gamma_{rc}})\big) \quad (62)$$

where the lower bound $Pe_{S_{KT}^{'''}}(x_0) > Pr(\|\underline{\tilde{n}}_{\text{ex}}\| \ge r_{\text{eff}}(x_0))$ was proven in [9]. We calculate the following lower bound

$$\Pr\left(\|\underline{\tilde{n}}_{\text{ex}}\| \ge r_{\text{eff}}(\overline{\gamma_{\text{rc}}})\right) > \int_{r_{\text{eff}}^2}^{r_{\text{eff}}^2+\sigma^2} \frac{r^{KT-1}e^{-\frac{r}{2\sigma^2}}}{\sigma^{2KT}2^{KT}\Gamma(KT)}dr \ge \frac{r_{\text{eff}}^{2KT-2}e^{-\frac{r_{\text{eff}}^2}{2\sigma^2}}}{\sigma^{2KT-2}2^{KT}\Gamma(KT)\sqrt{e}} \quad (63)$$

By assigning $r_{\text{eff}}^2 = (\frac{\Gamma(KT+1)}{\overline{\gamma_{rc}}\pi^{KT}})^{\frac{1}{KT}}$ we get

$$sup_{x\in S_{KT}^{'''}} Pe_{S_{KT}^{'''}}(x) > \overline{C}(KT)\cdot e^{-\frac{\overline{\gamma_{\text{rc}}}^{-\frac{1}{KT}}}{2\pi e\sigma^2}\overline{A}(KT)+(KT-1)\ln(\frac{\overline{\gamma_{\text{rc}}}^{-\frac{1}{KT}}}{2\pi e\sigma^2})}. \quad (64)$$

Hence, for certain ϵ_1 and ϵ_2 we get

$$sup_{x\in S_{KT}^{'''}} Pe_{S_{KT}^{'''}}(x) > \overline{C}(KT)\cdot e^{-\mu_{rc}\overline{A}(KT)+(KT-1)\ln(\mu_{rc})} \quad (65)$$

where $\mu_{rc} = \frac{\overline{\gamma_{\text{rc}}}^{-\frac{1}{KT}}}{2\pi e\sigma^2}$. For b large enough we get $(1-\epsilon^*)(1+\epsilon(b)) < 1$, and so (65) contradicts (61). As a result we get contradiction of the initial assumption in (53). This contradiction also holds for any $\overline{Pe}(H,\rho) < \frac{(1-\epsilon^*)\overline{C}(KT)}{4}e^{-\mu_{\text{rc}}\cdot\overline{A}(KT)+(KT-1)\ln(\mu_{\text{rc}})}$. Hence, we get that

$$\overline{Pe}(H,\rho) > \frac{\overline{C}(KT)}{4}e^{-\mu_{\text{rc}}\cdot\overline{A}(KT)+(KT-1)\ln(\mu_{\text{rc}})}. \quad (66)$$

Note that the lower bound holds for any $0 < \epsilon_1, \epsilon_2, \epsilon^* < 1$ and also that the expressions in (53), (66) are continuous. As a result we can also set $\epsilon_1 = \epsilon_2 = \epsilon^* = 0$ and get the desired lower bound. Finally, note that we are interested in a lower bound on the error probability of any IC for a given channel realization. Hence, we are free to choose different values for b and $b^{'}$ for each channel realization. and $b^{'}$.

Appendix B
Proof of the optimization problem in Theorem 2

We would like to solve the optimization problem in (15) for any value of $K = B + \beta \le L$, where $B \in \mathbb{N}$ and $0 < \beta \le 1$. First we consider the case of $0 < K \le 1$, i.e. the case where $B = 0$. In this case the constraint boils down to $\alpha_L = 1 - \frac{r}{K}$. By assigning $\alpha_1 = \cdots = \alpha_L = 1 - \frac{r}{K}$ we get that $d_{KT}(r) \le MN(1-\frac{r}{K})$. Next we analyze the case where $K > 1$. Due to the constraint, the minimal value must satisfy $\alpha_1 = \cdots = \alpha_{L-B}$. From the constraint we also know that $\alpha_L = K - r - \sum_{i=1}^{B-1}\alpha_{L-i} - \beta\alpha_{L-B}$. By assigning in (15) we get

$$\min_{\underline{\alpha}>0}(K-r)(N+M-1) + \big((M-B)(N-B) - \beta(N+M-1)\big)\alpha_{L-B} - \sum_{i=1}^{B-1} 2i\cdot\alpha_{L-i} \quad (67)$$

where $\alpha > 0$ signifies $\alpha_1 \ge \cdots \ge \alpha_L \ge 0$. We would like to consider 2 cases. The case where $\big((M-B)(N-B) - \beta(N+M-1)\big) > \sum_{i=1}^{B-1} 2i$ and the case where $\big((M-B)(N-B) - \beta(N+M-1)\big) \le \sum_{i=1}^{B-1} 2i$. The first case, where $\big((M-B)(N-B) - \beta(N+M-1)\big) > B(B-1)$, is achieved for $K < \frac{MN}{N+M-1}$. In this case we use the following Lemma in order to find the optimal solution

Lemma 3. *Consider the optimization problem*

$$\min_{\underline{c}} B_1c_1 - \sum_{i=2}^{D} B_ic_i$$

where: (1). $c_1 \ge \cdots \ge c_D \ge 0$*;* (2). $B_1 > \sum_{i=2}^{D} B_i$ *and* $B_2 > \cdots > B_D > 0$*;* (3). $\beta c_1 + \sum_{i=2}^{D} c_i = \delta > 0$*, where* $0 < \beta \le 1$*. The minimal value is achieved for* $c_1 = \cdots = c_D = \frac{\delta}{D-1+\beta}$*.*

Proof: We prove by induction. First let us consider the case where $D = 2$. In this case we would like to find

$$\min_{\underline{c}} B_1c_1 - B_2c_2. \quad (68)$$

where $c_1 \ge c_2 \ge 0$, $\beta c_1 + c_2 = \delta > 0$, $B_1 > B_2 > 0$ and $0 < \beta \le 1$. It is easy to see that for this case the minimum is achieved for $c_1 = c_2$, as increasing c_1 while decreasing c_2 to satisfy $\beta c_1 + c_2 = \delta$ will only increase (68).

Now let us assume that for D elements, the minimum is achieved for $c_1 = \cdots = c_D = \frac{\delta}{D-1+\beta}$. Let us consider $D+1$ elements with constraint $\beta c_1 + \sum_{i=2}^{D+1} c_i = \delta$. If we take $c_1 = \cdots = c_{D+1} = \frac{\delta}{D+\beta}$ we get

$$(B_1 - \sum_{i=2}^{D+1} B_i)\frac{\delta}{D+\beta}. \quad (69)$$

We would like to show that it is the minimal possible value for this problem. Let us take $c_{D+1}^{'} = \frac{\delta}{D+\beta} - \epsilon \ge 0$. In this case we get $\beta c_1^{'} + \sum_{i=2}^{D} c_i^{'} = \frac{(D-1+\beta)\delta+(D+\beta)\epsilon}{D+\beta}$ in order to satisfy $\beta c_1^{'} + \sum_{i=2}^{D+1} c_i^{'} = \delta$. According to our assumption $B_1c_1^{'} - \sum_{i=2}^{D} B_ic_i^{'}$ is minimal for $c_1^{'} = \cdots = c_D^{'} = \frac{\delta}{D+\beta} + \frac{\epsilon}{D-1+\beta}$. By assigning these values we get

$$(B_1 - \sum_{i=2}^{D+1} B_i)\frac{\delta}{D+\beta} + (B_1 - \sum_{i=2}^{D} B_i)\frac{\epsilon}{D-1+\beta} + B_{D+1}\epsilon$$

which is greater than (69). This concludes the Lemma proof. ■

For the case $\big((M-B)(N-B) - \beta(N+M-1)\big) > B(B-1)$, the optimization problem coincides with Lemma 3 as it fulfils the condition $B_1 > \sum_{i=2}^{D} B_i$ in the lemma. Hence, the optimization problem solution for $K < \frac{MN}{N+M-1}$ is

$\alpha_1 = \cdots = \alpha_{L-1} = \frac{K-r-\alpha_L}{K-1} = \alpha$. The minimum is achieved when $\alpha_L = \alpha$, i.e. the maximal value α_L can receive under the constraint $\alpha_1 \geq \cdots \geq \alpha_L \geq 0$. We get that $\alpha = 1 - \frac{r}{K}$, and the optimization problem solution of (15) for the case $K < \frac{MN}{M+N-1}$ is $d_{KT}(r) \leq MN(1-\frac{r}{K})$, .

For the case $\big((M-B)(N-B)-\beta(N+M-1)\big) \leq B(B-1)$, or equivalently $K \geq \frac{MN}{N+M-1}$, we would like to show that the optimal solution must fulfil $\alpha_L = 0$. It result from the fact that for the optimal solution, the term $\big((M-B)(N-B)-\beta(N+M-1)\big)\alpha_{L-B} - \sum_{i=1}^{B-1} 2i \cdot \alpha_{L-i}$ in (67) must be negative. This is due to the fact that taking $\alpha_1 = \cdots = \alpha_{L-1}$ gives negative value. Hence, for the optimal solution we would like to maximize $\sum_{i=1}^{B-1} \alpha_{L-i} - \beta\alpha_{L-B} = K - r - \alpha_L$. By taking $\alpha_L = 0$ the sum is maximized. Hence, the optimal solution for $K \geq \frac{MN}{M+N-1}$ must have $\alpha_L = 0$.

Now let us consider the general case. Assume that for $K \geq \frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1$ the optimal solution must have $\alpha_L = \cdots = \alpha_{L-l+1} = 0$. First we consider the case where $1 \leq l \leq B-1$. For this case the constraint is $\sum_{i=l}^{B-1} \alpha_{L-i} + \beta\alpha_{L-B} = K - r$, i.e. the constraint contains at least 2 singular values. We can rewrite (15) as follows

$$\min_{\underline{\alpha}>0}(K-r)(N+M-1-2\cdot l) + \big((M-B)(N-B) - \beta(N+M-1-2\cdot l)\big)\alpha_{L-B} - \sum_{i=l+1}^{B-1} 2(i-l)\cdot\alpha_{L-i}. \tag{70}$$

For the case $\big((M-B)(N-B)-\beta(N+M-1-2\cdot l)\big) > (B-1-l)(B-l)$ we get that $K < \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$ and we also assumed that $K \geq \frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1$. For this case we can use Lemma 3 and get that the optimization problem solution is $\alpha_{L-l-1} = \cdots = \alpha_{L-B} = \frac{K-r-\alpha_{L-l}}{K-l-1} = \alpha$. The minimum is achieved for $\alpha_{L-l} = \alpha$. We get that $\alpha_L = \cdots = \alpha_{L-l+1} = 0$ and $\alpha_1 = \cdots = \alpha_{L-l} = \frac{K-r}{K-l}$. Hence, for the case $\frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1 \leq K < \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$ the optimization problem solution is $d_{KT}(r) \leq (N-l)(M-l)\frac{K-r}{K-l}$.

For the case $\big((M-B)(N-B)-\beta(N+M-1-2\cdot l)\big) \leq (B-1-l)(B-l)$, or equivalently $K \geq \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$, the term $\big((M-B)(N-B)-\beta(N+M-1-2\cdot l)\big)\alpha_{L-B} - \sum_{i=l+1}^{B-1} 2(i-l)\cdot\alpha_{L-i}$ in (70) must be negative for the optimal solution. This is due to the fact that by taking $\alpha_1 = \cdots = \alpha_{L-l-1}$ we get negative value. Hence we would like to maximize the sum $\sum_{i=l+1}^{B-1} \alpha_{L-i} + \beta\alpha_{L-B} = K - r - \alpha_{L-l}$. The sum is maximized by taking $\alpha_{L-l} = 0$. Hence the optimal solution for the case $K \geq \frac{(M-l)(N-l)}{N+M-1-2\cdot l} + l$ must have $\alpha_{L-l} = \cdots = \alpha_L = 0$. Note that for the case $l = B-1$ we have only 2 terms in the constraint $\alpha_{L-B+1} + \beta\alpha_{L-B} = K - r$. However, the solution remains the same.

For the case $K \geq \frac{(M-l+1)(N-l+1)}{N+M-1-2(l-1)} + l - 1$ and $l = B$ the constraint is of the form $\alpha_{L-B} = \frac{K-r}{K-l}$. Again we assume that $\alpha_{L-B+1} = \cdots = \alpha_L = 0$. In this case the optimization problem solution is $\alpha_1 = \cdots = \alpha_{L-l} = \frac{K-r}{K-l}$ and the optimization problem solution gives us $d_{KT}(r) \leq (M-l)(N-l)\frac{K-r}{K-l}$. This concludes the proof.

Appendix C
Proof of Lemma 1

We begin by proving the case $N \geq M$. From the construction of G_l it can be seen that a set of columns $\{\underline{h}_j, \ldots, \underline{h}_i\}$ may occur in $N-i+j$ blocks at most. It results from the fact that we can only subtract $M-i$ columns to the right of $\underline{h}_i$ (18), and $j-1$ columns to the left of $\underline{h}_j$ (19), and still get a block that contains $\{\underline{h}_j, \ldots, \underline{h}_i\}$ (or even more specifically a block that contains $\{\underline{h}_j, \underline{h}_i\}$). In addition, columns $\{\underline{h}_j, \ldots, \underline{h}_i\}$ must occur in the first $N-M+1$ blocks, as these blocks equal to H (17). Hence, we can upper bound the number of occurrences by $N-M+1+j-1+M-i = N-i+j$.

Next we prove for the case $M > N$. In case $0 \leq i-j < N$, the set of columns $\{\underline{h}_j, \ldots, \underline{h}_i\}$ may occur in $N-i+j$ blocks at most. We divide the proof into four cases.

1) First consider the case where $i \leq N$ and $j \geq M-N+1$. In this case the set of columns $\{\underline{h}_j, \ldots, \underline{h}_i\}$ occurs in all the first $M-N+1$ blocks (20). As for the additional $N-1-l$ pairs of columns, the set of columns belongs both to the set $\{\underline{h}_1, \ldots, \underline{h}_N\}$ and $\{\underline{h}_{M-N+1}, \ldots, \underline{h}_M\}$. Hence, in the additional column pairs we can subtract $N-i$ columns to the right of $\underline{h}_i$ (21) and $j-M+N-1$ columns to the left of $\underline{h}_j$ (22). If we add it together we get that the number of occurrences can not exceed $N-i+j$.
2) For the case $i \leq N$ and $j < M-N+1$ the set of columns can have only j occurrences in the first $M-N+1$ blocks. In this case the set $\{\underline{h}_j, \ldots, \underline{h}_i\}$ occurs within $\{\underline{h}_1, \ldots, \underline{h}_N\}$ but does not occur within $\{\underline{h}_{M-N+1}, \ldots, \underline{h}_M\}$. Hence, the transmission scheme only subtracts columns to the right of $\underline{h}_i$ (21). In this case we can have $N-i$ subtractions and together we get $N-i+j$ occurrences at most.
3) For the case $i > N$ and $j \geq M-N+1$ we have $M-i+1$ occurrences in the first $M-N+1$ blocks. In this case the set $\{\underline{h}_j, \ldots, \underline{h}_i\}$ occurs within $\{\underline{h}_{M-N+1}, \ldots, \underline{h}_M\}$ but does not occur within $\{\underline{h}_1, \ldots, \underline{h}_N\}$. Hence we can subtract up to $j-M+N-1$ columns to the left of $\underline{h}_j$ (22). Together we get $N-i+j$ occurrences at most.
4) For the last case we have $i > N$ and $j < M-N+1$. In this case the set of columns can only occur in the first $M-N+1$ blocks. In this case there are exactly $N-i+j$ occurrences in the first $M-N+1$ blocks.

In case $i-j \geq N$, the set of columns does't occur in any block as each column of G_l doesn't have more than N non-zero entries.

Appendix D
Proof of Lemma 2

We know that

$$c(j) = \sum_{s=0}^{\min(j,L)-2} a(s, \underline{\xi}_j) + (N - \min(j,L) + 1)a(\min(j,L)-1, \underline{\xi}_j)$$

where

$$a(k,\underline{\xi}_j)=\min_{s\in\{k+1,\dots,N\}}\xi_{s,j}\qquad 0\le k\le \min(j,L)-1$$

and by definition

$$a(\min(j,L)-1,\underline{\xi}_j)\ge\cdots\ge a(0,\underline{\xi}_j)\ge 0.$$

In order to prove the Lemma we begin with $a(\min(j,L)-1,\underline{\xi}_j)$. We know that

$$\sum_{s=\min(j,L)}^{N}\xi_{s,j}\ge(N-\min(j,L)+1)\cdot\min_s\xi_{s,j}\qquad(71)$$

where $s\in\{\min(j,L),\dots,N\}$. We can also see that

$$\xi_{k+1,j}\ge\min_{s\in\{k+1,\dots,N\}}\xi_{s,j}\qquad(72)$$

for $0\le k\le\min(j,L)-2$. Hence we get

$$c(j)-\sum_{i=1}^{N}\xi_{i,j}\le 0.$$

This concludes the proof.

Appendix E
Proof of Theorem 5

We prove that there exists a sequence of $2K_lT_l$-real dimensional lattices (as a function of ρ) that attains the same diversity order as in Theorem 4. By using the *Minkowski-Hlawaka-Siegel* Theorem [9],[12], we upper bound the error probability of the ensemble of lattices, for each channel realization. This upper bound equals to the upper bound derived in Theorem 3. Then we average the upper bound over all channel realizations, and receive the desired diversity order.

We consider a $2K_lT_l$-real dimensional ensemble of lattices, transmitted using the transmission scheme defined in subsection IV-A. We spread the first K_lT_l dimensions of the lattice on the real part of the non-zero entries of G_l, and the other K_lT_l dimensions of the lattice on the imaginary part of the non-zero entries of G_l. Each lattice in the ensemble has transmitter density $\gamma_{tr}=\rho^{rT_l}$, i.e. multiplexing gain r. We begin by analyzing the performance of the ensemble of lattices in the receiver, for each channel realization. We assume a certain channel realization that induces receiver VNR $\mu_{rc}=\rho^{1-\frac{r}{K_l}-\sum_{i=1}^{K_lT_l}\frac{\eta_i}{K_lT_l}}$, where $\underline{\eta}\ge 0$. For each lattice in the ensemble we get that the channel realization induces a new lattice in the receiver, $H_{eff}^{(l)}\cdot\underline{x}$, with density γ_{rc} in accordance with (3) and subsection IV-B. For lattices with regular lattice decoding, the error probability is equal among all codewords. Hence, it is sufficient to analyze the lattice zero codeword error probability. We define the indication function

$$I_{Ball(0,2R)}(\underline{x})=\begin{cases}1, & \|\underline{x}\|\le 2R\\ 0, & else\end{cases}.$$

In a similar manner to (24) we can state that for each lattice induced in the receiver, Λ_{rc}, the lattice zero codeword error probability is upper bounded by

$$\sum_{\underline{x}\in\Lambda_{\mathrm{rc}},\underline{x}\neq0}I_{Ball(0,2R_{\mathrm{eff}})}(\underline{x})\cdot Pr(\|\underline{\tilde{n}}_{\mathrm{ex}}\|>\|\underline{x}-\underline{\tilde{n}}_{\mathrm{ex}}\|)+Pr(\|\underline{\tilde{n}}_{\mathrm{ex}}\|\ge R_{\mathrm{eff}})\qquad(73)$$

where $\frac{R_{\mathrm{eff}}^2}{2K_lT_l\sigma^2}=\mu_{rc}$, and $\underline{\tilde{n}}_{\mathrm{ex}}$ is the effective noise in the K_lT_l-complex hyperplane where Λ_{rc} resides in. By defining $f_{rc}(\underline{x})=I_{Ball(0,2R_{\mathrm{eff}})}(\underline{x})\cdot Pr(\|\underline{\tilde{n}}_{\mathrm{ex}}\|>\|\underline{x}-\underline{\tilde{n}}_{\mathrm{ex}}\|)$, we can rewrite the upper bound on the error probability from (73)

$$\sum_{\underline{x}\in\Lambda_{\mathrm{rc}},\underline{x}\neq0}f_{\mathrm{rc}}(\underline{x})+Pr(\|\underline{\tilde{n}}_{\mathrm{ex}}\|\ge R_{\mathrm{eff}}).\qquad(74)$$

Note that

$$\gamma_{rc}\int_{\mathbb{R}^{2K_lT_l}}f_{rc}(\underline{x})d\underline{x}+Pr(\|\underline{\tilde{n}}_{\mathrm{ex}}\|\ge R_{\mathrm{eff}})\qquad(75)$$

is equal to the expression in (27), where γ_{rc} is the density of the lattice induced in the receiver Λ_{rc}, as defined above.

We need to show that there exists a single probability measure for all channel realizations, that gives average decoding error probability over the ensemble, which is upper bounded by (75). Hence, we consider the ensemble of lattices in the transmitter which is fixed for each channel realization. For this reason we define

$$\underline{y}'_{\mathrm{ex}}=\left(H_{\mathrm{eff}}^{(l)\dagger}\cdot H_{\mathrm{eff}}^{(l)}\right)^{-1}H_{\mathrm{eff}}^{(l)\dagger}\cdot\underline{y}_{\mathrm{ex}}.\qquad(76)$$

Note that the operation in (76) does not change the error probability of the lattice when we use regular lattice decoding. Each lattice in the ensemble has density $\gamma_{tr}=\rho^{rT_l}$. Now we define the following indication function

$$I_{ellipse(H,2R)}(\underline{x})=\begin{cases}1, & \|H\cdot\underline{x}\|\le 2R\\ 0, & else\end{cases},$$

that is the function is one if $\underline{x}$ is within the ellipse and zero otherwise. Let us denote the error probability of a lattice in the ensemble for certain channel realization $\underline{\eta}$ by $P_e^{(\nu)}(\underline{\eta},\rho)$, where ν is a random variable that represents a certain lattice in the ensemble. Using regular lattice decoding, we get the following upper bound on the error probability for each lattice codeword

$$P_e^{(\nu)}(\underline{\eta},\rho)\le Pr(\|A\cdot\underline{\hat{n}}_{\mathrm{ex}}\|\ge R_{\mathrm{eff}})+\sum_{\underline{x}\in\Lambda_{\mathrm{tr}},\underline{x}\neq0}I_{ellipse(H_{\mathrm{eff}}^{(l)},2R_{\mathrm{eff}})}(\underline{x})\cdot Pr\big(\|A\cdot\underline{\hat{n}}_{\mathrm{ex}}\|>\|A\cdot(\underline{x}-\underline{\hat{n}}_{\mathrm{ex}})\|\big)\qquad(77)$$

where A is a K_lT_lxK_lT_l matrix that satisfies $A^\dagger A=H_{\mathrm{eff}}^{(l)\dagger}H_{\mathrm{eff}}^{(l)}$, Λ_{tr} is the lattice from the ensemble that corresponds to ν and $\underline{\hat{n}}_{\mathrm{ex}}\sim CN\big(0,(H_{\mathrm{eff}}^{(l)\dagger}H_{\mathrm{eff}}^{(l)})^{-1}\big)$. Note that (77) is equal to (74), and the corresponding terms in the expressions are also equal.

Let us define $g_{rc}(\underline{x})=I_{ellipse(H_{\mathrm{eff}}^{(l)},2R_{\mathrm{eff}})}(\underline{x})\cdot Pr\big(\|A\underline{\hat{n}}_{\mathrm{ex}}\|>\|A(\underline{x}-\underline{\hat{n}}_{\mathrm{ex}})\|\big)$. We get that

$$\gamma_{\mathrm{tr}}\int_{\mathbb{R}^{2K_lT_l}}g_{\mathrm{rc}}(\underline{x})d\underline{x}=\gamma_{\mathrm{rc}}\int_{\mathbb{R}^{2K_lT_l}}f_{\mathrm{rc}}(\underline{x})d\underline{x}.\qquad(78)$$

Next we show that by averaging the upper bound in (77) over the ensemble of lattices in the transmitter, with the correct probability measure, we get

$$E_\nu\{P_e^{(\nu)}(\underline{\eta},\rho)\}\leq\gamma_{rc}\int_{\mathbb{R}^{2K_lT_l}}f_{rc}(\underline{x})d\underline{x}+Pr(\|\underline{\tilde{n}}_{\text{ex}}\|\geq R_{\text{eff}}). \tag{79}$$

We prove (79) by using the *Minkowski-Hlawaka-Siegel* theorem, [9], presented on Theorem 6.

Theorem 6. *On the set of all the lattices of density γ in $\mathbb{R}^{2K_lT_l}$, there exists a probability measure ν such that for any Riemann integrable function $f(\underline{x})$ which vanishes outside some bounded region we have*

$$E_\nu\{\sum_{\underline{x}\in\Lambda}g(\underline{x})\}=\gamma\int_{\mathbb{R}^{2K_lT_l}}g(\underline{x})d\underline{x} \tag{80}$$

where $E_\nu\{\cdot\}$ represents the expectation with respect to the measure ν.

Note that considering a $2K_lT_l$-real dimensional lattices enables us to use this theorem. Hence, by choosing $\gamma=\gamma_{\text{tr}}$, $g(\underline{x})=g_{\text{rc}}(\underline{x})$, and considering (77), (78) we get the desired upper bound (79). As a result, we can upper bound the ensemble average decoding error probability for each channel realization by the upper bound from Theorem 3 (34).

Now we are ready to lower bound the diversity order. According to Theorem 6 there exists a single probability measure that satisfies (80), for any Riemann integrable function that vanishes outside some bounded region. Based on (47) and Lemma 2, we get for the set $\{\xi_{\underline{i},\underline{j}}|\sum_{i=1}^N\sum_{j=1}^M\xi_{i,j}\leq T_l(K_l-r);\xi_{\underline{i},\underline{j}}\geq 0\}$ a set of functions, $g_{\text{rc}}(\underline{x})$, which are bounded. As a result we can upper bound the ensemble average decoding error probability for this set by the expression from (34). For the set of events $\{\xi_{\underline{i},\underline{j}}|\sum_{i=1}^N\sum_{j=1}^M\xi_{i,j}>T_l(K_l-r);\xi_{\underline{i},\underline{j}}\geq 0\}$ we upper bound the ensemble average decoding error probability by 1. This bounds are the exact same bounds we used in order to average over the channel realizations on Theorem 4. Hence, by averaging over the channel realizations we get for the ensemble the same lower bound on the diversity order as in Theorem 4. This concludes the proof.

APPENDIX F
PROOF OF COROLLARY 5

The proof of this corollary relies heavily on Theorem 3. We begin by describing the L ensembles of IC's and how they are transmitted. Then we use averaging arguments in order to show that there exists a singe sequence of IC's that attains the optimal DMT.

We begin by considering a sequence of K_0T_0-complex dimensional IC's with multiplexing gain $r=0$, i.e. the transmitter density $\gamma_{tr}=1$ for any ρ. In a similar manner to Theorem 3, we first consider an ensemble of finite constellations drawn uniformly within $\text{cube}_{K_0T_0}(b)\subset\mathbb{C}^{K_0T_0}$. Each code-book contains $\lfloor\gamma_{\text{tr}}b^{2K_0T_0}\rfloor=\lfloor b^{2K_0T_0}\rfloor$ points, where each point is drawn uniformly within $cube_{K_0T_0}(b)$. Let us denote a certain finite constellation in the ensemble by $C_{FC}(\rho,K_0T_0,b)\subset\text{cube}_{K_0T_0}(b)$. We extend each finite constellation in the ensemble into an IC in a similar manner to (32)

$$IC(\rho,K_0T_0)=C_{FC}(\rho,K_0T_0,b)+(b+b^{'})\cdot\mathbb{Z}^{2K_0T_0}. \tag{81}$$

By choosing $b=\sqrt{\frac{K_0T_0}{\pi e}}\rho^{\frac{K_0T_0}{2}+2\epsilon}$ and $b^{'}=\sqrt{\frac{K_0T_0}{\pi e}}\rho^{\frac{K_0T_0}{2}+\epsilon}$, we get a sequence of ensembles of IC's with multiplexing gain $r=0$. For a certain channel realization $\underline{\eta}\geq 0$ we get in accordance with Theorem 3

$$\overline{Pe}(\rho,\underline{\eta},K_0T_0)\leq D(K_0T_0)\rho^{-T_0K_0+\sum_{i=1}^{K_0T_0}\eta_i} \tag{82}$$

where $\overline{Pe}(\rho,\underline{\eta},K_0T_0)$ is the average decoding error probability of the K_0T_0-complex dimensional ensemble of IC's. From Theorem 4 we know that by transmitting the ensemble of IC's over the transmission matrix G_0, and averaging over the channel realizations, we get diversity order $d_{K_0}=MN$. Transmitting over G_0 gives us a K_0T_0-complex dimensional ensemble of IC's within $\mathbb{C}^{MT_0}$.

Next we derive from the K_0T_0-complex dimensional ensemble of IC's, another K_lT_l-complex dimensional ensemble of IC's, where $l=1,\ldots,L-1$. For each IC, $IC(\rho,K_0T_0)$, in the ensemble we take the first $\lfloor b^{2K_lT_l}\rfloor$ points in $C_{FC}(\rho,K_0T_0,b)$. We take the components of these points inside $\text{cube}_{K_lT_l}(b)$, and denote this new finite constellation as $C_{FC}(\rho,K_lT_l,b)$. Then we replicate these points in a similar manner to (81). In this case we get a new K_lT_l-complex dimensional IC

$$IC(\rho,K_lT_l)=C_{FC}(\rho,K_lT_l,b)+(b+b^{'})\cdot\mathbb{Z}^{2K_lT_l}. \tag{83}$$

By doing it to each IC in the ensemble, we get a new K_lT_l-complex dimensional ensemble of IC's. This new ensemble is equivalent to ensemble of IC's generated by drawing uniformly $\lfloor b^{2K_lT_l}\rfloor$ points inside $\text{cube}_{K_lT_l}(b)$, and then replicate these points according to $(b+b^{'})\mathbb{Z}^{2K_lT_l}$. Each IC sequence in this ensemble has multiplexing gain $r=0$. Since $b>\sqrt{\frac{K_lT_l}{\pi e}}\rho^{\frac{K_lT_l}{2}+2\epsilon}$ and $b^{'}>\sqrt{\frac{K_lT_l}{\pi e}}\rho^{\frac{K_lT_l}{2}+\epsilon}$, we get in accordance with Theorem 3 that for a certain channel realization $\underline{\eta}\geq 0$

$$\overline{Pe}(\rho,\underline{\eta},K_lT_l)\leq D(K_lT_l)\rho^{-T_lK_l+\sum_{i=1}^{K_lT_l}\eta_i} \tag{84}$$

where $\overline{Pe}(\rho,\underline{\eta},K_lT_l)$ is the average decoding error probability of the K_lT_l-complex dimensional ensemble of IC's. By transmitting this ensemble of IC's on the transmission matrix G_l, and averaging over the channel realizations, we get diversity order $d_{K_l}=(M-l)(N-l)+l(N+M-2\cdot l-1)$. Transmitting over G_l gives us a K_lT_l-complex dimensional ensemble of IC's within $\mathbb{C}^{MT_l}$.

From the sequential structure of the transmission scheme we get that omitting the $2\cdot l$ rightmost columns of G_0 yields G_l. Hence we can derive from the K_0T_0-complex dimensional ensemble of IC's, that attains diversity order d_{K_0}, another K_lT_l-complex dimensional ensemble of IC's the attains diversity order d_{K_l}, where $l=1,\ldots,L-1$. We attain it by diluting the points of each K_0T_0-complex dimensional IC in the ensemble in the aforementioned manner, and then reducing its dimensionality by dropping the $2 \cdot l$ rightmost columns of G_0.

So far we have shown the connection between the ensembles. Now we would like to show that there exists a certain sequence of K_0T_0-complex dimensional IC's, that gives us the desired diversity orders by diluting its points and adapting its dimensionality. We denote the average decoding error probability of the K_lT_l-complex dimensional ensemble of IC's by $A_l(\rho)\rho^{-d_{K_l}}$, where $\lim_{\rho\to\infty} \frac{\log(A_l(\rho))}{\log(\rho)} = 0$. We also define $I_{l,\rho}$ as the event where a K_lT_l-complex dimensional IC in the ensemble has average decoding error probability which is smaller or equal to $(L+1)A_l(\rho)\rho^{-d_{K_l}}$, where $l = 0, \dots, L-1$. From averaging arguments we know that $Pr(I_{l,\rho}) \geq \frac{L}{L+1}$. We wish to show that the probability of the event $\{I_{0,\rho} \cap I_{1,\rho} \cdots \cap I_{L-1,\rho}\}$ is bounded away from zero. From averaging arguments we know that

$$Pr(I_{0,\rho} \cap I_{1,\rho} \cdots \cap I_{L-1,\rho}) \geq 1 - \sum_{i=0}^{L-1} Pr(I_{i,\rho}) \geq \frac{1}{L+1}.$$

Hence there must exist a sequence of K_0T_0-complex dimensional IC's that attains diversity order d_{K_0} and has multiplexing gain $r = 0$, from which we can derive for each $l = 1, \dots, L-1$, a sequence of K_lT_l-complex dimensional IC's with multiplexing gain $r = 0$ and diversity order d_{K_l}.

Next we show that these L sequences attain the optimal DMT. Consider a sequence of K_lT_l-complex dimensional IC's, that has multiplexing gain $r = 0$ and attains diversity order d_{K_l}. From Corollary 3 we know that scaling this sequence by a scalar $\rho^{-\frac{r}{2K_l}}$ yields a new sequence of IC's with multiplexing gain r and diversity order

$$d_{K_l}(r) = (M-l)(N-l) - (r-l)(N+M-2\cdot l-1)$$

where $0 \leq r \leq K_l$ and $l = 0, \dots, L-1$. Each of the L straight lines $d_{K_l}(r)$, $l = 0, \dots, L-1$, coincides with a different segment out of the L segments of the optimal DMT. This concludes the proof.