# The Pros and Cons of Democracy

A. R. Calderbank, *Fellow, IEEE,* and I. Daubechies, *Fellow, IEEE*

*Invited*

***Abstract*—We introduce the concept of "democracy," in which the individual bits in a coarsely quantized representation of a signal are all given "equal weight" in the approximation to the original signal. We prove that such democratic representations cannot achieve the same accuracy as optimal nondemocratic schemes.**

***Index Terms*—Democratic decoding, sigma–delta quantization.**

## I. Introduction and Motivation

The problem discussed in this correspondence arose in a study of the mathematical properties of so-called sigma–delta ($\Sigma\Delta$) converters, used extensively in analog to digital (A/D) and digital to analog (D/A) conversion, and of the accuracy that can be achieved by the most widely used decoders for the $\Sigma\Delta$ bitstreams, often carried out by simple convolution. There exist many different types of useful $\Sigma\Delta$ converters, adapted to a wide range of applications; for more information, see [1], [17]. For the mathematical discussion in this correspondence, we need only the following: given a band-limited function $f$ bounded by 1, i.e., $|f(t)| < 1$, sampled at rate $\lambda\nu$ (throughout this correspondence, $\nu$ stands for the Nyquist sampling rate $\pi/\Omega$ corresponding to the (fixed) band limit $\Omega$; we always consider $\lambda$ significantly larger than 1), 1-bit $\Sigma\Delta$-algorithms produce sequences $(q_n)_{n\in\mathbb{Z}}$, with $q_n \in \{-1, 1\}$, such that, for appropriate finite filter sequences $h$, one can construct a good approximation to the original sample values by convolving $h$ and $q$; moreover, the error

$$\left| f\left(\frac{k}{\lambda\nu}\right) - \sum_l h_{k-l} q_l \right| \tag{1}$$

can be bounded uniformly in $k$ and independently of the choice of $f$ within our constraints. Typically, the filters $h$ depend on $\lambda$, and their length increases linearly with $\lambda$; the approximation error (1) decreases as $\lambda$ increases, with a decay rate that depends on the chosen algorithm. For stable $M$th-order $\Sigma\Delta$-algorithms, and for filters $h^{[M](\lambda)}$ that satisfy certain technical conditions, one derives (see [2])

$$\left| f\left(\frac{k}{\lambda\nu}\right) - \sum_l h^{[M](\lambda)}_{k-l} q_l \right| \le C_{[M]} \lambda^{-M} \tag{2}$$

for all band-limited functions $f$ bounded uniformly by $A < 1$, with $C_{[M]}$ independent of $f$ and of $\lambda$. The estimate in [2] for the constant $C_{[M]}$ in (2) grows like $\alpha 2^{\beta M^2}$ with $M$. As far as we know, this is the only rigorous result for arbitrary bounded band-limited functions and arbitrary order. For the first-order $\Sigma\Delta$-algorithm, Güntürk proved the following stronger result [7]–[9]:

$$\left| f\left(\frac{k}{\lambda\nu}\right) - \sum_l h^{(\lambda)}_{k-l} q_l \right| \le C \lambda^{-4/3+\epsilon} \tag{3}$$

except near sample points where the derivative $f'$ of $f$ vanishes. (For an exact statement of this result and others, and for proofs, see [8], [9].) It is believed, and simulations bear out, that, if one averages over "time," the optimal result for first-order $\Sigma\Delta$ is slightly stronger than (3); more precisely

$$\lim_{K\to\infty} \frac{1}{2K} \sum_{k=-K}^{K} \left| f\left(\frac{k}{\lambda\nu}\right) - \sum_l h^{(\lambda)}_{k-l} q_l \right|^2 \le C\lambda^{-3}; \tag{4}$$

this has not been proved for general band-limited functions. Heuristic arguments, which replace the difference between $f(\frac{k}{\lambda\nu})$ and an auxiliary sequence $u_k$ that arises in the $\Sigma\Delta$-algorithms, by a uniformly distributed "white-noise" sequence, do lead to an estimate of type (4). For the very restricted case where $f$ is a constant, and after averaging over the possible values of this constant, the $\lambda^{-3}$ decay of (4) has been proved rigorously in [5]; see also [11] for another analysis. This result was also extended to other special cases for $f$, such as a sinusoid in [6].

For constant input, (2) can be improved for higher order $\Sigma\Delta$ schemes as well; see, e.g., [12]–[15] and [18]; these refinements give a small gain in the power of $\lambda$ and typically correspond to estimates of average errors (mean square error (MSE) estimates) as in the left-hand side of (4) rather than pointwise errors as in (2).

In all these studies, and even for constant input, the bounds on (1) decay like powers of $\lambda$. By optimizing over $M$ in (2) one achieves slightly more; for $M_\lambda = (\log\lambda)/2\beta$, one finds

$$\left| f\left(\frac{k}{\lambda\nu}\right) - \sum_l h^{[M_\lambda],(\lambda)}_{k-l} q_l \right| \le C\lambda^{-(\log\lambda)/4\beta}. \tag{5}$$

This approximation rate seems far from optimal. To illustrate this, let us consider the case where $f(t) = x$ is a constant. Since the filter length in (2) for an $M$th-order scheme is proportional to $M\lambda$, and the optimal $M_\lambda$ in (5) is of order $O(\log\lambda)$, formula (5) requires $O(\lambda\log\lambda)$ bits to reconstruct $x$ with accuracy $O(2^{-\gamma(\log\lambda)^2})$. This is a dismal performance when we compare with ordinary binary encoding; if $x \in (-1, 1)$ is written as $x = 2w - 1$, with

$$w = \sum_{k=1}^{\infty} b_k 2^{-k} \in (0, 1), \qquad b_k \in \{0, 1\}$$

then the choice $q_k = 2b_k - 1 \in \{-1, 1\}$ leads to

$$\left| x - \sum_{k=1}^{K} q_k 2^{-k} \right| \le 2^{-K} \tag{6}$$

or the familiar accuracy of order $O(2^{-\lambda})$ for $O(\lambda)$ bits.

There is one important difference between the approximations used in (6) and (1). The bits $q_k$ in the binary expansion in (6) play *unequal* roles: the first bit gives the largest contribution, and the importance of the $q_k$ diminishes exponentially as $k$ increases. In the convolution reconstruction in (1), the relative importance is spread more evenly over the different $q_l$: for one fixed $k$, the role of the $q_l$ in the reconstruction of $f(\frac{k}{\lambda\nu})$ is measured by $h_{k-l}$, and may of course vary with $l$; however, when one reconstructs a large family of consecutive $f(\frac{k}{\lambda\nu})$, say for $-K \le k \le K$, where $K$ exceeds the length of the filter $h$, i.e., $K > L := \#\{l;\, h_l \neq 0\}$, then the $q_l$ with $-K + L \le l \le K - L$

Manuscript received September 28, 2000; revised August 10, 2001. This work was supported in part by the National Science Foundation and the Air Force Office for Scientific Research.

A. R. Calderbank is with the AT&T Labs, Florham Park, NJ 07932 USA (e-mail: rc@research.att.com).

I. Daubechies is with the Mathematics Department and Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544 USA (e-mail: ingrid@math.princeton.edu).

Communicated by S. Shamai, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04036-1.

are used, on the average, with the same weight. We shall describe this "equal influence" of the $q_l$ in (1) by saying that these bits are *democratic* in the convolutional approximation used in (1). (Note that other, nonlinear, formulas have been proposed to compute an approximation to the sample values of $f$ from the $\Sigma\Delta$ sequence $(q_l)_{l\in\mathbb{Z}}$, in which the $q_l$ are not treated democratically, since nonlinear relationships with the neighboring $q_{l+j}$ play a role. In general, these more sophisticated techniques succeed in improving the polynomial decay rate of the approximation error, but the approximation rate is still only polynomial. In [11], for instance, it is shown that no formula, linear or not, can achieve better than $O(\lambda^{-2})$ approximation for a first-order $\Sigma\Delta$-scheme.)

The question that concerns us here is whether and to what extent the democracy of the $q_l$ in (1) contributes to the less-than-optimal accuracy of these convolutional approximations, as shown in estimate (5). This correspondence will give a partial answer only, because we shall restrict ourselves to considering constant functions, instead of the much larger class of bounded band-limited functions. Even for this much smaller family, we shall see that a decoder that treats bits democratically cannot achieve the $O(2^{-K})$ approximation rate, given $K$ bits, that is customary for binary expansions; this is our main result. However, there remains a large gap, as we shall see, between the theoretical limitations imposed by democracy, as proved by our Theorem 2.3, and the approximation rate of (5).

In Section II, we define more precisely what we mean, in the framework of this correspondence, by *democracy* for the encoding and decoding of numbers in the interval $(-1, 1)$, and we prove our main theorem on the incompatibility of the optimal accuracy achieved by (6) with this notion of democracy. Although our original question was motivated by work of one of us on $\Sigma\Delta$ schemes, it is to be noted that Theorem 2.4 applies to arbitrary decoding schemes, whether they are associated with $\Sigma\Delta$ or not. In Section III, we discuss our result and its shortcomings, and we outline some open questions in this context.

## II. Democracy Versus Accuracy

For the sake of simplicity, we transform our problem from the interval $(-1, 1)$ to $(0, 1)$ (as in the argument preceding (6)); we can also, without loss of generality, include the two endpoints.

Consider now two maps, the *$N$-bit encoder* $E$ from $[0, 1]$ to $\{0, 1\}^N$, and the *decoder* $D$ from $\{0, 1\}^N$ to $[0, 1]$. Given a number $w$ in $[0, 1]$, $E(w)$ gives us an $N$-bit representation for $w$; from an $N$-bit representation $a \in \{0, 1\}^N$, the decoder gives us a candidate number $D(a)$. We shall assume that, for all $N$-bit sequences $a$ in $\{0, 1\}^N$, $ED(a) = a$; however, for all but at most $2^N$ elements $w$ of $[0, 1]$, $DE(w) \neq w$, since the range of $D$ contains at most $2^N$ elements.

It is then natural to define the *accuracy* $\mathcal{A}(E, D)$ of the encoding+decoding process as the maximum distortion that can be incurred

$$\mathcal{A}(E, D) := \sup_{w\in[0, 1]} |w - DE(w)|.$$

For each $w \in [0, 1]$, we can define the *encoding neighborhood* $V_w$ of $w$ as the set of all elements of $[0, 1]$ that are encoded to the same $N$-bit sequence; we denote by $d_w$ the diameter of this set

$$V_w := E^{-1}[E(w)], \qquad d_w := \sup_{u,v\in V_w} |u - v|.$$

It follows from the definition of $\mathcal{A}(E, D)$ that, for all $w \in [0, 1]$, $d_w \leq 2\mathcal{A}(E, D)$. Note that the set $V_w$ is completely determined by $E(w)$; in particular, there are at most $2^N$ different sets $V_w$. Because the $V_w$ together cover $[0, 1]$, it follows that

$$1 \leq \sum_{a\in E([0, 1])} d_{D(a)} \leq 2^{N+1}\mathcal{A}(E, D).$$

The optimal accuracy $\mathcal{A}(E, D) = 2^{-(N+1)}$ is achieved by midpoint binary quantization, where

$$D(b_1, b_2, \ldots, b_N) = 2^{-(N+1)} + \sum_{k=1}^{N} 2^{-k} b_k$$

and where encoding of $w$ consists in picking the $N$-tuple $(b_1, b_2, \ldots b_N)$ corresponding to the nearest point to $w$ of this form.

Note that for any given decoder $D$ (that is, for any given set of $2^N$ points in $[0, 1]$), one can always use Voronoi regions to define the optimal encoder $E^{\mathrm{opt}}$ (i.e., the encoder $E$ that minimizes $\mathcal{A}(E, D)$ for that decoder) by encoding $w \in [0, 1]$ by the bit sequence $a$ for which $D(a)$ is closest to $w$. One can thus define an accuracy that is dependent on $D$ only

$$\begin{aligned}\mathcal{A}(D) &:= \mathcal{A}(E^{\mathrm{opt}}, D)\\ &= \max_{a\in\{0,1\}^N} [\min\{D(b);\, D(b) > D(a)\}\\ &\qquad - \max\{D(c);\, D(c) < D(a)\}]/2.\end{aligned}$$

Next, we want to formalize the concept of "democracy." It turns out [9] that there are many ways to do this, and that it matters whether one views democracy as a property of the encoder, of the decoder, or of the pair encoder+decoder. We shall here concentrate on the decoder. Intuitively, "democracy" means that all the entries in the $N$-tuple $a \in \{0, 1\}^N$ have an equal "influence" in the decoded outcome $D(a)$. To turn this into a quantitative statement, consider the decoding error that results from flipping the bit in $k$th position

$$e_D^k(a) := |D(a) - D(\varphi^k(a))|$$

where $\varphi^k(a)$ is defined by

$$\begin{aligned}(\varphi^k(a))_j &:= a_j, \qquad \text{if } j \neq k\\ (\varphi^k(a))_k &:= 1 - a_k.\end{aligned}$$

Democracy should ensure that none of these decoding errors can become overwhelmingly large, i.e., we should have a reasonable upper bound on the $e_D^k(a)$. The following easy argument shows that we cannot pick this upper bound too small. Consider an arbitrary encoder $E$ (it could be $E^{\mathrm{opt}}$), and take the two $N$-bit sequences $E(0)$ and $E(1)$; since one can transform one into the other by at most $N$ consecutive 1-bit flips, it follows that

$$\begin{aligned}1 - 2\mathcal{A}(E, D) &\leq |1-0| - |DE(1) - 1| - |DE(0) - 0|\\ &\leq |DE(1) - DE(0)| \leq N \sup_{a\in\{0,1\}^N, k=1,\ldots,N} e_D^k(a).\end{aligned}$$

It seems, therefore, reasonable to state that the decoding process is democratic only if all the $e_D^k(a)$ are bounded above by $C/N$; this becomes a meaningful statement only if we impose that $C$ be independent of $N$. This leads to the following definition.

*Definition 2.1:* A family of maps

$$\{D_N\colon \{0, 1\}^N \to [0, 1];\, N = 1, 2, \ldots\}$$

is called democratic if there exists a constant $C$, independent of $N$, such that

$$\max_{a\in\{0,1\}^N,\, k=1,\ldots,N} e_{D_N}^k(a) \leq C/N. \tag{7}$$

*Remarks:*

1) Note that in our definition we let $a$ roam over all of $\{0, 1\}^N$, not just over the range of $E$, which may be smaller. (For $\Sigma\Delta$ encoders, for instance, the range of $E$ is significantly smaller, because the $N$-bit sequences that are generated by $\Sigma\Delta$-encoding of constants satisfy very strong constraints.) This is why we stated above that the notion of democracy defined here is focused on the *decoder* only; if one brings in the *encoder* as well, matters become more complicated. For instance, a decoder could well fail to be democratic in our sense, yet still satisfy a restricted democracy with respect to the encoding, as in the requirement

$$\max_{a\in E([0,1])}\ \max_{k=1,\ldots,N,\,\varphi^k(a)\in E([0,1])} e_D^k(a) \le C/N.$$

In addition, one can also impose a lower bound of $O(N^{-1})$ on the *minimum* error. For a discussion of these and additional refinements, see [10].

2) For binary encoding and decoding, one has $e_{D_N}^k(a) = 2^{-k}$, independently of $N$, which is obviously not democratic.

An immediate consequence of (7) is that, for all $D_N$ in a democratic family of decoders, and for all $a, b \in \{0, 1\}^N$

$$|D_N(a) - D_N(b)| \le C d_H(a, b)/N$$

where $d_H$ stands for the Hamming distance

$$d_H(a,b) := \#\{j;\ a_j \ne b_j\}$$

and where $C$ is the same constant as in (7).

Let us also formalize the concept of optimal accuracy for a family of decoders.

*Definition 2.2:* A family of maps

$$\{D_N\colon \{0, 1\}^N \to [0, 1];\ N = 1, 2, \ldots\}$$

is called optimally accurate if there exists a constant $C$, independent of $N$, such that

$$\mathcal{A}(D_N) \le C2^{-(N+1)}. \tag{8}$$

We shall prove the following theorem.

*Theorem 2.3:* If $\{D_N\colon \{0, 1\}^N \to [0, 1];\ N = 1, 2, \ldots\}$ is a democratic family, then

$$\lim_{N\to\infty}\left(\inf_{v,\,w\in[0,1],\,v<w-1/8} 2^{-N}\#[D_N^{-1}([v, w])\right) = 0. \tag{9}$$

*Remark:* The value $1/8$ in the statement is completely arbitrary; it can be replaced by any $\delta$ in $(0, 1/2)$.

Our main result then follows as a corollary to this theorem.

*Theorem 2.4:* If $\{D_N\colon \{0, 1\}^N \to [0, 1];\ N = 1, 2, \ldots\}$ is an optimally accurate family of decoders, then it cannot be democratic.

*Proof:* Let us denote by $u_j^N$, $j = 1, 2, 3, \ldots, 2^N$ the linearly ordered set of elements of $[0, 1]$ that lie in the range of $D_N$. Because of (8), every point of $[0, 1]$ must be within a distance $C2^{-(N+1)}$ of a $u_j^N$, which implies

$$u_1^N \le C2^{-(N+1)}, \qquad u_{2^N}^N \ge 1 - C2^{-(N+1)}$$
$$|u_j^N - u_{j+1}^N| \le C2^{-N}.$$

It follows that, for $v, w \in [0, 1]$ and $v < w - C2^{-N}$, we must have

$$\#[D_N^{-1}([v, w])] = \#\{j;\ v \le u_j^N \le w\} \ge C^{-1}2^N|w - v| - 1.$$

If $v < w - 1/8$, then this implies

$$\begin{aligned}&\#\{D_N\colon \{0, 1\}^N \to [0, 1];\ N = 1, 2, \ldots\}\\ &\quad\ge 1/8|v - w|^{-1}\#\{D_N\colon \{0, 1\}^N \to [0, 1];\ N = 1, 2, \ldots\}\\ &\quad\ge C^{-1} - 2^{-N+3},\end{aligned}$$

contradicting (9). □

It remains to prove Theorem 2.3. This will be done using a theorem by Frankl and Füredi [4]. To give the statement of their theorem, we first need to introduce the notion of a Hamming sphere.

*Definition 2.5:* A Hamming $N$-sphere with center $c \in \{0, 1\}^N$ is a subset $S$ of $\{0, 1\}^N$ such that, for some $k \in \mathbb{N}$

$$\{a \in \{0, 1\}^N;\ d_H(a, c) \le k\} \subset S \subset \{a \in \{0, 1\}^N;\ d_H(a, c) \le k + 1\}.$$

In [4], Frankl and Füredi proved the following.

*Theorem 2.6:* For any two subsets $A, B$ of $\{0, 1\}^N$, there exist two Hamming $N$-spheres $S_0, S_1$ such that

- $S_0$ has center $(0, 0, \ldots, 0)$, $S_1$ has center $(1, 1, \ldots, 1)$
- $\#S_0 = \#A$, $\#S_1 = \#B$, and
- $d_H(S_0, S_1) \ge d_H(A, B) := \min\{d_H(a, b);\ a \in A, b \in B\}$.

We now show how this implies our Theorem 2.3.

*Proof of Theorem 2.3:* Suppose that

$$\{D_N\colon \{0, 1\}^N \to [0, 1];\ N = 1, 2, \ldots\}$$

is democratic, i.e., we have

$$|D_N(a) - D_N(b)| \le C_1 d_H(a, b)/N \tag{10}$$

and suppose that there exist a constant $C_2 > 0$, and a strictly increasing sequence $(N_l)_{l\in\mathbb{N}}$ such that, for all $v, w \in [0, 1]$ with $v + 1/8 < w$, and all $l \in \mathbb{N}$

$$\#[D_{N_l}^{-1}([v, w]) \ge C_2 2^{N_l}|w - v|. \tag{11}$$

In the remainder of this proof, we shall consider only values of $N$ in this sequence $(N_l)_{l\in\mathbb{N}}$. Define subsets $A^N$ and $B^N$ of $\{0, 1\}^N$ by

$$A^N = D_N^{-1}([0, 1/4]), \qquad B^N = D_N^{-1}([3/4, 1]).$$

Then we immediately have, as a consequence of (11)

$$\#A^N \ge C_2 2^{N-2}, \qquad \#B^N \ge C_2 2^{N-2}. \tag{12}$$

On the other hand, the distances between all the images, under the decoding map $D_N$, of points in $A^N$ and $B^N$, respectively, are at least $1/2$ apart, which by (10) implies that

$$d_H(A^N, B^N) \ge {C_1}^{-1}N/2. \tag{13}$$

Let us apply Theorem 2.6 to the sets $A^N$ and $B^N$; we have thus two Hamming $N$-spheres $S_0^N$ and $S_1^N$, centered around $(0, \ldots, 0)$ and $(1, \ldots, 1)$, respectively, with cardinality bounded below by the right-hand sides of (12), and Hamming distance bounded below by the right-hand side of (13). By the definition of Hamming $N$-spheres, this means that there exist $k_0^N$ and $k_1^N$ such that

$$N - k_0^N - k_1^N \ge {C_1}^{-1}N/2 \tag{14}$$

and which satisfy moreover

$$C_2 2^{N-2} \leq \sum_{m=0}^{\min(k_0^N,\, k_1^N)+1} \binom{N}{m}. \tag{15}$$

This will lead to a contradiction. Note that (14) implies that $1/(2C_1)$ must be strictly smaller than 1; with $\epsilon := 1/(2C_1) \in (0,\, 1)$, we have, therefore, for all $N$

$$\min(k_0^N,\, k_1^N) \leq [k_0^N + k_1^N]/2 \leq (1-\epsilon)N/2. \tag{16}$$

Define $k_N := \min(k_0^N,\, k_1^N)$; by using the standard Stirling upper and lower bounds for $M!$, we obtain

$$\begin{aligned}\sum_{m=0}^{k_N+1} \binom{N}{m} &\leq [k_N+1] \binom{N}{k_N+1} \\ &\leq CN \binom{N}{(1-\epsilon)N/2} \\ &\leq C'\sqrt{N}\, 2^N \left[(1-\epsilon)^{(1-\epsilon)}(1+\epsilon)^{(1+\epsilon)}\right]^{-N/2} \\ &\leq C'\sqrt{N}\, 2^N \rho^{-N} \end{aligned} \tag{17}$$

where $\rho := [(1-\epsilon)^{(1-\epsilon)}(1+\epsilon)^{(1+\epsilon)}]^{1/2} > 1$. Clearly, (15) and (17) cannot both be valid for arbitrarily large $N$. This contradiction concludes our proof. □

## III. Comments

Convolutional decoding is very convenient in D/A conversion; the possibility of using such a simple decoding is one of the reasons of the success of $\Sigma\Delta$ schemes. (There are others; see below.) This is the "pro" of (this method of) democratic decoding alluded to in our title; the "con" is, of course, the reduction in possible accuracy that democracy entails, as shown in Section II.

Theorem 2.4 showed that democracy makes it impossible to be *optimally* accurate. What about *sub*optimality? The argument given in the proof of Theorem 2.3 would no longer work if we deviate from optimal accuracy by relaxing the exponential decay rate, i.e., if we accept accuracy proportional to $2^{-\gamma N}$ (instead of $2^{-N}$), with $0 < \gamma < 1$. In fact, as pointed out by the following example by Güntürk [9], one can construct decoders that satisfy this suboptimality and that are democratic according to our definition. For $N = 2M$, define a decoder $D_{2M}$ by

$$D_{2M}(b_1,\, \ldots,\, b_{2M}) = N^{-1}\sum_{k=1}^{M} b_k + N^{-1}\sum_{k=1}^{M} 2^{-k} b_{M+k}.$$

One easily checks that

$$\mathcal{A}(D_{2M}) = 2^{-M} = 2^{-N/2}$$

and that

$$\max_{k=1,\, \ldots,\, 2M,\, a\in\{0,\,1\}^{2M}} e^k_{[D_{2M}]}(a) = 1/M = 2/N.$$

This example is, therefore, "democratic" in the sense of Definition 2.1 and has exponential accuracy, corresponding to $\gamma = 1/2$; it can be adapted to achieve other $\gamma \in (0,\, 1)$. Nevertheless, the two groups of bits in this example clearly have a different share of "influence" in the outcome; this suggests that Definition 2.1 is too weak, and that more detailed descriptions are required to exclude it and others of its type; Güntürk's work [10] addresses this issue and proves some interesting results on $\Sigma\Delta$ quantization in this context. It is not clear, at this point, what type of restrictions on the accuracy follow from such stronger notions of "democracy"; this is a subject for future work.

It is important to note that for constant input, the encoding schemes encountered in $\Sigma\Delta$ modulation, which were the motivation for this correspondence, have a particular "embedding" structure, not exploited in this correspondence, in that the encoders can be viewed as finite projections of a single infinite encoding stream. More specifically, one can define a map $E$ from $[0,\, 1]$ to $\{0,\, 1\}^{\mathbb{N}}$, from which other maps $E_{M,\,N}$ from $[0,\, 1]$ to $\{0,\, 1\}^N$ are derived by $[E_{M,\,N}(a)]_j = [E(a)]_{M+j}$, with $j$ ranging from 1 to $N$; the decoders $D_N$ mentioned in the Introduction are then independent of $M$ (a consequence, for the particular case of constant input, of the "convolutional" structure of the reconstruction formulas valid for more general band-limited functions, and of the finite length of the reconstruction filters $h$—see [7], [9]); one is interested in bounds on $\mathcal{A}(E_{M,\,N},\, D_N)$ that do not depend on $M$. It is not known whether suboptimal but exponential accuracy (with $\gamma < 1$) is possible if we impose this extra "embedded" structure. It therefore remains an open problem how much, if any, improvement can be made over (5) for 1-bit quantization of band-limited functions.

Another comment concerns an improvement on a slightly different type of $\lambda$-dependent reconstruction for $\Sigma\Delta$ schemes with constant input. One easy (albeit impractical) way to choose the filters $h$ in (1) is to pick a function $H$ such that its infinitely differentiable Fourier transform $\hat{H}$ is supported in $[-2\nu\pi,\, 2\nu\pi]$, such that $\hat{H}(\xi) = 1$ for $|\xi| \leq \nu\pi$, and to set $h_k^{(\lambda)} := \lambda^{-1}H(k/\lambda\nu)$. These filters are no longer finite, but because of the smoothness of $\hat{H}$ they decay faster than any inverse polynomial. If convolution with such *infinite* filters is allowed, then a construction by Konyagin [16] improves on $\Sigma\Delta$ schemes by choosing a sequence $q_n^{(\lambda)}$ in such a way that, uniformly for $a \in (-1,\, 1)$ and $k \in \mathbb{Z}$, and keeping $H$ fixed, one has the following stronger decay estimate:

$$\left| a - \sum_l h_{k-l}^{(\lambda)} q_l^{(\lambda)} \right| \leq C e^{-\beta\sqrt{\lambda}}. \tag{18}$$

Because the filters are no longer finite, this does not quite fit in the framework described in this correspondence; however, the $h^{(\lambda)}$ in Konyagin's construction do "spread" linearly in $\lambda$, like the finite filters considered earlier. Note also that Konyagin's estimate holds for all $k$, which means that his $q_l^{(\lambda)}$ are democratic. The bound (18) is much better than (5), and it holds out hope that much better than (5) can be achieved even for finite filters, although it is not clear at this point how to use Konyagin's construction with finite filters. In addition, it is not clear either whether his method can be extended to band-limited rather than constant input.

Finally, it should be noted that $\Sigma\Delta$-encoders have many other virtues than the possibility of decoding their output democratically. One of their amazing advantages is their robustness: imperfections in the quantizer do not affect the rate of convergence in $\lambda$ given in (2), whereas the same imperfections would lead to a strictly positive lower bound on the error in a binary encoding procedure, regardless of $\lambda$ (see [2], [3]). It turns out that one can devise encoding schemes that have robustness properties similar to $\Sigma\Delta$ schemes, but that nevertheless exhibit (suboptimal) exponential accuracy for the representation of bounded band-limited functions [3]; the schemes constructed in [3] cannot, however, be decoded democratically. It would be extremely interesting to know whether the convenience of democratic (convolutional) decoding is compatible with exponential accuracy, albeit suboptimal, for bounded band-limited functions. This is another question for future work.

### *Note Added in Proof*

In February 2002, C. Sinan Güntürk constructed sigma–delta encoders and decoders that give exponential precision for all bounded

band-limited functions. The decoders are still convolutional, and thus democratic; in his construction, the rate of the exponential decay of the error, shown here to be necessarily strictly inferior to 1, is smaller than 1 by several orders of magnitude.

## ACKNOWLEDGMENT

The authors would like to thank Ron DeVore and C. Sinan Güntürk for helpful discussions concerning the topic of this paper. They are also grateful to the referees for helpful comments. I. Daubechies would also like to thank the Institute for Advanced Study in Princeton for its hospitality during the writing of this correspondence.

## REFERENCES

[1] J. C. Candy and G. C. Temes, Eds., *Oversampling Delta–Sigma Data Converters Theory, Design and Simulation*. New York: IEEE Press, 1992.

[2] I. Daubechies and R. DeVore, "Reconstructing a bandlimited function from very coarsely quantized data: A family of stable sigma–delta modulators of arbitrary order," *Ann. Math.*, to be published.

[3] I. Daubechies, R. DeVore, C. S. Güntürk, and V. Vaishampayan, "Exponential precision in A/D conversion with an imperfect quantizer," paper, submitted for publication.

[4] P. Frankl and Z. Füredi, "A short proof for a theorem of Harper about Hamming-spheres," *Discr. Math.*, vol. 34, pp. 311–313, 1981.

[5] R. M. Gray, "Spectral analysis of quantization noise in single-loop sigma–delta modulator with dc input," *IEEE Trans. Commun.*, vol. COM-35, pp. 481–489, 1987.

[6] R. M. Gray, W. Chou, and P. W. Wong, "Quantization noise in single-loop sigma–delta modulation with sinusoidal inputs," *IEEE Trans. Commun.*, vol. 37, pp. 956–968, Sept. 1989.

[7] S. Güntürk, "Improved error estimates for first order sigma–delta modulation," in *Sampling Theory and Applications, SampTA'99*, Leon, Norway, Aug. 1999.

[8] ——, "Reconstructing a bandlimited function from very coarsely quantized data: Improving the error estimate for first order sigma–delta modulators," in preparation.

[9] ——, "Harmonic analysis of two problems in signal compression," Ph.D. dissertation, Program in Applied and Computational Mathematics, Princeton Univ., Princeton, NJ, Sept. 2000.

[10] ——, "On democratic encoding and decoding," in preparation.

[11] C. S. Güntürk, J. C. Lagarias, and V. Vaishampayan, "Robustness of single loop sigma–delta modulation for constant inputs," *IEEE Trans. Inform. Theory*, submitted for publication.

[12] S. Hein, K. Ibrahim, and A. Zakhor, "New properties of sigma–delta modulators with dc inputs," *IEEE Trans. Commun.*, vol. 40, pp. 1375–1387, Aug. 1992.

[13] S. Hein and A. Zakhor, "Optimal decoding for data acquisition applications of sigma delta modulators," *IEEE Trans. Signal Processing*, vol. 41, pp. 602–616, Feb. 1993.

[14] ——, *Sigma Delta Modulators: Nonlinear Decoding Algorithms and Stability Analysis*. Dordrecht, The Netherlands: Kluwer, 1993.

[15] ——, "Reconstruction of oversampled band-limited signals from $\Sigma\Delta$ encoded binary sequences," *IEEE Trans. Signal Processing*, vol. 42, pp. 799–811, Apr. 1994.

[16] S. Konyagin, private communication, 1998.

[17] S. R. Norsworthy, R. Schreier, and G. C. Temes, Eds., *Delta–Sigma Data Converters Theory, Design and Simulation*. New York: IEEE Press, 1997.

[18] A. Zakhor, "Lower bounds on the MSE of the single loop sigma delta modulator," in *Proc. 23rd Asilomar Conf. Signals, Systems and Computers*, Oct. 1989, pp. 849–853.

# Error Exponents of Expander Codes

Alexander Barg, *Senior Member, IEEE*, and Gilles Zémor

***Abstract*—We show that expander codes attain the capacity of the binary-symmetric channel under iterative decoding. The error probability has a positive exponent for all rates between zero and channel capacity. The decoding complexity grows linearly with code length.**

***Index Terms*—Expander code, iterative decoding, Ramanujan graph.**

## I. INTRODUCTION

Constructing families of codes of growing length $N$ with low error probability and large minimum distance is one of the main problems of coding theory. The first result that gave us codes together with a polynomial decoding algorithm that achieves an exponentially small error probability goes back to Forney [3], and relies on a concatenated construction. Specifically, it states that for any $\varepsilon > 0$ there exists an infinite family of easily constructible codes of rate $R$ that are decodable with complexity $O(N^2)$ and error probability $P_e \leq 2^{-Nf_1(R,\, p)}$, where $p$ is the bit transition probability of the binary symmetric channel and where

$$f_1(R,\, p) = \max_{R \leq R_0 < 1 - H(p)} E(R_0,\, p)(1 - R/R_0) - \varepsilon. \qquad (1)$$

Here $E(R_0,\, p)$ is the "random coding exponent" [4] and $H(\cdot)$ is the binary entropy function. Thus, $f_1(R,\, p) > 0$ for all rates $R$ up to the channel capacity. A similar idea was used by Zyablov in [10] to construct codes with polynomial decoding complexity and relative distance arbitrarily close to

$$\delta(R) = (1 - R/R_0)H^{-1}(1 - R_0) \qquad (R \leq R_0 \leq 1). \qquad (2)$$

Thus, for these codes we have $\delta(R) > 0$ for any value $R$ of the code rate $0 \leq R < 1$. These results underwent a number of improvements (surveyed, for instance, in [2]), but until recently no codes were known with lower decoding complexity and nonvanishing relative distance and/or error exponent for nontrivial values of the code rate.

The first result of this kind [7] gives us a family of codes, based on a graph-theoretic approach of [8] and constructions of Ramanujan graphs [5], [6]. The result is as follows.

*Theorem 1 [7]:* For any $\varepsilon > 0$, there exists a polynomial-time constructible family of codes with distance $\delta - \varepsilon$ and rate $1 - 2H(\sqrt{\delta})$ for which any $\alpha < \delta/48$ fraction of errors can be corrected by a circuit of size $O(N \log N)$ and depth $O(\log N)$. The complexity of a sequential implementation of this decoding is $O(N)$.

This result was a remarkable novelty because not only did it decrease the complexity of decoding but it was also the first time that concatenation was not used to construct families of binary codes with a nontrivial rate and a nonzero relative minimal distance. It implies the existence of linear-time decodable codes with $R > 0$ for $0 \leq \delta < 0.0121$.

Sipser–Spielman's decoding algorithm was modified in [9] where the factor 48 was improved to 4. In the present correspondence, we study the error probability for iterative decoding of expander codes and

Manuscript received September 29, 2000; revised February 12, 2001.

A. Barg is with Bell Labs, Lucent Technologies, Murray Hill, NJ 07974 USA (e-mail: abarg@research.bell-labs.com).

G. Zémor is with École Nationale Supérieure des Télécommunications, 75 634 Paris 13, France (e-mail: zemor@infres.enst.fr).

Communicated by S. Shamai, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04034-8.