LOW-COMPLEXITY MULTI-RATE LATTICE VECTOR QUANTIZATION WITH APPLICATION TO WIDEBAND TCX SPEECH CODING AT 32 KBIT/S

Stéphane Ragot, Bruno Bessette and Roch Lefebvre

University of Sherbrooke, Dept. of Electrical Eng., Sherbrooke, Québec, J1K 2R1 Canada
E-mail: {Stephane.Ragot,Bruno.Bessette,Roch.Lefebvre}@Usherbrooke.ca

ABSTRACT

We present a new method, called *Voronoi extension*, for the design of low-complexity multi-rate lattice vector quantization (VQ). With this technique, lattice codebooks of arbitrarily large bit rates can be generated algorithmically and the problem of lattice codebook overload can be bypassed. We describe a practical multi-rate quantization system based on Voronoi extension and derived from the lattice RE_8. This system is applied to the TCX coding model using pitch prediction, so as to extend AMR-WB speech coding at high bit rates (in particular 32 kbit/s).

1. INTRODUCTION

The AMR-WB (Adaptive Multi-Rate-WideBand) speech coding standard [1] has been designed for speech applications in 3G networks. The underlying algorithm comprises 9 codecs with bit rates ranging from 6.6 to 23.85 kbit/s, and relies essentially on the ACELP model [1]. We investigate in this paper how to extend the performance of AMR-WB at high bit rates (e.g. 32 kbit/s), so as to reach transparent speech quality. This application scenario was considered in the early development of AMR-WB for high-rate 3G channels. The main problem for extending AMR-WB in bit rate is the complexity and the suboptimality of ACELP fixed codebook search with increasing bit allocation. We propose here to employ at high bit rates the algebraic TCX coding model of [2], in lieu of the ACELP model.

The main contribution of this paper is to extend the coding framework of [2] so that the algebraic TCX coding model becomes less complex and scalable in bit rate. In [2], a block of transform coefficients, which is the spectrum of the so-called TCX target signal, is split into equal divisions and the same set of lattice codebooks is used to represent each sub-vector. This approach simplifies the problem of lattice codebook design, assuming the sub-vectors are uncorrelated source vectors with similar statistics. We extend the work of [2] so as to implement arbitrarily large lattice codebooks, while reducing the storage requirement of indexing tables and avoiding the complex operation of lattice codebook overload.

This paper is organized as follows. The multirate lattice VQ of [2] is reviewed and discussed in Section 2. In Section 3, we introduce a general method which enables to extend lattice codebooks in bit rate and support region. This method is called Voronoi extension, for the underlying mechanism relies on the Voronoi coding of [3]. An efficient multi-rate lattice VQ system using Voronoi extension is described in Section 4, before concluding in Section 5.

This work was supported by the NSERC and VoiceAge Corp.

2. BACKGROUND : EMBEDDED ALGEBRAIC VECTOR QUANTIZATION (EAVQ)

Embedded algebraic vector quantization (EAVQ) has been introduced in [2]. In this technique, an 8-dimensional vector $\mathbf{x}$ is encoded into one of 6 codebooks, denoted by Q_n where $n = 0, 1, \cdots, 5$. The rate of Q_n is $4n$ bits per vector, $n/2$ bits per dimension. The (integer) number n of Q_n is called here a *codebook number*. The multi-rate codebooks Q_n are mainly based on the integer lattice $RE_8 = 2D_8^+$ [2]:

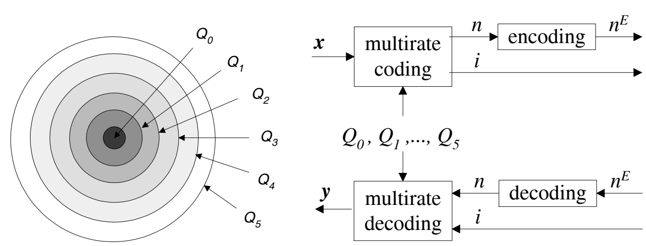
$$RE_8 = 2D_8 \cup \{2D_8 + (1, \cdots, 1)\}, \tag{1}$$

where

$$D_8 = \left\{ (x_1, \cdots, x_8) \in \mathbb{Z}^8 \ : \ \sum_{i=1}^{8} x_i \text{ is even} \right\}. \tag{2}$$

Furthemore, the codebooks $Q_1, Q_2, \cdots, Q_5$ are embedded (i.e. $Q_1 \subset Q_2 \cdots \subset Q_5$), as illustrated in Figure 1 (a). This constraint simplifies the encoding of codebook numbers [2].

The encoding in Q_n is shown in Figure 1 (b). The 8-dimensional input vector $\mathbf{x}$ is encoded into an index i of $4n$ bits, as well as an integer codebook number n represented in binary format by n^E. Assuming the encoding of n into n^E is lossless, and that n^E and i are received as such, the codevector $\mathbf{y}$ in Q_n of index i is reconstructed at the decoder. The coder minimizes the squared error $\|\mathbf{x}-\mathbf{y}\|^2$ over Q_n, $n = 0, 1, \cdots, 5$. Note that the bit allocation is not controlled explicitly: the choice of n depends on $\mathbf{x}$. Therefore, the RE_8 coder may be viewed as *self-scalable*.



(a) codebook structure (b) coder/decoder

Figure 1: Embedded Algebraic VQ.

This quantization technique proved efficient for TCX coding of wideband speech signals at 16 kbit/s [2]. However it has several limitations for general transform audio coding:

- Source mismatch: The codebooks Q_n are designed in [2] by spherical shaping of RE_8, which assumes a zero-mean i.i.d. Gaussian distribution for $\mathbf{x}$. The real source distribution of transform subvectors often deviates from this assumption, and tends to be more Laplacian than Gaussian.
- Handling of outliers: The technique of [2] does not allow to allocate more than 20 bits per vector in dimension 8. In transform coding, subvectors with the highest energies (eager to cause an overload) shall be quantized with a small distortion to maximize quality. The constraint on maximal bit allocation shall therefore be relaxed.
- Complexity : A saturation procedure is required to handle the overload in the largest codebook Q_5 [2]. This operation is computationnaly intensive.

  With the technique of [2], the storage requirement and search complexity are related to the number of so-called *absolute leaders* specifying the codebooks Q_n. This complexity explodes as the bit rate increases. Furthermore, the use of Q_5 implies to store some indexing tables in 32-bit format (instead of 16-bit).

We present in this paper a technique which provides a solution to all these problems.

3. VORONOI EXTENSION

We introduce in this section a general method, called Voronoi extension, for extending fixed-rate lattice codebooks in bit rate and support region. For the sake of generality, we consider an arbitrary lattice Λ in dimension N. The method is illustrated in 2-D using the hexagonal lattice $\Lambda = A_2$. The Voronoi extension is used in the next section to extend the technique of [2] for $\Lambda = RE_8$.

3.1. Preliminaries: partitioning of Λ by $m\Lambda$

The Voronoi extension is based on the self-similarity property of lattices (in the sense of fractals) and the theory of *coset codes* of [4]. For a (full-rank) lattice Λ in dimension N, the underlying fondation can be summarized by the formula:

$$\Lambda = m\Lambda + V_m = \bigcup_{\mathbf{c}\in\Lambda, \mathbf{v}\in V_m} m\mathbf{c} + \mathbf{v}, \tag{3}$$

where

$$V_m = \Lambda \cap (mV(\Lambda) + \mathbf{a}) \tag{4}$$

is a *Voronoi code* [3] derived from Λ of $\log_2 m$ bits per dimension, with m integer ≥ 2. The region $V(\Lambda)$ in Eq. 4 is the Voronoi region of Λ related to the origin. The offset $\mathbf{a}$ is usually selected to fix ties, i.e. ensure that no point of Λ falls on the truncation boundary of $mV(\Lambda) + \mathbf{a}$.

Eq. 3 shows that Λ can be generated by m^N translations (or cosets) of its sublattice $m\Lambda$. Voronoi coding is one way to represent the translation vectors [5], which are called *coset leaders*.

3.2. Definition of Voronoi extension

We assume a codebook C of R bits per dimension, comprising 2^{NR} codevectors, is constructed from a lattice Λ in dimension N ($C \subset \Lambda$). We define the *Voronoi extension* $C^{(r)}$ of C of *order* r, where r is an integer ≥ 1, as:

$$C^{(r)} = 2^r C + V_{2^r} = \bigcup_{\mathbf{c}\in C, \mathbf{v}\in V_{2^r}} 2^r\mathbf{c} + \mathbf{v}. \tag{5}$$

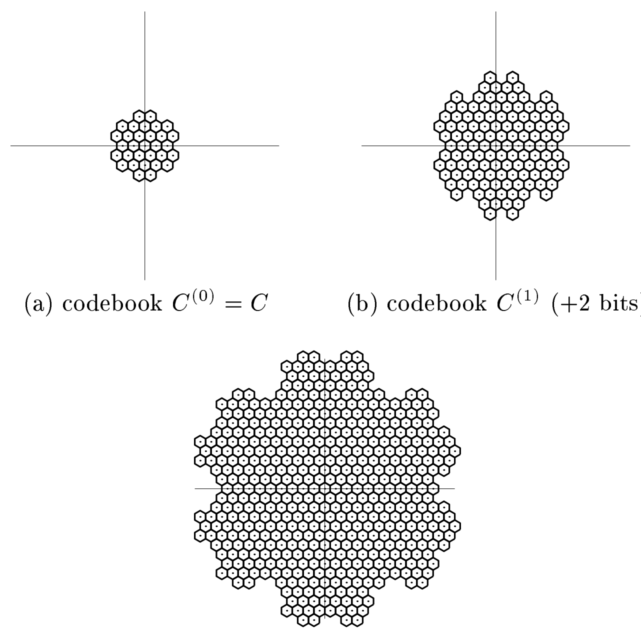
The offset $\mathbf{a}$ defining the Voronoi codes V_{2^r} is the same for all r. By convention, we also define $C^{(0)} = C$. The codebooks $C^{(r)}$ are subsets of Λ. They require NR bits per vector for C as well as Nr bits per vector for V_{2^r}. Consequently, the bit rate per dimension of $C^{(r)}$ is $R + r$.

3.3. 2-D Example

The Voronoi extension is illustrated in Figure 2 for the hexagonal lattice A_2 used in [3]. The codebook C is obtained by spherical truncation (or shaping) of $\Lambda = A_2$:

$$C = \left\{\mathbf{y} \in A_2 \ : \ \|\mathbf{y}\|^2 \leq 7\right\}. \tag{6}$$

This codebook, comprising 31 codevectors, is indexed with 5 bits per vector. The granular region of C is shown in Figure 2 (a). The extended codebooks $C^{(1)}$ and $C^{(2)}$, comprising respectively 124 and 496 codevectors, are shown in Figure 2 (b) and (c) for $\mathbf{a} = (0.3, 0)$. They are indexed with respectively 2 and 4 more bits per vector than C.



(a) codebook $C^{(0)} = C$ (b) codebook $C^{(1)}$ (+2 bits)

(c) codebook $C^{(2)}$ (+4 bits)

Figure 2: Example of Voronoi extension for the lattice A_2.

From Figure 2, one can understand intuitively several important properties of the Voronoi extension:

- The Voronoi cells (in the granular region) of all $C^{(r)}$ have the same size. Provided there is no codebook overload, the mean square error is identical for all $C^{(r)}$.

- The "shapes" of $C^{(r)}$ are all similar. If the shape of C is optimized for a given source distribution, the codebooks $C^{(r)}$ inherit from this optimization.
- Under certain conditions, $C^{(+\infty)} = \Lambda$ and the Voronoi extension allows to index the infinite lattice Λ – for instance, this property is valid for the codebook C used in the 2-D example used here.

4. MULTIRATE RE_8 QUANTIZATION BASED ON VORONOI EXTENSION

We present here a technique of multirate lattice VQ using Voronoi extension, which extends the system of [2]. This technique is illustrated in Figure 3. An 8-dimensional input vector $\mathbf{x}$ is rounded to its nearest neighbor $\mathbf{y}$ in the *infinite* lattice RE_8. The multirate indexing produces a codebook number n (represented by a binary code n^E) and an index i of $4n$ bits.

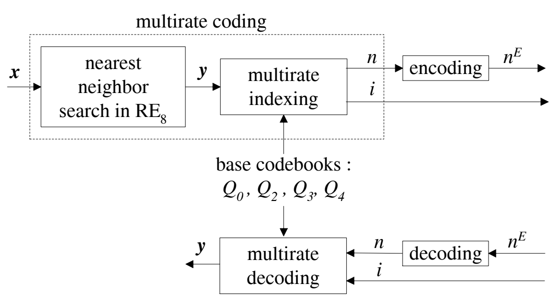


Figure 3: Multirate RE_8 vector quantization.

As in [2], the system employs multirate codebooks, denoted by Q_n, of $4n$ bits per vector, $n/2$ bits per dimension. The codebooks Q_n are subsets of the lattice RE_8. The codebook number n is defined in the set $\{0, 2, 3, 4, \cdots, n_{max}\}$, with $n_{max} = 36$. The case $n = 1$ is forbidden because it corresponds to an allocation of 4 bits in dimension 8: at such a low rate, lattice quantization is not efficient compared to, for instance, noise injection.

The codebooks Q_n are divided in two categories :

- Low-rate (base) codebooks ($n = 0, 2, 3, 4$): These are classical leader-based codebooks described physically by a set of indexing tables. They are specified in terms of absolute leaders in Table 1. These leaders were selected based on the frequencies of occurence of RE_8 absolute leaders, obtained with several minutes of speech and music signals and a specific transform coding model. Note that $Q_2 \subset Q_3$ for Voronoi extension to work properly.

  For each absolute leader $\mathbf{y} = (y_1, \cdots, y_8)$, the parameter P of Table 1 is defined as:

$$P = |y_1| + \cdots + |y_8|, \qquad (7)$$

  and specifies RE_8 pyramidal shells. Thus, the low-rate codebooks comprise successive pyramidal shells (and not spherical shells, unlike [2]), which implies that these codebooks are mainly optimized for a memoryless Laplacian distribution.
- High-rate codebooks ($n > 4$): They are generated algorithmically by employing Voronoi extension. The Voronoi extension is applied alternatively to Q_3 and Q_4 – instead of a single codebook C – to reach a rate increment of 1/2 bit per dimension. The extended codebook Q_n with $n = n' + 2r > 4$ is generated by a rth-order extension of $Q_{n'}$ where $n' = 3$ or 4. For instance, $Q_5 = Q_3^{(1)}$, $Q_6 = Q_4^{(1)}$, $Q_7 = Q_3^{(2)}$, and so on. Note that the offset $\mathbf{a}$ used in Voronoi coding is set to $\mathbf{a} = (2, 0, 0, 0, 0, 0, 0, 0)$.

Table 1: Specification of Q_0, Q_2, Q_3 and Q_4 (with size of permutation classes).

absolute leader	P	size	Q_0	Q_2	Q_3	Q_4
0, 0, 0, 0, 0, 0, 0, 0	0	1	•			
2, 2, 0, 0, 0, 0, 0, 0	4	112		•	•	
4, 0, 0, 0, 0, 0, 0, 0	4	16		•	•	
1, 1, 1, 1, 1, 1, 1, 1	8	128		•	•	
2, 2, 2, 2, 0, 0, 0, 0	8	1120			•	
4, 2, 2, 0, 0, 0, 0, 0	8	1344			•	
4, 4, 0, 0, 0, 0, 0, 0	8	112			•	
6, 2, 0, 0, 0, 0, 0, 0	8	224			•	
8, 0, 0, 0, 0, 0, 0, 0	8	16			•	
3, 1, 1, 1, 1, 1, 1, 1	10	1024			•	
4, 2, 2, 2, 2, 2, 2, 0	10	7168				•
2, 2, 2, 2, 2, 2, 0, 0	12	1792				•
3, 3, 1, 1, 1, 1, 1, 1	12	3584				•
4, 2, 2, 2, 2, 0, 0, 0	12	8960				•
4, 4, 2, 2, 0, 0, 0, 0	12	6720				•
4, 4, 4, 0, 0, 0, 0, 0	12	448				•
5, 1, 1, 1, 1, 1, 1, 1	12	1024				•
6, 2, 2, 2, 0, 0, 0, 0	12	4480				•
6, 4, 2, 0, 0, 0, 0, 0	12	2688				•
6, 6, 0, 0, 0, 0, 0, 0	12	112				•
8, 2, 2, 0, 0, 0, 0, 0	12	1344				•
8, 4, 0, 0, 0, 0, 0, 0	12	224				•
10, 2, 0, 0, 0, 0, 0, 0	12	224				•
12, 0, 0, 0, 0, 0, 0, 0	12	16				•
5, 3, 1, 1, 1, 1, 1, 1	14	7168				•
7, 1, 1, 1, 1, 1, 1, 1	14	1024				•
2, 2, 2, 2, 2, 2, 2, 2	16	256				•
3, 3, 3, 3, 1, 1, 1, 1	16	8960				•
8, 8, 0, 0, 0, 0, 0, 0	16	112				•
9, 1, 1, 1, 1, 1, 1, 1	16	1024				•
10, 6, 0, 0, 0, 0, 0, 0	16	224				•
12, 4, 0, 0, 0, 0, 0, 0	16	224				•
14, 2, 0, 0, 0, 0, 0, 0	16	224				•
16, 0, 0, 0, 0, 0, 0, 0	16	16				•
3, 3, 3, 1, 1, 1, 1, 1	17	7168				•
10, 10, 0, 0, 0, 0, 0, 0	20	112				•
12, 8, 0, 0, 0, 0, 0, 0	20	224				•

The multi-rate encoding algorithm is described below:

1) Find the nearest neighbor $\mathbf{y}$ of $\mathbf{x}$ in RE_8.
2) If $\mathbf{y} \in Q_n$ with $n \leq 4$, compute the index i of $\mathbf{y}$ in Q_n (if $n > 0$) and stop. Otherwise:
   (i) Initialization : set $r = 1$
   (ii) Find $\mathbf{v} \in V_{2^r}$ such that $\mathbf{y} \in 2^r RE_8 + \mathbf{v}$, where V_{2^r} is defined in Eq. 4 (here $\Lambda = RE_8$, $m = 2^r$).
   (iii) Compute $\mathbf{c} = \frac{1}{2^r}(\mathbf{y} - \mathbf{v})$ – by definition of $\mathbf{v}$, $\mathbf{c} \in RE_8$.
   (iv) If $\mathbf{c} \notin Q_{n'}$ with $n' = 3$ or 4, increment r by 1 and go to step (ii). Otherwise: set $n = n' + 2r$, compute the index j of $\mathbf{c} \in Q_{n'}$, compute the

Voronoi index $\mathbf{k}$ of $\mathbf{v} \in V_{2^r}$ using the algorithm of [3], and multiplex j and $\mathbf{k}$ to form the index i of $\mathbf{y} \in Q_n$.

Note that with the codebooks Q_3 and Q_4 of Table 1 the unbounded loop over r can be reduced to 2 iterations by pre-selecting properly r in step (i) and updating r by ± 1 in step (iv).

The corresponding multi-rate decoding algorithm is simple. If $n \leq 4$, the index i is decoded directly into $\mathbf{y} \in Q_n$ as in [2]. Otherwise, if $n > 4$, the Voronoi order r is given by the quotient of $(n-3)/2$. From this value, we compute $n' = n - 2r$, demultiplex the indices j and $\mathbf{k}$, decode j into $\mathbf{c} \in Q_{n'}$ and $\mathbf{k}$ into $\mathbf{v} \in V_{2^r}$. The reconstruction $\mathbf{y}$ is given by $\mathbf{y} = 2^r\mathbf{c} + \mathbf{v}$.

The integer codebook number n is encoded in binary format into n^E by *unary coding* using the mapping: $Q_0 \rightarrow 0$, $Q_2 \rightarrow 10$, $Q_3 \rightarrow 110$, $Q_4 \rightarrow 1110$, etc.

Important remarks:

1. The Voronoi extension is applied only when the nearest neigbhor $\mathbf{y}$ of $\mathbf{x}$ in RE_8 is not in the base codebooks Q_0, Q_2, Q_3 or Q_4.
2. The limitation $n_{max} = 36$ is found assuming the components of the Voronoi index $\mathbf{k}$ are represented with 16-bit unsigned integer – this value is found for $Q_4^{(16)}$ with $n' = 4$ and $r = 16$. In practice, this allows to represent any set of transform coefficients *without overload*, if the dynamic range does not exceed 90 dB.

4.1. Application to wideband TCX speech coding

The experiments are based here on the AMR-WB speech coding algorithm [1]. The AMR-WB modes at 15.85 and 23.05 kbit/s are taken as a reference. An additional mode at 32 kbit/s is defined by replacing the ACELP fixed codebook search in AMR-WB with a direct quantization of the target signal. The target is encoded in Discrete Fourier Transform (DFT) domain by split RE_8 quantization, following the approach of [2]. The sub-frame length of AMR-WB is 5 ms (64 samples at the 12.8 kHz sampling frequency). Therefore the transformed target comprises 8 subvectors of dimension 8. These sub-vectors are quantized using the multi-rate system described previously. As in [2], a single scale factor is applied prior to algebraic VQ, so as to control bit allocation and distortion. The bit allocation to coding parameters is detailed in Table 2 (a). The encoding of AMR-WB parameters is unchanged (ISF parameters, pitch delay and filter, VAD flag, gains). Note that the ACELP fixed codebook gain is replaced here by the TCX global gain. Comfort noise is injected in unquantized subbands (which are here 8-dimensional sub-vectors coded in Q_0). The noise injection level is not quantized, but set relatively to the TCX global gain.

The objective results can be found in Table 2 (b). The segmental signal-to-noise ratio (segSNR) is measured using 20-ms frames in the 0–6400 Hz band. A wideband speech signal comprising 4 sentences in English (2 males, 2 females) is used. Table 2 (b) shows that the performance of AMR-WB can be extended at 32 kbit/s, and the difference between the 23.05 and 32 kbit/s modes is similar to the difference between the 15.85 and 23.05 kbit/s reference.

Some statistics on the use of RE_8 codebooks are shown in Table 2 (c). The Voronoi extension appears to be used to encode 19 % of all sub-vectors (this corresponds to the case $n > 4$). This result validates the importance of this method for wideband TCX speech coding at 32 kbit/s. In practice a peak codebook number n of 11 was measured.

Table 2: Bit allocation and results.

(a) Bit allocation to TCX parameters at 32 kbit/s

Parameter	Bits/frame
ISF parameters	46
Pitch delay	9+6+9+6
Pitch filter	4×1
VAD flag	1
Pitch gain & TCX global gain	4×7
RE_8 VQ (8 sub-vectors)	4×132

(b) segSNR results (in dB) for 20-ms segments

AMR-WB		TCX
15.85 kbit/s	23.05 kbit/s	32 kbit/s
11.88	14.47	17.21

(c) statistics on codebook numbers (in %)

Q_0	Q_2	Q_3	Q_4	Q_n ($n > 4$)
3	29	26	23	19

5. CONCLUSION

We described in this paper a lattice codebook extension method, called Voronoi extension, and its application to the design of (self-scalable) multirate VQ. The method is algorithmic and requires virtually no storage. It was used to extend the technique of [2], so as to design a *low-complexity* multirate RE_8 quantization system. This new tool was applied to extend the performance of AMR-WB at 32 kbit/s.

ACKNOWLEDGMENT

The basic algorithm of Voronoi extension is due to Prof. J.-P. Adoul (retired since 1998).

REFERENCES

[1] B. Bessette et al., "The Adaptive Multirate Wideband Speech Codec (AMR-WB)," *IEEE Trans. on Speech and Audio Proc.*, vol. 10, no. 8, pp. 620–637, Nov. 2002.

[2] M. Xie and J.-P. Adoul, "Embedded algebraic vector quantization (EAVQ) with application to wideband audio coding," in *Proc. ICASSP*, 1996, pp. 240–243.

[3] J.H. Conway and N.J.A. Sloane, "A fast encoding method for lattice codes and quantizers," *IEEE Trans. Inform. Theory*, vol. 29, no. 6, pp. 820–824, Nov. 1983.

[4] G.D. Forney, "Coset codes. I. Introduction and geometrical classification," vol. 34, no. 5, pp. 1123–1151, Sep. 1988.

[5] G.D. Forney, "Multidimensional Constellations - Part II: Voronoi Constellations," *IEEE Trans. Select. Areas in Commun.*, vol. 7, no. 6, pp. 941–958, Aug. 1989.