A Note on the Dispersion of Network Problems

Eli Haim
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: elih@eng.tau.ac.il

Yuval Kochman*
School of CSE, HUJI
Jerusalem, Israel
Email: yuvalko@cs.huji.ac.il

Uri Erez†
Dept. of EE-Systems, TAU
Tel Aviv, Israel
Email: uri@eng.tau.ac.il

***Abstract*—Recently there has been significant interest in the analysis of finite-blocklength performance in different settings. Specifically, there is an effort to extend the performance bounds, as well as the Gaussian approximation (dispersion) beyond point-to-point settings. This proves to be a difficult task, as the performance may be governed by multiple dependent constraints. In this work we shed light on these difficulties, using the multiple-access channel as a test case. We show that a local notion of dispersion is more informative than that of dispersion regions sought after thus far. On the positive side, we show that for channels posessing certain symmetry, the dispersion problem reduces to the single-user one. Furthermore, for such channels, linear codes enable to translate single-user achievability bounds to the multiple-access channel.**

I. Introduction

The work of Polyanskiy et al. [1] revived the interest in analyzing the optimal performance subject to blocklength constraints. The work considers the point-to-point channel coding problem, and the results can be roughly divided into two parts. First, upper and lower bounds are given on the optimal codebook rate R as a function of the average probability of error[1] P_e and the blocklength n. Then, for asymptotic analysis, P_e is taken to be some constant ϵ while n is taken to infinity. It turns out that both the upper and lower bounds have the same asymptotic behavior, giving rise to: [2]

$$R = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right). \tag{1}$$

In this celebrated expression, $Q^{-1}(\cdot)$ is the inverse of the complementary Gaussian CDF, C is the channel capacity, and V is the channel *dispersion*, given by the variance of the information density:

$$i(x;y) \triangleq \log \frac{P_{Y|X}(y|x)}{P_Y(y)}. \tag{2}$$

with respect to P_X, the channel capacity-achieving distribution, and $P_{Y|X}$, the channel law. Here and in the sequel, P_X is assumed to be unique, and the resulting dispersion positive.[2] Roughly, this result can be understood as follows. Define J as the event that $i(X,Y)$, averaged over the transmission block, is below R. To the approximation level needed, a decoding error event amounts to the event J. Now define G to be the event that a Gaussian variable Z with mean $C - R$ and variance V/n takes a negative value. By the central-limit theorem (CLT), the probabilities of G and J are approximately equal. Equation (1) gives the probability of J, with the correction term corresponding to the transitions from the true probability of a decoding error, to that of J and then of G.

The elegance of both the results and the derivation have motivated many to work on extensions and refinements. Among these works, we note those that aim to extend the results to network settings, and in particular to the multiple-access (MAC) channel [3]–[5]. Although these works differ, we may roughly describe the basic approach as follows. Consider a two-user MAC channel, and define the random vector formed by the information densities

$$\begin{bmatrix} i_1 \\ i_2 \\ i_{1,2} \end{bmatrix} \triangleq \begin{bmatrix} i(X_1;Y|X_2) \\ i(X_2;Y|X_1) \\ i(X_1,X_2;Y) \end{bmatrix}, \tag{3}$$

averaged over the transmission block. Here, the conditional information densities are the obvious extensions of (2). Now define J_1, J_2 and $J_{1,2}$ as the events that the respective averaged densities fall below the corresponding element of rate vector

$$\tilde{\mathbf{R}} \triangleq \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix}.$$

It is shown, that (up to the correction term),

$$\max\left\{\mathbb{P}(J_1), \mathbb{P}(J_2), \mathbb{P}(J_{1,2})\right\} \le P_e \le \mathbb{P}\left(\bigcup \{J_1, J_2, J_{1,2}\}\right). \tag{4}$$

Finally, a normal approximation is applied to the information density. That is, a Gaussian vector

$$\tilde{\mathbf{Z}} \triangleq \begin{bmatrix} Z_1 \\ Z_2 \\ Z_{1,2} \end{bmatrix} \tag{5}$$

* This work was supported in part by the Israel Science Foundation under grant 956/12.

† This work was supported in part by the U.S. - Israel Binational Science Foundation under grant 2008/455.

[1] Bounds are also given for maximal probability of error, and in particular the bounds we use in this work, as well as the Gaussian approximation, are valid under both definitions. However, except for MAC channels with triangular capacity region which is achievable without time sharing (see Section III), we only consider average error probability, as it yields more tractable results in network settings.

[2] Detailed treatment is given in [1] to other cases, but the main ideas of this work are better understood in the more regular case considered.

is defined, with mean

$$\begin{bmatrix} I(X_1;Y|X_2) \\ I(X_2;Y|X_1) \\ I(X_1,X_2;Y) \end{bmatrix} - \tilde{\mathbf{R}} \tag{6}$$

and a covariance matrix that is equal to the covariance of the information density (3). Let G_1, G_2 and $G_{1,2}$ be the events that the corresponding element of $\mathbf{z}/\sqrt{n}$ falls below zero. Then, up to correction terms, the events J_1, J_2 and $J_{1,2}$ in (4) can be replaced by G_1, G_2 and $G_{1,2}$.

While this analysis is valid, we show in this work that it has two important limitations. First, using the Gaussian statistics of $\tilde{\mathbf{Z}}$ to define global "dispersion regions" is problematic, since it does not capture the nature of convergence to the boundary of the capacity region. And second, proving tight finite-blocklength (non-asymptotic) achievable bounds turns out to be a very difficult task. We demonstrate this using two important special cases:

1) Orthogonal MAC channels. Here, there are two independent single-user channels, coupled only via the joint error-probability constraint. Thus, in principle, any single-user performance bound directly translates to a bound on the MAC rate region. We show that even in this case, a *local* approach is more appropriate for describing the asymptotics.
2) Symmetric MAC channels. In this case (an extension of a symmetric single-user channel, to be defined precisely in the sequel) the capacity region is triangular and the dispersion analysis reduces to a single-user one. We show, that the reduction goes beyond asymptotics, and that using *linear codes*, single-user finite-blocklength achievability bounds can be utilized.

Sections II, III are devoted to these cases. Then in Section IV, we derive the dispersion of a general MAC channel.

II. Rectangular Rate Region: The Difficulty of Defining Global Dispersion

In this section we demonstrate the inherent problematic in deriving meaningful "dispersion regions". We do so using the very degenerate case of an orthogonal MAC channel, where the decoder observes the outputs of two independent point-to-point channels. We further assume for simplicity that these two channels are identical. That is, $Y = (Y_1, Y_2)$ and

$$P_{Y|X_1,X_2}(y_1, y_2|x_1, x_2) = P_{Y_1|X_1}(y_1|x_1)P_{Y_2|X_2}(y_2|x_2)$$

for all x_1, x_2, y_1, y_2. In this setting, the two independent point-to-point channels are coupled only through the definition of the MAC error event: if at least one of the messages was decoded incorrectly, we declare an error.

Let $\epsilon(n, R)$ be the minimum error probability for the single-user channel, using a code of blocklength n and rate R. Since the error events of the sub-channels are independent, we have that a rate-pair (R_1, R_2) is achievable at blocklength n if and only if:

$$\epsilon(n, R_1) \vee \epsilon(n, R_2) \le \epsilon, \tag{7}$$

where

$$\epsilon_1 \vee \epsilon_2 \stackrel{\triangle}{=} \epsilon_1 + \epsilon_2 - \epsilon_1\epsilon_2$$

is the probability of the union of independent events. Thus, the finite-blocklength performance of this MAC channel is completely dictated by the single-user performance $\epsilon(n, R)$, i.e., the right hand side of (4) holds with equality. We concentrate, then, on translating (7) into an asymptotic (dispersion) expression.

In this simple MAC channel, since the error events of the users are independent, the asymptotic analysis greatly simplifies. The events J_1 and J_2 are just single-user events, and their union subsumes the event $J_{1,2}$. Thus, in the Gaussian approximation as well, it is sufficient to take two independent Gaussian variables Z_1 and Z_2, where Z_i has mean $C - R_i$ and variance V/n. Then, the approximated rate region is given by the rate pairs for which $\mathbb{P}(G_1 \cup G_2) \le \epsilon$, yielding the rate region:

$$\begin{bmatrix} C - R_1 \\ C - R_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} O\left(\frac{\log n}{n}\right) \in \sqrt{\frac{V}{n}}\Sigma(\epsilon), \tag{8}$$

where

$$\Sigma(\epsilon) = \left\{ \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} : Q(t_1) \vee Q(t_2) \le \epsilon \right\}.$$

In order to show the explicit dependence upon V, we use t which is a standard Gaussian, rather than z. The achievable rate-regions in [3]–[5] all reduce to (8) in this degenerate case.

At first glance, the expression (8) seems quite pleasing, and indeed resembles (1), where the border of the two-dimensional region $\Sigma(\epsilon)$ plays the role of a vector extension of the inverse-Q function. However, the region contracts in both axes simultaneously. That is, for any choice of t_1, t_2 on the face of $\Sigma(\epsilon)$, the trajectory (as a function of n) of the corresponding point on the face of (8) always tends to the origin, i.e., both rates approach capacity simultaneously. This raises two questions. First, maybe we can describe the convergence to the origin in a simpler manner, without resorting to the rather complicated region $\Sigma(\epsilon)$? And second, what happens if we wish to approach a different point on the face of the capacity region, where one of the users operates below capacity even in the infinite-blocklength limit? In order to answer these questions, we take a different, localized approach.[3]

Without loss of generality, assume that we wish to approach a rate-pair $(R_1, R_2) = (C, R^*)$, where $0 \le R^* \le C$. We can describe the trajectory in a parametric way: $(R_1, R_2) = (R, f(R))$, where $f(R)$ is some function satisfying $f(C) = R^*$. In order to avoid complications, we only take smooth trajectories; thus we assume that $f(\cdot)$ is continuously differentiable. We further denote the asymptotic slope of approach by

$$\lambda = \left.\frac{df(R)}{dR}\right|_{R=C}.$$

[3]A somewhat similar analysis is performed by Tan and Kosut [5] for a more general case (in the context of the Slepian-Wolf problem), in addition to deriving the global region. However, Tan and Kosut do not draw from this analysis the conclusions that we reach here.

Then for each such trajectory, we can derive a scalar expression for the asymptotic behavior.

Proposition 1 (Dispersion of the orthogonal MAC channel): When approaching a rate-pair (C, R^*) with asymptotic slope λ, the dispersion is given by the following.

1) If $R^* < C$ then R_1 follows the single-user dispersion (1).
2) If $R^* = C$ and $\lambda = 0$ then
$$R_1 = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right).$$
(for infinite λ we have the obvious dual)
3) If $R^* = C$ and $\lambda \notin \{0, \pm\infty\}$ then R_1 and R_2 follow (1) with ϵ replaced by ϵ_1, ϵ_2 respectively. These probabilities are given by the solution to:
$$\begin{aligned} \epsilon_1 \vee \epsilon_2 &= \epsilon \\ \frac{Q^{-1}(\epsilon_1)}{Q^{-1}(\epsilon_2)} &= \lambda. \end{aligned} \tag{9}$$

Before outlining the proof, the following remarks shed light on this result.

Remark 1: The key to this result, is understanding the behavior of the error probability of the second user, along the trajectory, see Figure 1. For $R^* < C$, the second user is in the large-deviations regime, where its error probability decays to zero fast, and $\epsilon_1 \approx \epsilon$. Approaching the corner point with a finite positive slope, the error probability is split between the users. Thus, they are both in the CLT regime with $\epsilon_1, \epsilon_2 < \epsilon$. In the intermediate case, where the corner is approached with a trajectory tangent to the axis, the error probability of the second user decays to zero, but may do so slowly.

Remark 2: Unlike (8) which exhibits a rounded dispersion region, Proposition 1 gives straight lines with a singularity at the corner points. Technically speaking, both are valid asymptotic approximations of the region (7). The difference between the rounded region and the straight lines is of a lower order than the correction term (non-uniformly around the corner points), a fact that is hidden by the rather implicit form of $\Sigma(\epsilon)$.

Remark 3: The analysis above holds, with obvious modifications, to orthogonal MAC channels where the channels differ. For example, the second equation in (9) is multiplied by the square root of the ratio of dispersions. Clearly, it can be extended to more than two users as well.

Remark 4: As in the single-user case, the analysis also holds for $\epsilon > 1/2$. The approach to non-corner points will always be from the outside. As for the corner point, if $\epsilon > 1 - \sqrt{2}/2$ then necessarily at least one of ϵ_1, ϵ_2 is greater than $1/2$. Thus, the approach is from the outside. For intermediate values of ϵ, it may be from the inside or outside, depending on the slope λ.

Remark 5: One may claim that approaching non-corner points is not interesting, as they are all dominated by the optimum (C, C). However, the approach presented here extends to general MAC channels, where the capacity region may have a sum-rate face; see Section IV in the sequel.

Remark 6: Approaching the corner point with a slope λ has an interesting engineering interpretation: it is the result of optimizing a linear combination of the rates, $R_1 + \lambda R_2$.

Remark 7: It should be noted that the straight-line approximation only holds in the asymptotic (dispersion) sense. For any finite blocklength n, there is indeed a tradeoff between the rates.

Proof outline: For the case $R^* < C$, consider the error probability of the second user, $P_{e2,n}$. By the smoothness of the trajectory,
$$\lim_{n\to\infty} \frac{-\log p_{e2,n}}{n} \geq \liminf_{n\to\infty} \frac{-\log \epsilon(n, R')}{n}. \tag{10}$$
for some $R' < C$. By large-deviation analysis on the information density, one may show that this quantity is positive (indeed, such analysis yields the random-coding exponent). In other words, when the second user does not try to approach capacity, large-deviations analysis applies to that user, and $P_{e2,n}$ decays exponentially. Substituting back in (7), we see that any R satisfying
$$\epsilon(n, R) = \epsilon - \delta(n)$$
is achievable, where $\delta(n)$ is exponentially small. Then it can be shown that such a small perturbation on ϵ does not change the dispersion behavior, i.e., it inflicts a rate penalty that is below the correction term in (1). Consequently, for any $R^* < C$, the dispersion is equal to the single-user dispersion.

For the approach to the corner point $R^* = C$, any trajectory with $0 < \lambda < \infty$ is tangent to the straight trajectory corresponding to taking the corresponding (ϵ_1, ϵ_2) in $\Sigma(\epsilon)$ as defined via (9). The difference between the dispersion along these tangent trajectories is (by the smoothness assumption) only of order $1/n$. Thus, the dispersion is given by (1), with the error probability given by (9).

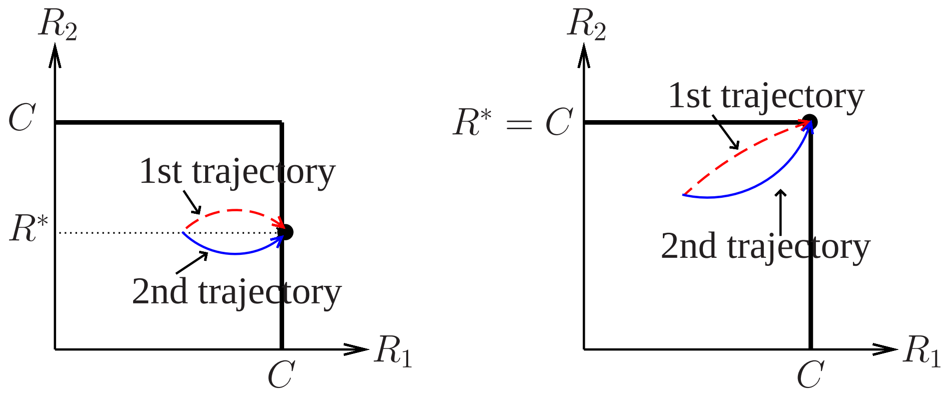
Approaching the corner point with $\lambda = 0$ or $\lambda = \infty$ lies somewhere between these cases. Delicate analysis is needed, taking into account the exact trajectory. In some cases, the second user is in the moderate-deviations regime [6]. ■

III. Triangular Rate Region: Equivalence to Single User

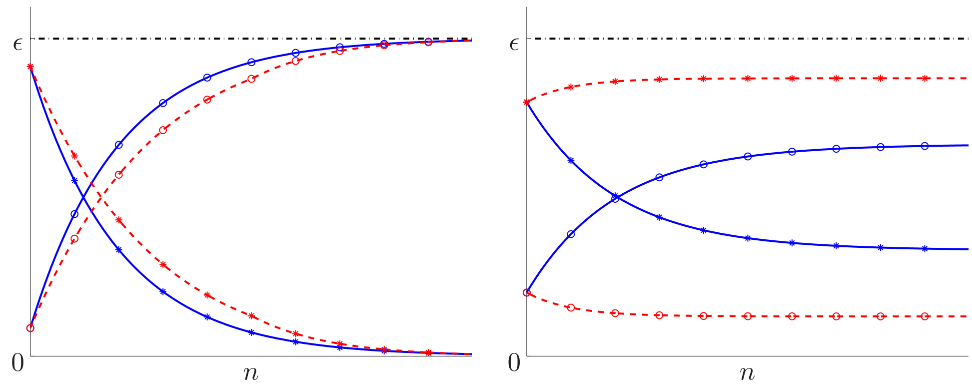
Whereas in the previous section the error events of the users were statistically independent, in this section we discuss the other extreme case, where one of the error events dominates the others.

A. Achievable Dispersion for Triangular Rate Regions

Consider a MAC channel $P_{Y|X_1,X_2}$ with input alphabets $\mathcal{X}_1, \mathcal{X}_2$ and output alphabet $\mathcal{Y}$. Further, consider an input distribution $P_{X_1}(x_1) \cdot P_{X_2}(x_2)$ such that the capacity region under this distribution is triangular. This implies that $I(Y; X_1) = I(Y; X_2) = 0$. Therefore
$$P_{Y|X_1}(y|x_1) = P_{Y|X_2}(y|x_2) = P_Y(y) \tag{11}$$



(a) $R^* < C$ (b) $R^* = C$



(c) P_{e1}, P_{e2} along the two trajectories in Figure 1(a). (d) P_{e1}, P_{e2} along the two trajectories in Figure 1(b).

Fig. 1. Schematic view of convergence along different trajectories. Figures 1(a) and 1(b) show the capacity region of an orthogonal MAC channel, and each sub-figure shows two different trajectories, starting from the same rate-pair and approaching the same point (C, R^*). Figures 1(c) and 1(d) compare P_{e1} (circles) and P_{e2} (asterisk) along the two trajectories in Figures 1(a) and 1(b) respectively

for any $x_1 \in \mathcal{X}_1, x_2 \in \mathcal{X}_2, y \in \mathcal{Y}$. Now consider the information-density vector (3). In the present case, all the elements are equal. Then, since the rate threshold for the last event $J_{1,2}$ is higher than the ones in J_1 and J_2, we can always ignore the other events, as they are subsets of the last event, and therefore dispersion is a scalar quantity, similar to the single-user case (1).

In particular, If this input distribution achieves the capacity region of this MAC channel (i.e., the capacity region is triangular and time sharing is not required), then an achievable dispersion of a MAC channel is given by

$$R_1 + R_2 = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right). \qquad (12)$$

In this expression, C is the maximal sum-rate of the channel, and V is the variance of $i_{1,2}$. This agrees with the expressions of [3], [5].[4]

B. Dispersion of Linear-Symmetric Channels

We start with formally defining some special classes of MAC channels. Then we show that these definitions lead to a triangular capacity region, achievable by a single input distribution (i.e., time sharing is not required). Moreover, we show that these channels are equivalent (in terms of error probability) to a single-user channel, and hence the dispersion is known.

Definition 1: Consider a MAC channel $P_{Y|X_1,X_2}$. For any random variable X, where the following Markov chain

$$(X_1, X_2) \leftrightarrow X \leftrightarrow Y \qquad (13)$$

holds, the channel $P_{Y|X}$, is called the *associated single-user channel* of the MAC channel.

Notice that when comparing the MAC channel to its associated single-user one,

$$\epsilon_{\mathrm{MAC}}(n, R_1, R_2) \geq \epsilon_{\mathrm{SU}}(n, R_1 + R_2), \qquad (14)$$

since the single-user channel can emulate the MAC channel. This gives outer bounds on the dispersion; we now concentrate on a case where we can obtain a meaningful bound.

In order to define a symmetric MAC channel, we first recall the definition of a symmetric single-user DMC:

Definition 2 ([7]): A DMC is defined to be symmetric if its probability transition matrix (using inputs as rows and outputs of the subset as columns) can be divided column-wise into sub-matrices, such that for each sub-matrix, the rows are

[4]The derivation in [4] uses a union bound to simplify the analysis around corner points; this bound cannot use the special structure of this channel, where the region has no corner points.

equal up to permutations, and the columns are equal up to permutations.

Definition 3: A *symmetric* MAC channel is defined as a channel that has an associated single-user channel which is a symmetric DMC.

Definition 4: A MAC channel with input alphabet $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{F}$, where $\mathbb{F}$ is a finite field, is called *linear* if there exists X in (13) such that X is a linear combination of X_1 and X_2 over the field.

The class of linear and symmetric MAC channels includes the important case of modulo-additive channels over prime-alphabets, i.e., $Y = (X_1 + X_2 + N) \bmod m$, where the alphabets of the input and output are equal to $\{0, 1, \ldots, m-1\}$, and m is a prime.

For all linear and symmetric channels, the rate region is triangular, and it is achieved by a uniform distribution over the input of both users. Other channels with triangular region may also be transformed to this form. However, there exist channels with triangular rate region which do not belong to this class, e.g., modulo-additive channels over non-prime alphabet.

We conclude this subsection by showing that the right hand side of (12) is also an upper bound on the sum-rate, thus characterizing the dispersion of a linear and symmetric MAC channels. This can be seen by the symmetric DMC $P_{Y|X}$ with X equal to the linear combination of X_1 and X_2 over a field: on one hand $i(x_1, x_2; y) = i(x; y)$ for all x_1, x_2, y, and on the other hand the dispersion of the single-user channel is known.

C. Finite-Blocklength Achievability Bounds

Seeing that (12) amounts to single-user dispersion, one hopes that single-user finite-blocklength achievability bounds can also be translated to this special class of MAC channels. However, this does *not* follow from the analysis above, as it requires a stronger sense of equivalence to a single-user channel. In the sequel we do show that the following bound can be extended to MAC channels that are linear and symmetric. It is known as the "dependence testing" (DT) bound, and the interested user is referred to [1].

Proposition 2 (DT bound for single user DMC [1], Th.18): Consider a DMC $P_{Y|X}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$. For any distribution P_X on $\mathcal{X}$, there exists a code with blocklength n and average probability of error:

$$\begin{aligned} P_e \le & \Pr\left[n \cdot i(X;Y) \le \log \frac{M-1}{2}\right] \\ & + \frac{M-1}{2} \Pr\left[n \cdot i(X;\bar{Y}) > \log \frac{M-1}{2}\right], \end{aligned} \tag{15}$$

where $M = \lfloor \exp\{nR\} \rfloor$ is the codebook size, and $\bar{Y}$ is statistically independent of (X, Y), with $P_{\bar{Y}}(y) = P_Y(y)$.

In the context of symmetric point-to-point channels, *linear codes* achieve capacity [8], as well as the channel dispersion and the DT bound [1]. We define a linear code $\mathcal{C}$ (over a finite field $\mathbb{F}$) via a $k \times n$ full-rank generating matrix G, by

$$\mathcal{C} = \{\mathbf{c} : \mathbf{c} = \mathbf{u}G,\ \mathbf{u} \in \mathbb{F}^k\}, \tag{16}$$

The codebook size is equal to $M = |\mathbb{F}|^k$ (the rate is equal to $R = {}^{k}\!/_{n} \cdot \log|\mathbb{F}|$). Clearly, every rate is possible asymptotically as $n \to \infty$.

A codebook pair for a linear MAC channel can be generated from a linear code for its associated single-user channel, by splitting the generating matrix into two sub-matrices (one for each user). The sum of codewords is indistinguishable from a codeword of the single-user code with $R = R_1 + R_2$. This leads to the following proposition.

Proposition 3 ([9]): Consider a linear MAC channel. Any error probability which is achievable by the associated single-user channel using a linear code, is also achievable for the MAC channel.

For any linear and symmetric MAC channel, the capacity region is triangular, and it is achieved by a uniform input distribution. Using linear codes and by Proposition 3, we can obtain achievable bounds for finite blocklength.

Corollary 1: For any linear and symmetric MAC channel, linear codes achieve the DT bound (15) with $R = R_1 + R_2$.

We now compare this result to previous MAC finite-blocklength achievability bounds. Existing literature includes the following.

1) In [10], Slepian and Wolf show that for a given input distribution and an achievable rate pair with this distribution, the MAC error probability is upper-bounded by the sum of three terms, each decaying exponentially with n. Applying this to triangular MAC channels, one of these exponents corresponds to the "right" random-coding error exponent of the single-user channel, but the others add to the error probability and prevent the bound from being tight, unless staying far from the face of the capacity region.
2) In [3], [5], the summation problem is avoided by considering the different error events jointly. Although no explicit bounds for finite blocklength are given, one may obtain such bounds by carefully considering the correction terms in the proofs of the asymptotic regions. However, these works rely on the method of types, yielding large finite-blocklength penalties compared to the DT bound.
3) In [4], the authors use a simplified approach, which allows to apply the DT bound. However, the simplification also implies that correction terms of all MAC error events are summed.

It is possible, that with some more efforts these obstacles can be overcome, and the achievability of the single-user DT bound be also shown using the results of [3]–[5]. However, we can point out two main advantages of the linear-codes approach, beyond its simplicity. First, *any* single-user achievable rate using linear codes applies to the MAC channel. Indeed, we do not believe that the DT bound is tight. And second, in a linear code the average error probability also equals the *maximal* probability of error.

We conclude by pointing out that the linear-codes approach allows to derive finite-blocklength achievability bounds also for general MAC channels. We can use a transformation of

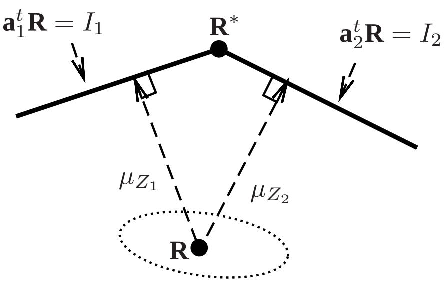


Fig. 2. For a K-user MAC channel, an illustration of approaching the point R^* which lies on the intersection of two faces of an achievable rate region.

the channel into a linear and symmetric MAC channel, as described in [9]. While this transformation is lossy in terms of capacity, and thus is necessarily sub-optimal in terms of asymptotics (dispersion), it may still be better than other bounds for some finite blocklengths, when the channel is "almost" linear and symmetric.

IV. The Dispersion of a General MAC Channel

Consider a two-user MAC channel. For any given fixed distribution, the dispersion is set by three events G_1, G_2 and $G_{1,2}$, corresponding to the statistics of the corresponding information densities. In the previous two sections, for a fixed input distribution, the dispersion of a MAC channel when approaching a point on the boundary of the achievable region was shown to have a similar asymptotic behavior as (1), for any trajectory that is not tangent to the faces of the capacity region. In Section IV-A we show that this behavior also extends to general MAC channels. Then, in Section IV-B we consider the Slepian-Wolf problem, where dual results can be obtained, and also meaningful outer bounds on the dispersion can be given. Finally in Section IV-C we go back to the MAC channel, and discuss the problem of finding the optimal dispersion when time-sharing is taken into account.

A. Achievable Dispersion Using a Single Input Distribution

In this section we consider a fixed input distribution $P_{X_1}(x_1) \cdot P_{X_2}(x_2)$. Recalling (4), the inner bound on the finite-blocklength region is given by the rate pairs such that the probability that one of the information densities i_1, i_2 and $i_{1,2}$ is below R_1, R_2 and $R_{1,2}$, respectively, is below ϵ. We apply now a local asymptotic analysis, similar to that presented in the previous sections for special cases. We prefer to present the problem in more general terms, that allow to see that the form of the approach is not attached necessarily to the two-user MAC problem, but can also be applied to other network problems.

We start with a K-dimensional rate region, defined by linear inequalities, such that the surface is made of faces (hyperplanes). These faces are given by

$$\bar{A}\mathbf{R} \leq \bar{\mathbf{I}}. \tag{17}$$

Here, $\mathbf{R} = [R_1, \ldots, R_K]^T$, $\bar{A}$ is an $\bar{L} \times K$ matrix, where each row corresponds to a constraint (thus, $\bar{\mathbf{I}}$ gives the values of the constraints). In the context of the two-user MAC channel, $K = 2$, $\bar{L} = 3$ and

$$\bar{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

These $\bar{L}$ constraints correspond to random variables $i_1, \ldots, i_{\bar{L}}$ and error events $J_1, \ldots, J_{\bar{L}}$, similar to those defined the Introduction. Let the covariance matrix of these variables be $\bar{V}$. Similar to (4), an achievable finite-blocklength rate region is given by:

$$\left\{ \mathbf{R} : \mathbb{P}\left(\bigcup_{l=1,\ldots,L} J_l \right) \leq \epsilon \right\}. \tag{18}$$

One could reformulate the Gaussian approximation for this general presentation. However, we choose to do so directly for the localized analysis. To this end, suppose that we wish to approach a point $\mathbf{R}^*$ on that surface; for such a point, some $L \leq K$ constraints are active. Since, as we saw, to the dispersion approximation, other constraints do not matter (they only "take" an exponentially small error probability from the active ones), we concentrate on the active ones only. Therefore, define A to be the sub-matrix of $\bar{A}$ by taking only the L rows, which correspond to the active constraints, and define $\mathbf{I}$ as the sub-vector of $\bar{\mathbf{I}}$ which corresponds to the active constraints. Then the following equation holds:

$$A\mathbf{R}^* = \mathbf{I}. \tag{19}$$

Note that this equation does not uniquely determine $\mathbf{R}^*$, but rather characterizes the hyperplane determined by the intersection of the constraints. We also define a normalized version of A:

$$B = \left[\frac{\mathbf{a}_1^T}{\|\mathbf{a}_1\|} \middle| \cdots \middle| \frac{\mathbf{a}_L^T}{\|\mathbf{a}_L\|} \right]^T \tag{20}$$

We are interested in approaching $\mathbf{R}^*$ in the direction defined by the K-dimensional unit vector $\mathbf{e}$. That is, we approach $\mathbf{R}^*$ in such a way that:[5]

$$\frac{\mathbf{R}^* - \mathbf{R}}{\|\mathbf{R}^* - \mathbf{R}\|} = \mathbf{e}. \tag{21}$$

We assume that $\mathbf{e}$ is not in the direction of any of the active constraints.[6]

For the Gaussian approximation, define the random vector $\mathbf{Z} = [Z_1, \ldots, Z_L]^T$, which has mean

$$\mu_{\mathbf{Z}} = \mathbf{I} - A\mathbf{R} = A(\mathbf{R}^* - \mathbf{R}) \tag{22}$$

and covariance matrix V that equals the sub-matrix of $\bar{V}$ corresponding to the active constraints. In these terms, we can formulate the local dispersion (see Figure 2).

Proposition 4: Consider an achievable finite-blocklength rate region given by (18). Then, asymptotic achievable rates

[5] The actual trajectory does not have to be a straight line, and the results hold for any continuously-differentiable trajectory which is tangent to $\mathbf{e}$ at $\mathbf{R}^*$, as in Section II.

[6] Otherwise, more careful analysis in needed, see the second case of Proposition 1.

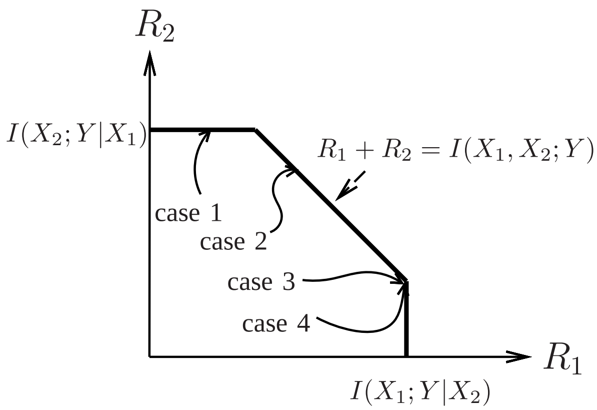


Fig. 3. Different cases which are considered in Section IV.

when approaching $\mathbf{R}^*$ in the direction $\mathbf{e}$ are given by:

$$A\mathbf{R} = \mathbf{I} - \frac{\mu_{\mathbf{Z}}}{\sqrt{n}} + \mathbf{1}\cdot O\left(\frac{\log n}{n}\right). \tag{23}$$

The vector $\mu_{\mathbf{Z}}$ is the solution of the following two equations:

$$\frac{\mu_{\mathbf{Z}}}{\|\mu_{\mathbf{Z}}\|} = B\mathbf{e} \tag{24}$$

where B is given by (20), and

$$\mathbb{P}\left(\bigcup_{l=1,\dots,L}\{Z_l \le 0\}\right) = \epsilon. \tag{25}$$

The first equation gives $(L-1)$ constraints, defining the direction of $\mu_{\mathbf{Z}}$. It is derived by multiplying both sides of (21) by A, substituting (22) and re-normalizing. The second equation gives the missing constraint, setting the length of $\mu_{\mathbf{Z}}$ using the Gaussian statistics of $\mathbf{Z}$.

An important case is when $L=1$, i.e., when approaching a point where only one constraint is active ("non-corner" point). In that case, (23) reduces to scalar dispersion, and

$$\mu_Z = \sqrt{V}Q^{-1}(\epsilon),$$

regardless of the direction $\mathbf{e}$. For other points, if one is interested in the scalar dispersion of a specific rate, or of a linear combination, it can be derived using linear operations on (23). Rates or combinations that are not determined this way, have zero dispersion (they correspond to inactive constraints at $\mathbf{R}^*$, thus they can be approached arbitrarily fast, without affecting the other rates).

We now go back to the specific problem at hand, the two-user MAC channel. Further, we consider the generic case where the rate-region for the chosen input distribution is a pentagon. We use the representation above in order to find the asymptotic approach to different cases, as shown in Figure 3. We use here the notation I_1, I_2 and $I_{1,2}$ for $I(X_1;Y|X_2)$, $I(X_2;Y|X_1)$ and $I(X_1,X_2;Y)$, respectively. Also, i_1, i_2 and $i_{1,2}$ denote the corresponding information densities. Due to symmetry, we can restrict attention to the following cases.

1) Approaching (R_1^*, I_2) for $R_1^* < I_1$. The only active constraint is given by $A=[0\ 1]$, resulting in scalar dispersion, i.e., R_2 behaves according to (1), substituting C, R and V with I_2, R_2 and the variance of i_2, respectively. Note that this is slope-independent, as long as the approaching trajectory is not on the face. This is identical to the behavior of non-corner points of a rectangular region, presented in Section II.
2) Approaching a non-vertex point on the diagonal face, i.e., (R_1^*, R_2^*) for $R_1^*+R_2^* = I_{1,2}$, $R_1^* < I_1$ and $R_2^* < I_2$. The only active constraint is given by $A=[1\ 1]$. Therefore, the dispersion is again scalar and slope-independent (unless parallel to an active constraint). It is given by (1), substituting C, R and V with $I_{1,2}$, R_1+R_2 and the variance of $i_{1,2}$, respectively. This is identical to the behavior in the case of a triangular rate-region, presented in Section III.
3) Approaching a vertex point $(I_1, I_{1,2})$ with slope $\mathbf{e}$ not proportional to $[0\ 1]^T$ or $[1\ -1]^T$. Here
$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$
Thus,
$$\frac{\mu_{\mathbf{Z}}}{\|\mu_{\mathbf{Z}}\|} = B\mathbf{e} = \begin{bmatrix} e_1 \\ \frac{e_1+e_2}{\sqrt{2}} \end{bmatrix}.$$
From here, one may derive the explicit solution; after projecting in the directions R_1 and R_2, it coincides with the scalar dispersions calculated in [5, Section V]. While in principal this analysis is similar to that applied to a corner point in Section II, there are two differences that make it more involved: there is statistical dependence between the dominating error events, and the faces meeting at the corner are not perpendicular.
4) Approaching a vertex point $(I(Y;X_1|X_2), I(Y;X_2))$ with slope $\mathbf{e}$ proportional to $[0\ 1]^T$. As discussed in Section II, this is an intermediate case between the ones discussed above. That is, we have that R_1 approaches I_1 with scalar dispersion according to the variance of i_1, but the correction term may be larger than in (1). A similar situation arises when the approach direction parallels the sum-rate face.

B. *Dispersion of the Slepian-Wolf Problem*

The Slepian-Wolf problem [11] is a source-coding setting, where two correlated sources should be (almost) losslessly conveyed to a single decoder. The rate region is dual to that of the MAC problem: each rate should be at least the entropy of the corresponding source conditioned on the other, and the sum of rates should be at least the joint entropy. However, unlike the MAC problem, there is no choice of distributions, thus no time-sharing is needed. This simplifies the analysis, and allows to also give a simple *outer bound* on the distortion region.

Indeed, define the entropy density vector as

$$\begin{bmatrix} h_1 \\ h_2 \\ h_{1,2} \end{bmatrix} \triangleq \begin{bmatrix} h(X_1|X_2) \\ h(X_2|X_1) \\ h(X_1,X_2) \end{bmatrix} = -\begin{bmatrix} \log P_{X_1|X_2}(x_1|x_2) \\ \log P_{X_2|X_1}(x_2|x_1) \\ \log P_{X_1,X_2}(x_1,x_2) \end{bmatrix}. \tag{26}$$

Further, define J_1, J_2 and $J_{1,2}$ as the events that, when averaged over a source block, $h_1 > R_1$, $h_2 > R_2$ and $h_{1,2} > R_1 + R_2$, respectively. Then, the error probability is bounded (to the asymptotic approximation of interest for dispersion) by (4). Since there is no choice of distributions, this leads directly to inner and outer bounds on the dispersion, when approaching a point (R_1, R_2) on the surface of the capacity region. The inner bound is given by Proposition 4, substituting information by entropy. It is left to derive the outer bound, and see when they coincide.

For the outer bound, we can follow the same strategy as in the derivation of Proposition 4. First we ignore all constraints not active in the target $\mathbf{R}^*$. Between the remaining L constraints, we need to find the limiting one. Since the expectation vector $\mu_{\mathbf{Z}}$ is proportional to $B\mathbf{e}$, and the Gaussian error probability is monotonously decreasing in the ratio between the expected value and the square root of the variance, the index of the limiting constraint is given by:

$$l^* = \arg\min_l \frac{[B\mathbf{e}]_l}{\sqrt{V_l}}. \tag{27}$$

Here, V_l is the variance of i_L. Then, an outer bound is given by the scalar dispersion (1) corresponding to the l^* constraint (in case of multiple minimizers, all of them give a valid bound).

Comparing to the inner bound, we have two distinct cases. For a non-corner point, as (to the required approximation) there is only one active constraint, the inner and the outer bounds coincide. For a vertex point, on the other hand, the outer bound is loose, as for jointly-Gaussian variables of positive variance, the probability of the union of error events is strictly higher than the maximal probability.

C. Discussion: Converse for General MAC Channels

We have thus far avoided the issue of time sharing between input distributions. In general, such time sharing may be needed in order to achieve the MAC capacity region itself. We now discuss the effect of time sharing on finite-blocklength performance.

First consider the class of symmetric and linear MAC channels, defined in Section III. For these channels, we showed that the dispersion is reduced to that of the associated single-user channel, which is optimal, by using a single (uniform) input distribution. We now note that capacity may also be achieved by means of time sharing. In the most extreme case, one can achieve the points $(0, C)$ and $(C, 0)$ by one of the users using an optimal point-to-point code, while the other user transmits a deterministic symbol. Time sharing amounts to splitting the block and reversing the roles. While this strategy is capacity achieving, it clearly suffers in terms of error probability. Namely, the blocklength is shortened in proportion to the fraction of the block allocated to each user. Indeed, if one calculates the achievable dispersion for this strategy using the expressions in [3], the region obtained is strictly sub-optimal. This is also the case for time sharing between other distributions.

As we see in the example above, time sharing comes at a price.[7] It seems plausible that whenever capacity may be achieved using a single input distribution, the optimal dispersion is also achieved that way, see [3, Example 1]. We conjecture that this is true; however, we do not see an immediate proof.

A very different situation arises, when time sharing is needed in order to achieve some points of the capacity region. Clearly, for asymptotically long blocklength, a non capacity-approaching strategy cannot be dispersion optimal. Thus, for these points, for achieving the optimal dispersion it is necessary to perform time-sharing between at least two codebooks. Typically, the working point for each of the distributions is a corner point. Therefore, for such MAC channels, the behavior at corner points plays a major role. Finding the dispersion-optimal strategy is beyond the scope of this work.

For finite blocklength, as opposed to asymptotics, the price of time sharing may be too high. That is, for short enough blocklength, a single input distribution will likely outperform any time-sharing strategy, even if it cannot achieve the capacity region.

ACKNOWLEDGEMENT

The authors thank Anatoly Khina for helpful comments on the manuscript.

REFERENCES

[1] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[2] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third Prague Conf. Information Theory*, 1962, pp. 679–723.

[3] Y. W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *Proc. Int. Symp. Info. Theory*, July 2012.

[4] E. MolavianJazi and J. N. Laneman, "Discrete memoryless multiple access channel in the finite blocklength regime," in *Proc. Int. Symp. Info. Theory*, July 2012.

[5] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Information Theory*, submitted. online available: http://arxiv.org/abs/1201.3901.

[6] Y. Altuğ and A. B. Wagner, "Moderate deviation analysis of channel coding: discrete memoryless case," in *Proc. Int. Symp. Info. Theory*, June 2012.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[8] R. L. Dobrushin, "Asymptotic optimality of group and systematic codes for some channels," *Theor. Probab. Appl.*, vol. 8, pp. 52–66, 1963.

[9] E. Haim, Y. Kochman, and U. Erez, "Distributed structure: Joint expurgation for the multiple-access channel," *IEEE Trans. Information Theory*, submitted, 2012. online available: http://arxiv.org/abs/1207.1345.

[10] D. Slepian and J. K. Wolf, "A coding theorem for multiple access channels with correlated sources," *Bell System Tech. J.*, vol. vol. 52, pp. 1037–1076, Sept. 1973.

[11] ——, "Noiseless coding of correlated information sources," *IEEE Trans. Information Theory*, vol. IT-19, pp. 471–480, July 1973.

[12] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, "Rate-splitting multiple access for discrete memoryless channels," *IEEE Trans. Information Theory*, vol. 47, no. 3, pp. 873–890, Mar. 2001.

[7] An alternative strategy to time-sharing is rate splitting [12]. However, we note that it cannot improve the dispersion performance, as it cannot outperform random coding with the induced input probabilities. Thus, rate splitting can only serve as a means to reduce complexity.