Minimum Expected Length of Fixed-to-Variable Lossless Compression Without Prefix Constraints

Wojciech Szpankowski, *Fellow, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—The minimum expected length for fixed-to-variable length encoding of an n-block memoryless source with entropy H grows as $nH + O(1)$, where the term $O(1)$ lies between 0 and 1. However, this well-known performance is obtained under the implicit constraint that the code assigned to the whole n-block is a prefix code. Dropping the prefix constraint, which is rarely necessary at the block level, we show that the minimum expected length for a finite-alphabet memoryless source with known distribution grows as**

$$nH - \frac{1}{2}\log n + O(1)$$

unless the source is equiprobable. We also refine this result up to $o(1)$ for those memoryless sources whose log probabilities do not reside on a lattice.

***Index Terms*—Analytic information theory, fixed-to-variable lossless compression, memoryless sources, one-to-one codes, Shannon theory, source coding.**

I. Introduction

LOSSLESS symbol-by-symbol compressors are required to satisfy the condition of "unique decodability" whereby different input strings are assigned different compressed versions. Uniquely decodable nonprefix codes do not offer any advantages over prefix codes since any uniquely decodable code must assign lengths to the various symbols that satisfy Kraft's inequality, while a prefix code is guaranteed to exist with those symbol lengths. Achieved by the Huffman code, an exact expression for the minimum average length of a prefix symbol-by-symbol binary code is unknown. It is upper bounded by the entropy (in bits) of the probability distribution of the symbols plus one bit (this follows by the analysis of the suboptimal Shannon code in [28], which, incidentally, Shannon devised to encode blocks of data). Macmillan [19] showed that the minimum average length of a prefix symbol-by-symbol binary code is lower bounded by the entropy-a result which is frequently wrongly attributed to Shannon, who never addressed the fundamental limits of prefix codes. Further improvements on the upper bound (as a function of the distribution) were reported in [3], [4], [12], [21], [27]. However, the paradigm of symbol-by-symbol compression is severely suboptimal even for memoryless sources. For example, symbol-by-symbol compression is unable to exploit the redundancy of biased coin flips. Algorithmically, at the expense of a slight penalty in average encoding length, this inefficiency is dealt with stream codes such as arithmetic coding. To approach the minimum average encoding length one can partition the source string of length n into blocks of length k and apply the symbol-by-symbol approach at the block level. The resulting average compressed length per source symbol is equal to the entropy of each symbol, $H(X)$, plus at most $1/k$ bits if the source is memoryless, or more generally, equal to the entropy of k consecutive symbols divided by k plus at most $1/k$ bits. Thus, to achieve the best average efficiency without regard to complexity, we can let $k = n$, apply a Huffman code to the whole n-tuple and the resulting average compressed length behaves as

$$L_n = nH(X) + O(1). \tag{1}$$

The $O(1)$ term in (1) belongs to $[0, 1]$, and has been investigated in detail for biased coin flips in [27], [29]. In particular,[1] when $\log_2 \frac{1-p}{p}$ is irrational (where p is the bias)

$$L_n = nh(p) + \frac{3}{2} - \log_2 e + o(1)\,\text{bits} \tag{2}$$

where $h(p)$ is the binary entropy function.

As argued in [32], [35], it is possible to attain average compressed length lower than (1). The reason is that it is often unnecessary, and in fact wasteful, to impose the prefix condition on a code that operates at the level of the whole file to be compressed. Applying prefix codes to n-block supersymbols is only optimal in terms of the linear growth with n (it attains the entropy rate for stationary ergodic sources); however, as far as sublinear terms, this conventional approach incurs loss of optimality. The optimal fixed-to-variable length code performs no blocking on the source output; instead the optimal length-n compressor chooses an encoding table that lists all source realizations of length n in decreasing probabilities (breaking ties using a lexicographical ordering on the source symbols) and assigns, starting with the most probable, the binary strings of increasing lengths[2]

$$\{\emptyset, 0, 1, 00, 01, 10, 11, 000, 001, 010, 011, 100, \ldots\}.$$

The fact that the length of the compressed file is unknown a priori is immaterial since the decompressor receives as input the compressed file, including where the file "starts and ends." For example, files stored in a random-access medium (such as a hard

Manuscript received July 30, 2009; revised January 17, 2011; accepted January 17, 2011. Date of current version June 22, 2011. This work was supported by the Science and Technology NSF Center for Science of Information under Grant CCF-0939370 and by the NSF under Grants CCF-0830140, CCF-0939370, and CCF-1016625.

W. Szpankowski is with the Department of Computer Science, Purdue University, West Lafayette, IN 47907 USA (e-mail: spa@cs.purdue.edu).

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: verdu@princeton.edu).

Communicated by E. Ordentlich, Associate Editor for Source Coding.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2011.2145590

[1] We omit the more involved formula given in [29] for the rational case.

[2] Including the empty string is convenient but has no impact on our results.

disk) do not satisfy the prefix condition: a directory (organized as a so-called *inode pointer structure*, e.g., [20]) contains the starting and ending locations of the sequence of blocks occupied by each file in the storage medium.

The foregoing code is optimal not just in the sense of average length but in the sense that the cumulative distribution function of its length is larger than or equal to that of any other code. Such optimal codes have been previously considered under the rubric of *one-to-one* codes, but because of their misguided standing as nonuniquely decodable *symbol-by-symbol* codes, they have failed to attract much attention.

In the rest of this paper, Section II deals with the nonasymptotic analysis of one-to-one codes. Section III summarizes previous results on the minimum average length achievable for biased coin flips. Section IV states our results on asymptotic analysis of the minimum average length of fixed-to-variable length codes for memoryless sources with known distributions. For equiprobable distributions we can save on average between 1.914 and 2 bits (from the logarithm of the number of equiprobable realizations) plus an exponentially vanishing term. For nonequiprobable distributions, we can save $\frac{1}{2}\log_2 n$ from the entropy of the n-tuple plus an $O(1)$ term. If the log probabilities of the source do not reside on a lattice, we show that the $O(1)$ term is in fact

$$\frac{1}{2}\log_2(8\pi e\sigma^2)$$

plus a vanishing term, where σ^2 is the variance of $\log_2 P_X(X)$. Proofs are given in Section V.

II. Nonasymptotic Analysis of Optimal Variable-Length Codes

Consider a probability distribution P_X on a set of ordered elements $\mathcal{X}$. Define $\pi_X : \mathcal{X} \mapsto \{1, \ldots, |\mathcal{X}|\}$ by $\pi_X(a) < \pi_X(b)$ if $P_X(a) > P_X(b)$ or if $P_X(a) = P_X(b)$ and $a < b$. Thus, $\pi_X(x) = \ell$ if x is the ℓ-th most probable element in $\mathcal{X}$ according to distribution P_X, with ties broken according to the ordering in $\mathcal{X}$. It is easy to verify that

$$P_X(x)\pi_X(x) \leq 1 \tag{3}$$

for all $x \in \mathcal{X}$: if (3) failed to be satisfied for $x_0 \in \mathcal{X}$, there would be at least $\pi_X(x_0)$ masses strictly larger than $1/\pi_X(x_0)$.

The one-to-one code assigns to x the shortest (possibly empty) binary string (ties broken with the ordering $0 < 1$) not assigned to any element y with $\pi_X(y) < \pi_X(x)$. Thus, we obtain the simple but important conclusion that the length of the encoding of x is $\lfloor \log_2 \pi_X(x) \rfloor$. Finding an exact expression for the minimum average length

$$L(X) = \mathbb{E}[\lfloor \log_2 \pi_X(X) \rfloor] \tag{4}$$

as a function of P_X appears to be challenging. For X equiprobable on a set of $M = |\mathcal{X}|$ elements, it can be shown that the average length of the one-to-one code is (cf. [17])

$$\begin{aligned} L(X) &= \frac{1}{M}\sum_{i=1}^{M} \lfloor \log_2 i \rfloor & (5)\\ &= \lfloor \log_2 M \rfloor + \frac{1}{M}\left(2 + \lfloor \log_2 M \rfloor - 2^{\lfloor \log_2 M \rfloor + 1}\right) & (6)\end{aligned}$$

which simplifies to

$$\frac{1}{M}\sum_{i=1}^{M} \lfloor \log_2 i \rfloor = \frac{(M+1)\log_2(M+1)}{M} - 2 \tag{7}$$

when $M + 1$ is a power of 2.

A simple upper bound first noticed in [34] is obtained as

$$\begin{aligned} L(X) &= \mathbb{E}[\lfloor \log_2 \pi_X(X) \rfloor] & (8)\\ &\leq \mathbb{E}[\log_2 \pi_X(X)] & (9)\\ &\leq \mathbb{E}\left[\log_2 \frac{1}{P_X(X)}\right] & (10)\\ &= H(X) & (11)\end{aligned}$$

where (10) follows from (3). Note that dropping the prefix condition makes the entropy an upper bound to the minimum average length, rather than a lower bound. Various lower bounds on $L(X)$ have been proposed in [1], [2], [5], [10], [18], [23], [33]. Distilling the main ideas in [1], the following result gives the tightest known bound.

Theorem 1: Define the monotonically increasing function $\psi : \mathbb{R}^+ \mapsto \mathbb{R}^+$ by

$$\psi(x) = x + (1+x)\log_2(1+x) - x\log_2 x. \tag{12}$$

Then

$$\psi^{-1}(H(X)) \leq L(X). \tag{13}$$

Proof: For brevity, denote $Y = \lfloor \log_2 \pi_X(X) \rfloor$, and $Z = Y + 1$

$$\begin{aligned} H(X) &= H(X \,|\, Y) + H(Y) & (14)\\ &\leq \mathbb{E}[Y] + H(Y) & (15)\\ &= \mathbb{E}[Y] + H(Z) & (16)\\ &= \mathbb{E}[Y] + \mathbb{E}[Z]h(1/\mathbb{E}[Z]) - D(P_Z \| G_{1/\mathbb{E}[Z]}) & (17)\\ &\leq \psi(\mathbb{E}[Y])) & (18)\end{aligned}$$

where:

- (14) $\Longleftarrow$ Y is a deterministic function of X;
- (15) $\Longleftarrow$ $H(X \,|\, Y = k) \leq k$ bits;
- (17) follows by writing out the relative entropy on the right side where the reference measure is the geometric (positive) distribution $G_p(k) = p(1-p)^{k-1}$;
- (18) $\Longleftarrow$ the relative entropy $D(\cdot \| \cdot) \geq 0$.

■

Weakening the bound in (13) by

$$\psi(x) \leq x + \log_2 e + \log_2(1+x) \tag{19}$$

and using the upper bound (11), we obtain the bound in [1]

$$H(X) - \log_2(H(X)+1) - \log_2 e \leq \mathbb{E}[\lfloor \log_2 \pi_X(X) \rfloor] \tag{20}$$

Another way of weakening (13) is to use the monotonic increasing nature of $(1+x)\log(1+x) - x\log x$ and (11) to conclude with (21) and (22), as shown at the bottom of the page, which is the bound found in [2].

In the remainder of the paper, we turn attention to the asymptotic behavior of the minimum average length of the encoding of an n-tuple of a memoryless stationary source with marginal distribution P_{X}

$$L_n^* = L(X^n). \tag{23}$$

Note that all the results obtained in this section apply to that case by letting X^n and $nH(\mathsf{X})$ play the role of X and $H(X)$, respectively.

III. Asymptotic Minimum Average Length: Coin Flips

A. *Fair Coin Flips*

For fair coin flips ($p = \frac{1}{2}$), the exact result can be obtained from (6) letting $M = 2^n$

$$L_n^* = n - 2 + 2^{-n}(n+2) \tag{24}$$

in contrast to

$$L_n = n \tag{25}$$

obtained with the Huffman code operating on n-tuples (or single bits).

B. *Biased Coin Flips*

The minimum average length for a binary memoryless source with bias $p \neq \frac{1}{2}$ has been investigated in great detail (up to $o(1)$ term) in [31], which shows that

$$L_n^* = nh(p) - \frac{1}{2}\log_2 n + O(1) \tag{26}$$

and in fact [31] characterizes the $O(1)$ term up to vanishing terms. If $\log_2 \frac{1-p}{p}$ is irrational and positive ($p < 1/2$), then we get (27), as shown at the bottom of the page. If

$$\log_2 \frac{1-p}{p} = \frac{N}{J} \tag{28}$$

where (28) is an irreducible fraction, we need to add (29), shown at the bottom of the page, divided by J to (27) where $\langle x \rangle = x - \lfloor x \rfloor$ and U is standard normal.

IV. Asymptotic Minimum Average Length: Memoryless Sources

We assume henceforth that the source is memoryless with distribution P_{X} on a finite alphabet $\mathcal{A}$, i.e.,

$$P_{X^n} = P_{\mathsf{X}} \times \cdots \times P_{\mathsf{X}}. \tag{30}$$

The proofs of the following asymptotic results are given in Section V.

Theorem 2: For a nonredundant source (i.e., memoryless and equiprobable) with finite alphabet $\mathcal{A}$, the minimum expected length of a lossless binary encoding of X^n is given by

$$L_n^* = n\log_2|\mathcal{A}| - 2 + \tau(n\log_2|\mathcal{A}|) + o(\rho^n). \tag{31}$$

where $\frac{1}{|\mathcal{A}|} < \rho < 1$ and

$$\tau(x) = \lfloor x \rfloor - x - 2^{\lfloor x \rfloor + 1 - x} + 2 \tag{32}$$

which satisfies

$$0 \leq \tau(x) \leq 1 - \log_2 e + \log_2\log_2 e = 0.086 \tag{33}$$

Definition 1: A discrete real-valued random variable is of *lattice-type* if there is a pair of real numbers (α, β), such that the random variable has zero mass outside the lattice $\alpha + k\beta, k = \ldots, -1, 0, 1, \ldots$

Theorem 3: If $\log_2 P_{\mathsf{X}}(\mathsf{X})$ is a nonlattice random variable then, the minimum expected length of a lossless binary encoding of X^n is given by

$$L_n^* = nH(\mathsf{X}) - \frac{1}{2}\log_2(8\pi e\sigma^2 n) + o(1) \tag{34}$$

where

$$\sigma^2 = \mathrm{var}\left(\log_2 P_{\mathsf{X}}(\mathsf{X})\right). \tag{35}$$

Note that in the cases in which X is either equiprobable or binary valued, $\log_2 P_{\mathsf{X}}(\mathsf{X})$ only takes one or two values, respectively, and therefore it is a lattice distribution, outside the purview of Theorem 3. The complexity of the $O(1)$ term solution in the binary case, particularly (29), illustrates that a general expression for the lattice case may be challenging. Furthermore,

$$\begin{aligned} L(X) &\geq H(X) - (1 + L(X))\log_2(1 + L(X)) - L(X)\log_2 L(X) && (21)\\ &\geq H(X) - (1 + H(X))\log_2(1 + H(X)) - H(X)\log_2 H(X) && (22)\end{aligned}$$

$$L_n^* = nh(p) - \frac{1}{2}\log_2 n - \frac{1}{2}\log_2\frac{e^3}{\pi} + \frac{p}{1-2p} + \log_2\frac{1}{1-2p} + \frac{1}{2(1-2p)}\log_2\frac{1-p}{p} + o(1) \tag{27}$$

$$\frac{1}{2} + \frac{1-4p}{1-2p}\log_2 e + \frac{-3+7p}{1-2p}\mathbb{E}\left[\left\langle \frac{J}{2}(n+1)\log_2\frac{1}{1-p} - \frac{1}{2}U^2\log_2 e - \frac{J}{2}\log_2 2\pi p(1-2p)^2 \right\rangle\right] \tag{29}$$

for any fixed n, one can modify P_{X} so slightly that $\log_2 P_{\mathsf{X}}(\mathsf{X})$ becomes nonlattice and the change in L_n^* is as small as desired. Therefore, pursuing the modification to the $O(1)$ term for sources with lattice-type $\log_2 P_{\mathsf{X}}(\mathsf{X})$ does not in fact improve the usefulness of the asymptotic results as approximations to finite-n fundamental limits.

At the expense of a weaker conclusion, the following result is more general than Theorem 3.

Theorem 4: If P_{X} is not equiprobable, the minimum expected length of a lossless binary encoding of X^n is given by

$$L_n^* = nH(\mathsf{X}) - \frac{1}{2}\log_2 n + O(1). \tag{36}$$

The dominant sublinear term is, thus, independent of the distribution of the source (as long as it is not equiprobable). It is tempting to conjecture that the same behavior holds for finite-alphabet Markov chains and other sources whose memory decays sufficiently rapidly.

V. Proofs

Proof of Theorem 2: Substituting

$$M = |\mathcal{A}|^n \tag{37}$$

in (6), we obtain

$$L_n^* = \lfloor n\log_2|\mathcal{A}|\rfloor + \frac{1}{|\mathcal{A}|^n}\left(2 + \lfloor n\log_2|\mathcal{A}|\rfloor - 2^{\lfloor n\log_2|\mathcal{A}|\rfloor+1}\right) \tag{38}$$

$$= \lfloor n\log_2|\mathcal{A}|\rfloor - 2^{\lfloor n\log_2|\mathcal{A}|\rfloor+1-n\log_2|\mathcal{A}|} + \frac{1}{|\mathcal{A}|^n}\left(2 + \lfloor n\log_2|\mathcal{A}|\rfloor\right) \tag{39}$$

$$= n\log_2|\mathcal{A}| - 2 + \tau(n\log_2|\mathcal{A}|) + o(\rho^n) \tag{40}$$

where $\rho > \frac{1}{|\mathcal{A}|}$ and the nonnegative function $\tau(x)$ is maximized when $x - \lfloor x\rfloor = 1 - \log_2\log_2 e$ at which point it attains the value

$$\tau(n+1-\log_2\log_2 e) = 1 + \log_2\log_2 e - \log_2 e = 0.086 \tag{41}$$

■

Proof of Theorem 3: As shown in [32], the analysis of the minimal length of the optimal fixed-to-variable nonprefix code is intimately connected to the analysis of error probability in fixed-to-fixed data compression: the minimum error probability of an n-to-k fixed-to-fixed code $\epsilon^*(n,k)$ is equal to the probability that the minimum length of the fixed-to-variable code is greater than or equal to k, i.e.,

$$\epsilon^*(n,k) = \mathbb{P}\left[\lfloor\log_2 \pi_{X^n}(X^n)\rfloor \geq k\right]. \tag{42}$$

To verify (42), note that the optimum n-to-k fixed-to-fixed code assigns a unique k-bit string to each of the most likely $2^k - 1$ realizations of X^n, and uses the string with k 1s to signal error; thus, an error obtains if $\pi_{X^n}(X^n) \geq 2^k$, which happens with probability

$$\mathbb{P}\left[\log_2 \pi_{X^n}(X^n) \geq k\right] = \mathbb{P}\left[\lfloor\log_2 \pi_{X^n}(X^n)\rfloor \geq k\right] \tag{43}$$

and (42) is established. This enables us to analyze the minimum average length of fixed-to-variable coding through the analysis of the fixed-to-fixed error probability

$$L_n^* = \sum_{k=1}^{\infty}\mathbb{P}\left[\lfloor\log_2 \pi_{X^n}(X^n)\rfloor \geq k\right] \tag{44}$$

$$= \sum_{k=1}^{\infty}\epsilon^*(n,k) \tag{45}$$

$$= -\epsilon^*(n,1) + \sum_{k=1}^{\infty}(k+1)\left(\epsilon^*(n,k) - \epsilon^*(n,k+1)\right) \tag{46}$$

$$= -1 + \int_{\epsilon^*(n,1)}^{1} dt + \sum_{k=1}^{\infty}\int_{\epsilon^*(n,k+1)}^{\epsilon^*(n,k)}(k+1)\,dt \tag{47}$$

$$= -1 + \int_0^1 n\,R^*(n,\epsilon)\,d\epsilon \tag{48}$$

where

$$R^*(n,\epsilon) = \frac{k}{n} \text{ if } \epsilon^*(n,k) \leq \epsilon < \epsilon^*(n,k-1) \tag{49}$$

is the smallest rate of an n-to-k code with error probability not exceeding ϵ. Based on the refined central limit theorem (e.g., [22]), Strassen [26] showed that for a memoryless source with a nonlattice distribution, we have (50)–(52), as shown at the bottom of the page, where Q^{-1} is the inverse of the complementary cumulative Gaussian distribution function

$$Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-t^2/2}\,dt \tag{53}$$

$$\sigma^2 = \mathbb{E}\left[\log_2^2\frac{1}{P_{\mathsf{X}}(\mathsf{X})}\right] - H^2 > 0 \tag{54}$$

$$\mu_3 = \mathbb{E}\left[\left(\log_2\frac{1}{P_{\mathsf{X}}(\mathsf{X})} - H\right)^3\right] \tag{55}$$

which are both finite since the alphabet is finite.

$$R^*(n,\epsilon) = \bar{R}(n,\epsilon) + \Delta(n,\epsilon) \tag{50}$$

$$\bar{R}(n,\epsilon) = H + \frac{\sigma}{\sqrt{n}}Q^{-1}(\epsilon) - \frac{1}{2n}\log_2\left(2\pi\sigma^2 n e^{(Q^{-1}(\epsilon))^2}\right) + \frac{\mu_3}{6\sigma^2 n}\left((Q^{-1}(\epsilon))^2 - 1\right) \tag{51}$$

$$\Delta(n,\epsilon) = o\left(\frac{1}{n}\right) \tag{52}$$

In order to integrate (50) with respect to ϵ, note that if U is uniform on $[0,1]$, $Q^{-1}(U)$ is a standard Gaussian distribution. Therefore

$$\int_0^1 Q^{-1}(\epsilon)\, d\epsilon = 0 \tag{56}$$

$$\int_0^1 \left(Q^{-1}(\epsilon)\right)^2 \, d\epsilon = 1. \tag{57}$$

Using (56) and (57), we obtain (58)–(59), shown at the bottom of the page.

We now proceed to deal with the integration of $n\Delta(n,\epsilon)$. Denote the zero-mean unit-variance random variable

$$Z_n = \frac{1}{\sqrt{n}\sigma}\left(\sum_{i=1}^{n} \log_2 \frac{1}{P_{\mathsf{X}}(X_i)} - nH\right). \tag{60}$$

Letting α_3 be the third (noncentered) moment of $\log_2 \frac{1}{P_{\mathsf{X}}(\mathsf{X})}$, [22, Theorem 5.22] states that the function

$$\phi_n(x) = \mathbb{P}[Z_n \le x] - 1 + Q(x) - \frac{\alpha_3}{6\sigma^3\sqrt{2\pi n}}(1-x^2)e^{-x^2/2} \tag{61}$$

is such that for any $\tau > 0$, there exists n_0, such that for all $n > n_0$

$$|\phi_n(x)| < \frac{\tau}{\sqrt{n}} \tag{62}$$

for all real scalars x. Letting

$$\rho_n = \exp(\sqrt{n}\,\sigma\, Q^{-1}(\epsilon)) \tag{63}$$

and in view of [26], we can bound for any $\tau > 0$ and all sufficiently large n, we get (64)–(65), shown at the bottom of the page. By monotone convergence, the integral of (65) with respect to ϵ on $[0,1]$ vanishes. Together with (48) and (59) the desired result is established. ■

Proof of Theorem 4: Henceforth, we assume that the source is not equiprobable. We abbreviate $|\mathcal{A}| = m$, denote by $p_1, \dots p_m$ the atoms of P_{X} such that

$$p_1 \le p_2 \le \cdots \le p_{m-1} \le p_m \tag{66}$$

and we denote

$$B_i = \log \frac{p_m}{p_i} \tag{67}$$

for $i = 1, \dots, m-1$. Note that the entropy of P_{X} can be expressed as

$$H(\mathsf{X}) = \log \frac{1}{p_m} + \sum_{i=1}^{m-1} p_i B_i. \tag{68}$$

Let

$$\mathbf{k} = (k_1, \dots, k_m) \tag{69}$$

such that $k_1 + \cdots + k_m = n$ denote the *type* of an n-string; the probability of each such string is equal to

$$p^{\mathbf{k}} = p_1^{k_1} \dots p_m^{k_m}. \tag{70}$$

Denote the set of all types of n-strings drawn from an alphabet of m elements by

$$\mathcal{T}_{n,m} = \{(k_1, \dots, k_m) \in \mathbb{N}^m, k_1 + \cdots + k_m = n\}. \tag{71}$$

We introduce an order among types:

$$\mathbf{j} \preceq \mathbf{k} \quad \text{iff} \quad p^{\mathbf{j}} \ge p^{\mathbf{k}}.$$

and we sort all types from the smallest index (largest probability) to the largest. This can be accomplished by observing that $p^{\mathbf{j}} \ge p^{\mathbf{k}}$ is equivalent to

$$j_1 B_1 + \cdots + j_{m-1}B_{m-1} \le k_1 B_1 + \cdots + k_{m-1}B_{m-1}. \tag{72}$$

Therefore, to sort types $\mathbf{k}$ one needs to sort the function $S : \mathbb{R}^{m-1} \mapsto \mathbb{R}^+$

$$S(\mathbf{k}) = k_1 B_1 + \cdots + k_{m-1}B_{m-1} \tag{73}$$

from the smallest value $S(00\cdots 0) = 0$ to the largest.

There are

$$\binom{n}{\mathbf{k}} = \binom{n}{k_1, \dots, k_m} = \frac{n!}{k_1! \cdots k_m!} \tag{74}$$

$$\begin{aligned}\int_0^1 n\,\bar{R}^*(n,\epsilon)\, d\epsilon &= n\,H - \frac{1}{2}\log_2\left(2\pi\sigma^2 n\right) \\ &\quad + \int_0^1 \sigma\sqrt{n}\,Q^{-1}(\epsilon) - \frac{\log_2 e}{2}(Q^{-1}(\epsilon))^2 + \frac{\mu_3}{6\sigma^2}\left((Q^{-1}(\epsilon))^2 - 1\right)\, d\epsilon && (58)\\ &= n\,H - \frac{1}{2}\log_2\left(2\pi\sigma^2 ne\right). && (59)\end{aligned}$$

$$\begin{aligned} n\Delta(n,\epsilon) &= \log\left(1 + \exp\left(-n\bar{R}(n,\epsilon)\right)\int_0^{\rho_n} \phi_n\left(Q^{-1}(\epsilon)\right) - \phi_n\left(\frac{\log z}{\sqrt{n}\sigma}\right)\, dz\right) && (64)\\ &\le \log\left(1 + \exp\left(-n\bar{R}(n,\epsilon)\right)\frac{2\rho_n\tau}{\sqrt{n}}\right). && (65)\end{aligned}$$

sequences of type $\mathbf{k}$ and we list them in lexicographic order. Then, the optimum code assigns length $\lfloor \log i \rfloor$ to the ith sequence $(1 \leq i \leq m^n)$ in this list. Denote the number of sequences more probable than or equal to type $\mathbf{k}$ as

$$A_{\mathbf{k}} := \sum_{\mathbf{j} \preceq \mathbf{k}} \binom{n}{\mathbf{j}}. \tag{75}$$

Using somewhat informal, but intuitive, notation, $\mathbf{k}+1$ and $\mathbf{k}-1$ denote the *next* and *previous* types, respectively, in the sorted list of the elements of $\mathcal{T}_{n,m}$. Clearly, starting from position $A_{\mathbf{k}}$ the next $\binom{n}{\mathbf{k}+1}$ sequences have probability $p^{\mathbf{k}+1}$. Thus, the average code length can be computed as follows:

$$\begin{aligned} L_n^* &= \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} p^{\mathbf{k}} \sum_{i=A_{\mathbf{k}-1}+1}^{A_{\mathbf{k}}} \lfloor \log_2 i \rfloor && (76)\\ &= \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} p^{\mathbf{k}} \sum_{i=1}^{\binom{n}{\mathbf{k}}} \lfloor \log_2 (A_{\mathbf{k}} - i + 1) \rfloor && (77)\\ &= \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} \binom{n}{\mathbf{k}} p^{\mathbf{k}} \log_2 A_{\mathbf{k}} + O(1) && (78)\\ &= \log_2 A_{n\mathbf{p}} + O(1) && (79) \end{aligned}$$

where $\mathbf{p} = (p_1, \ldots, p_m)$ with $p_m = 1 - p_1 - \cdots - p_{m-1}$. We now proceed to justify (78) and (79). Noticing that for $1 \leq i \leq \binom{n}{\mathbf{k}}$

$$\log_2 \left(A_{\mathbf{k}} - \binom{n}{\mathbf{k}} + 1 \right) \leq \lfloor \log_2 (A_{\mathbf{k}} - i + 1) \rfloor \leq \log_2 (A_{\mathbf{k}} + 1) \tag{80}$$

we conclude that We first estimate the second sum on the left side of (81). In (101) and (103) below, we establish that

$$\frac{\binom{n}{n\mathbf{p}}}{A_{\mathbf{p}}} = O\left(n^{-(m-2)/2} \right) \tag{83}$$

which along with $\log(1-x) = -x + O(x^2)$ enables us to conclude that the second sum in (81) is of order $O(n^{-(m-2)/2})$.

In order to verify (78), we shall use a *multinomial sum* paradigm of the following form:

$$S_f(n) := \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} \binom{n}{\mathbf{k}} p^{\mathbf{k}} f(\mathbf{k}) \tag{84}$$

where $f(\mathbf{k})$ is a function of at most polynomial growth. In our case, $f(\mathbf{k}) = \log A_{\mathbf{k}} = \Theta(n)$, where $n = k_1 + \cdots + k_m$. In [11] and [16], it is proven that such a sum grows asymptotically as $f(n\mathbf{p})$. For the reader's convenience, we offer a streamlined justification for functions of polynomial growth; in particular when $f(\mathbf{k})$ has an analytic continuation to a complex cone around the real positive axis [16], [30].

In general, Taylor's expansion of f around $n\mathbf{p}$ is as shown in (85), at the bottom of the page, for some $\mathbf{x}'$ in the vicinity of $n\mathbf{p}$, where we use the same simplified notations as before. Observe now that

$$\begin{aligned} S_f(n) &= \mathbb{E}[f(\mathbf{X})] && (86)\\ &= f(n\mathbf{p}) + O(n \max_{\mathbf{x}', ij} f''_{ij}(\mathbf{x}')) && (87)\\ &= f(n\mathbf{p}) + O(n\xi(n)) && (88) \end{aligned}$$

where $\mathbf{X}$ is a multinomial distribution with parameters n and $\mathbf{p}$ and $f''_{ij}(\mathbf{x})$ is the second derivative with respect to x_i and x_j. Observe that in (87), we use the fact that variance of $\mathbf{X}$ is of order $O(n)$. The above asymptotic result is useful as long as the first term dominates the second term $O(n\xi(n))$, as is the case in our situation. One can argue that f has an analytic continuation in a cone around the real positive axis and polynomial growth (cf. (105) below). By [15, Lemma 3] or [30], we conclude that $n\xi(n) = O(1/n)$ and $f''(\mathbf{k}) = O(1/n)$. Thus, (78)–(79) follow.

Let now

$$j_i = np_i + x_i \tag{89}$$

for $i = 1, \ldots, m-1$. Then, by (72) $p^{\mathbf{j}} \geq p^{n\mathbf{p}}$, is equivalent to

$$B_1 x_1 + \cdots + B_{m-1} x_{m-1} \leq 0. \tag{90}$$

Thus

$$\begin{aligned} A_{n\mathbf{p}} &= \sum_{p^{\mathbf{j}} \geq p^{n\mathbf{p}}} \binom{n}{\mathbf{j}} && (91)\\ &= \sum_{\mathbf{x}: \mathbf{b}^T \mathbf{x} \leq 0} \binom{n}{n\mathbf{p} + \mathbf{x}} && (92) \end{aligned}$$

where

$$\begin{aligned} \mathbf{x}^T &= [x_1, \ldots, x_{m-1}] && (93)\\ \mathbf{b}^T &= [B_1, \ldots, B_{m-1}]. && (94) \end{aligned}$$

$$\begin{aligned} \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} \binom{n}{\mathbf{k}} p^{\mathbf{k}} \log_2 A_{\mathbf{k}} + \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} \binom{n}{\mathbf{k}} p^{\mathbf{k}} \log \left(1 - \frac{\binom{n}{\mathbf{k}} - 1}{A_{\mathbf{k}}} \right) &\leq L_n^* && (81)\\ &\leq \sum_{\mathbf{k} \in \mathcal{T}_{n,m}} \binom{n}{\mathbf{k}} p^{\mathbf{k}} \log_2 (A_{\mathbf{k}} + 1). && (82) \end{aligned}$$

$$f(\mathbf{x}) = f(n\mathbf{p}) + (\mathbf{x} - \mathbf{p}) \nabla f(n\mathbf{p}) + \frac{1}{2} (\mathbf{x} - n\mathbf{p}) \nabla^2 f(\mathbf{x}') (\mathbf{x} - n\mathbf{p}) \tag{85}$$

The next step is to use Stirling's formula

$$n! = \sqrt{2\pi n} \cdot n^n e^{-n}(1 + O(1/n)) \tag{95}$$

to estimate the summands in (92). This leads to (96), also shown at the bottom of the page. Applying now Taylor's expansion to (96), we get (97)–(99), as shown at the bottom of the page, and we arrive at (100)–(101), as shown at the bottom of the page, where $\mathbf{\Sigma}$ is an appropriately chosen invertible covariance matrix.

We are now in the position to evaluate the sum (92). We need to sum over $\mathbf{b}^T\mathbf{x} \leq 0$ which we split by summing over hyperplanes $\mathbf{b}^T\mathbf{x} = -d$ for $d \geq 0$ of dimension $m-2$. We denote such a hyperplane by $\mathcal{D}^{m-2} = \{\mathbf{x} : \mathbf{b}^T\mathbf{x} = -d\}$. Noting that the Gaussian kernel of (101) when summed over the hyperplane $\mathcal{D}^{m-2}$ is of order $O(n^{(m-2)/2})$ we arrive at our final result. More precisely, plugging (101) into (92) yields

$$A_{n\mathbf{p}} = \frac{C2^{nH(\mathsf{X})}}{n^{(m-1)/2}} \left(\sum_{\mathbf{b}^T\mathbf{x} \leq 0} \exp\left(\mathbf{b}^T\mathbf{x} - \frac{1}{2n}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} \right) \right) \tag{102}$$

$$= \sum_{d \geq 0} \exp(-d) \sum_{\mathbf{b}^T\mathbf{x}=-d} \exp\left(-\frac{1}{2n}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} \right). \tag{103}$$

Noting now that [13]

$$\sum_{\mathbf{x}\in\mathcal{D}^{m-2}} \exp\left(-\frac{1}{2n}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} \right) = C(d)n^{(m-2)/2} \tag{104}$$

$$\begin{aligned}
\binom{n}{n\mathbf{p}+\mathbf{x}} &= \frac{n!}{(np_1+x_1)!\cdots(np_{m-1}+x_{m-1})!(np_m - x_1 - \cdots - x_{m-1})!} = \\
&\frac{\sqrt{2\pi n}n^n e^{-n} e^{np_1+x_1}\cdots e^{np_m-x_1-\cdots-x_{m-1}}(1+O(1/n))}{\sqrt{2\pi(np_1+x_1)}(np_1+x_1)^{np_1+x_1}\cdots\sqrt{2\pi(np_m-x_1-\cdots-x_{m-1})}(np_m-x_1-\cdots-x_{m-1})^{np_m-x_1-\cdots-x_{m-1}}} \\
&= \frac{1}{(2\pi)^{(m-1)/2}}\frac{1}{\sqrt{p_1\cdots p_m}}\frac{1}{n^{(m-1)/2}}\left(1+O(1/\sqrt{n})\right)\cdot \\
&\cdot \frac{n^n}{(np_1)^{np_1+x_1}\left(1+\frac{x_1}{np_1}\right)^{np_1+x_1}\cdots(np_m)^{np_m-x_1\cdots x_{m-1}}\left(1-\frac{x_1+\cdots+x_{m-1}}{np_m}\right)^{np_m-x_1-\cdots-x_{m-1}}} \\
&= \frac{1}{(2\pi)^{(m-1)/2}}\frac{1}{\sqrt{p_1\cdots p_m}}\frac{1}{n^{(m-1)/2}}\frac{1}{p_1^{np_1}\cdots p_m^{np_m}}\left(\frac{p_m}{p_1}\right)^{x_1}\cdots\left(\frac{p_m}{p_{m-1}}\right)^{x_{m-1}}\left(1+O(1/\sqrt{n})\right) \\
&\cdot\left(1+\frac{x_1}{np_1}\right)^{-(np_1+x_1)}\cdots\left(1-\frac{x_1+\cdots x_{m-1}}{np_m}\right)^{-(np_m-x_1\cdots-x_{m-1})}
\end{aligned} \tag{96}$$

$$\begin{aligned}
\left(1+\frac{x}{np}\right)^{-(np+x)} &= \exp\left(-(np+x)\ln\left(1+\frac{x}{np}\right)\right) && (97)\\
&= \exp\left(-(np+x)\left(\frac{x}{np}-\frac{x^2}{2(np)^2}+O(n^{-3})\right)\right) && (98)\\
&= \exp\left(-\frac{x^2}{2np}\right)(1+O(1/n)) && (99)
\end{aligned}$$

$$\begin{aligned}
\binom{n}{n\mathbf{p}+\mathbf{x}} &= \frac{1}{(2\pi)^{(m-1)/2}}\frac{1}{\sqrt{p_1\cdots p_m}}\frac{1}{n^{(m-1)/2}}2^{nH(\mathsf{X})} \\
&\cdot\left(\frac{p_m}{p_1}\right)^{x_1}\cdots\left(\frac{p_m}{p_{m-1}}\right)^{x_{m-1}}\left(1+O(1/\sqrt{n})\right) \\
&\cdot\exp\left(-\frac{x_1^2}{2np_1}-\cdots-\frac{x_{m-1}^2}{2np_{m-1}}-\frac{(x_1+\cdots+x_{m-1})^2}{2np_m}\right) && (100)\\
&= \left(1+O(1/\sqrt{n})\right)C\frac{2^{nH(\mathsf{X})}}{n^{(m-1)/2}} \\
&\cdot\exp\left(B_1x_1+\cdots+B_{m-1}x_{m-1}\right) \\
&\cdot\exp\left(-\frac{1}{2n}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right) && (101)
\end{aligned}$$

where $C(d)$ is of at most polynomial growth of d (in fact, $C(d) = O(d^2)$). Combining (103) and (104), we finally arrive at

$$\begin{aligned}\log_2 A_{n\mathbf{p}} &= \log_2 \left(C' \frac{2^{nH(\mathsf{X})}}{n^{(m-1)/2}} n^{(m-2)/2} \right) \\ &= nH(\mathsf{X}) - \frac{1}{2}\log_2 n + O(1) \qquad (105)\end{aligned}$$

where C' is a constant. Observe that the right order of $A_{n\mathbf{p}}$ can be obtained by considering only the hyperplane $\mathbf{b}^T\mathbf{x} = 0$. In view of (79), this completes the proof of Theorem 4. ■

Example: To illustrate our methodology, we explain it first for $m = 2$ and then we give some details for the case of $m = 3$ symbols with probability $p_1 < p_2 < p_3$. For $m = 2$, we have $(p < 1 - p)$, shown in (106) and (107) at the bottom of the page. Observe again that the order of growth of A_{np} is determined by $x = 0$. The summation of the geometric series contributes to the constant.

Let's now focus on the case $m = 3$. With $B_1 = \log(p_3/p_1)$ and $B_2 = \log_2(p_3/p_2)$, we need to evaluate

$$A_{np_1,np_2} = \sum_{k_1B_1+k_2B_2 \le np_1B_1+np_2B_2} \binom{n}{k_1, k_2}. \qquad (108)$$

As before, we denote $k_1 = np_1 + x$ and $k_2 = np_2 + y$ to arrive at (109), also shown at the bottom of the page. In Fig. 1, we show the behavior of the above multinomial coefficient on the critical line $k_1B_1 + k_2B_2 = 0$ and below it. On the critical line the coefficient is well approximated by the normal distribution (curve labelled "normal") around the point (np_1, np_2), while for (k_1, k_2) (or equivalently for (x, y)) away from the critical line the coefficient decays exponentially. This leads to (110)–(113), as shown at the bottom of the page, where (112) follows from the normal approximation on the line $B_1x + B_2y = 0$.

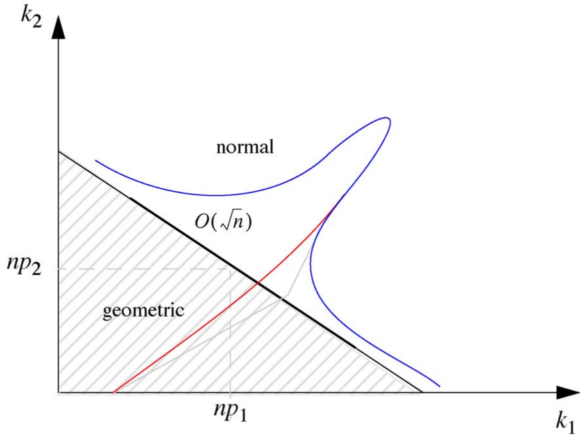

Fig. 1. Illustration for $m = 3$. The value of the multinomial coefficient (109) is shown as the third dimension: The normal distribution is along the line $k_1B_1 + k_2B_2 = np_1B_1 + np_2B_2$, while away from this line the multinomial coefficient decays exponentially.

$$\binom{n}{np - x} = \frac{2^{nH(\mathsf{X})}}{\sqrt{2\pi p(1-p)n}} \left(\frac{p}{1-p}\right)^x \exp\left(-\frac{x^2}{2np(1-p)}\right)(1 + O(1/n)) \qquad (106)$$

$$A_{np} = \sum_{x \ge 0} \binom{n}{np - x} = \frac{1}{1 - \frac{p}{(1-p)}} \frac{2^{nH(\mathsf{X})}}{\sqrt{2\pi p(1-p)n}}(1 + O(1/n)) \qquad (107)$$

$$\begin{aligned}\binom{n}{np_1 + x, np_2 + y} = &\frac{1}{\sqrt{2\pi p_1p_2p_3 n}} 2^{nH(\mathbf{p})} \left(\frac{p_3}{p_1}\right)^x \left(\frac{p_3}{p_2}\right)^y \\ &\cdot \exp\left(-\frac{x^2}{2np_1} - \frac{y^2}{2np_2} - \frac{(x+y)^2}{2np_3}\right)\left(1 + O(1/\sqrt{n})\right) \qquad (109)\end{aligned}$$

$$\begin{aligned}A_{n\mathbf{p}} &= \sum_{B_1x+B_2y \le 0} \binom{n}{np_1 + x, np_2 + y} && (110) \\ &\sim \frac{2^{nH(\mathsf{X})}}{n\sqrt{2\pi p_1p_2p_3}} \sum_{B_1x+B_2y=0} \exp\left(-\frac{x^2}{2np_1} - \frac{y^2}{2np_2} - \frac{(x+y)^2}{2np_3}\right) && (111) \\ &= O(\sqrt{n})\frac{2^{nH(\mathsf{X})}}{n} && (112) \\ &= C\frac{2^{nH(\mathsf{X})}}{\sqrt{n}} && (113)\end{aligned}$$

ACKNOWLEDGMENT

The detailed reading and useful suggestions by Associate Editor E. Ordentlich are gratefully acknowledged.

REFERENCES

[1] N. Alon and A. Orlitsky, "A lower bound on the expected length of one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 40, no. 5, pp. 1670–1672, Sep. 1994.

[2] C. Blundo and R. de Prisco, "New bounds on the expected length of one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 246–250, Jan. 1996.

[3] R. Capocelli and A. de Santis, "Tight upper bounds on the redundancy of Huffman codes," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, pp. 1084–1091, Sep. 1989.

[4] R. Capocelli and A. de Santis, "New bounds on the redundancy of Huffman codes," *IEEE Trans. Inf. Theory*, vol. 37, pp. 1095–1104, Jul. 1991.

[5] J. Cheng, T. K. Huang, and C. Weidmann, "New bounds on the expected length of optimal one-to-one codes," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, pp. 1884–1895, May 2007.

[6] M. Drmota, "A bivariate asymptotic expansion of coefficients of powers of generating functions," *Eur. J. Combinator.*, vol. 15, pp. 139–152, 1994.

[7] M. Drmota, H-K. Hwang, and W. Szpankowski, "Precise average redundancy of an idealized arithmetic coding," in *Proc. Data Compression Conf.*, Snowbird, UT, 2002, pp. 222–231.

[8] M. Drmota and R. Tichy, *Sequences, Discrepancies, and Applications*. Berlin: Springer-Verlag, 1997.

[9] M. Drmota and W. Szpankowski, "Precise minimax redundancy and regret," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2686–2707, Nov. 2004.

[10] J. G. Dunham, "Optimal noiseless coding of random variables," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, p. 345, May 1980.

[11] P. Flajolet, "Singularity analysis and asymptotics of Bernoulli sums," *Theoret. Comput. Sci.*, vol. 215, pp. 371–381, 1999.

[12] R. Gallager, "Variations on a theme by Huffman," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 6, pp. 668–674, Nov. 1978.

[13] B. Gnedenko, *The Theory of Probability and Elements of Statistics*. New York: Chelsea Publishing Company, 1991.

[14] P. Henrici, *Applied and Computational Complex Analysis*. New York: Wiley, 1977, vol. 2.

[15] P. Jacquet and W. Szpankowski, "Analytical depoissonization and its applications," *Theoret. Comput. Sci.*, vol. 201, pp. 1–62, 1998.

[16] P. Jacquet and W. Szpankowski, "Entropy computations via analytic depoissonization," *IEEE Trans. Inf. Theory*, vol. 45, no. 4, pp. 1072–1081, May 1999.

[17] D. Knuth, *The Art of Computer Programming: Fundamental Algorithms*, 3rd ed. Reading, MA: Addison-Wesley, 1997, vol. 1.

[18] S. K. Leung-Yan-Cheong and T. Cover, "Some equivalences between Shannon entropy and Kolmogorov complexity," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 331–338, May 1978.

[19] B. MacMillan, "Two inequalities implied by unique decipherability," *IRE Trans. Inf. Theory*, vol. 2, pp. 115–116, Dec. 1956.

[20] M. K. McKusick, W. N. Joy, S. J. Leffler, and R. S. Fabry, "A fast file system for UNIX," *ACM Trans. Comput. Syst.*, vol. 2, no. 3, pp. 181–197, 1984.

[21] D. Manstetten, "Tight upper bounds on the redundancy of Huffman codes," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 144–151, Jan. 1992.

[22] V. V. Petrov, *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford, U.K.: Oxford Science Publications, 1995.

[23] J. Rissanen, "Tight lower bounds for optimum code length," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 348–349, Mar. 1982.

[24] J. Rissanen, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, pp. 629–636, Jul. 1984.

[25] S. Savari, "On one-to-one codes for memoryless cost channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 367–379, Jan. 2008.

[26] V. Strassen, "Asymptotische Abschäzungen in Shannons informationstheorie," in *Trans. 3rd Conf. Information Theory on Statistics, Decision Functions, Random Processes*, Prague, Czechoslovakia, 1964, pp. 689–723.

[27] P. Stubley, "On the redundancy of optimum fixed-to-variable length codes," in *Proc. Data Compression Conf.*, Snowbird, UT, 1994, pp. 90–97.

[28] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pt. I, II, pp. 379–423, 623–656, Jul. and Oct. 1948, 1948.

[29] W. Szpankowski, "Asymptotic average redundancy of Huffman (and other) block codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2434–2443, Nov. 2000.

[30] W. Szpankowski, *Average Case Analysis of Algorithms in Sequences*. New York: Wiley, 2000.

[31] W. Szpankowski, "A one-to-one code and its anti-redundancy," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4762–4766, Oct. 2008.

[32] S. Verdú, ""teaching it," XXVIII Shannon Lecture," in *Proc. 2007 IEEE Int. Symp. Information Theory*, Nice, France, Jun. 28, 2007.

[33] E. I. Verriest, "An achievable bound for optimal noiseless coding of a random variable," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 4, pp. 592–594, Jul. 1986.

[34] A. D. Wyner, "An upper bound on the entropy series," *Inf. Control*, vol. 20, pp. 176–181, 1972.

[35] "IEEE Information Theory Society Newsletter," Dec. 2007.

Wojciech Szpankowski (F'04) received the M.S. and Ph.D. degrees in electrical and computer engineering from Gdansk University of Technology.

He is the Saul Rosen Professor of Computer Science and (by courtesy) Electrical and Computer Engineering at Purdue University, West Lafayette, IN, where he teaches and conducts research in analysis of algorithms, information theory, bioinformatics, analytic combinatorics, random structures, and stability problems of distributed systems. He held several Visiting Professor/Scholar positions, including McGill University, INRIA, France, Stanford, Hewlett-Packard Labs, Universite de Versailles, University of Canterbury, New Zealand, Ecole Polytechnique, France, and the Newton Institute, Cambridge, U.K. In 2008, he launched the interdisciplinary Institute for Science of Information, and in 2010 he became the Director of the newly established NSF Science and Technology Center for Science of Information. He is author of the book *Average Case Analysis of Algorithms on Sequences* (New York: Wiley, 2001).

Dr. Szpankowski is an Erskine Fellow. He received the Humboldt Research Award in 2009. He has been a Guest Editor and an Editor of technical journals, including *Theoretical Computer Science*, the *ACM Transaction on Algorithms*, the IEEE TRANSACTIONS ON INFORMATION THEORY, *Foundation and Trends in Communications and Information Theory*, *Combinatorics, Probability, and Computing*, and *Algorithmica*.

Sergio Verdú (F'93) is the Eugene Higgins Professor of Electrical Engineering at Princeton University, Princeton, NJ.

A member of the National Academy of Engineering, he is the recipient of the 2007 Claude Shannon Award and the 2008 IEEE Richard Hamming Medal. He was awarded a Doctorate Honoris Causa from the Universitat Politècnica de Catalunya in 2005.

He is a recipient of the 1992 Donald Fink Paper Award, the 1998 Information Theory Outstanding Paper Award, an Information Theory Golden Jubilee Paper Award, the 2002 Leonard Abraham Award, the 2006 Joint Communications/Information Theory Paper Award, the and the 2009 Stephen Rice Prize.

He is currently Editor-in-Chief of *Foundations and Trends in Communications and Information Theory*.