SIAM J. Matrix Anal. Appl.
Vol. 36, No. 3, pp. 1164–1179

A LOG-DET INEQUALITY FOR RANDOM MATRICES*

MEIK DÖRPINGHAUS†, NORBERT GAFFKE‡, LORENS A. IMHOF§, AND RUDOLF MATHAR¶

Abstract. We prove a new inequality for the expectation $E\left[\log\det\left(\mathbf{WQ}+\mathbf{I}\right)\right]$, where $\mathbf{Q}$ is a nonnegative definite matrix and $\mathbf{W}$ is a diagonal random matrix with identically distributed nonnegative diagonal entries. A sharp lower bound is obtained by substituting $\mathbf{Q}$ by the diagonal matrix of its eigenvalues $\mathbf{\Gamma}$. Conversely, if this inequality holds for all $\mathbf{Q}$ and $\mathbf{\Gamma}$, then the diagonal entries of $\mathbf{W}$ are necessarily identically distributed. From this general result, we derive related deterministic inequalities of Muirhead- and Rado-type. We also present some applications in information theory: We derive bounds on the capacity of parallel Gaussian fading channels with colored additive noise and bounds on the achievable rate of noncoherent Gaussian fading channels.

Key words. expectation inequality, rearrangement inequality, Muirhead-type inequality, Rado-type inequality, L-superadditive function, conditional entropy, channel capacity

AMS subject classifications. 15B52, 15A45, 26D15, 60E15, 94A15

DOI. 10.1137/140954647

1. Introduction.

In this paper, we present a lower bound on the expected value

$$E\left[\log\det\left(\mathbf{WQ}+\mathbf{I}\right)\right], \tag{1}$$

where $\mathbf{Q}$ is a deterministic nonnegative definite Hermitian $N\times N$ matrix, $\mathbf{W}$ is an $N\times N$ diagonal random matrix with identically distributed but not necessarily independent, nonnegative diagonal entries $w_1,\dots,w_N$, and $\mathbf{I}$ is the $N\times N$ identity matrix. Expectations of this form occur in information theory, e.g., in the analysis of the capacity of Gaussian channels. The evaluation of an expectation as in (1) is usually difficult, because it requires the computation of an N-dimensional integral. Our bound is, in general, much simpler to evaluate, because the bound involves only N one-dimensional integrals.

An upper bound on (1) can easily be derived: Since the determinant is a log-concave function on the set of positive definite matrices, it follows from Jensen's inequality that

$$\begin{aligned} E\left[\log\det\left(\mathbf{WQ}+\mathbf{I}\right)\right] &\le \log\det\left(E\left[\mathbf{W}\right]\mathbf{Q}+\mathbf{I}\right) \\ &= \log\det\left(E\left[\mathbf{W}\right]\mathbf{\Gamma}+\mathbf{I}\right), \end{aligned} \tag{2}$$

*Received by the editors January 28, 2014; accepted for publication (in revised form) by M. P. Friedlander May 22, 2015; published electronically August 11, 2015.
http://www.siam.org/journals/simax/36-3/95464.html

†Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, 01062 Dresden, Germany (meik.doerpinghaus@tu-dresden.de). The work of the first author was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant DO1568/1-1 when he was with the Institute for Theoretical Information Technology, RWTH Aachen University, and by the DFG in the framework of the Collaborative Research Center 912 "Highly Adaptive Energy-Efficient Computing" at TU Dresden.

‡Faculty of Mathematics, University of Magdeburg, 39016 Magdeburg, Germany (norbert.gaffke@mathematik.uni-magdeburg.de).

§Department of Statistics and Hausdorff Center for Mathematics, Bonn University, 53113 Bonn, Germany (limhof@uni-bonn.de).

¶Institute for Theoretical Information Technology, RWTH Aachen University, 52056 Aachen, Germany (mathar@ti.rwth-aachen.de).

where $\mathbf{\Gamma} = \operatorname{diag}(\gamma_1, \ldots, \gamma_N)$, the γ_j are the eigenvalues of $\mathbf{Q}$, and it is assumed that $E[w_j] < \infty$. The equality in (2) follows from the fact that $E[\mathbf{W}]$ is a scalar multiple of the identity matrix, since $w_1, \ldots, w_N$ are identically distributed. In the present paper, we derive a simple lower bound on (1) under general assumptions, and we discuss various applications.

The rest of the paper is organized as follows. Section 2 contains the main contribution of the paper, the new inequality in Theorem 1. Its proof is based on the rearrangement inequality of Lorentz [18]. As a corollary, we obtain a set of deterministic Muirhead- and Rado-type inequalities. In section 3, we give a more elementary proof of Theorem 1 under more restrictive assumptions using the concept of *L-superadditivity*, and we discuss the tightness of the new inequality. To illustrate our inequality, we present in section 4 several applications in information theory: First, we derive bounds on the capacity of parallel Gaussian fading channels with colored additive Gaussian noise. These new bounds quantify the increase in capacity due to correlation in the noise. Second, we derive bounds on the achievable data rate of noncoherent Rayleigh fading channels with Gaussian input symbols. In contrast to most bounds in the existing literature, which are tailored to specific signal-to-noise ratio (SNR) regimes, see, e.g., [17], [26], our bounds are uniformly tight in the sense that the gap between the upper and the lower bound is small for all SNRs. In particular, our bounds give a relatively accurate estimate of the achievable rate in the practically relevant mid-SNR regime.

2. New determinantal inequalities.

THEOREM 1. (a) *Let $\mathbf{W}$ be an $N \times N$ diagonal matrix whose diagonal entries $w_1, \ldots, w_N$ are identically distributed nonnegative random variables. Let $\mathbf{Q}$ be a nonnegative definite Hermitian $N \times N$ matrix with eigenvalues $\gamma_1, \ldots, \gamma_N$ and let $\mathbf{\Gamma} = \operatorname{diag}(\gamma_1, \ldots, \gamma_N)$. Then*

$$E[\log\det(\mathbf{WQ} + \mathbf{I})] \geq E[\log\det(\mathbf{W\Gamma} + \mathbf{I})]. \tag{3}$$

(b) *Conversely, if $\mathbf{W} = \operatorname{diag}(w_1, \ldots, w_N)$ is a nonnegative random diagonal matrix with $E[\log(w_j + 1)] < \infty$ for at least one $j \in \{1, \ldots, N\}$ and if inequality* (3) *holds whenever $\mathbf{Q}$ and $\mathbf{\Gamma}$ are nonnegative $N \times N$ diagonal matrices having the same eigenvalues, then $w_1, \ldots, w_N$ are identically distributed.*

The proof is postponed to the end of this section. An important feature of Theorem 1(a) is that no assumptions are made on the joint distribution of the w_j—only the marginal distributions have to be the same. Inequalities (3) and (2) can be written in the form

$$E[\log\det(w_1\mathbf{\Gamma} + \mathbf{I})] \leq E\left[\log\det\left(\mathbf{WQ} + \mathbf{I}\right)\right] \leq \log\det\left(E\left[w_1\right]\mathbf{\Gamma} + \mathbf{I}\right), \tag{4}$$

showing that the lower bound and the upper bound depend only on the one-dimensional marginal distribution.

We note that in part (b) of the theorem (contrary to part (a)), the assumption $E[\log(w_j + 1)] < \infty$ for at least one $j \in \{1, \ldots, N\}$ is needed. In fact, suppose that $w_1, \ldots, w_N$ are nonnegative random variables with $E[\log(w_j + 1)] = \infty$ for all $j \in \{1, \ldots, N\}$. Then for $\mathbf{W} = \operatorname{diag}(w_1, \ldots, w_N)$, both sides of (3) are equal to infinity (in case $\mathbf{Q} \neq \mathbf{0}$) or equal to zero (in case $\mathbf{Q} = \mathbf{0}$) for all pairs $\mathbf{Q}, \mathbf{\Gamma}$ considered in Theorem 1(b). That is, the remaining condition in part (b) is satisfied, but the conclusion that $w_1, \ldots, w_N$ are identically distributed need not hold.

Theorem 1(a) implies as a special case a set of inequalities of classical Muirhead- and Rado-type [13, pp. 44–48], [23], which are inequalities on sums of the form

$\Sigma_{\pi\in\mathcal{P}} F(c_{\pi(1)},\ldots,c_{\pi(N)})$, where $\mathcal{P}$ is a suitable set of permutations of $\{1,\ldots,N\}$. We will call a set $\mathcal{P}$ of permutations *balanced* if

$$(5)\qquad |\{\pi\in\mathcal{P} : \pi(i)=j\}| = \frac{|\mathcal{P}|}{N} \qquad \text{for all } i,j=1,\ldots,N.$$

Examples of balanced sets $\mathcal{P}$ are the set of *all* permutations, the set of all circular shifts, the set of all even permutations provided $N\neq 2$, and the set of all permutations that have exactly one fixed point. Moreover, if $\mathcal{P}$ is balanced, then the complement of $\mathcal{P}$, that is, the set of permutations of $\{1,\ldots,N\}$ that are not contained in $\mathcal{P}$, is balanced as well. The set of all permutations without fixed points is *not* balanced, unless $N=1$.

If $\mathcal{P}$ is a subgroup of the group of all permutations, it can be shown that the balancedness of $\mathcal{P}$ is equivalent to the transitivity of $\mathcal{P}$. A subgroup $\mathcal{P}$ is called transitive if for each pair $i,j\in\{1,\ldots,N\}$, there exists a $\pi\in\mathcal{P}$ such that $\pi(i)=j$; for examples of transitive permutation groups, see [5].

If $\mathbf{C}$ is a diagonal matrix with diagonal entries $c_1,\ldots,c_N$ and π is a permutation, we denote by $\mathbf{C}_\pi$ the diagonal matrix with diagonal entries $c_{\pi(1)},\ldots,c_{\pi(N)}$.

Corollary 1. *Let $\mathcal{P}$ be a nonempty set of permutations of $\{1,\ldots,N\}$. The inequality*

$$(6)\qquad \sum_{\pi\in\mathcal{P}} \log\det(\mathbf{U}\mathbf{C}_\pi\mathbf{U}^H+\mathbf{D}) \geq \sum_{\pi\in\mathcal{P}} \log\det(\mathbf{C}_\pi+\mathbf{D})$$

holds for every nonnegative $N\times N$ diagonal matrix $\mathbf{C}$, every positive $N\times N$ diagonal matrix $\mathbf{D}$, and every unitary $N\times N$ matrix $\mathbf{U}$ if and only if $\mathcal{P}$ is balanced.

Proof. To show that (5) is sufficient for (6), let σ be a random permutation that is uniformly distributed on $\mathcal{P}$. Set $\mathbf{W}=\mathbf{C}_\sigma$ and $\mathbf{Q}=\mathbf{U}^H\mathbf{D}^{-1}\mathbf{U}$. The balancedness of $\mathcal{P}$ implies that the diagonal entries of $\mathbf{W}$ are identically distributed. Hence, inequality (6) follows from (3).

To show that (5) is necessary, suppose that (6) holds for all $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{U}$. Let $j\in\{1,\ldots,N\}$. For each $i=1,\ldots,N$, let $\mathcal{P}_i=\{\pi\in\mathcal{P}:\pi(i)=j\}$. The sets $\mathcal{P}_1,\ldots,\mathcal{P}_N$ are pairwise disjoint and $\bigcup_i \mathcal{P}_i=\mathcal{P}$. We will show that $|\mathcal{P}_{i_1}|=|\mathcal{P}_{i_2}|$ for all i_1,i_2, which implies that indeed $|\mathcal{P}_i|=|\mathcal{P}|/N$ for every i. Let $i_1\neq i_2$. Let $\mathbf{U}$ be the permutation matrix that corresponds to the transposition that interchanges i_1 and i_2. Let $\lambda>0$. Let $\mathbf{C}=\operatorname{diag}(c_1,\ldots,c_N)$, where $c_j=\lambda$ and $c_i=0$ for $i\neq j$. Let $\mathbf{D}=\operatorname{diag}(d_1,\ldots,d_N)$, where $d_{i_2}=1$ and $d_i=\lambda^2$ for $i\neq i_2$. Observing that for a diagonal matrix $\mathbf{A}=\operatorname{diag}(a_1,\ldots,a_N)$ one has $\mathbf{U}\mathbf{A}\mathbf{U}^H=\operatorname{diag}(a_1',\ldots,a_N')$, where $a_{i_1}'=a_{i_2}$, $a_{i_2}'=a_{i_1}$, and $a_i'=a_i$ for all $i\in\{1,\ldots,N\}\setminus\{i_1,i_2\}$, we get from (6) that

$$1\leq \prod_{\pi\in\mathcal{P}} \frac{\det(\mathbf{U}\mathbf{C}_\pi\mathbf{U}^H+\mathbf{D})}{\det(\mathbf{C}_\pi+\mathbf{D})} = \prod_{\pi\in\mathcal{P}} \frac{(c_{\pi(i_2)}+\lambda^2)(c_{\pi(i_1)}+1)}{(c_{\pi(i_1)}+\lambda^2)(c_{\pi(i_2)}+1)} = \prod_{\pi\in\mathcal{P}} r(\pi),$$

where $r(\pi)=\lambda$ if $\pi\in\mathcal{P}_{i_1}$, $r(\pi)=1/\lambda$ if $\pi\in\mathcal{P}_{i_2}$, and $r(\pi)=1$ otherwise. Hence, $1\leq\lambda^{|\mathcal{P}_{i_1}|-|\mathcal{P}_{i_2}|}$. As $\lambda>0$ was arbitrary, it follows that $|\mathcal{P}_{i_1}|=|\mathcal{P}_{i_2}|$. ∎

We now set out to prove Theorem 1. The proof of part (a) uses the following version of Lorentz's inequality for rearrangements [18].

Proposition 1. *If $\Phi:[0,\infty)^N\to\mathbb{R}$ has continuous second-order partial derivatives with respect to all variables, $\partial^2\Phi/(\partial v_j\partial v_k)\geq 0$ for all $j\neq k$, and if $w_1,\ldots,w_N$ are identically distributed bounded nonnegative random variables, then*

$$E[\Phi(w_1,\ldots,w_N)]\leq E[\Phi(w_1,\ldots,w_1)].$$

To see that Proposition 1 is a special case of Lorentz's inequality assume, without loss of generality, that the underlying probability space is the interval $(0,1)$ endowed with the Lebesgue measure [9, Theorem 11.7.5]. Let $w_j^* : (0,1) \to [0,\infty)$ be the decreasing rearrangement of w_j, that is, $w_j^*(s) = \sup\{t \in [0,\infty) : P(w_j \geq t) \geq s\}$; see [13, pp. 276–277]. Then, according to [18], $E[\Phi(w_1,\ldots,w_N)] \leq E[\Phi(w_1^*,\ldots,w_N^*)]$. But if $w_1,\ldots,w_N$ are identically distributed, then $w_1^* = \cdots = w_N^*$ and each w_j^* has the same distribution as w_1. Thus, $E[\Phi(w_1^*,\ldots,w_N^*)] = E[\Phi(w_1,\ldots,w_1)]$.

The proof of Theorem 1(b) rests on the following uniqueness result for distributions.

PROPOSITION 2. *If x and y are nonnegative random variables with $E[\log(x+1)] < \infty$ and*

$$E[\log(\lambda x + 1)] = E[\log(\lambda y + 1)] \qquad \text{for all } \lambda > 0,$$

then x and y have the same distribution.

Proof. The conditions imply that $E[\log(x+\lambda)] = E[\log(y+\lambda)] < \infty$ for all $\lambda > 0$. The derivatives $(\partial/\partial\lambda)\log(x+\lambda) = (x+\lambda)^{-1}$ and $(\partial/\partial\lambda)\log(y+\lambda) = (y+\lambda)^{-1}$ are nonnegative and bounded from above by $1/\lambda$. It follows that the derivatives of the expectations with respect to λ can be computed by differentiation inside the expectation signs; see, e.g., [3, p. 212]. Therefore, $E[(x+\lambda)^{-1}] = E[(y+\lambda)^{-1}]$ for all $\lambda > 0$. Thus, the Stieltjes transforms of the distributions of x and y coincide, and it follows that the distributions coincide; see [29, p. 336]. □

Proof of Theorem 1. (a) Let $\mathbf{U}$ be a unitary matrix such that $\mathbf{U}^H\mathbf{Q}\mathbf{U} = \mathbf{\Gamma}$. Fix $\epsilon \in (0,1]$ and set $\mathbf{\Gamma}_\epsilon = \mathbf{\Gamma} + \epsilon\mathbf{I}$ and

$$\Phi(\mathbf{v}) = -\log\det(\mathbf{\Gamma}_\epsilon^{1/2}\mathbf{U}^H\mathbf{V}\mathbf{U}\mathbf{\Gamma}_\epsilon^{1/2} + \mathbf{I}),$$

where $\mathbf{v} \in [0,\infty)^N$ and $\mathbf{V} = \operatorname{diag}(\mathbf{v})$. Recall that for a differentiable positive definite matrix function $\mathbf{A}(t)$, $(\partial/\partial t)\log\det\mathbf{A} = \operatorname{tr}\mathbf{A}^{-1}(\partial\mathbf{A}/\partial t)$ and $(\partial/\partial t)\mathbf{A}^{-1} = -\mathbf{A}^{-1}(\partial\mathbf{A}/\partial t)\mathbf{A}^{-1}$. Thus, for $j,k = 1,\ldots,N$,

$$\begin{aligned}\frac{\partial^2\Phi(\mathbf{v})}{\partial v_j\partial v_k} &= -\frac{\partial^2}{\partial v_j\partial v_k}\log\det(\mathbf{V} + \mathbf{U}\mathbf{\Gamma}_\epsilon^{-1}\mathbf{U}^H)\\ &= [\mathbf{e}_j^T(\mathbf{V} + \mathbf{U}\mathbf{\Gamma}_\epsilon^{-1}\mathbf{U}^H)^{-1}\mathbf{e}_k]^2 \geq 0,\end{aligned}$$

where $\mathbf{e}_j$ is the jth column of $\mathbf{I}$.

For $K \in \mathbb{N}$, let $\tilde{w}_j(K) = w_j 1_{\{w_j \leq K\}}$ and $\tilde{\mathbf{W}}(K) = \operatorname{diag}(\tilde{w}_1(K),\ldots,\tilde{w}_N(K))$. Then, by Proposition 1,

$$E[\log\det(\mathbf{\Gamma}_\epsilon^{1/2}\mathbf{U}^H\tilde{\mathbf{W}}(K)\mathbf{U}\mathbf{\Gamma}_\epsilon^{1/2} + \mathbf{I})] \geq E[\log\det(\tilde{w}_1(K)\mathbf{\Gamma}_\epsilon + \mathbf{I})].$$

Letting $\epsilon \to 0$ and using the bounded convergence theorem, we see that this inequality remains true when $\mathbf{\Gamma}_\epsilon$ is replaced by $\mathbf{\Gamma}$. Then letting $K \to \infty$ and using monotone convergence, we obtain

$$\begin{aligned}E[\log\det(\mathbf{\Gamma}^{1/2}\mathbf{U}^H\mathbf{W}\mathbf{U}\mathbf{\Gamma}^{1/2} + \mathbf{I})] &\geq E[\log\det(w_1\mathbf{\Gamma} + \mathbf{I})]\\ &= E[\log\det(\mathbf{W}\mathbf{\Gamma} + \mathbf{I})].\end{aligned}$$

Using that $\det(\mathbf{AB} + \mathbf{I}) = \det(\mathbf{BA} + \mathbf{I})$ for all matrices $\mathbf{A}$ and $\mathbf{B}$ of appropriate dimensions [14, Theorem 1.3.20], we obtain

$$\begin{aligned}E[\log\det(\mathbf{\Gamma}^{1/2}\mathbf{U}^H\mathbf{W}\mathbf{U}\mathbf{\Gamma}^{1/2} + \mathbf{I})] &= E[\log\det(\mathbf{U}\mathbf{\Gamma}\mathbf{U}^H\mathbf{W} + \mathbf{I})]\\ &= E[\log\det(\mathbf{Q}\mathbf{W} + \mathbf{I})],\end{aligned}$$

which completes the proof of (a).

(b) Let $E[\log(w_j+1)] < \infty$. Let $k \in \{1,\dots,N\} \setminus \{j\}$. We will show that w_k has the same distribution as w_j. Let $\lambda > 0$. Let $\mathbf{Q}$ be the $N \times N$ matrix with jth diagonal entry equal to λ and all other entries equal to 0. Let $\mathbf{\Gamma}$ be the $N \times N$ matrix with kth diagonal entry equal to λ and all other entries equal to 0. Then, by (3), $E[\log(\lambda w_j + 1)] \geq E[\log(\lambda w_k + 1)]$. Interchanging the roles of $\mathbf{Q}$ and $\mathbf{\Gamma}$ yields the reverse inequality, and so $E[\log(\lambda w_j + 1)] = E[\log(\lambda w_k + 1)]$ for every $\lambda > 0$. Thus, by Proposition 2, w_k has the same distribution as w_j. $\square$

3. Variations on the theme. The short proofs of Theorem 1(a) in its general form and of the deterministic version in Corollary 1 relied on deep results from the theory of rearrangements. It is instructive to look at a more elementary proof of the theorem under more restrictive assumptions. The alternative approach uses the concept of L-superadditivity.

A function $F : [0,\infty)^N \to \mathbb{R}$ is called *L-superadditive* if

$$F(\mathbf{c} + h\mathbf{e}_i + k\mathbf{e}_j) - F(\mathbf{c} + h\mathbf{e}_i) - F(\mathbf{c} + k\mathbf{e}_j) + F(\mathbf{c}) \geq 0$$

for all $\mathbf{c} \in [0,\infty)^N$, all i,j with $1 \leq i < j \leq N$, and all $h,k \in (0,\infty)$; see [19, Definition C.2, p. 213], [24, p. 58]. We note that if F has continuous second-order partial derivatives, then L-superadditivity of F is equivalent to $\partial^2 F/(\partial c_i \partial c_j) \geq 0$ for all $i \neq j$; see [19, p. 218].

If F is L-superadditive and $\mathcal{P}$ is a nonempty balanced set of permutations of $\{1,\dots,N\}$, then for all nonnegative numbers $c_1,\dots,c_N$,

$$\frac{1}{|\mathcal{P}|} \sum_{\pi \in \mathcal{P}} F(c_{\pi(1)},\dots,c_{\pi(N)}) \leq \frac{1}{N} \sum_{j=1}^{N} F(c_j,\dots,c_j). \tag{7}$$

This follows from the elementary discrete version of Lorentz's inequality given in [19, Chapter 6, Proposition E.1(2)] applied to the N vectors $(c_{\pi(i)})_{\pi \in \mathcal{P}}$, $i = 1,\dots,N$. To obtain the expression on the right-hand side of (7), note that, by (5), each of these vectors contains each c_j exactly $|\mathcal{P}|/N$ times.

3.1. A direct proof of inequality (6). Let

$$F(\mathbf{c}) = -\log\det(\mathbf{U}\mathbf{C}\mathbf{U}^H + \mathbf{D}),$$

where $\mathbf{c} = (c_1,\dots,c_N) \in [0,\infty)^N$ and $\mathbf{C} = \operatorname{diag}(\mathbf{c})$. As in the proof of Theorem 1(a) in section 2, one can verify that the second-order partial derivatives $\partial^2 F/(\partial c_i \partial c_j)$, $i \neq j$, are nonnegative, which implies L-superadditivity of F. However, we will give a direct proof of L-superadditivity of F. Fix $\mathbf{c} \in [0,\infty)^N$, $i < j$, and $h,k > 0$. Set $\mathbf{B} = \mathbf{C} + \mathbf{U}^H\mathbf{D}\mathbf{U}$. Then

$$\begin{aligned}
&F(\mathbf{c} + h\mathbf{e}_i + k\mathbf{e}_j) - F(\mathbf{c} + h\mathbf{e}_i) - F(\mathbf{c} + k\mathbf{e}_j) + F(\mathbf{c}) \\
&\qquad = \log \frac{\det(\mathbf{B} + h\mathbf{e}_i\mathbf{e}_i^T)}{\det(\mathbf{B} + h\mathbf{e}_i\mathbf{e}_i^T + k\mathbf{e}_j\mathbf{e}_j^T)} \frac{\det(\mathbf{B} + k\mathbf{e}_j\mathbf{e}_j^T)}{\det \mathbf{B}} \\
&\qquad = \log \frac{1 + k\mathbf{e}_j^T\mathbf{B}^{-1}\mathbf{e}_j}{1 + k\mathbf{e}_j^T(\mathbf{B} + h\mathbf{e}_i\mathbf{e}_i^T)^{-1}\mathbf{e}_j},
\end{aligned}$$

where we have used that $\det(\mathbf{A} + k\mathbf{e}_j\mathbf{e}_j^T)/\det\mathbf{A} = 1 + k\mathbf{e}_j^T\mathbf{A}^{-1}\mathbf{e}_j$. By [14, Corollary 7.7.4], $\mathbf{e}_j^T\mathbf{B}^{-1}\mathbf{e}_j \geq \mathbf{e}_j^T(\mathbf{B} + h\mathbf{e}_i\mathbf{e}_i^T)^{-1}\mathbf{e}_j$, and it follows that F is L-superadditive.

Therefore, by (7),

$$\sum_{\pi\in\mathcal{P}} \log\det(\mathbf{U}\mathbf{C}_\pi\mathbf{U}^H + \mathbf{D}) = -\sum_{\pi\in\mathcal{P}} F(c_{\pi(1)},\dots,c_{\pi(N)}) \geq -\frac{|\mathcal{P}|}{N}\sum_{j=1}^N F(c_j,\dots,c_j).$$

Moreover, denoting the diagonal entries of $\mathbf{D}$ by $d_1,\dots,d_N$, we have, by (5),

$$\begin{aligned}\sum_{\pi\in\mathcal{P}} \log\det(\mathbf{C}_\pi + \mathbf{D}) &= \sum_{\pi\in\mathcal{P}} \log\prod_{i=1}^N (c_{\pi(i)} + d_i)\\ &= \sum_{i=1}^N \log\prod_{\pi\in\mathcal{P}} (c_{\pi(i)} + d_i) = \sum_{i=1}^N \log\prod_{j=1}^N (c_j + d_i)^{|\mathcal{P}|/N}\\ &= \frac{|\mathcal{P}|}{N}\sum_{j=1}^N \log\det(c_j\mathbf{I} + \mathbf{D}) = -\frac{|\mathcal{P}|}{N}\sum_{j=1}^N F(c_j,\dots,c_j).\end{aligned}$$

3.2. Near equivalence of Theorem 1(a) and Corollary 1. Interestingly, Theorem 1(a) can be concluded from Corollary 1 under the additional assumption that $\mathbf{Q}$ is *positive* definite and the nonnegative random variables $w_1,\dots,w_N$ are *exchangeable*, that is, the distribution of $(w_1,\dots,w_N)$ is invariant under permutations. The proof goes as follows.

Let $\mathcal{P}$ be the set of all permutations of $\{1,\dots,N\}$. Replace $\mathbf{C}$ in (6) by the random matrix $\mathbf{W} = \operatorname{diag}(w_1,\dots,w_N)$ and apply expectations on the right- and left-hand side. By exchangeability, $\mathbf{W}_\pi$ has the same distribution as $\mathbf{W}$ for every permutation π, and so $E[\log\det(\mathbf{U}\mathbf{W}\mathbf{U}^H + \mathbf{D})] \geq E[\log\det(\mathbf{W} + \mathbf{D})]$. Replacing $\mathbf{D}$ by $\mathbf{\Gamma}^{-1}$ and choosing $\mathbf{U}$ so that $\mathbf{Q} = \mathbf{U}^H\mathbf{\Gamma}\mathbf{U}$, one obtains $E[\log\det(\mathbf{W} + \mathbf{Q}^{-1})] \geq E[\log\det(\mathbf{W} + \mathbf{\Gamma}^{-1})]$, which is equivalent to (3).

3.3. Tightness of the inequalities. Recall that $\mathbf{W} = \operatorname{diag}(w_1,\dots,w_N)$, where $w_1,\dots,w_N$ are identically distributed nonnegative random variables. The tightness of inequality (3) strongly depends on the *joint* distribution of $w_1,\dots,w_N$.

The lower bound (3) and the upper bound (2) together yield

$$\sum_{j=1}^N E\left[\log\left(w_1\gamma_j + 1\right)\right] \leq E\left[\log\det\left(\mathbf{W}\mathbf{Q} + \mathbf{I}\right)\right] \leq \sum_{j=1}^N \log\left(E[w_1]\gamma_j + 1\right). \tag{8}$$

Both inequalities become equalities in case of a Dirac distribution, i.e., if $P(w_1 = c) = 1$ for some constant c. This shows that the bounds are sharp. There is also equality in (3) if $P(w_1 = w_2 = \cdots = w_N) = 1$ and if $\mathbf{Q}$ is a diagonal matrix.

4. Applications in information theory. We now set out to demonstrate the usefulness of the key results from section 2 in information theory. Some of these results have been applied in recent publications in the context of bounding the achievable rate of noncoherent Gaussian fading channels [8] and in the context of demonstrating the convergence of the conditional per symbol entropy [6]. Another information theoretic application of rearrangements is given in [28], where a lower bound on the differential entropy of a sum of independent random vectors is derived.

4.1. Parallel Gaussian fading channels with colored additive Gaussian noise. We consider a parallel Gaussian fading channel consisting of N individual subchannels which can be described by the input-output relation

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n},$$

where $\mathbf{y} \in \mathbb{C}^N$ is an N-dimensional vector describing the complex channel output, $\mathbf{x} \in \mathbb{C}^N$ is the channel input, $\mathbf{n} \in \mathbb{C}^N$ is additive Gaussian noise, and $\mathbf{H} = \mathrm{diag}(\mathbf{h})$ is a diagonal matrix with a vector $\mathbf{h} \in \mathbb{C}^N$ containing the channel fading weights of the individual subchannels.

The noise $\mathbf{n}$ is assumed to be zero-mean proper complex Gaussian, cf. [22], which we denote by $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_n})$, where $\mathbf{R_n} = E[\mathbf{n}\mathbf{n}^H]$ is the covariance matrix of $\mathbf{n}$. We also assume that $\mathbf{R_n}$ is positive definite, as a singular $\mathbf{R_n}$ yields infinite capacity. Hence, the probability density of $\mathbf{n}$ is given by

$$p(\mathbf{n}) = \pi^{-N} \det(\mathbf{R_n})^{-1} \exp(-\mathbf{n}^H \mathbf{R_n}^{-1} \mathbf{n}).$$

Moreover, the fading weights of the individual subchannels are also zero-mean proper complex Gaussian. They are assumed to be independent and identically distributed (i.i.d.) such that the covariance matrix is given by

$$\mathbf{R_h} = E\left[\mathbf{h}\mathbf{h}^H\right] = \sigma_h^2 \mathbf{I}.$$

Finally, we assume $\mathbf{x}, \mathbf{h}$, and $\mathbf{n}$ to be mutually independent.

Parallel Gaussian channels with colored noise have been studied in various publications. The basic setup without fading has, e.g., been studied in [4, Chapter 9.5]. If the individual subchannels of the parallel channel are assumed to be a series of channel uses over time with $\mathbf{R_n}$ being a Toeplitz matrix, this setup corresponds to a stationary channel with colored Gaussian noise, as it has for example been considered by Kim [16] when studying stationary additive Gaussian noise channels with feedback. While these papers do consider parallel Gaussian channels with colored noise, we extend this setup with fading. For this channel, we study the capacity and the outage probability. To the best of our knowledge, there are no results on the capacity and outage probability of parallel Gaussian channels with fading.

4.1.1. Capacity. We focus on the case that the receiver knows the realization of the channel $\mathbf{h}$, and so we consider the mutual information $I(\mathbf{x}; \mathbf{y}, \mathbf{h})$ between the channel input $\mathbf{x}$ on the one hand and the output $\mathbf{y}$ and the channel $\mathbf{h}$ on the other. The *ergodic capacity*, also referred to as *Shannon capacity*, is defined by

$$C = \max_{p(\mathbf{x}) \,:\, E[\mathbf{x}^H \mathbf{x}] \le N P_{\mathrm{avg}}} I(\mathbf{x}; \mathbf{y}, \mathbf{h}), \tag{9}$$

where the maximization is over all input densities $p(\mathbf{x})$ fulfilling the average power constraint $E[\mathbf{x}^H \mathbf{x}] \le N P_{\mathrm{avg}}$. That is, $N P_{\mathrm{avg}}$ is the maximum average sum input power over all subchannels.

To the best of our knowledge, no closed-form expression for the capacity is available, but the inequalities (4) can be applied to derive useful bounds. We will show that

$$\sum_{k=1}^{N} E[\log(P_{\mathrm{avg}} \psi_k^{-1} z + 1)] \le C \le \sum_{k=1}^{N} \log\left(\psi_k^{-1}\left[P_{\mathrm{avg}} \sigma_h^2 + \frac{1}{N} \mathrm{tr}(\mathbf{R_n})\right]\right), \tag{10}$$

where $\psi_1, \ldots, \psi_N$ are the eigenvalues of $\mathbf{R_n}$ and z is an exponential random variable with mean σ_h^2.

By the convexity of $\log(P_{\text{avg}}\psi^{-1}z+1)$ in the variable $\psi > 0$, one easily concludes that among all noise covariance matrices $\mathbf{R_n}$ with $\frac{1}{N}\operatorname{tr}(\mathbf{R_n}) = \sigma_n^2$, the lower bound in (10) is minimized if $\mathbf{R_n} = \sigma_n^2\mathbf{I}$, that is, if the Gaussian noise on the different subchannels is independent. For this specific case, the capacity is known and is equal to the minimum $NE[\log(P_{\text{avg}}\sigma_n^{-2}z+1)]$. Hence, the new bounds in (10) quantify the possible increase in capacity when the noise on the individual subchannels is correlated.

In the following, we prove inequalities (10). After using the chain rule for mutual information [4, Theorem 2.5.2], we can express $I(\mathbf{x};\mathbf{y},\mathbf{h})$ based on the differential entropy as

$$I(\mathbf{x};\mathbf{y},\mathbf{h}) = I(\mathbf{x};\mathbf{y}|\mathbf{h}) + I(\mathbf{x};\mathbf{h}) \\ = h(\mathbf{y}|\mathbf{h}) - h(\mathbf{y}|\mathbf{x},\mathbf{h}), \tag{11}$$

where (11) follows from the independence of $\mathbf{x}$ and $\mathbf{h}$ yielding $I(\mathbf{x};\mathbf{h}) = 0$, and with the differential entropies

$$h(\mathbf{y}|\mathbf{h}) = E\left[-\log p(\mathbf{y}|\mathbf{h})\right], \qquad h(\mathbf{y}|\mathbf{x},\mathbf{h}) = E\left[-\log p(\mathbf{y}|\mathbf{x},\mathbf{h})\right].$$

As $p(\mathbf{y}|\mathbf{x},\mathbf{h})$ is proper complex Gaussian with covariance matrix $\mathbf{R_n}$, it follows that

$$h(\mathbf{y}|\mathbf{x},\mathbf{h}) = \log\det\left(\pi e\mathbf{R_n}\right). \tag{12}$$

To bound $h(\mathbf{y}|\mathbf{h})$, we use that, according to [22, Theorem 2], proper complex Gaussian random vectors maximize the differential entropy among all continuous complex random vectors with a given covariance matrix. Since

$$E[(\mathbf{y}-E[\mathbf{y}|\mathbf{h}])(\mathbf{y}-E[\mathbf{y}|\mathbf{h}])^H|\mathbf{h}] = \mathbf{H}\mathbf{R_x}\mathbf{H}^H + \mathbf{R_n},$$

where $\mathbf{R_x} = E[(\mathbf{x}-E[\mathbf{x}])(\mathbf{x}-E[\mathbf{x}])^H]$, it follows that

$$h(\mathbf{y}|\mathbf{h}) \le E_{\mathbf{h}}\left[\log\det\left(\pi e\left(\mathbf{H}\mathbf{R_x}\mathbf{H}^H + \mathbf{R_n}\right)\right)\right] \tag{13}$$

with equality if $\mathbf{x} \sim \mathcal{CN}(\mathbf{0},\mathbf{R_x})$. Hence, by (11) and (12),

$$C = \max_{\operatorname{tr}(\mathbf{R_x})\le NP_{\text{avg}}} E_{\mathbf{h}}\left[\log\det\left(\mathbf{H}\mathbf{R_x}\mathbf{H}^H\mathbf{R_n}^{-1} + \mathbf{I}\right)\right]. \tag{14}$$

The following lemma allows us to conclude that the maximum in (14) is achieved by a diagonal matrix $\mathbf{R_x}$, i.e., by a channel input $\mathbf{x}$ with independent elements. Its proof is given in Appendix A.

LEMMA 1. *If $\mathbf{W} \in \mathbb{C}^{N\times N}$ is a random diagonal matrix with independent zero-mean proper complex Gaussian diagonal elements and $\mathbf{B} \in \mathbb{C}^{N\times N}$ is a positive definite matrix, then the maximization*

$$\max_{\operatorname{tr}(\mathbf{Q})\le c} E\left[\log\det\left(\mathbf{W}\mathbf{Q}\mathbf{W}^H\mathbf{B}^{-1} + \mathbf{I}\right)\right]$$

over all nonnegative definite matrices $\mathbf{Q} \in \mathbb{C}^{N\times N}$ with $\operatorname{tr}(\mathbf{Q}) \le c$ and c a positive constant has a diagonal matrix $\mathbf{Q}$ as solution.

With Lemma 1, the capacity in (14) is given by

$$C = \max_{\mathrm{tr}(\mathbf{\Lambda_x}) \leq NP_{\mathrm{avg}}} E_{\mathbf{h}} \left[\log\det\left(\mathbf{H}\mathbf{\Lambda_x}\mathbf{H}^H\mathbf{R_n}^{-1} + \mathbf{I}\right)\right] = \max_{\mathrm{tr}(\mathbf{\Lambda_x}) \leq NP_{\mathrm{avg}}} E_{\mathbf{z}} \left[\log\det\left(\mathbf{Z}\mathbf{\Lambda_x}\mathbf{R_n}^{-1} + \mathbf{I}\right)\right], \tag{15}$$

where $\mathbf{Z} = \mathbf{H}\mathbf{H}^H = \mathrm{diag}(\mathbf{z})$ and the maximum is taken over all nonnegative diagonal $N \times N$ matrices $\mathbf{\Lambda_x}$ with $\mathrm{tr}(\mathbf{\Lambda_x}) \leq NP_{\mathrm{avg}}$. The special choice $\mathbf{\Lambda_x} = P_{\mathrm{avg}}\mathbf{I}$ and the lower bound in (4) give the lower bound in (10).

For every nonnegative diagonal $N \times N$ matrix $\mathbf{\Lambda_x}$, the upper bound in (4) gives

$$E_{\mathbf{z}} \left[\log\det\left(\mathbf{Z}\mathbf{\Lambda_x}\mathbf{R_n}^{-1} + \mathbf{I}\right)\right] = E_{\mathbf{z}} \left[\log\det\left(\mathbf{Z}\mathbf{\Lambda_x}^{1/2}\mathbf{R_n}^{-1}\mathbf{\Lambda_x}^{1/2} + \mathbf{I}\right)\right] \leq \log\det\left(\sigma_h^2\mathbf{\Lambda_x}^{1/2}\mathbf{R_n}^{-1}\mathbf{\Lambda_x}^{1/2} + \mathbf{I}\right) = \log\det\left(\sigma_h^2\mathbf{\Lambda_x} + \mathbf{R_n}\right) - \log\det\mathbf{R_n},$$

and if $\mathrm{tr}(\mathbf{\Lambda_x}) \leq NP_{\mathrm{avg}}$, then, by the arithmetic-geometric mean inequality,

$$\log\det(\sigma_h^2\mathbf{\Lambda_x} + \mathbf{R_n}) \leq N\log\left(\frac{1}{N}\mathrm{tr}(\sigma_h^2\mathbf{\Lambda_x} + \mathbf{R_n})\right) \leq N\log\left(\sigma_h^2 P_{\mathrm{avg}} + \frac{1}{N}\mathrm{tr}(\mathbf{R_n})\right).$$

The upper bound in (10) now follows from (15).

4.1.2. Capacity achieving input distribution. A distribution at which the maximum in (9) is attained, that is, a capacity achieving input distribution, is given by the $\mathcal{CN}(\mathbf{0}, \mathbf{\Lambda_x})$-distribution, where $\mathbf{\Lambda_x}$ solves the maximization problem in (15). An explicit description of the solution is not known for general $\mathbf{R_n}$. We now address the question when the input distribution $\mathcal{CN}(\mathbf{0}, P_{\mathrm{avg}}\mathbf{I})$, generating i.i.d. Gaussian input symbols, is capacity achieving. This question leads to the case of a "balanced matrix" $\mathbf{R_n}$.

We say that an $N \times N$ matrix $\mathbf{B} = (b_{kl})$ is *balanced* if there is a nonempty balanced set $\mathcal{P}$ of permutations of $\{1, \ldots, N\}$, i.e., a set $\mathcal{P}$ satisfying (5), such that

$$b_{kl} = b_{\pi(k),\pi(l)} \text{ for all } k, l = 1, \ldots, N \text{ and all } \pi \in \mathcal{P}. \tag{16}$$

For example, taking the balanced set $\mathcal{P}$ of all circular shifts, one sees that every circulant matrix is balanced.

From the next lemma, whose proof is given in Appendix B, it follows that if $\mathbf{R_n}$ is balanced, then $\mathbf{\Lambda_x} = P_{\mathrm{avg}}\mathbf{I}$ achieves the maximum in (15). The case where $\mathbf{R_n}$ is a circulant matrix turns out to be relevant for bounding the capacity of a fading channel with stationary colored noise; see section 4.1.3.

LEMMA 2. *If $\mathbf{W} \in \mathbb{C}^{N \times N}$ is a random diagonal matrix with nonnegative exchangeable random variables on the diagonal and $\mathbf{B} \in \mathbb{C}^{N \times N}$ is a positive definite balanced matrix, then*

$$\max_{\mathrm{tr}(\mathbf{\Lambda}) \leq c} E\left[\log\det\left(\mathbf{W}\mathbf{\Lambda}\mathbf{B}^{-1} + \mathbf{I}\right)\right] = E\left[\log\det\left(\frac{c}{N}\mathbf{W}\mathbf{B}^{-1} + \mathbf{I}\right)\right],$$

where the maximization is over all nonnegative diagonal $N \times N$ matrices $\mathbf{\Lambda}$ with $\mathrm{tr}(\mathbf{\Lambda}) \leq c$ and c a positive constant. Hence, the maximum is attained at $\mathbf{\Lambda} = \frac{c}{N}\mathbf{I}$.

Thus, if $\mathbf{R_n}$ is balanced, then by (15) and Lemma 2, $C = E[\log\det(P_{\mathrm{avg}}\mathbf{Z}\mathbf{R_n}^{-1} + \mathbf{I})]$. For this case, the upper bound on C in (10) can be improved: By (8),

$$\sum_{k=1}^{N} E[\log(P_{\mathrm{avg}}\psi_k^{-1}z + 1)] \leq C \leq \sum_{k=1}^{N} \log(P_{\mathrm{avg}}\psi_k^{-1}\sigma_h^2 + 1). \tag{17}$$

We note that for every positive definite $\mathbf{R_n}$, not necessarily balanced, and for the fixed input distribution $\mathbf{x} \sim \mathcal{C}N(\mathbf{0}, P_{\rm avg}\mathbf{I})$, the bounds in (17) are bounds on the achievable date rate, i.e., on the mutual information $I(\mathbf{x}; \mathbf{y}, \mathbf{h})$.

Expressing the expectations in the lower bound in (17) in terms of the exponential integral $\mathrm{Ei}(\xi) = \int_{-\infty}^{\xi} e^t/t\, dt$ and using that

$$0 \le \log(x+1) + e^{1/x}\mathrm{Ei}\left(-\frac{1}{x}\right) \le \gamma \tag{18}$$

with $\gamma \approx 0.57721$ being the Euler constant, one can show that the difference between the upper bound and the lower bound in (17) is itself upper-bounded by $N\gamma$ [nat/channel use]; cf. [8, section III-E]. Hence, for every positive definite matrix $\mathbf{R_n}$, we have found narrow bounds on the achievable rate with i.i.d. Gaussian input symbols; and the bounds on the capacity are guaranteed to be narrow when $\mathbf{R_n}$ is balanced.

4.1.3. Capacity of a fading channel with stationary colored noise. Now we assume that the parallel subchannels are individual channel uses of a Gaussian fading channel in temporal domain, having stationary colored additive Gaussian noise and independent fading realizations. The additive Gaussian noise process $\{n_k\}$ has the covariance function $r_n(l) = E[n_{k+l}n_k^*]$. We assume that $\sum_{l=-\infty}^{\infty} |r_n(l)| < \infty$. This implies that $\{n_k\}$ has a power spectral density (PSD) given by

$$S_n(f) = \sum_{l=-\infty}^{\infty} r_n(l)e^{-j2\pi l f}, \quad |f| < 0.5,$$

where f is the normalized frequency and $j = \sqrt{-1}$. We also assume that S_n is positive.

For this scenario, the capacity is defined by

$$C_\infty = \lim_{N\to\infty} \frac{1}{N} \max_{p(\mathbf{x})\,:\,E[\mathbf{x}^H\mathbf{x}]\le N P_{\rm avg}} I(\mathbf{x}; \mathbf{y}, \mathbf{h}), \tag{19}$$

i.e., we normalize the mutual information by the number of channel uses N. To evaluate (19), we substitute the sequence of Hermitian Toeplitz matrices $\mathbf{R_n} = (r_n(k-l))_{k,l=0}^{N-1}$, $N = 1, 2, \dots$, by an asymptotically equivalent sequence of circulant matrices having the same asymptotic eigenvalue distribution; see [11, equation (4.32)]. We can then apply the bounds for the parallel channel in (17) and get by [11, Lemma 11, Theorem 4],

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} E\left[\log\left(\frac{P_{\rm avg} z}{\psi_k} + 1\right)\right] \le C_\infty \le \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} \log\left(\frac{P_{\rm avg}\sigma_h^2}{\psi_k} + 1\right). \tag{20}$$

Note that all the eigenvalues $\psi_1, \dots, \psi_N$ of $\mathbf{R_n}$ depend on N and that, by [11, Lemma 6], they are bounded and bounded away from zero.

Using Szegö's theorem on the asymptotic eigenvalue distribution of Hermitian Toeplitz matrices [12, pp. 64–65], [11, Theorem 9], the bounds in (20) become

$$\int_{f=-\frac{1}{2}}^{\frac{1}{2}} E\left[\log\left(\frac{P_{\rm avg} z}{S_n(f)} + 1\right)\right] df \le C_\infty \le \int_{f=-\frac{1}{2}}^{\frac{1}{2}} \log\left(\frac{P_{\rm avg}\sigma_h^2}{S_n(f)} + 1\right) df. \tag{21}$$

The difference between the upper and the lower bound in (21) is bounded (in nats) by the Euler constant γ.

Similar to the case of finite N discussed in sections 4.1.1 and 4.1.2, the new bounds in (21) allow one to evaluate the possible increase in capacity in case the additive Gaussian noise is colored in comparison to white noise, where for the latter case the capacity is known and is given by the left-hand side of (21) with $S_n(f)$ being constant for $f \in [-1/2, 1/2]$.

4.1.4. Outage probability. We can also apply the new inequality given in Theorem 1 for bounding the outage probability of the parallel Gaussian fading channel with colored noise and input vector $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, P_{\text{avg}}\mathbf{I})$. The outage probability $p_{\text{out}}(t)$ is the probability that the instantaneous mutual information falls below a given value $t > 0$, that is, cf. (11) to (13),

$$p_{\text{out}}(t) = P\left(\log\det(P_{\text{avg}}\mathbf{H}\mathbf{H}^H\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I}) \le t\right) = P\left(\log\det(P_{\text{avg}}\mathbf{Z}\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I}) \le t\right)$$

with $\mathbf{Z} = \mathbf{H}\mathbf{H}^H = \text{diag}(z_1, \ldots, z_N)$.

While we are not aware of any closed-form expression for $p_{\text{out}}(t)$, it can be bounded as follows. Markov's inequality and (8) yield the following lower bound:

$$p_{\text{out}}(t) \ge 1 - \frac{1}{t}E\left[\log\det(P_{\text{avg}}\mathbf{Z}\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I})\right] \ge 1 - \frac{1}{t}\sum_{k=1}^{N}\log\left(P_{\text{avg}}\psi_k^{-1}E[z_1] + 1\right).$$

With the Paley–Zygmund inequality, see, e.g., [15, p. 8], and Theorem 1, an upper bound on $p_{\text{out}}(t)$ is given by

$$\begin{aligned} p_{\text{out}}(t) &\le 1 - \frac{\left(E\left[\log\det(P_{\text{avg}}\mathbf{Z}\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I})\right] - t\right)^2}{E\left[\left(\log\det(P_{\text{avg}}\mathbf{Z}\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I})\right)^2\right]} \\ &\le 1 - \frac{\left(\sum_{k=1}^{N} E[\log(P_{\text{avg}}\psi_k^{-1}z_1 + 1)] - t\right)^2}{E\left[\left(\log\det(P_{\text{avg}}\mathbf{Z}\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I})\right)^2\right]} \end{aligned} \tag{22}$$

for $0 < t \le \sum_{k=1}^{N} E[\log(P_{\text{avg}}\psi_k^{-1}z_1 + 1)]$.

The bound in (22) can further be estimated from above using

$$\begin{aligned} E\left[\left(\log\det(P_{\text{avg}}\mathbf{Z}\mathbf{R}_{\mathbf{n}}^{-1} + \mathbf{I})\right)^2\right] &= E\left[\left(\log\det(P_{\text{avg}}\mathbf{H}\mathbf{U}\mathbf{\Psi}_{\mathbf{n}}^{-1}\mathbf{U}^H\mathbf{H}^H + \mathbf{I})\right)^2\right] \\ &\le E\left[\left(\log\det(P_{\text{avg}}\psi_{\min}^{-1}\mathbf{Z} + \mathbf{I})\right)^2\right] \\ &= E\left[\left(\sum_{k=1}^{N}\log\left(P_{\text{avg}}\psi_{\min}^{-1}z_k + 1\right)\right)^2\right] \\ &= NE\left[\left(\log\left(P_{\text{avg}}\psi_{\min}^{-1}z_1 + 1\right)\right)^2\right] \\ &\quad + N(N-1)\left(E\left[\log\left(P_{\text{avg}}\psi_{\min}^{-1}z_1 + 1\right)\right]\right)^2, \end{aligned}$$

where we have used the spectral decomposition $\mathbf{R}_{\mathbf{n}} = \mathbf{U}\mathbf{\Psi}_{\mathbf{n}}\mathbf{U}^H$ with $\mathbf{U}$ being unitary, $\mathbf{\Psi}_{\mathbf{n}} = \text{diag}(\psi_1, \ldots, \psi_N)$, and where $\psi_{\min}$ is the minimal eigenvalue of $\mathbf{R}_{\mathbf{n}}$.

4.2. Achievable rate of noncoherent Gaussian fading channels. Expectations of the form (1) occur also in the analysis of the achievable rate of noncoherent

Gaussian fading channels where the channel state information is unknown to the receiver [2], [27, Chapter 2]. In this scenario, the complex channel output $\mathbf{y} \in \mathbb{C}^N$ is given by

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{n}. \tag{23}$$

Here, $\mathbf{h} \in \mathbb{C}^N$ is a vector containing the complex channel fading weights, whose probability distribution is again proper complex Gaussian $\mathcal{CN}(\mathbf{0}, \mathbf{R_h})$ with the nonnegative definite Hermitian Toeplitz matrix $\mathbf{R_h}$ describing the covariance of the stationary channel fading process. Moreover, $\mathbf{X}$ is a diagonal matrix whose diagonal entries are provided by the components of the random channel input $\mathbf{x} \in \mathbb{C}^N$. Finally, $\mathbf{n} \in \mathbb{C}^N$ is additive white Gaussian noise such that $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The vector size N denotes the number of channel uses. The random quantities $\mathbf{x}$, $\mathbf{n}$, and $\mathbf{h}$ are mutually independent.

In contrast to the scenario discussed in section 4.1, now the realization of $\mathbf{h}$ is not known to the receiver (*noncoherent setting*). Thus, the channel in (23) is described by the conditional probability density $p(\mathbf{y}|\mathbf{x})$, which is the density of the $\mathcal{CN}\left(\mathbf{0}, \mathbf{X}\mathbf{R_h}\mathbf{X}^H + \mathbf{I}\right)$-distribution. This distribution governs the complex random channel output $\mathbf{y} \in \mathbb{C}^N$, given the random channel input $\mathbf{x} \in \mathbb{C}^N$.

The maximum achievable data rate of this channel, i.e., the Shannon capacity, is given by

$$C_\infty = \lim_{N\to\infty} \left\{ \frac{1}{N} \max_{p(\mathbf{x})\,:\, E[\mathbf{x}^H\mathbf{x}] \le N P_{\text{avg}}} I(\mathbf{x};\mathbf{y}) \right\}, \tag{24}$$

where the maximization is over all input densities $p(\mathbf{x})$ fulfilling the power constraint.

The capacity of the class of noncoherent channels, which is particularly important as it applies to many realistic mobile communication systems, has received a lot of attention in the literature. However, as its treatment is rather difficult, all prior works consider different simplifications. In [20], a block-fading channel model is considered, where the channel is assumed to be constant over a block of symbols and changes independently from block to block. The capacity of stationary fading channels, cf. (24), has been investigated in [17] for an asymptotic high-SNR when P_{avg} is large. In [10], the approximate behavior of the capacity for different SNR regimes has been studied. Moreover, in [25] and [26], the low-SNR behavior of the capacity with a peak power constraint has been investigated, i.e., $|x_k|^2 \le P_{\text{peak}}$, where x_k are the components of $\mathbf{x}$. The works [1], [7], and [21] focus on the achievable rate with input distributions containing training sequences. The training sequences are known at the receiver and are used for channel estimation and coherent detection. In general, however, the capacity of noncoherent stationary fading channels is not known so far.

In the following, as the evaluation of the capacity in (24) turns out to be very difficult, we focus on the achievable data rate with i.i.d. zero-mean proper complex Gaussian input symbols, i.e., instead of considering the maximization in (24), we fix the input distribution to $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, P_{\text{avg}}\mathbf{I})$ for every N. Clearly, this simplification does not yield the capacity. However, this choice is interesting as it is capacity-achieving in case the fading realization $\mathbf{h}$ is known to the receiver (*coherent setting*). As an application of inequalities (4), we can derive bounds on $I(\mathbf{x};\mathbf{y})$ for every N and get bounds on the achievable rate and, in particular, a lower bound on C_∞. It can be shown that the gap between the upper and the lower bound on the achievable rate is less than 2γ [nat/channel use] uniformly for all SNRs with $\gamma \approx 0.57721$ being the Euler constant.

Let N be given. The mutual information $I(\mathbf{x};\mathbf{y})$ can be expressed as

$$I(\mathbf{x};\mathbf{y}) = h(\mathbf{y}) - h(\mathbf{y}|\mathbf{x}) \tag{25}$$

with the differential entropies $h(\mathbf{y}) = E[-\log p(\mathbf{y})]$ and $h(\mathbf{y}|\mathbf{x}) = E[-\log p(\mathbf{y}|\mathbf{x})]$.

The differential entropy $h(\mathbf{y})$ can be bounded as

$$NE\left[\log\left(\pi e(|h_1|^2 P_{\text{avg}} + 1)\right)\right] \le h(\mathbf{y}) \le N \log\left(\pi e(\sigma_h^2 P_{\text{avg}} + 1)\right), \tag{26}$$

where h_1 is the first component of $\mathbf{h}$ and $\sigma_h^2 = E\left[|h_1|^2\right]$. See [8], [22] for the upper bound. The lower bound follows from

$$\begin{aligned} h(\mathbf{y}) \ge h(\mathbf{y}|\mathbf{h}) &= I(\mathbf{x};\mathbf{y}|\mathbf{h}) + h(\mathbf{y}|\mathbf{x},\mathbf{h}) \\ &= NE\left[\log\left(|h_1|^2 P_{\text{avg}} + 1\right)\right] + N\log(\pi e). \end{aligned}$$

As the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ is a $\mathcal{CN}\left(\mathbf{0}, \mathbf{X}\mathbf{R}_{\mathbf{h}}\mathbf{X}^H + \mathbf{I}\right)$-distribution,

$$h(\mathbf{y}|\mathbf{x}) = E\left[\log\det\left(\pi e\left(\mathbf{X}\mathbf{R}_{\mathbf{h}}\mathbf{X}^H + \mathbf{I}\right)\right)\right] = E\left[\log\det\left(\pi e\left(\mathbf{Z}\mathbf{R}_{\mathbf{h}} + \mathbf{I}\right)\right)\right], \tag{27}$$

where $\mathbf{Z} = \mathbf{X}^H\mathbf{X}$. It is rarely possible to compute $h(\mathbf{y}|\mathbf{x})$ explicitly, but since the channel input symbols are identically distributed, expression (27) matches the general form (1), so that we can apply inequalities (4) to obtain the following simple bounds:

$$N\log(\pi e) + \sum_{j=1}^{N} E[\log(|x_1|^2\gamma_j + 1)] \le h(\mathbf{y}|\mathbf{x}) \le N\log(\pi e) + \sum_{j=1}^{N} \log(\gamma_j P_{\text{avg}} + 1), \tag{28}$$

where $\gamma_1, \ldots, \gamma_N$ are the eigenvalues of $\mathbf{R}_{\mathbf{h}}$. Observing that $|x_1|^2$ has an exponential distribution with mean P_{avg}, we can express the expectations in the lower bound in terms of the exponential integral and obtain

$$h(\mathbf{y}|\mathbf{x}) \ge N\log(\pi e) - \sum_{j:\gamma_j>0} e^{1/(\gamma_j P_{\text{avg}})}\text{Ei}\left(-\frac{1}{\gamma_j P_{\text{avg}}}\right). \tag{29}$$

From (28), (29) together with (25), (26), we get

$$\begin{aligned} &-Ne^{1/(\sigma_h^2 P_{\text{avg}})}\text{Ei}\left(-\frac{1}{\sigma_h^2 P_{\text{avg}}}\right) - \sum_{j=1}^{N}\log(\gamma_j P_{\text{avg}} + 1) \\ &\qquad \le I(\mathbf{x};\mathbf{y}) \le N\log\left(\sigma_h^2 P_{\text{avg}} + 1\right) + \sum_{j:\gamma_j>0} e^{1/(\gamma_j P_{\text{avg}})}\text{Ei}\left(-\frac{1}{\gamma_j P_{\text{avg}}}\right). \end{aligned} \tag{30}$$

By (18), the gap between the upper bound and the lower bound in (30) is not larger than $2N\gamma$.

As an example, consider a stationary 1-dependent channel fading process, so that the covariance matrix $\mathbf{R}_{\mathbf{h}}$ is a nonnegative definite tridiagonal matrix with common diagonal entries 1, say, and common superdiagonal entries $\rho \neq 0$. The eigenvalues of $\mathbf{R}_{\mathbf{h}}$ are

$$\gamma_j = 1 - 2|\rho|\cos\frac{j\pi}{N+1}, \qquad j = 1, \ldots, N;$$

see, e.g., [12, p. 67].

If $\rho \to 0$, then $\gamma_j \to 1$ and the difference between $h(\mathbf{y}|\mathbf{x})$ and the lower bound in (29) converges to 0. If $P_{\text{avg}} \to 0$, then the upper bound and the lower bound on $h(\mathbf{y}|\mathbf{x})$ converge to $N \log(\pi e)$, so that the difference between $h(\mathbf{y}|\mathbf{x})$ and either bound converges to 0. The latter holds also for the bounds on $I(\mathbf{x};\mathbf{y})$ in (30). Note that for the bounds on $h(\mathbf{y}|\mathbf{x})$, no assumption on the dependence structure of the x_j is required.

Using a Riemann sum argument, we now obtain the following bounds on the capacity and the achievable rate:

$$C_\infty \geq \lim_{N\to\infty} \frac{1}{N} I(\mathbf{x};\mathbf{y}) \geq -e^{1/P_{\text{avg}}} \text{Ei}\left(-\frac{1}{P_{\text{avg}}}\right) - \int_0^1 \log(\{1-2|\rho|\cos(u\pi)\}P_{\text{avg}}+1)\, du$$

and

$$\lim_{N\to\infty} \frac{1}{N} I(\mathbf{x};\mathbf{y}) \leq \log\left(P_{\text{avg}}+1\right) + \int_0^1 \int_{-\infty}^0 \frac{e^t}{t - [\{1-2|\rho|\cos(u\pi)\}P_{\text{avg}}]^{-1}}\, dt\, du,$$

provided $|\rho| < \frac{1}{2}$.

In conclusion, based on the new inequality (3), we have been able to derive bounds on the achievable rate with i.i.d. zero-mean proper complex Gaussian input symbols which are tight in the sense that the gap between the upper and the lower bound is bounded by 2γ independent of P_{avg}, i.e., the SNR. Thus, these bounds hold in particular in the practically relevant mid-SNR regime. By contrast, most existing literature focuses on specific SNR regimes, mostly either on the low- or on the high-SNR regime.

Appendix A. Proof of Lemma 1. We have to show that the maximum of

$$\Psi(\mathbf{Q}) = E\left[\log\det\left(\mathbf{W}\mathbf{Q}\mathbf{W}^H\mathbf{B}^{-1} + \mathbf{I}\right)\right]$$

over all nonnegative definite matrices $\mathbf{Q}$ with $\text{tr}(\mathbf{Q}) \leq c$ is achieved by a diagonal matrix $\mathbf{Q}$.

Let $\mathcal{S} = \{\text{diag}(\mathbf{s}) : \mathbf{s} \in \{-1,1\}^N\}$. Obviously, if $\mathbf{Q}$ is nonnegative definite, then so is $\mathbf{SQS}$ for every $\mathbf{S} \in \mathcal{S}$. The function Ψ is concave on the set of nonnegative definite matrices $\mathbf{Q}$. Moreover, the following invariance property holds:

$$\Psi(\mathbf{SQS}) = \Psi(\mathbf{Q}) \quad \text{for all } \mathbf{S} \in \mathcal{S} \text{ and all nonnegative definite } \mathbf{Q}$$

as

$$\mathbf{W}(\mathbf{SQS})\mathbf{W}^H = (\mathbf{WS})\mathbf{Q}(\mathbf{WS})^H$$

and $\mathbf{WS}$ is diagonal with independent diagonal elements having the same zero-mean proper complex Gaussian distribution as the corresponding diagonal elements of $\mathbf{W}$. That is, $\mathbf{WS}$ and $\mathbf{W}$ have the same distribution and

$$\begin{aligned}\Psi(\mathbf{SQS}) &= E\left[\log\det\left((\mathbf{WS})\mathbf{Q}(\mathbf{WS})^H\mathbf{B}^{-1} + \mathbf{I}\right)\right]\\ &= E\left[\log\det\left(\mathbf{W}\mathbf{Q}\mathbf{W}^H\mathbf{B}^{-1} + \mathbf{I}\right)\right]\\ &= \Psi(\mathbf{Q}).\end{aligned}$$

For every nonnegative definite $\mathbf{Q}$ with $\text{tr}(\mathbf{Q}) \leq c$, let $\bar{\mathbf{Q}}$ be the diagonal matrix with the diagonal elements of $\mathbf{Q}$ on its diagonal. Clearly, $\bar{\mathbf{Q}}$ is again nonnegative definite with $\text{tr}(\bar{\mathbf{Q}}) \leq c$. Moreover,

$$\bar{\mathbf{Q}} = 2^{-N} \sum_{\mathbf{S}\in\mathcal{S}} \mathbf{SQS},$$

and with the concavity and invariance of Ψ, it follows that

$$\Psi(\bar{\mathbf{Q}}) \geq 2^{-N} \sum_{\mathbf{S}\in\mathcal{S}} \Psi(\mathbf{SQS}) = \Psi(\mathbf{Q}).$$

Hence, $\Psi(\mathbf{Q})$ is maximized by a diagonal $\mathbf{Q}$ on the set of nonnegative definite matrices with $\mathrm{tr}(\mathbf{Q}) \leq c$. ▯

Appendix B. Proof of Lemma 2. We have to show that the maximum of

$$\begin{aligned}\Psi(\mathbf{\Lambda}) &= E\left[\log\det\left(\mathbf{W\Lambda B}^{-1} + \mathbf{I}\right)\right]\\ &= E\left[\log\det\left(\mathbf{W\Lambda} + \mathbf{B}\right)\right] - \log\det\left(\mathbf{B}\right)\end{aligned}$$

over all nonnegative diagonal matrices $\mathbf{\Lambda}$ with $\mathrm{tr}(\mathbf{\Lambda}) \leq c$ is achieved by $\mathbf{\Lambda} = \frac{c}{N}\mathbf{I}$.

Since $\mathbf{B}$ is a balanced matrix, there exists a nonempty balanced set $\mathcal{P}$ of permutations satisfying (16). For every $\pi \in \mathcal{P}$ and nonnegative diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, let $\mathbf{\Lambda}_\pi = \mathrm{diag}(\lambda_{\pi(1)}, \ldots, \lambda_{\pi(N)})$. Using (16) and the fact that the diagonal elements of the diagonal matrix $\mathbf{W}$ are exchangeable random variables, it can easily be verified that

$$\Psi(\mathbf{\Lambda}_\pi) = \Psi(\mathbf{\Lambda}) \quad \text{for all } \pi \in \mathcal{P}.$$

Using this invariance property, the concavity of $\log\det$ on the set of positive definite matrices, and (5), it follows that

$$\Psi(\mathbf{\Lambda}) = \frac{1}{|\mathcal{P}|}\sum_{\pi\in\mathcal{P}} \Psi(\mathbf{\Lambda}_\pi) \leq \Psi\left(\frac{1}{|\mathcal{P}|}\sum_{\pi\in\mathcal{P}} \mathbf{\Lambda}_\pi\right) = \Psi\left(\frac{\mathrm{tr}(\mathbf{\Lambda})}{N}\mathbf{I}\right) \leq \Psi\left(\frac{c}{N}\mathbf{I}\right)$$

if $\mathrm{tr}(\mathbf{\Lambda}) \leq c$. This concludes the proof. ▯

Acknowledgment. We thank the referees for very useful comments.

REFERENCES

[1] J. Baltersee, G. Fock, and H. Meyr, *An information theoretic foundation of synchronized detection*, IEEE Trans. Commun., 49 (2001), pp. 2115–2123.

[2] E. Biglieri, J. Proakis, and S. Shamai (Shitz), *Fading channels: Information-theoretic and communications aspects*, IEEE Trans. Inform. Theory, 44 (1998), pp. 2619–2692.

[3] P. Billingsley, *Probability and Measure*, 3rd ed., Wiley, New York, 1995.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, New York, 2006.

[5] J. D. Dixon and B. Mortimer, *Permutation Groups*, Springer, New York, 1996.

[6] M. Dörpinghaus, N. Gaffke, L. A. Imhof, and R. Mathar, *Convergence of the conditional per symbol entropy for stationary Gaussian fading channels for almost all input sequences*, in Proceedings of the 2013 Information Theory and Applications Workshop, San Diego, CA, 2013.

[7] M. Dörpinghaus, A. Ispas, and H. Meyr, *On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels*, IEEE Trans. Inform. Theory, 58 (2012), pp. 2963–2982.

[8] M. Dörpinghaus, H. Meyr, and R. Mathar, *On the achievable rate of stationary Rayleigh flat-fading channels with Gaussian inputs*, IEEE Trans. Inform. Theory, 59 (2013), pp. 2208–2220.

[9] R. M. Dudley, *Real Analysis and Probability*, Wadsworth & Brooks/Cole, Pacific Grove, CA, 1989.

[10] R. H. Etkin and D. N. C. Tse, *Degrees of freedom in some underspread MIMO fading channels*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1576–1608.

[11] R. M. Gray, *Toeplitz and circulant matrices: A review*, Found. Trends Commun. Inform. Theory, 2 (2006), pp. 155–239.
[12] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, University of California Press, Berkeley, CA, 1958.
[13] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, Cambridge, 1952.
[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
[15] J.-P. Kahane, *Some Random Series of Functions*, 2nd ed., Cambridge University Press, Cambridge, 1985.
[16] Y.-H. Kim, *Feedback capacity of stationary Gaussian channels*, IEEE Trans. Inform. Theory, 56 (2010), pp. 57–85.
[17] A. Lapidoth, *On the asymptotic capacity of stationary Gaussian fading channels*, IEEE Trans. Inform. Theory, 51 (2005), pp. 437–446.
[18] G. G. Lorentz, *An inequality for rearrangements*, Amer. Math. Monthly, 60 (1953), pp. 176–179.
[19] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., Springer, New York, 2011.
[20] T. L. Marzetta and B. M. Hochwald, *Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading*, IEEE Trans. Inform. Theory, 45 (1999), pp. 139–157.
[21] M. Médard, *The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel*, IEEE Trans. Inform. Theory, 46 (2000), pp. 933–946.
[22] F. D. Neeser and J. L. Massey, *Proper complex random processes with applications to information theory*, IEEE Trans. Inform. Theory, 39 (1993), pp. 1293–1302.
[23] R. Rado, *An inequality*, J. London Math. Soc., 27 (1952), pp. 1–6.
[24] L. Rüschendorf, *Solution of a statistical optimization problem by rearrangement methods*, Metrika, 30 (1983), pp. 55–61.
[25] V. Sethuraman, B. Hajek, and K. Narayanan, *Capacity bounds for noncoherent fading channels with a peak constraint*, in Proceedings of the IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia, 2005, pp. 515–519.
[26] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, *Low-SNR capacity of noncoherent fading channels*, IEEE Trans. Inform. Theory, 55 (2009), pp. 1555–1574.
[27] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, 2005.
[28] L. Wang and M. Madiman, *Beyond the entropy power inequality, via rearrangements*, IEEE Trans. Inform. Theory, 60 (2014), pp. 5116–5137.
[29] D. V. Widder, *The Laplace Transform*, Princeton University Press, Princeton, NJ, 1941.