



tests is given by<sup>3</sup>

$$\beta_\alpha(P, Q) = \min_{w \in \mathbf{W}} Q(w) P_{Z|W}(1|w), \quad (5)$$

where the minimum is over all probability distributions  $P_{Z|W}$  satisfying

$$P_{Z|W} : \sum_{w \in \mathbf{W}} P(w) P_{Z|W}(1|w) \geq \alpha. \quad (6)$$

The minimum in (5) is guaranteed to be achieved by the Neyman-Pearson lemma. Thus,  $\beta_\alpha(P, Q)$  gives the minimum probability of error under hypothesis  $Q$  if the probability of error under hypothesis  $P$  is not larger than  $1 - \alpha$ . It is easy to show that (e.g. [4]) for any  $\gamma > 0$

$$\alpha \leq \mathbb{P} \left[ \frac{dP}{dQ} \geq \gamma \right] + \gamma \beta_\alpha(P, Q). \quad (7)$$

On the other hand,

$$\beta_\alpha(P, Q) \leq \frac{1}{\gamma_0}, \quad (8)$$

for any  $\gamma_0$  that satisfies

$$\mathbb{P} \left[ \frac{dP}{dQ} \geq \gamma_0 \right] \geq \alpha. \quad (9)$$

Each per-codeword cost constraint can be defined by specifying a subset  $\mathbf{F} \subset \mathbf{A}$  of permissible inputs. For an arbitrary  $\mathbf{F} \subset \mathbf{A}$ , we define a related measure of performance for the composite hypothesis test between  $Q_Y$  and the collection  $\{P_{Y|X=x}\}_{x \in \mathbf{F}}$ :

$$\kappa_\tau(\mathbf{F}, Q_Y) = \inf_{P_{Z|Y} : \inf_{x \in \mathbf{F}} P_{Z|X}(1|x) \geq \tau} \sum_{y \in \mathbf{B}} Q_Y(y) P_{Z|Y}(1|y). \quad (10)$$

Typically we will take  $\mathbf{A}$  and  $\mathbf{B}$  as  $n$ -fold Cartesian products of alphabets  $\mathcal{A}$  and  $\mathcal{B}$ . To emphasize dependence on  $n$  we will write  $\beta_\alpha^n$  and  $\kappa_\tau^n$ .

### B. Achievability and Converse Bounds

Our main tool in showing the achievability part of (3) is the following result (Theorem 4 of [3]):

*Theorem 1 ( $\kappa_\tau$  bound):* For any  $0 < \epsilon < 1$ , there exists an  $(M, \epsilon)$  code with codewords chosen from  $\mathbf{F} \subset \mathbf{A}$ , satisfying

$$M \geq \sup_{0 < \tau < \epsilon} \sup_{Q_Y} \frac{\kappa_\tau(\mathbf{F}, Q_Y)}{\sup_{x \in \mathbf{F}} \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)}. \quad (11)$$

Virtually all known converse results for channel coding can be shown to be applications of the following result (or its variant for average probability of error [5]) by a judicious choice of  $Q_{Y|X}$  and a lower bound on  $\beta$ , see [1].

*Theorem 2 (meta-converse, [1]):* Consider two different abstract channels  $P_{Y|X}$  and  $Q_{Y|X}$  defined on the same input

<sup>3</sup>We write summations over alphabets for simplicity; however, all of our general results hold for arbitrary probability spaces.

and output spaces. For a given code (possibly with a randomized decoder) with codewords belonging to  $\mathbf{F} \subset \mathbf{A}$ , let

$$\begin{aligned} \epsilon &= \text{maximal error probability with } P_{Y|X} \\ \epsilon' &= \text{maximal error probability with } Q_{Y|X} \end{aligned}$$

Then,

$$\inf_{x \in \mathbf{F}} \beta_{1-\epsilon}(P_{Y|X=x}, Q_{Y|X=x}) \leq 1 - \epsilon'. \quad (12)$$

### III. THE AWGN CHANNEL

For the real-valued AWGN channel we set  $\mathbf{A} = \mathbb{R}^n$ ,  $\mathbf{B} = \mathbb{R}^n$ ,  $P_{Y^n|X^n=x^n} = \mathcal{N}(x^n, \mathbf{I}_n)$  and codewords are subject to one of three types of power constraints:

- *equal-power constraint:*  $M_e^*(n, \epsilon, P)$  denotes the maximal number of codewords, such that each codeword  $c_i \in X^n$  satisfies

$$\|c_i\|^2 = nP. \quad (13)$$

- *maximal power constraint:*  $M_m^*(n, \epsilon, P)$  denotes the maximal number of codewords, such that each codeword  $c_i \in X^n$  satisfies

$$\|c_i\|^2 \leq nP. \quad (14)$$

- *average power constraint:*  $M_a^*(n, \epsilon, P)$  denotes the maximal size  $M$  of a codebook that satisfies

$$\frac{1}{M} \sum_{i=1}^M \|c_i\|^2 \leq nP. \quad (15)$$

*Theorem 3:* For the AWGN channel with SNR  $P$ ,  $0 < \epsilon < 1$  and under equal-power, maximal-power or average-power constraints, we all have

$$\log M^*(n, \epsilon, P) = nC_1(P) - \sqrt{nV_1(P)} Q^{-1}(\epsilon) + \rho_n, \quad (16)$$

where

$$\rho_n = O(\log n), \quad (17)$$

$$C_1(P) = \frac{1}{2} \log(1 + P), \quad (18)$$

$$V_1(P) = \frac{P}{2} \frac{P+2}{(P+1)^2} \log^2 e. \quad (19)$$

More precisely, for equal-power and maximal-power constraints, the  $O(\log n)$  term in (16) can be bounded by

$$O(1) \leq \rho_n \leq \frac{1}{2} \log n + O(1), \quad (20)$$

whereas for average-power constraint we have

$$O(1) \leq \rho_n \leq \frac{3}{2} \log n + O(1). \quad (21)$$

A proof of this result is sketched in the appendix. For full details see [1]. It is interesting to note that in [1] we have found that the  $\rho_n$  term is equal to  $1/2 \log n + O(1)$  for the BSC and  $O(1)$  for the BEC.

Proceeding heuristically from the reliability function in [7], the expansion in (16) was put forward in [6] with  $\rho_n = o(n^{-1/2})$ , for the case of per-codeword power constraint.

#### IV. THE PARALLEL AWGN CHANNEL

For the real-valued  $L$ -parallel AWGN channel we set  $\mathbf{A} = \mathbb{R}^{L \times n}$ ,  $\mathbf{B} = \mathbb{R}^{L \times n}$  and  $P_{Y|X}$  is defined by

$$Y_{i,j} = X_{i,j} + \sigma_i Z_{i,j}, \quad i = 1 \dots L, j = 1 \dots n, \quad (22)$$

where  $Z_{i,j}$  are independent  $\mathcal{N}(0, 1)$  random variables. Also, codewords  $\mathbf{c}$  are subject to a (maximal) power constraint:

$$\|\mathbf{c}\|^2 = \sum_{j=1}^n \sum_{i=1}^L |c_{i,j}|^2 \leq nP. \quad (23)$$

*Theorem 4:* For a parallel AWGN channel and  $\epsilon \in (0, 1)$  we have

$$\log M^*(n, \epsilon, P) = nC_L(P) - \sqrt{nV_L(P)}Q^{-1}(\epsilon) + O(\log n), \quad (24)$$

with

$$C_L(P) = \sum_{i=1}^L C_1\left(\frac{W_i}{\sigma_i^2}\right), \quad \text{and} \quad (25)$$

$$V_L(P) = \sum_{i=1}^L V_1\left(\frac{W_i}{\sigma_i^2}\right), \quad (26)$$

where  $C_1$  and  $V_1$  are defined in (18) and (19) and  $\{W_j\}$  are the usual waterfilling powers

$$W_i = [\lambda - \sigma_i^2]^+ \quad (27)$$

and  $\lambda$  is the solution of

$$\sum_{i=1}^L W_i = P. \quad (28)$$

The proof is outlined in the appendix.

#### V. THE GAUSSIAN CHANNEL WITH ISI

Consider the Gaussian channel with ISI:

$$Y_j = (h * X)_j + Z_j, \quad (29)$$

where  $h * X$  denotes convolution with a deterministic sequence  $h$ , whose Fourier transform is  $H(f) = \sum_{k=-\infty}^{\infty} h_k e^{-2i\pi f k}$  and  $Z_j$  are independent  $\mathcal{N}(0, 1)$  random variables.

*Theorem 5:* For the Gaussian channel with ISI, the channel dispersion is given by

$$V_{\text{ISI}} = \frac{\log^2 e}{2} \int_{-1/2}^{1/2} \left[1 - \frac{1}{|H(f)|^4 P^2 \xi^2}\right]^+ df, \quad (30)$$

where  $\xi$  is the solution of

$$\int_{-1/2}^{1/2} \left[\xi - \frac{1}{P|H(f)|^2}\right]^+ df = 1. \quad (31)$$

If the noise process,  $Z_j$ , is not white, then we simply need to replace  $|H(f)|^2$  by  $|H(f)|^2/N(f)$ , where  $N(f)$  is the power spectral density of  $Z_j$ .

We cannot include the proof of Theorem 5, but to partly motivate the expression (30) one simply needs to notice that (26) can be written as

$$V_L(P) = \left(\frac{\log^2 e}{2}\right) \sum_{i=1}^L \left[1 - \left(\frac{\sigma_i^2}{\lambda}\right)^2\right]^+. \quad (32)$$

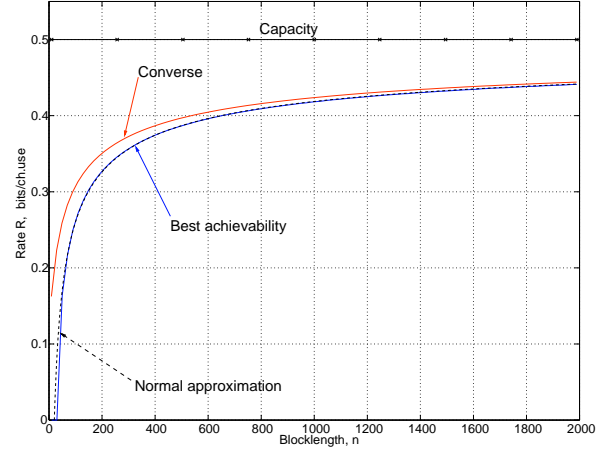


Fig. 1. Upper and lower bounds on the rate  $R = \frac{\log M^*(n, \epsilon)}{n}$  of the best code vs. the normal approximation for the AWGN channel with maximal power constraint SNR = 0dB,  $\epsilon = 10^{-3}$ .

#### VI. DISCUSSION

As shown in [1], the normal approximation

$$\log M^*(n, \epsilon, P) \approx nC - \sqrt{nV}Q^{-1}(\epsilon) + \alpha \log n \quad (33)$$

is rather tight for the BSC ( $\alpha = 1/2$ ) and the BEC ( $\alpha = 0$ ). Similarly, in [1], we computed various upper and lower bounds on  $\log M^*(n, \epsilon, P)$  for the AWGN channel and found that for many practical scenarios (33) with  $\alpha = 1/2$  is valid to within a few bits. An example of such comparison is given in Fig. 1. The converse bound is obtained from Theorem 2 and the tightest achievability in this case is the one found in [7]<sup>4</sup>.

A good analytical approximation to the maximal rate achievable with a given blocklength and error probability opens a range of practical applications. We now give some examples.

First, an interesting figure of merit for the AWGN is the excess energy per bit,  $\Delta E_b(n)$ , over that predicted by channel capacity incurred as a function of blocklength for a given required bit rate and block error rate:

$$\Delta E_b(n) = 10 \log_{10} \frac{P(n, R, \epsilon)}{\exp(2R) - 1}, \quad (34)$$

where, according to the normal approximation,  $P(n, R, \epsilon)$  is the solution to

$$C_1(P) - \sqrt{\frac{V_1(P)}{n}}Q^{-1}(\epsilon) + \frac{1}{2n} \log n = R. \quad (35)$$

In [1] we evaluate  $\Delta E_b(n)$  and compare it against the best upper and lower bounds. Again, we find that (34) and (35) yield good precision e.g., for  $n = 200$ ,  $R = 1/2$  and  $\epsilon = 10^{-4}$  the difference between an approximate value of  $\Delta E_b(n)$  and the true upper bound is only 0.04 dB.

Second, we consider a basic automatic repeat request (ARQ) transmission scheme in which a packet is retransmitted until the receiver acknowledges successful decoding (which the receiver determines using a variety of known highly reliable

<sup>4</sup>Note that while the  $\kappa\beta$  bound is more useful for proving Theorem 3, numerically [7] is unsurpassed.

TABLE I  
OPTIMAL BLOCK ERROR RATE FOR PACKET SIZE  $k = 1000$  BITS

Channel	Optimal $\epsilon^*(k)$	Optimal $R/C$	Throughput
BEC(0.5)	$8.1 \cdot 10^{-3}$	0.95	0.94
BSC(0.11)	$16.7 \cdot 10^{-3}$	0.91	0.90
AWGN, SNR = 0dB	$15.5 \cdot 10^{-3}$	0.92	0.90
AWGN, SNR = 20dB	$6.2 \cdot 10^{-3}$	0.96	0.95

hashing methods). Typically, the size  $k$  of the information packets is determined by the particular application, and both the blocklength  $n$  and the block error probability  $\epsilon$  are degrees of freedom. A natural objective is to maximize the average throughput (or, equivalently, minimize the average delivery delay) given by

$$T(k) = \max_{n, \epsilon} \frac{k}{n} (1 - \epsilon), \quad (36)$$

assuming decoding errors are independent for different retransmissions. The maximization in (36) is over those  $(n, \epsilon)$  such that

$$\log_2 M^*(n, \epsilon) = k. \quad (37)$$

Note that the number of required retransmissions is geometrically distributed, with mean equal to  $\frac{k}{T(k)}$ . In view of the tightness of the approximation in (3), it is sensible to maximize

$$\tilde{T}(k) = \max_n \frac{k}{n} \left[ 1 - Q \left( \frac{nC - k}{\sqrt{nV}} \right) \right], \quad (38)$$

where  $C$  and  $V$  are the channel capacity and channel dispersion, respectively. Table I shows the results of the optimization for the channel examples we have discussed above. Additionally, some plots can be found in [1].

Of particular note is that, for 1000 information bits, the optimum block error rate is as high as  $10^{-2}$ . In fact, this optimum  $\epsilon$  does not depend too much on either the channel or the packet size. A tight approximation to it is given by

$$\tilde{\epsilon}(k) = \left( \frac{kC}{V} \ln \frac{kC}{2\pi V} \right)^{-1/2} \left( 1 - \frac{1}{\ln \frac{kC}{2\pi V}} \right), \quad (39)$$

which is obtained by retaining only the dominant terms in the asymptotic solution as  $k \rightarrow \infty$ .

#### REFERENCES

- [1] Y. Polyanskiy, H. V. Poor and S. Verdú, "Channel coding rate in the finite blocklength regime," submitted to *IEEE Trans. Inform. Theory*, Nov. 2008.
- [2] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informationstheorie," *Trans. Third Prague Conf. Information Theory*, 1962, Czechoslovak Academy of Sciences, Prague, pp. 689-723.
- [3] Y. Polyanskiy, H. V. Poor and S. Verdú, "New channel coding achievability bounds," *Proc. 2008 IEEE Int. Symp. Information Theory (ISIT)*, Toronto, Canada, July 2008.
- [4] S. Verdú, *EE528-Information Theory, Lecture Notes*, Princeton University, Princeton, NJ, 2007.
- [5] Y. Polyanskiy, H. V. Poor and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *Proc. 2009 IEEE Int. Symp. Information Theory (ISIT)*, Seoul, Korea, July 2009.
- [6] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "How quickly can we approach channel capacity?" *Proc. 38th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, November 2004.

- [7] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Tech. Journal*, vol. 38, pp. 611-656, 1959.
- [8] J. N. Laneman, "On the distribution of mutual information," *Proc. 2006 Workshop on Inform. Theory and its Appl.*, San Diego, CA, Feb. 2006.

#### APPENDIX

*Proof of Theorem 3: Achievability:* To analyze  $\log M_e^*(n, \epsilon)$  we plan to use Theorem 1 with  $\mathbf{F}_n = \{x^n \in \mathbb{R}^n : \|x^n\|^2 = nP\}$  and

$$Q_{Y^n} = \mathcal{N}(0, (1+P)\mathbf{I}_n). \quad (40)$$

By the spherical symmetry we notice that  $\beta_\alpha^n(P_{Y^n|X^n=x^n}, Q_{Y^n})$  does not depend on  $x^n \in \mathbf{F}_n$ . A convenient choice of  $x^n$  is

$$x^n = (\sqrt{P}, \dots, \sqrt{P}). \quad (41)$$

Indeed, with this choice we can show that (an alternative analysis of the distribution of  $i(X^n; Y^n)$  is found in [8])

$$\log \frac{dP_{Y^n|X^n=x^n}}{dQ_{Y^n}} = \sum_{j=1}^n L_j, \quad (42)$$

where  $L_j$  are independent random variables distributed as

$$L_j = \frac{1}{2} \log(1+P) + \frac{\log e}{2} \frac{P}{(1+P)} \left( 1 - Z_i^2 + \frac{2}{\sqrt{P}} Z_i \right) \quad (43)$$

with  $Z_j \sim \mathcal{N}(0, 1)$ . Notice that

$$\mathbb{E}[L_j] = C_1(P) \text{ and } \text{Var}[L_j] = V_1(P). \quad (44)$$

Therefore, an application of (7), (8) and the Berry-Esseen inequality to (42) allows us to prove

$$\log \beta_\alpha^n = -nC_1(P) + \sqrt{nV_1(P)} Q^{-1}(1-\alpha) + O(\log n). \quad (45)$$

To compute  $\kappa_\tau(\mathbf{F}_n, Q_{Y^n})$ , notice that by spherical symmetry the optimal test in (10) will also be spherically symmetric. But then the quantity

$$S_n = \sum_{j=1}^n |Y_j|^2 \quad (46)$$

is a sufficient statistic for a composite HT problem (10). Moreover, the distribution of  $S_n$  is the same under all  $P_{Y^n|X^n=x^n}$  provided that  $x^n \in \mathbf{F}_n$ . Then we conclude that

$$\kappa_\tau(\mathbf{F}_n, Q_{Y^n}) = \beta_\tau(P_{S_n}, Q_{S_n}), \quad (47)$$

where

$$\text{under } P_{S_n}: S_n \sim \sum_{j=1}^n (\sqrt{P} + Z_j)^2, \quad (48)$$

$$\text{under } Q_{S_n}: S_n \sim \sum_{j=1}^n (1+P)(Z_j)^2. \quad (49)$$

An application of the (local) central limit theorem shows that for some constants  $K_1, K_2 > 0$  and all  $\tau \in [0, 1]$  we have

$$\kappa_\tau(\mathbf{F}_n, Q_{Y^n}) \geq K_1 \tau + O(e^{-K_2 n}) \quad (50)$$

(the main reason for this type of behavior of  $\kappa$  is the equality of expectations of (48) and (49)).

To conclude the proof of achievability we apply Theorem 1 with  $\tau_n = \frac{1}{\sqrt{n}}$  and find that

$$\log M_e^*(n, \epsilon) \geq \log \kappa_{\tau_n}(\mathbf{F}_n, Q_{Y^n}) - \log \beta_{1-\epsilon+\tau_n}^n \quad (51)$$

$$\geq nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon - \tau_n) + \log \kappa_{\tau_n} + O(\log n) \quad (52)$$

$$= nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon) + \log \kappa_{1/\sqrt{n}} + O(\log n) \quad (53)$$

$$\geq nC_1 - \sqrt{nV_1}Q^{-1}(\epsilon) + O(\log n), \quad (54)$$

where (52) is by (45), (53) is by Taylor's theorem and because  $\tau_n = 1/\sqrt{n}$ , and (54) is by (50).

*Converse:* Take the code satisfying an equal-power constraint with the largest possible cardinality  $M_e^*(n, \epsilon)$ . We apply Theorem 2 to this code with  $Q_{Y^n|X^n} = Q_{Y^n}$ , where  $Q_{Y^n}$  is defined in (40). Since under the  $Q$ -channel the input and output are independent, we easily find

$$1 - \epsilon' \leq \frac{1}{M_e^*(n, \epsilon)}. \quad (55)$$

Putting this into (12) we find

$$\log M_e^*(n, \epsilon) \leq -\log(1 - \epsilon') \leq -\log \beta_{1-\epsilon}^n \quad (56)$$

$$\leq nC_1(P) - \sqrt{nV_1(P)}Q^{-1}(\epsilon) + O(\log n), \quad (57)$$

where (57) is by (45). We omit the proof of similar statements about  $\log M_a^*(n, \epsilon)$  and  $\log M_m^*(n, \epsilon)$ . ■

*Proof of Theorem 4: Achievability:* We choose

$$\mathbf{F}_n = \{\mathbf{x} \in \mathbb{R}^{L \times n} : \|\mathbf{x}_{i,\cdot}\|^2 = nW_i, i = 1, \dots, L\} \quad (58)$$

and

$$Q_{\mathbf{Y}} = \prod_{j=1}^n \prod_{i=1}^L \mathcal{N}(0, \sigma_i^2 + W_i). \quad (59)$$

Again, by spherical symmetry we can see that  $\beta_\alpha(P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}, Q_{\mathbf{Y}})$  does not depend on the choice of  $\mathbf{x} \in \mathbf{F}_n$  and for convenience we choose

$$\mathbf{x} : x_{i,j} = \sqrt{W_i}, \quad i = \overline{1, L}, j = \overline{1, n}. \quad (60)$$

Then, by analyzing  $\log \frac{dP_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}{dQ_{\mathbf{Y}}}$  and similar to (45) we obtain

$$\log \beta_\alpha^n = -nC_L(P) + \sqrt{nV_L(P)}Q^{-1}(1 - \alpha) + O(\log n). \quad (61)$$

Again, analogously to (50) it can be shown that for some  $K_1, K_2 > 0$

$$\kappa_\tau(\mathbf{F}_n, Q_{\mathbf{Y}}) \geq K_1\tau + O(e^{-K_2n}). \quad (62)$$

The proof of achievability is concluded by following the steps (51)-(54).

*Converse:* Proving the converse for the parallel AWGN is not as simple as for the AWGN. The main obstacle is that we can not apply (61) since it was derived under assumption  $\mathbf{x} \in \mathbf{F}_n$ , whereas according to the power constraint codeword  $\mathbf{x}$  can belong to a larger set:

$$\mathbf{F}'_n = \{\mathbf{x} \in \mathbb{R}^{L \times n} : \|\mathbf{x}\|^2 \leq nP\}. \quad (63)$$

To each codeword  $\mathbf{x} \in \mathbf{F}'_n$  we associate a *power allocation vector*

$$\mathbf{v}(\mathbf{x}) \in \mathbb{R}^L : v_i(\mathbf{x}) = \frac{1}{n} \|\mathbf{x}_{i,\cdot}\|^2. \quad (64)$$

We choose the following  $Q$ -channel in Theorem 2:

$$Q_{\mathbf{Y}|\mathbf{X}=\mathbf{x}} = \prod_{j=1}^n \prod_{i=1}^L Q_{Y_{i,j}|\mathbf{X}=\mathbf{x}}, \quad (65)$$

where

$$Q_{Y_{i,j}|\mathbf{X}=\mathbf{x}} = \mathcal{N}(0, \sigma_i^2 + v_i(\mathbf{x})). \quad (66)$$

Again, by spherical symmetry  $\beta_\alpha^n(P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}, Q_{\mathbf{Y}|\mathbf{X}=\mathbf{x}})$  depends on  $\mathbf{x}$  only through  $\mathbf{v}$ . Then, analogously to (45) and (61) we find

$$\begin{aligned} \log \beta_\alpha^n(\mathbf{x}) &= -n \sum_{i=1}^L C_1 \left( \frac{v_i(\mathbf{x})}{\sigma_i^2} \right) \\ &+ \sqrt{n \sum_{i=1}^L V_1 \left( \frac{v_i(\mathbf{x})}{\sigma_i^2} \right)} Q^{-1}(1 - \alpha) + O(\log n). \end{aligned} \quad (67)$$

According to (12), we must take the infimum of (67) over all codewords  $\mathbf{x} \in \mathbf{F}'_n$ . Since the latter restriction is equivalent to requiring  $\sum_{i=1}^L v_i(\mathbf{x}) \leq P$ , there exists a unique minimizer  $v_i = W_i$  so that

$$\inf_{\mathbf{x} \in \mathbf{F}'_n} \log \beta_\alpha^n(\mathbf{x}) = -nC_L + \sqrt{nV_L(P)}Q^{-1}(1 - \alpha) + O(\log n). \quad (68)$$

Finally, assume that we could establish the following converse bound for the  $Q$ -channel:

*Lemma 6:* There exists a constant  $K_3 > 0$  such that for any code with  $M$  codewords the maximum probability of error  $\epsilon'$  over a  $Q$ -channel satisfies

$$1 - \epsilon' \leq \frac{K_3 n^{L/2}}{M}. \quad (69)$$

The proof of the theorem then follows by the same argument as (55)-(57) with (45) replaced by (68), and the  $n^{L/2}$  factor affecting only the  $O(\log n)$  term. ■

*Proof of Lemma 6:* According to (66) the output  $\mathbf{Y}$  depends only on  $\mathbf{V} = \mathbf{v}(\mathbf{X})$  and moreover  $\mathbf{U} = \mathbf{v}(\mathbf{Y})$  is a sufficient statistic of  $\mathbf{Y}$  for  $\mathbf{X}$ . Therefore, an equivalent channel  $Q_{\mathbf{U}|\mathbf{V}}$  is defined as

$$U_i = (\sigma_i^2 + V_i) \frac{1}{n} \sum_{j=1}^n Z_{i,j}^2, \quad i = \overline{1, L}, \quad (70)$$

where  $Z_{i,j} \sim \mathcal{N}(0, 1)$ . Note that  $\mathbf{V}$  is required to belong to a certain ball in  $\mathbb{R}^L$ , and that with overwhelming probability  $\mathbf{U}$  belongs to a slightly larger ball. Therefore, we can assume that the output space has a bounded volume  $K_4$ . Then at least for one codeword  $\mathbf{v}_0$  the decoding set  $D_0$  must have a volume smaller than  $\frac{K_4}{M}$ . But  $Q_{\mathbf{U}|\mathbf{V}=\mathbf{v}_0}$  is a product of  $L$  copies of a  $\chi^2$ -distribution and we can show that its density is bounded everywhere by  $K_5 n^{L/2}$ . Hence, we have

$$1 - \epsilon' \leq Q_{\mathbf{U}|\mathbf{V}=\mathbf{v}_0}[D_0] \leq \frac{K_4 K_5 n^{L/2}}{M}. \quad (71)$$

■