Error Exponents in Distributed Hypothesis Testing of Correlations

Uri Hadar, Jingbo Liu, Yury Polyanskiy and Ofer Shayevitz

***Abstract*—We study a distributed hypothesis testing problem where two parties observe i.i.d. samples from two ρ-correlated standard normal random variables X and Y. The party that observes the X-samples can communicate R bits per sample to the second party, that observes the Y-samples, in order to test between two correlation values. We investigate the best possible type-II error subject to a fixed type-I error, and derive an upper (impossibility) bound on the associated type-II error exponent. Our techniques include representing the conditional Y-samples as a trajectory of the Ornstein-Uhlenbeck process, and bounding the associated KL divergence using the subadditivity of the Wasserstein distance and the Gaussian Talagrand inequality.**

I. Introduction

We study the problem of distributively testing between two possible correlation values ρ_0 or ρ_1 of a bivariate Gaussian vector (X, Y) from i.i.d. samples, where the party observing the X-samples can send a message of rate R bits per sample to the party observing the Y-samples. We investigate the best possible type-II error exponent, under the constraint that the type-I error is small.

The problem of hypothesis testing with communication constraints was introduced by Berger in [1] and was addressed by several authors where the vast majority of works deals with achievability schemes, e.g., [2], [3], [4], [5], [6], [7]. Several extensions where proposed e.g. distributed hypothesis testing with interaction [8], feedback [9], security considerations [10] and noisy channels [11]. The first work that introduced nontrivial upper (converse) bounds to the error exponent of the general problem of distributed hypothesis testing was by Rahman and Wagner [12]. They applied their general result to the Gaussian case and showed (numerically) that the obtained upper bound is close to the lower bound of Ahlswede and Csiszár [2]. A related setup was addressed by the present authors in a recent work [13] in the context of distributed correlation estimation, which is reduced to hypothesis testing with vanishing $|\rho_1 - \rho_0|$. In that work we obtained lower bounds on the minimax mean squared error in a regime where the amount of communication is fixed (i.e. $R \to 0$) and interaction is allowed.

In this work we obtain a new upper bound on the error exponent that holds for all positive correlation values, and improves upon [12] for certain values of ρ_0, ρ_1, including regions where [12] is not defined. Loosely speaking, our approach is the following. We represent the Y-samples, conditioned on the message, as a trajectory of the Ornstein-Uhlenbeck process Y_t, where the process time-index t corresponds to a correlation value of $\rho = e^{-t}$. The error exponent can be upper bounded using the data processing inequality, by controlling the expectation of the KL divergence $D(Y_{t_0} \| Y_{t_1})$. We are therefore interested in relating $D(Y_{t_0} \| Y_{t_1})$ to the communication constraint, which is shown to be upper bounded by the expectation of $D(Y_0 \| Y_\infty)$. The aforementioned relation is established by first upper bounding the KL divergence using the quadratic Wasserstein distance $W_2(Y_{t_0}, Y_{t_1})$, then relating $W_2(Y_{t_0}, Y_{t_1})$ to $W_2(Y_0, Y_\infty)$ via subadditivity of the Wassertein distance, and finally by applying the Gaussian Talagrand inequality to bound the Wasserstein distance by the KL divergence. We further optimize the obtained bound using a device of shifting the correlation values by introducing common randomness, similarly to the method proposed in [12].

Let us proceed to formally define the problem. Alice observes X^n and Bob observes Y^n, where $\{X_k, Y_k\}_{k=1}^n$ are i.i.d. pairs of correlated Gaussian r.v.s. It is known that the correlation is either ρ_0 or ρ_1. Alice can send Bob a message W of rate R bits per sample, i.e., such that $W - X^n - Y^n$ and $H(W) \le nR$. We call such a message a *rate-R encoding*. Bob in turn decides between ρ_0 and ρ_1, based on (W, Y^n). His binary decision is denoted here by $U \in \{0, 1\}$, where $U - (Y^n, W) - X^n$. Below, we use P and Q to denote probabilities of r.v.s under correlations ρ_0 and ρ_1, respectively. Define

$$\gamma^*(n, R, \epsilon) \triangleq \inf Q_U(0) \tag{1}$$

where the infimum is taken over all rate-R encodings W and all binary decisions U satisfying $P_U(1) \le \epsilon$. Namely, γ^* is the minimal possible false alarm probability for a misdetect probability of at most ϵ.

We are interested in the *Stein exponent*, defined as

$$s(R) \triangleq \inf_{\epsilon > 0} \limsup_{n \to \infty} -\frac{1}{n} \log \gamma^*(n, R, \epsilon). \tag{2}$$

The case where $\rho_1 = 0$ corresponds to the so-called special setup of *testing against independence*, namely where X and

Order of authors is alphabetical. U.H. and O.S. {emails: urihadar@mail.tau.ac.il, ofersha@eng.tau.ac.il} are with the Department of Electrical Engineering–Systems, Tel Aviv University, Tel Aviv, Israel. J.L. and Y.P. {emails: jingbo@mit.edu, yp@mit.edu} are with the Institute for Data, Systems, and Society and the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. This material is based upon work supported by the National Science Foundation CAREER award under grant agreement CCF-12-53205, the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370, the European Research Council, under grant agreement 639573, the Israeli Science Foundation, under grant agreement 1367/14, and the Yitzhak and Chaya Weinstein Research Institute for Signal Processing.

Y are independent under the alternative hypothesis (with the same marginals under both hypotheses). This problem was introduced and completely solved by Ahlswede and Csiszár for general distributions [2] yielding

$$s(R) = \max I(U;Y) \tag{3}$$

where the maximization is taken over all U such that $U - X - Y$, and $I(X;U) \leq R$. This exact characterization, and in particular the upper (converse) bound, is facilitated by the fact that the KL divergence associated with the type-II error becomes a mutual information in this special case.

The general case of testing between two arbitrary distributions is notoriously difficult. In the same paper [2], Ahlswede and Csiszár obtained a lower (achievability) bound given by

$$s(R) \geq \max\{D(P_X \| Q_X) + D(P_{UY} \| Q_{UY})\} \tag{4}$$

where the maximization is over all r.v.s U satisfying $I(U;X) \leq R$ and $U - X - Y$. This was later improved in [3] and [4]. Specializing to our Gaussian case and using a Gaussian channel $P_{U|X}$ (this was not shown to be the maximizer) yields

$$\begin{aligned} s(R) \geq \frac{1}{2} \log \frac{1-\rho_1^2(1-2^{-2R})}{1-\rho_0^2(1-2^{-2R})} \\ - \frac{\rho_1(\rho_0-\rho_1)(1-2^{-2R}) \log e}{1-\rho_1^2(1-2^{-2R})}. \end{aligned} \tag{5}$$

Note that the rates and log in this paper have base 2. As mentioned in [12], one can use the lower bounds of [3], [4], but they are quite complicated, and Ahlswede and Csiszár's lower bound is close to the upper bound of [12] for some cases.

To the best of our knowledge, the only nontrivial upper bound in the general setting, as well as in the Gaussian setup we consider here, is by Rahman and Wagner [12] (building on an original converse technique of introducing conditional independence by Wagner and Anantharam [14] in the context of distributed source coding). In that work, the authors cleverly augment the probability space with an additional side-information Z supplied to both Alice and Bob, such that $X - Z - Y$ holds under the alternative hypothesis. They then proceed to solve this new problem of *testing against conditional independence*, again utilizing the fact that the KL divergence reduces to a mutual information, and finally they optimize over the choice of Z to obtain the tightest upper (impossibility) bound on the Stein exponent with no side-information. As an example, they applied their bound to the Gaussian case considered in this paper. In the case where both correlations have the same sign (say, positive), which is the case we handle in this paper, their bound is nontrivial in the region $0 \leq \rho_1 \leq \frac{\rho_0+1}{2}$, where it is given by

$$s(R) \leq \frac{1}{2} \log \frac{1}{1-\left(\frac{\rho_0-\rho_1}{1-\rho_1}\right)^2 (1-2^{-2R})}. \tag{6}$$

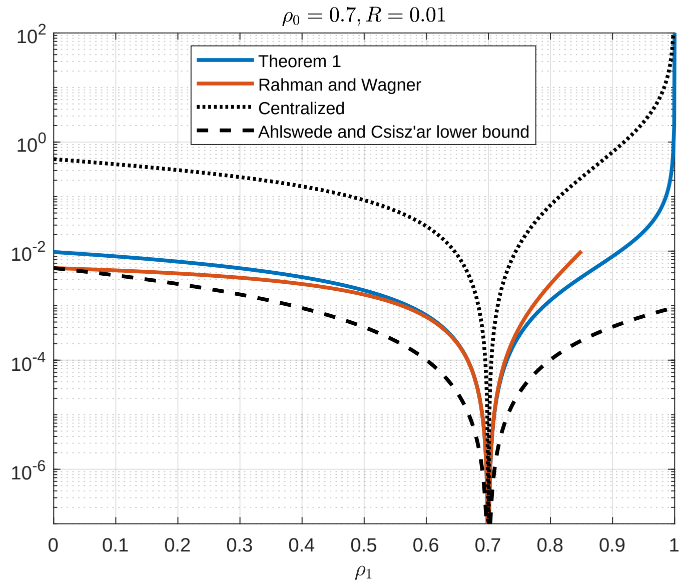

Fig. 1. Bounds for the error exponent. Depicted as a function of ρ_1 for fixed values of ρ_0 and R is our bound (8) along with the centralized (7), Rahman and Wagner's (6) (wherever it is defined) and the *lower* bound of Ahlswede and Csiszár (5).

Outside that region, the only known upper bound is the trivial centralized bound, i.e. where $R = \infty$, which is easily obtained using Stein's lemma to yield (see [12])

$$s(\infty) = \frac{1}{2} \log \frac{1-\rho_1^2}{1-\rho_0^2} - \frac{\rho_1(\rho_0-\rho_1) \log e}{1-\rho_1^2} \tag{7}$$

II. Main Result

Our main result is a new upper bound on the Stein exponent in the Gaussian setup. We stress that while tailored to the Gaussian case, our approach attacks the problem of general correlations without resorting to a direct reduction to a case of independent samples.

Theorem 1. *For any $\rho_0, \rho_1 \geq 0$, the Stein exponent satisfies*

$$s(R) \leq \frac{R}{\left(\frac{1-\min\{\rho_0,\rho_1\}}{\rho_1-\rho_0}\right)^2 - 1}. \tag{8}$$

We prove Theorem 1 in Section III. Our proof makes use of dimension-independent arguments, hence our bound is naturally nontrivial only for very small rates, but can nevertheless improve upon the known bounds in such regions. This is demonstrated in Figure 1, which depicts our upper bound for $R = 0.01$ as a function of ρ_1, along with the centralized upper bound as well as Rahman and Wagner's upper bound (wherever it holds), and also Ahlswede and Csiszár's lower bound. Figure 2 shows these bounds as a function of R for fixed correlations values.

The result we obtained for the Gaussian case can be applied to the case where X, Y are ρ'-correlated binary r.v.s, i.e. where

$$\Pr(X = x, Y = y) = \frac{1 + (-1)^{\mathbf{1}(x \neq y)} \rho'}{4}. \tag{9}$$

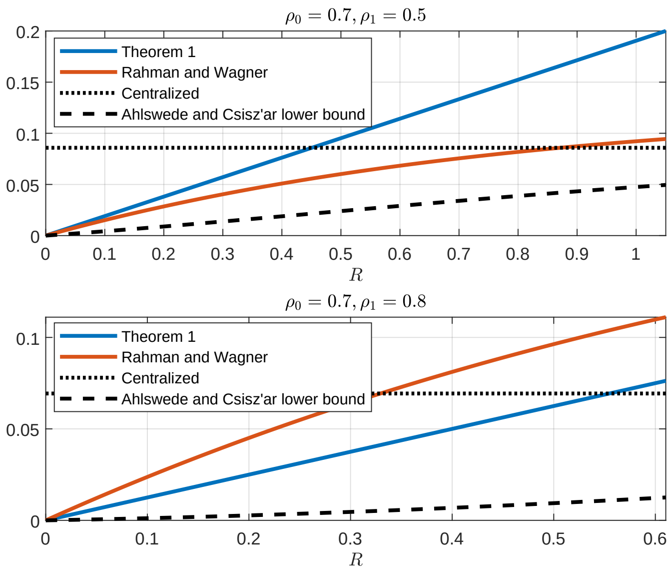

Fig. 2. Bounds for the error exponent. Depicted as a function of the rate R for fixed values of ρ_0 and ρ_1 (for which Rahman and Wagner's bound is defined) is our bound (8) along with the centralized (7), Rahman and Wagner's (6) and the *lower* bound of Ahlswede and Csiszár (5).

This is done by having both terminals take the signs of their local Gaussian measurements to obtain binary ones with correlation $\rho' = \frac{2}{\pi} \arcsin \rho$. This immediately implies the following.

Corollary 2. *Theorem 1 holds for the case where X, Y are ρ-correlated binary with (ρ_0, ρ_1) replaced by $(\sin \frac{\pi\rho_0}{2}, \sin \frac{\pi\rho_1}{2})$.*

III. Proof of Main Result

The following Lemma is key in proving Theorem 1.

Lemma 3. *For any $\rho_0, \rho_1 \geq 0$ and any encoding W of X^n,*

$$\mathbb{E}D(P_{Y^n|W}(\cdot|W) \| Q_{Y^n|W}(\cdot|W)) \tag{10}$$

$$\leq \frac{(\rho_0 + \rho_1)^2}{1 - \max\{\rho_0^2, \rho_1^2\}} \cdot H(W). \tag{11}$$

Before we prove Lemma 3 we need the following result. Recall that in one dimension, the Ornstein-Uhlenbeck process is a continuous-time Markov process with generator

$$\mathcal{L}f(x) = f''(x) - xf'(x) \tag{12}$$

and corresponding semigroup operator

$$P_t f(x) = \mathbb{E}f\left(e^{-t}x + \sqrt{1 - e^{-t}}Z\right), \quad Z \sim \mathcal{N}(0,1). \tag{13}$$

It admits $\mathcal{N}(0,1)$ as a stationary distribution. As can be seen, the connection to our setup is that the induced channel at time t is Gaussian $Y = e^{-t}X + \sqrt{1 - e^{-t}}Z$, with e^{-t} as the correlation parameter. The associated n-dimensional Ornstein-Uhlenbeck semigroup over $\mathbb{R}^n$ is the product semigroup $P_t^{\otimes n}$.

Lemma 4. *Let μ_t be the distribution at time t of the Markov process associated with the n-dimensional Ornstein-Uhlenbeck semigroup. Then for any $t, \tau \geq 0$,*

$$W_2(\mu_t, \mu_\tau) \leq (e^{-t} + e^{-\tau}) W_2(\mu_0, \mu_\infty). \tag{14}$$

Proof. In the following we consider the scalar case (i.e. $n = 1$), which extends to the vector case by applying it per coordinate. Let $X \sim \mu_0$, $Z \sim \mu_\infty = \mathcal{N}(0,1)$ and Z' an independent copy of Z. For any $0 \leq s \leq t$, using the triangle inequality we have

$$W_2(\mu_s, \mu_t) \tag{15}$$

$$= W_2(e^{-s}X + \sqrt{1 - e^{-2s}}Z, e^{-t}X + \sqrt{1 - e^{-2t}}Z) \tag{16}$$

$$\leq W_2(e^{-s}X + \sqrt{1 - e^{-2s}}Z, Z)$$

$$+ W_2(e^{-t}X + \sqrt{1 - e^{-2t}}Z, Z). \tag{17}$$

From the subadditivity of the Wasserstein distance (see e.g. [15, Proposition 7.17]), we can further bound

$$W_2(e^{-s}X + \sqrt{1 - e^{-2s}}Z, Z) \tag{18}$$

$$= W_2(e^{-s}X + \sqrt{1 - e^{-2s}}Z, e^{-s}Z + \sqrt{1 - e^{-2s}}Z') \tag{19}$$

$$\leq W_2(e^{-s}X, e^{-s}Z) \tag{20}$$

$$= e^{-s}W_2(X, Z). \tag{21}$$

Thus

$$W_2(\mu_s, \mu_t) \leq (e^{-s} + e^{-t}) W_2(X, Z). \tag{22}$$

□

Remark 5. We can also bound $W_2(\mu_t, \mu_\tau)$ using Otto-Villani's short-time estimates for gradient flows [16, eq. (24.21)],

$$W_2(\mu_t, \mu_\tau) \leq \min\left\{\frac{|\tau - t|}{\min\{t, \tau\}}, \sqrt{\frac{|\tau - t|}{\min\{t, \tau\}}}\right\} W_2(\mu_0, \mu_\infty). \tag{23}$$

However, this does not further improves our bound.

Proof of Lemma 3. We use the natural base for the divergence in this proof. Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_{n\times n})$, and let $(\mathbf{A}, \mathbf{B})$ be any jointly distributed vectors in $\mathbb{R}^n$, independent of $\mathbf{Z}$. The divergence between two normal distributions is given by

$$D(\mathcal{N}(\mathbf{a}, \sigma^2 I_{n\times n}) \| \mathcal{N}(\mathbf{b}, \sigma^2 I_{n\times n})) = \frac{\|\mathbf{a} - \mathbf{b}\|^2}{2\sigma^2}. \tag{24}$$

From the joint convexity of $D(P\|Q)$ in P and Q, we have

$$D(p_{\mathbf{A}+\mathbf{Z}} \| p_{\mathbf{B}+\mathbf{Z}}) \leq \mathbb{E}D(\mathcal{N}(\mathbf{A}, \sigma^2 I_{n\times n}) \| \mathcal{N}(\mathbf{B}, \sigma^2 I_{n\times n})) \tag{25}$$

$$= \frac{\mathbb{E}\|\mathbf{A} - \mathbf{B}\|^2}{2\sigma^2}. \tag{26}$$

Let $\mu_t(w)$ be the distribution at time t of the Markov process associated with the Ornstein-Uhlenbeck semigroup initialized with $\mu_0(w) = p_{\mathbf{X}|W=w}$. Recall that $p_{\mathbf{X}W}$ does not depend on the correlation, and that the process is ergodic, with $\mu_\infty = \mu$

where $\mu = \mathcal{N}(0,1)^{\otimes n}$ is the stationary distribution. Recall also that the correlation at time t is e^{-t}.

Now, set

$$t = -\ln \max\{\rho_0, \rho_1\} \tag{27}$$

$$\tau = -\ln \min\{\rho_0, \rho_1\} \tag{28}$$

and pick any $0 \leq \alpha < t$. Set the marginals $\mathbf{A} \sim \mu_{\tau-\alpha}$ and $\mathbf{B} \sim \mu_{t-\alpha}$, and the noise level $\sigma^2 = 1 - e^{-2\alpha}$. Then

$$D(\mu_\tau \| \mu_t) = D(p_{e^{-\alpha}\mathbf{A}+\mathbf{Z}} \| p_{e^{-\alpha}\mathbf{B}+\mathbf{Z}}) \tag{29}$$

$$\leq \frac{e^{-2\alpha}\mathbb{E}\|\mathbf{A}-\mathbf{B}\|^2}{2(1-e^{-2\alpha})}. \tag{30}$$

Minimizing over all the couplings $(\mathbf{A}, \mathbf{B})$ with these marginals, we obtain

$$D(\mu_\tau \| \mu_t) \leq \frac{e^{-2\alpha}}{2(1-e^{-2\alpha})} W_2^2(\mu_{t-\alpha}, \mu_{\tau-\alpha}). \tag{31}$$

where W_2 is the 2-Wasserstein metric induced by the Euclidean metric (Bound (30) was the basis of [17]).

Applying Lemma 4, followed by the well-known Gaussian Talagrand inequality (see e.g. [18, Theorem 8.10]), which states

$$W_2^2(\mu_0, \mu_\infty) \leq 2D(\mu_0 \| \mu_\infty), \tag{32}$$

yields

$$D(\mu_\tau \| \mu_t) \leq \frac{(e^{-(t-\alpha)} + e^{-(\tau-\alpha)})^2}{e^{2\alpha}-1} D(\mu_0 \| \mu_\infty), \tag{33}$$

which is minimized at $\alpha = t$.

Recalling that $p_{\mathbf{X}W}$ does not depend on the correlation, we have

$$I(\mathbf{X}; W) = D(p_{\mathbf{X}W} \| p_{\mathbf{X}} p_W) \tag{34}$$

$$= \mathbb{E}D(\mu_0(W) \| p_{\mathbf{X}}) \tag{35}$$

$$= \mathbb{E}D(\mu_0(W) \| \mu_\infty(W)). \tag{36}$$

Thus, using (33), we obtain

$$\mathbb{E}D(\mu_\tau(W) \| \mu_t(W)) \leq \frac{(e^{-t} + e^{-\tau})^2}{1 - e^{-2t}} I(\mathbf{X}; W). \tag{37}$$

The claim now follows due to the symmetry of (29) in t, τ, and since $I(\mathbf{X}; W) \leq H(W)$. $\square$

We can now prove our main result.

Proof of Theorem 1. Let W be a rate-R encoding of X^n on Alice's side, and U be a binary decision using (W, Y^n) on Bob's side, such that $P_U(1) = \epsilon' \leq \epsilon$. Write $\gamma = Q_U(0)$, and $d(\cdot\|\cdot)$ for the binary KL divergence. Then by the data processing inequality,

$$d(1-\epsilon' \| \gamma) = D(P_U \| Q_U) \tag{38}$$

$$\leq D(P_{Y^n W} \| Q_{Y^n W}). \tag{39}$$

We now introduce common randomness in order to shift the correlations, similar to the method proposed in [12], as follows. Let Z^n be i.i.d. $\mathcal{N}(0,1)$ independent of (X^n, Y^n), and let, for every $i \in [n]$ and some $\lambda \in [0,1]$,

$$X_i' = \lambda Z_i + \sqrt{1-\lambda^2} X_i \tag{40}$$

$$Y_i' = \lambda Z_i + \sqrt{1-\lambda^2} Y_i. \tag{41}$$

It follows that if (X_i, Y_i) are bivariate standard normal with correlation ρ then (X_i', Y_i') are bivariate standard normal with correlation $\lambda^2 + (1-\lambda^2)\rho$. For any $\lambda^2 \leq \min\{\rho_0, \rho_1\}$ denote by P^λ, Q^λ the respective distributions with correlations $\frac{\rho_0-\lambda^2}{1-\lambda^2}, \frac{\rho_1-\lambda^2}{1-\lambda^2}$. Then

$$d(1-\epsilon' \| \gamma) \leq \mathbb{E}D(P_{Y^n W|Z^n} \| Q_{Y^n W|Z^n}) \tag{42}$$

$$= D(P^\lambda_{Y^n W} \| Q^\lambda_{Y^n W}) \tag{43}$$

$$= \mathbb{E}D(P^\lambda_{Y^n|W} \| Q^\lambda_{Y^n|W}) \tag{44}$$

$$\leq \frac{(\frac{\rho_0-\lambda^2}{1-\lambda^2} + \frac{\rho_1-\lambda^2}{1-\lambda^2})^2}{1 - \max\{(\frac{\rho_0-\lambda^2}{1-\lambda^2})^2, (\frac{\rho_0-\lambda^2}{1-\lambda^2})^2\}} \cdot H(W) \tag{45}$$

by Lemma 3, which can be shown to be minimized by $\lambda^2 = \min\{\rho_0, \rho_1\}$ to yield

$$d(1-\epsilon' \| \gamma) \leq \frac{(\rho_1 - \rho_0)^2}{(1 - \min\{\rho_0, \rho_1\})^2 - (\rho_1 - \rho_0)^2} \cdot H(W). \tag{46}$$

On the other hand, we have

$$d(1-\epsilon' \| \gamma) \geq (1-\epsilon') \log(1/\gamma) - h(\epsilon'). \tag{47}$$

Hence

$$\log(1/\gamma) \leq \frac{\frac{(\rho_1-\rho_0)^2}{(1-\min\{\rho_0,\rho_1\})^2 - (\rho_1-\rho_0)^2} nR + h(\epsilon')}{1-\epsilon'} \tag{48}$$

and the claim follows. $\square$

ACKNOWLEDGEMENT

We thank Yuval Kochman, Yihong Wu, Himanshu Tyagi and Sahasranand KR for useful discussions.

REFERENCES

[1] T. Berger, "Decentralized estimation and decision theory," in *IEEE Seven Springs Workshop on Information Theory, Mt. Kisco, NY*, 1979.

[2] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE transactions on information theory*, vol. 32, no. 4, 1986.

[3] T. Han, "Hypothesis testing with multiterminal data compression," *IEEE transactions on information theory*, vol. 33, no. 6, pp. 759–772, 1987.

[4] H. Shimokawa, S. Amari, *et al.*, "Error bound of hypothesis testing with data compression," in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*, p. 114, IEEE, 1994.

[5] T. S. Han and S.-i. Amari, "Statistical inference under multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2300–2324, 1998.

[6] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via Körner-Marton coding," in *Information Theory Workshop (ITW), 2016 IEEE*, pp. 146–150, IEEE, 2016.

[7] K. Sahasranand and H. Tyagi, "Extra samples can reduce the communication for independence testing," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 2316–2320, IEEE, 2018.

[8] Y. Xiang and Y.-H. Kim, "Interactive hypothesis testing with communication constraints," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pp. 1065–1072, IEEE, 2012.

[9] D. Pados, K. W. Halford, D. Kazakos, and P. Papantoni-Kazakos, "Distributed binary hypothesis testing with feedback," *IEEE transactions on systems, man, and cybernetics*, vol. 25, no. 1, pp. 21–42, 1995.
[10] M. Mhanna and P. Piantanida, "On secure distributed hypothesis testing," in *Information Theory (ISIT), 2015 IEEE International Symposium on*, pp. 1605–1609, IEEE, 2015.
[11] S. Sreekumar and D. Gündüz, "Distributed hypothesis testing over noisy channels," in *Information Theory (ISIT), 2017 IEEE International Symposium on*, pp. 983–987, IEEE, 2017.
[12] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6282–6303, 2012.
[13] U. Hadar, J. Liu, Y. Polyanskiy, and O. Shayevitz, "Communication complexity of estimating correlations," in *Proceedings of the 51st Annual ACM SIGACT Symposium on the Theory of Computing (STOC 19)*, ACM, 2019.
[14] A. B. Wagner and V. Anantharam, "An improved outer bound for multiterminal source coding," *IEEE Trans. Inf. Theory*, vol. 54, pp. 1919–1937, May 2008.
[15] C. Villani, *Topics in optimal transportation*. No. 58, American Mathematical Soc., 2003.
[16] C. Villani, *Optimal transport: old and new*, vol. 338. Springer Science & Business Media, 2008.
[17] Y. Wu, "A simple proof of the Gaussian HWI inequality and extensions," *Preprint*, Sep. 2011.
[18] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.