Data Compression Conference (DCC 2015)
7 April 2015

A Partial Hstry of Losy Compression

Robert M Gray
Dept of Electrical Engineering, Stanford University
Dept of Electrical and Computer Engineering, Boston University
rmgray@stanford.edu

title decompresses as (see abstract)

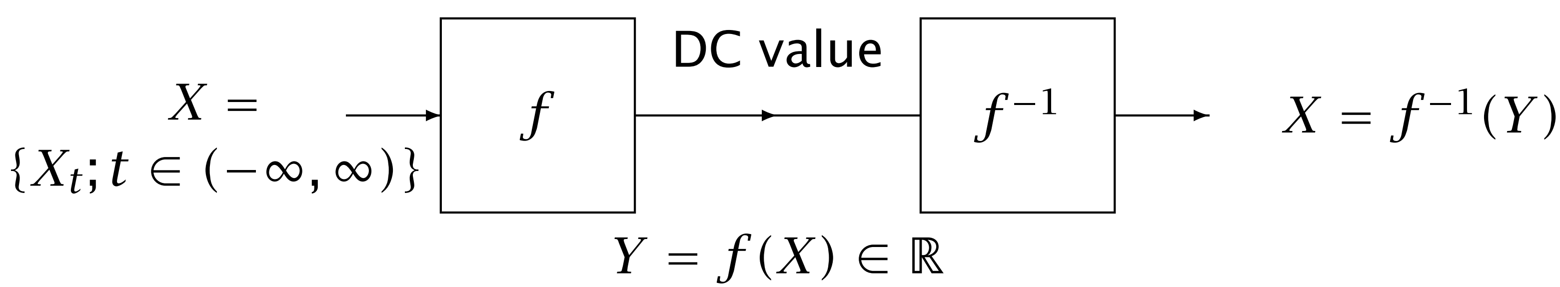
An incomplete biased history of lossy or lousy compression

What is compression?

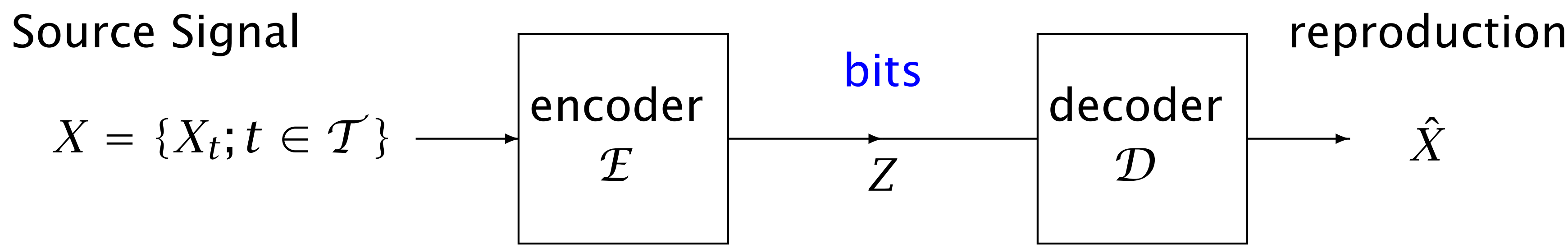
Consider *data compression* in the sense of *Shannon (1949,1959)* and not in the older sense of bandwidth compression.

To be *compressed* analog signals must be converted into bits. Dimension reduction alone is not compression—

mathematically there is an invertible correspondence (isomorphism) between all continuous time waveforms and the real numbers, so there is no compression of information going from one to the other. BW $= \infty \to$ BW $= 0$



Compression Basics in Hindsight



Key properties:

- **Distortion**: distortion measure, fidelity criteria

squared error dominates literature

average distortion $d(\mathcal{E}, \mathcal{D})$

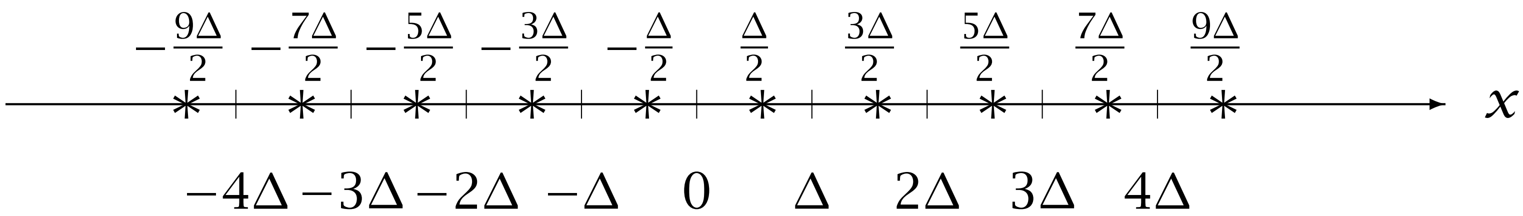
- **Rate**: actual bits or entropy rate $r(\mathcal{E})$

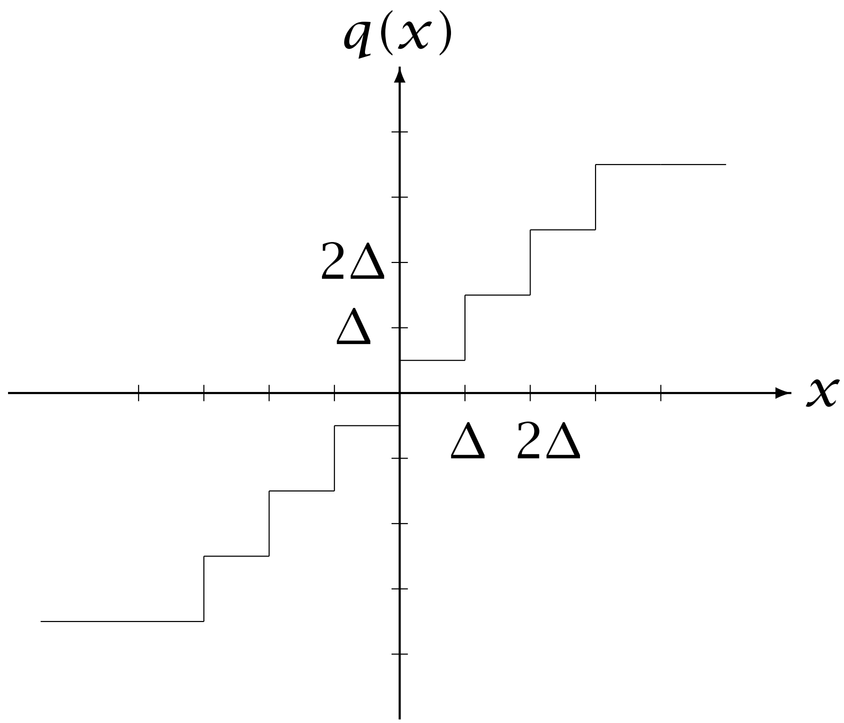
---

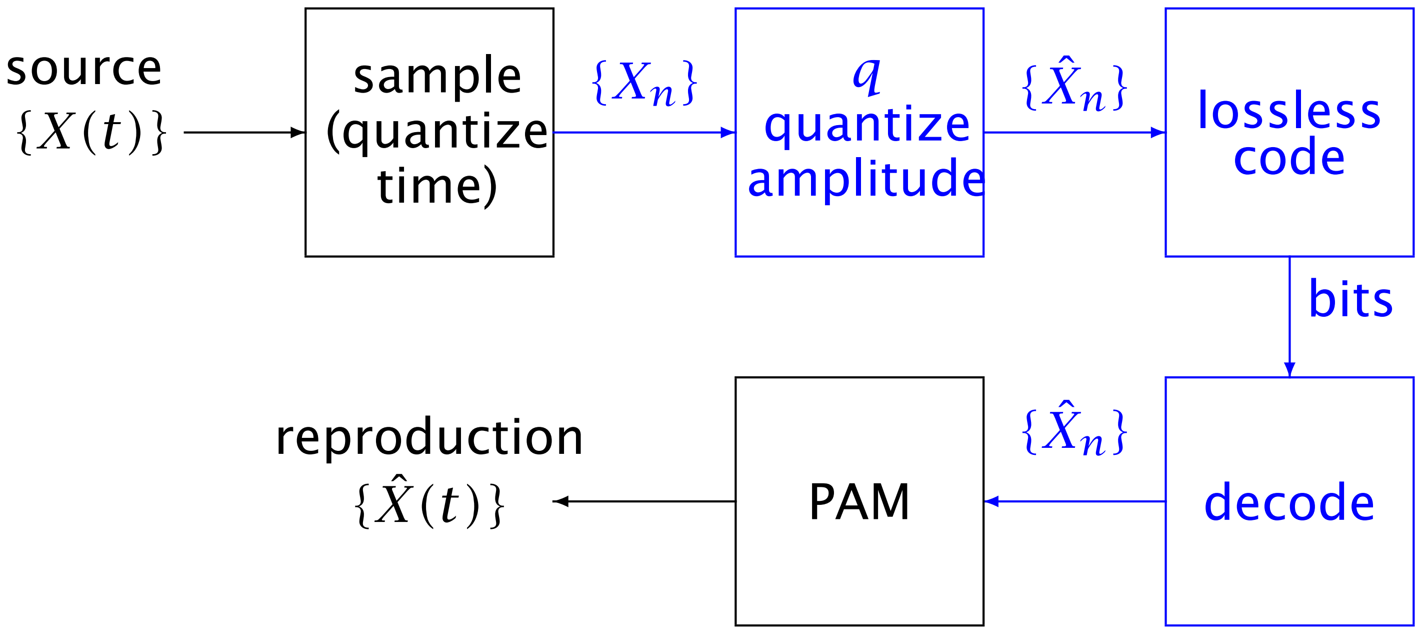
d vs r: beginnings, theories, highlights, and lowlights

Origins

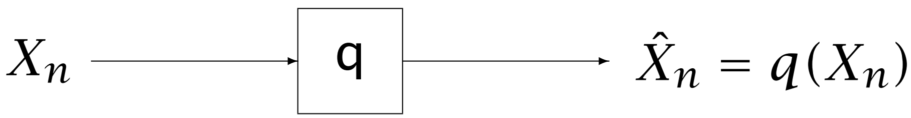
First genuine compression system for analog signals: PCM [Reeves (1938), Bennett (1948), Oliver, Pierce, Shannon (1948)] **uniform quantization** *A/D-conversion*







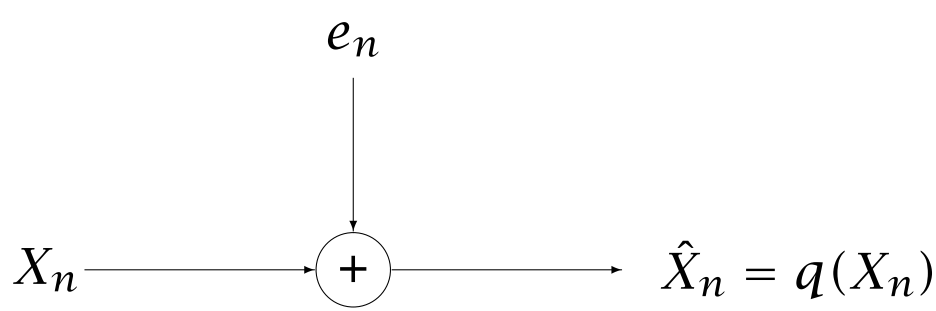
In digital age, input is often discrete time and do not need to sample. Compression *usually* treated as discrete time problem, quantization as *scalar memoryless nonlinear mapping*:



$$q(x) = \mathcal{D}(\mathcal{E}(x))$$

quantizer error $= e(x) \stackrel{\Delta}{=} q(x) - x \Rightarrow \boxed{q(x) = x + e(x)}$

suggests



☠ additive noise "model"

but not really a "model" unless *assume* specific behavior for e_n rather than derive from input statistics and quantizer definition — most common assumption e_n uniformly distributed, independent of the signal, and iid — often called the "white noise model" for uniform quantization

Theories of Compression

Exact Analysis

Some systems can be solved exactly via methods from nonlinear systems, especially “transform method” [Rice (1944)]

Clavier, Panter, and Grieg’s (1947) analysis of the spectra of the quantization error for uniformly quantized sinusoidal signals, Bennett’s (1948) series representation of the power spectral density of a uniformly quantized Gaussian random process.

Quantizer error for 10 bit uniform quantizer with sinusoidal input:

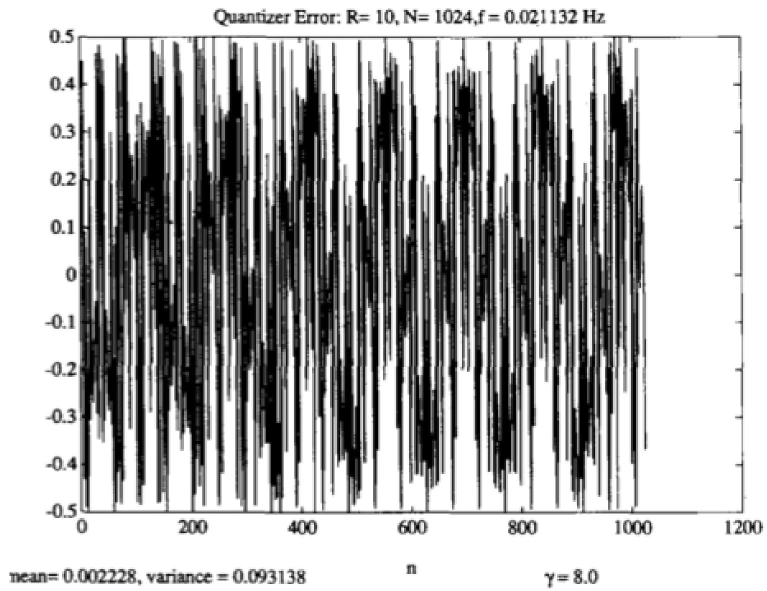


Fig. 5. PCM quantization error.

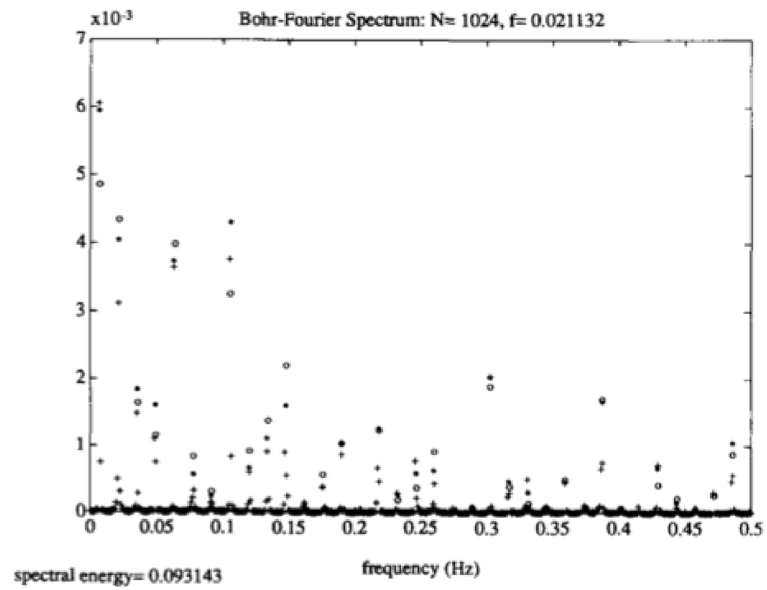


Fig. 6. PCM quantization error spectrum.

Widrow (1956,1960,1961) used ideas from sampling theory to show that if q has infinite levels and the characteristic function $\phi_{X_0}(u) = E(e^{juX_0})$ of the marginal input density has the "band-limited" property

$$\phi_{X_0}(u) = 0; |u| \geq 2\pi/\Delta \qquad \text{(Widrow)}$$

then the quantizer error distribution $\sim U(-\Delta/2, \Delta/2)$, $d = \Delta^2/12$. turns out *sufficient* but not *necessary*.

Fueled interest in white noise model: Quantizer error moments behave *as if* white noise model held.

Few interesting distributions satisfy this property, usefulness is minimal. *But characteristic function method proved very useful!*

Sripad and Snyder (1977) showed a *necessary **and** sufficient* condition is

$$\phi_{X_0}(\frac{j2\pi l}{\Delta}) = 0;\ l \neq 0 \qquad \text{(Sripad-Snyder)}$$

$f_{X_0} = U(-\Delta/2, \Delta/2)$ satisfies (Sripad-Snyder) but not (Widrow). Also showed that 2D version $\Leftrightarrow$ joint error pdf product of uniforms.

Problems with approach:

1. Requires an infinite level quantizer.
2. Extremely limited class of input distributions e.g., $U(-\Delta/2, \Delta/2)^{*k}$.
3. Original papers generated considerable confusion about appropriateness of white noise model.

High Rate Quantization Theory

Bennett (1948), Oliver, Pierce, and Shannon (1948): uniform quantization, input support of width A divided into $M = A/\Delta$ intervals of width $\Delta \downarrow 0$ has average MSE $\approx \Delta^2/12 \Rightarrow$ 6dB per bit — essentially same result appeared in Sheppard (1898)

Bennett (1948) also argued that quantizer error distribution was roughly **uniform** and that the power spectral density of quantization error was **approximately flat** in the signal bandwidth, **independent of the input signal.** Subsequently used to justify *white noise model*. *Does it?*

What does high rate regime actually imply about quantization error?

Long believed and finally proved by Lee & Neuhoff (1996), Viswanathan & Zamir (2001), Marco & Neuhoff (2005): If

a) input marginal distribution smooth *and fixed* with finite support

b) uniform quantizer on support of input density

c) large rate and small quantization intervals

d) all pairwise-joint input distributions are smooth *and fixed*

then the quantizer error e_n behaves approximately like additive noise that

1. is uncorrelated with the signal
2. has a uniform marginal distribution
3. has a flat power spectral density

} *weak* white noise model

Bennett (1948), Oliver, Pierce, and Shannon (1948), Panter and Dite (1951), Lloyd (1957): Asymptotic approximations for d and r for *nonuniform* quantization in the limit of many quantization levels ($R \to \infty, D \to 0$)

Zador (1963) extended high rate theory to vectors and entropy and Gersho (1979) developed an intuitive simplification and popularized the theory.

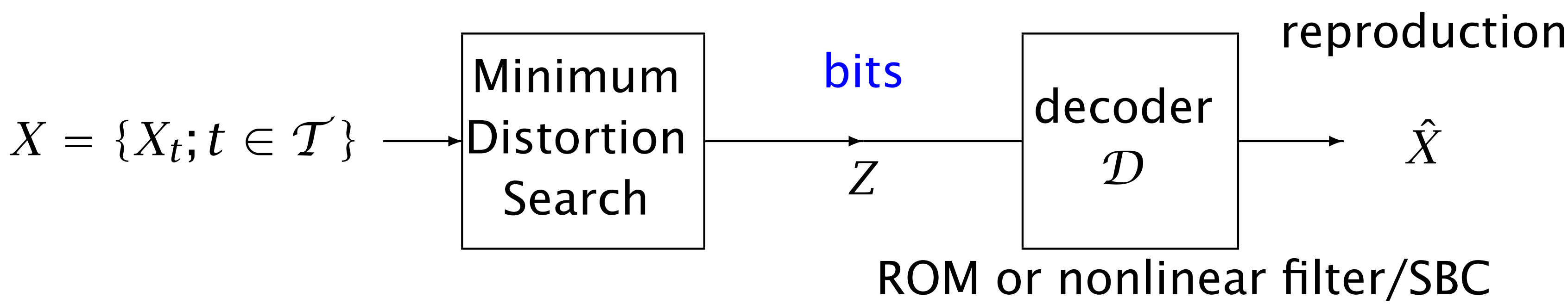
Gish & Pierce (1968) developed an intuitive high-rate theory for entropy.

Conway and Sloane (1982) applied to lattice VQ.

Shannon Rate Distortion Theory

Shannon (1948,1959), Gallager (1968), Jelinek (1968), Berger (1971): Shannon modeled lossy compression as block coding (=vector quantization (VQ)). Bounds on performance $R(D), D(R)$ for block codes which are theoretically achievable in limit as vector dimension $\to \infty$.

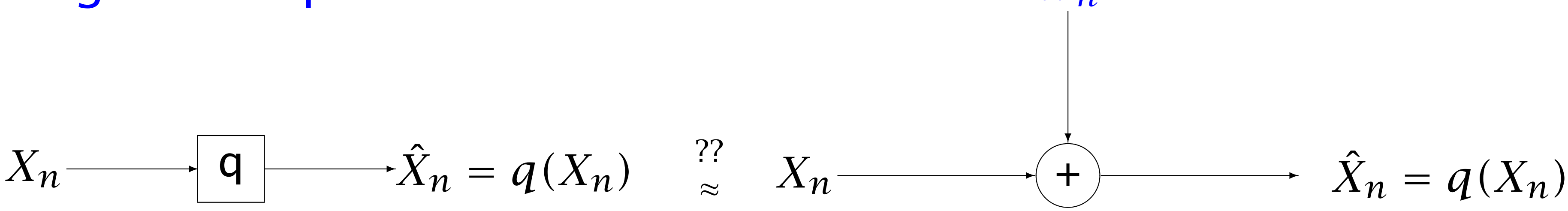
Implicit in Shannon: If decoder $\mathcal{D}$ is fixed, optimal encoder $\mathcal{E}$ performs minimum distortion search (nearest neighbor, dynamic programming, Viterbi algorithm)



Fundamental point: Shannon *assumed* $\mathcal{E}$ drives $\mathcal{D}$ so as to produce minimum distortion output $d(\mathcal{E}, \mathcal{D})$, $\mathcal{E}$ *optimal* for $\mathcal{D}$

“Model” Compression Noise

Replace hard part of analysis by simplifying assumption. “Wishful thinking” approach. Replace quantizer by an additive signal-independent white uniform noise W_n



Good news: Linearizes nonlinear system, simplifies a very difficult (maybe impossible) analysis of correlation and spectra, especially if q inside a feedback loop.

Most commonly used "theory" for ADC analysis.

but **Bad news**: It's wrong!!

- quantizer error is a deterministic function of the input, so input and quantizer error can not be independent.
- The quantizer error will not be iid unless input joint characteristic functions satisfy 2D (Sripad-Snyder) condition; e.g., input is iid.

Good news: High rate regime ⇒ weak white noise model gives a good approximation, but need to validate required underlying assumptions a)–d).

More bad news: Vast majority of literature simply assumes model without validating required conditions. If conditions violated, predictions of actual performance can be **wildly** inaccurate. E.g., discrete tones in audio compression.

For example, high-rate theory does not apply to feedback quantization systems like Delta Modulation, predictive quantization (DPCM), and $\Sigma - \Delta$ modulation because there are not fixed quantizer input distributions as rate increases.

E.g. problems with "theory" in VandeWeg's (1953) analysis of Delta modulation, Inose & Yasuda's, "A Unity Bit Coding Method by Negative Feedback" (1963) classic elaboration of Cutler's (1954) $\Sigma\Delta$ or $\Delta\Sigma$ ADC, O'Neal's "Differential Pulse-Code Modulation (PCM) with entropy coding" (1976)

Furthermore, often $M = 2$, not high rate!

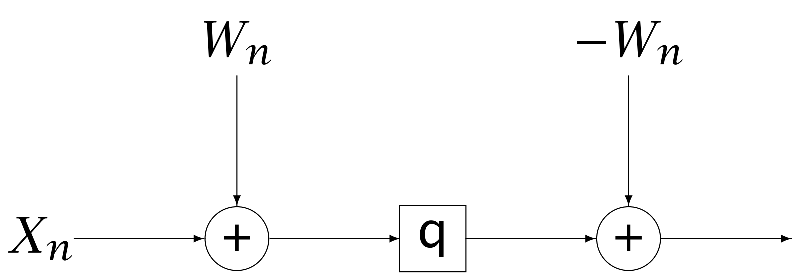
NO rigorous results justifying the white noise model — strong or weak — for feedback quantization!

Dither

There are perceptual reasons (eye and ear) to want quantizer error to be independent of the input signal and white.

Can force similar noise-like behavior of quantization error if use *subtractive dither*.

Roberts (1962) proposed adding a random or pseudo-random signal to input signal prior to quantization and subtracting it after decoding: W_n signal-independent uniform iid



$$\widetilde{X}_n = q(X_n + W_n) - W_n$$

$$\epsilon_n = \underbrace{\widetilde{X}_n - X_n}_{\text{quantization noise}} = \underbrace{q(X_n + W_n) - W_n - X_n}_{\text{quantization error}} = e_n,$$

Schuchman (1964) (a Widrow student) showed that if

$$\phi_W(\frac{j2\pi l}{\Delta}) = 0;\ l \neq 0, \qquad \text{(Schuchman)}$$

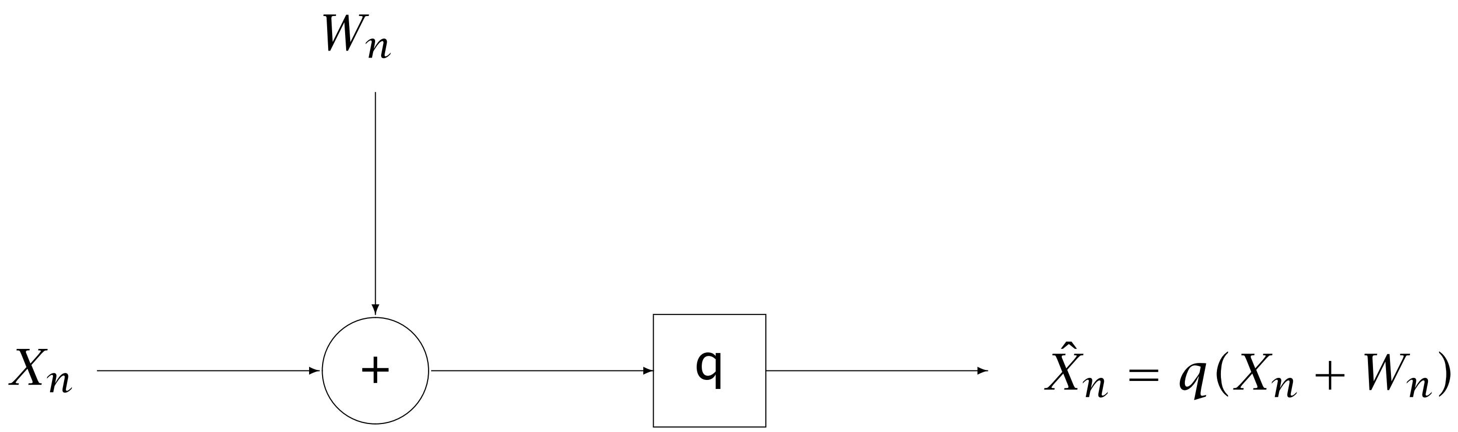
then the quantizer noise ϵ_n will be uniformly distributed and independent of the quantizer input. Sherwood (1985) showed also iid if the 2D extension of (Schuchman) holds.

Example: $W \sim U(-\Delta/2, \Delta/2)$

Here the condition is on the dither distribution and not the input distribution, the result holds for any input provided that $\Pr(q \text{ overloads }) = 0$.

Problem: Requires receiver know dither signal, hence either must also communicate perfect dither or use pseudorandom noise, in which case the theorem does not hold.

In real world usually use *non subtractive dither*



Schuchman's result remains valid for the quantization *error* $e_n = q(X_n + W_n) - (X_n + W_n)$, *but not for the overall quantization noise* $\epsilon_n = \hat{X}_n - X_n$ — an unfortunately common error in a major text [Jayant and Noll (1984) p. 170] and many papers.

Non subtractive dither *can not* make the quantizer noise ϵ_n independent of the signal or independent of $\epsilon_k, k \neq n$. It can force *conditional moments of the error given the signal to be functionally independent of the signal* and it can force uncorrelation of signal and error, *but Schuchman's condition is not enough.*

Wright (1979), Stockham (1979), Brinton (1984), Vanderkooy and Lipschitz (1984) showed that for a positive integer k

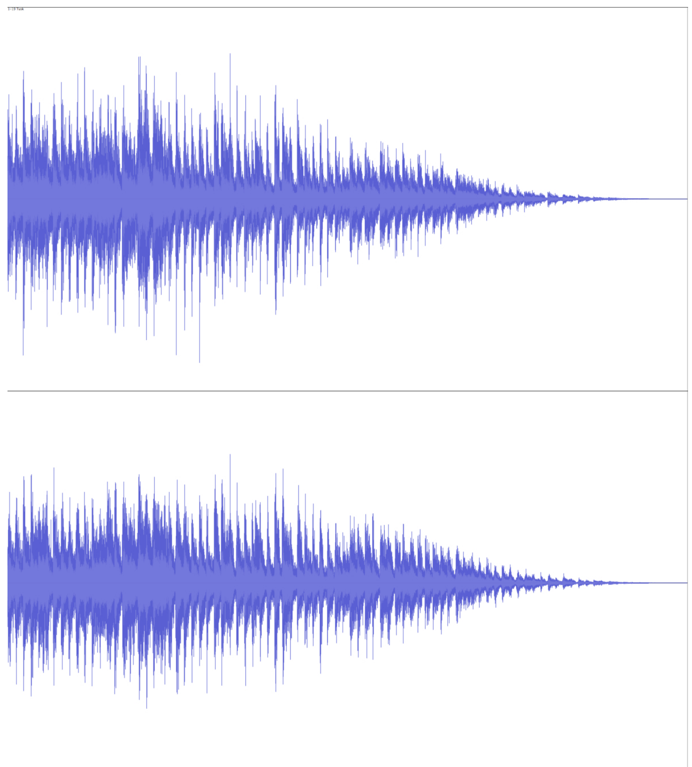
$$E[\epsilon^k|X] = E[\epsilon^k] = \frac{1}{j^k}\frac{d^k}{du^k}[\phi_W(u)\phi_U(u)]|_{u=0} = E[(W+U)^k] \quad \text{iff}$$

$$\frac{d^k}{du^k}[\phi_W(u)\phi_U(u)]|_{u=2\pi l/\Delta} = 0; \text{ all } l \neq 0$$

If true for $k = 1, 2$ then quantization noise and signal will be uncorrelated, quantization noise uncorrelated, &

$$E[\epsilon^2|X] = E[\epsilon^2] = E[W^2] + E[U^2] = \Delta^2/4.$$

An important example is the convolution of two uniform densities — the triangle pdf first used commercially in 1979 in the Fleetwood Mac *Tusk* album by Tom Stockham, the founder of Soundstream, the first digital audio recording company in the U.S. *uniform failed, triangular worked*



Lloyd Optimization

Constrain code structure and find *necessary conditions for optimality* for each component given the others for a given input distribution. Design code by sequentially optimizing each component for the other *a descent algorithm*

Original example in quantization: Lloyd (1957) “Least squares quantization in PCM,” Lloyd Method I for design of scalar quantizer for given input distribution and squared error. Alternate optimizing **encoder for decoder**:

minimum distortion mapping into reproduction codebook

(i.e., Shannon minimum distortion $\mathcal{E}$ for given $\mathcal{D}$)

and optimizing **decoder for encoder**:

replace decoder codebook by centroids wrt distortion

Lloyd et al.

Lloyd proved optimality properties using fundamental principles and *did not use variational methods requiring differentiability*. His results and his algorithm therefore extended immediately from scalar to vector, from squared error to any distortion measure possessing centroids, and from probability density functions to general distributions — including empirical distributions.

Steinhaus (1956) scooped Lloyd by using non variational techniques (= moment of inertia of an object about a vector is minimized if the vector is the centroid of the object) to demonstrate optimality conditions for a quantizer in 3D Euclidean space and a discrete distribution.

Several similar results were found earlier for statistical stratification or grouping: Lukaszewicz and Steinhaus (1955) considered an absolute magnitude distortion (centroid = median), Dalenius (1950), Dalenius and Gurney (1951), and Cox (1957) considered an equivalent problem in optimal stratification using squared error. All used variational methods and their results did not easily extend to general distortion measures and vector spaces. Cox further assumed normality and derived conditions for 2D.

Important theoretical point: For squared error distortion, centroid condition of optimal decoder ⇒ quantizer error is uncorrelated with *output*, but strongly correlated with *input* ⇒ white noise model breaks down for uniform quantization if use optimal (MMSE) codebook instead of interval midpoints.

Lloyd's Method II – later rediscovered by Max (1960). Became quite popular for scalar MMSE quantizer design, but does not generalize.

[Fine (1964)], Gish (1967)] developed similar ideas for feedback quantization like Δ-modulation

Ideas extended to entropy-constrained quantization by Berger (1972), Farvardin and Modestino (1984) using Lagrangian distortion $R + \lambda D$, now common in variable rate compression.

Vector Quantization

Lloyd's Method I (1957), Steinhaus (1956), Zador (1963) developed necessary conditions for optimality.

Basic ideas rediscovered in the statistical literature by Forgey (1965), Ball and Hall (1965), Jancey (1966), and MacQueen (1967) who named the method *k-means*. Steinhaus (1956) often credited with inventing k-means.

Can optimize multiple sets, e.g., gain-shape VQ in many vocoders [Sabin (1982,1984)] (almost always suboptimal in literature, needless division by gain)

Clustering ideas applied to LPC for speech [Chaffee & Omura (1974), Chaffee (1975)] and image coding and classification [Hilbert (1977)]

Other Theories of Data Compression

There are a variety of methods aimed at particular types of compression. See e.g. L.D. Davisson's review ["The theoretical analysis of data compression systems," *IEEE Proceedings*, (1968)] for a discussion of predictive, interpolative systems, and adaptive systems.

In addition to theories providing bounds, approximations, and formulas for optimization, much of the literature relies on **simulation** to validate heuristically derived approximations and models.

Shannon Theory vs. High Rate Theory

Simple entropy coded PCM only 1/4 bit from Shannon optimum *under suitable conditions* [Goblick & Holsinger (1967), Gish & Pierce (1968)]

So why bother to try harder?

Some extensions exist (e.g., Ziv's result using subtractive dither (1986)).

Often can do much better if can capture more **signal structure**, e.g., LPC and CELP/AbS

Furthermore, Shannon $D(R)$ computed for models, not for actual signals. $D(R)$ for model may be far from $D(R)$ for actual process (e.g., speech is NOT Gaussian, so $D(R)$ based on Gaussian is worst case)

Lastly, perceptual distortion measures may better reflect perceived quality than MSE (implicit in LPC, explicit in CELP) [Budrikis (1972), Stockham (1972), Safranek and Johnston (1988-9)]

A Few Historical Highlights

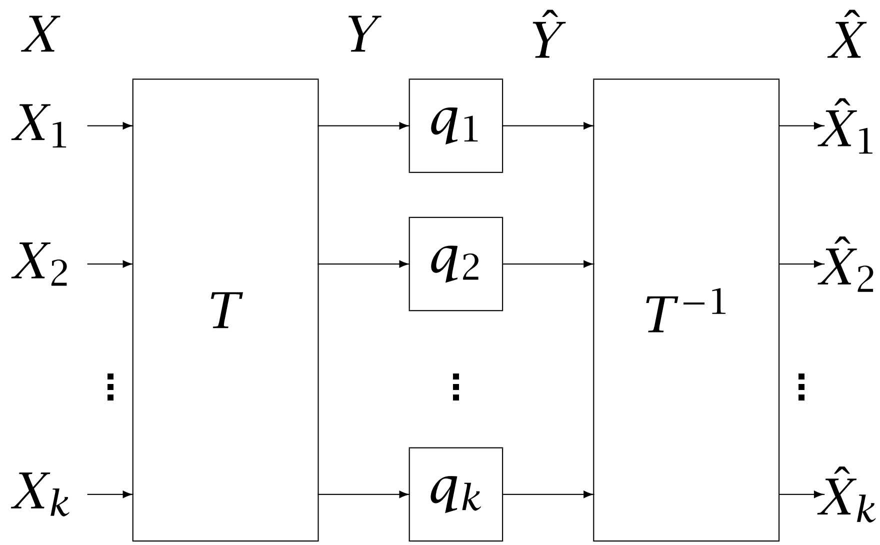
Combine scalar quantization with other signal processing

⋆ **Precede quantization by invertible linear transform/filter**

decorrelate vector components and concentrate energy in lower order coordinates, reduce dynamic range.
Scalar quantize output.

Can optimize *bit allocation* among quantizers.

Transform coding (VQ using scalar quantizers)



[Kramer & Mathews (1956), Huang & Schultheis (1962-3)] Fourier, cosine [Ahmed, Natarajan, Rao (1974), Chen and Pratt (1984)], JPEG [C-Cube et al. (1991)], wavelets [Antonini, Barlaud, Mathieu, & Daubechies (1992), Shapiro Zero-trees (1993), Said-Pearlman SPHIT (1996), Taubman EBCOT (1999, JPEG 2000)]

Dominates still image coding (JPEG, JPEG 2000). Included in most video coders

JPEG was a huge success, significantly helped by a specific market – *pornography*[1]

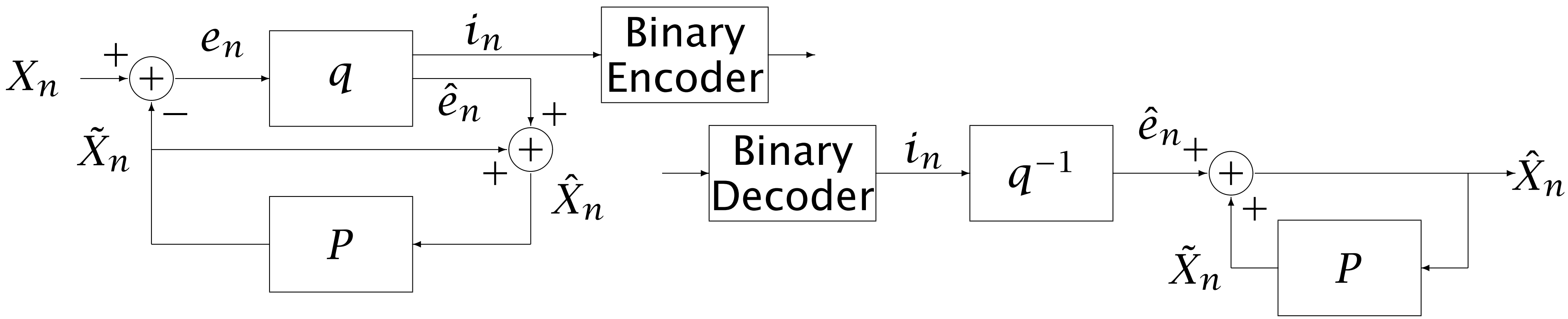
Subband coding (SBC). Variation on transform coding. Can incorporate perceptual weighting into distortion measure, various frequency decompositions. Dominates audio coding (mp3 is transform/subband coding + perceptual coding). Woods & O'Neil for images (1986)

---

[1]Jonathan Dotan, a co-producer of HBO series *Silicon Valley* about a lossless compression startup, is creating a website to go with a planned episode which will tell the story of the pornography industry's influence on imaging technology, including daguerreotypes, JPEG, VHS, Quicktime, DVDs. Stay tuned.

⋆Quantization inside a feedback loop (SBC)

Δ-modulation Derjavitch, Deloraine, and Van Mierlo (1946), Cutler's U.S. Patent 2,605,361 (1952), DeJager's Philips technical report on delta modulation (1952). More generally, Cutler's (1952) patent introduced **predictive quantization (DPCM).** Video coding dominated by motion compensation (predictive coding) + transform coded residuals (H.26*).



$\Sigma - \Delta$ **modulation** Cutler's U.S. Patent 2,927,962 (1960), Inose, Yasuda, and Murakami (1962), Inose & Yasuda (1963)

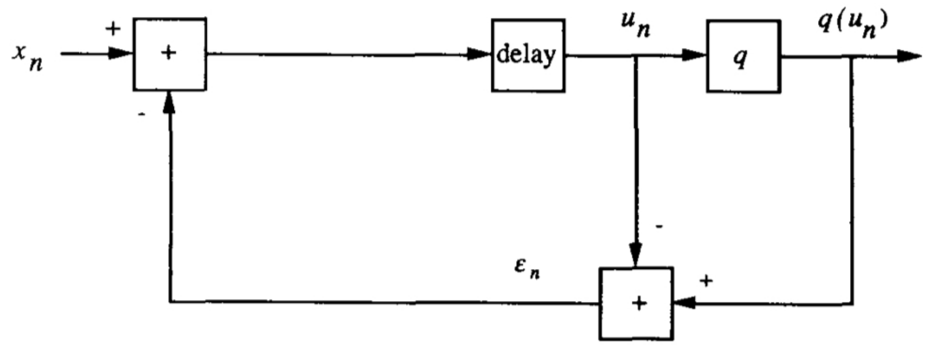


Fig. 8. Deterministic dithering.

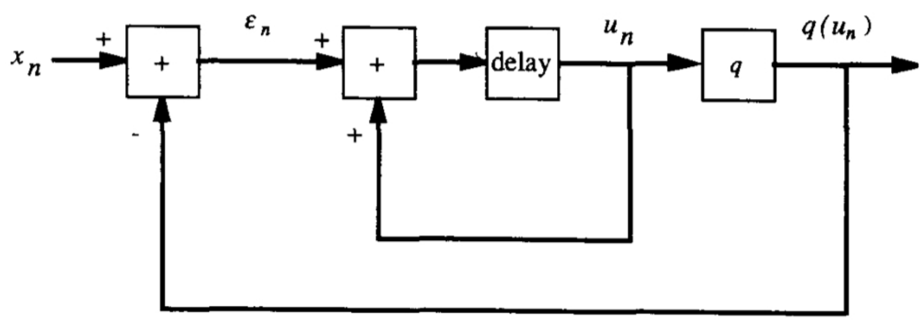


Fig. 9. Single-loop Sigma–Delta.

Cutler's (1960) patent also introduced **"noise shaping" quantization**.

Tree/Trellis Encoding

Origins in tree and trellis channel decoders and delayed-decision coding with DPCM: populate tree with DPCM outputs and encode by minimum distortion tree search.

Fano, (M, L), M algorithms:[Jelinek (1969), Jelinek and Anderson (1971) Anderson and Jelinek (1973), Dick, Berger, and Jelinek (1974), Anderson and Bodie (1975)]

Trellis encoding: [Viterbi and Omura (1974)] decoder is sliding-block code, encoder is Viterbi algorithm search of trellis

Trellis coded quantization improves on trellis encoding, lower complexity [Marcellin and Fischer (1990), M. W. Marcellin, T. R. Fischer, and J. D. Gibson (1990)]

Model Coding

Form synthesis model of input signal, quantize model parameters.

Receiver synthesizes local reproduction from model description, match spectra not waveforms

Primary example, speech coding/vocoding

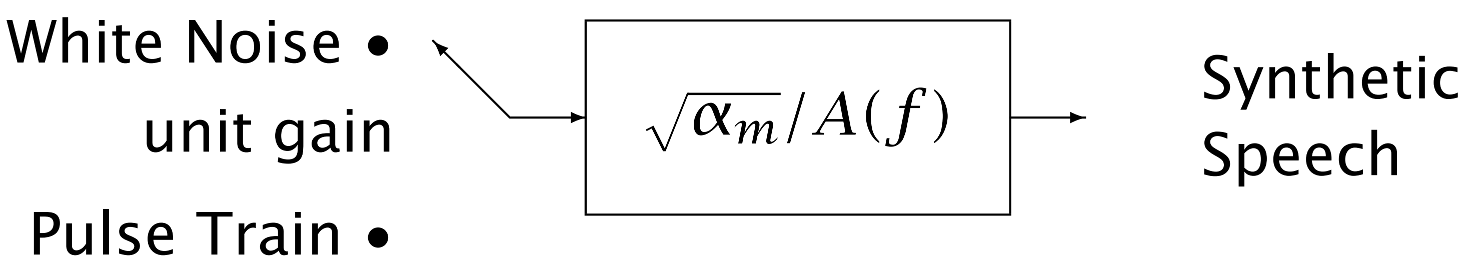
Linear predictive coding (LPC) [Itakura (1966), Atal and Hanauer (1971)]

Linear Prediction Observe data sequence $\{X_0, X_1, \ldots, X_{m-1}\}$, linear predictor $\hat{X}_m = -\sum_{l=1}^{m} a_l X_{m-l}$

$\epsilon_m = X_m - \hat{X}_m = \sum_{\ell=0}^{m} a_\ell X_{m-\ell}$ where $a_0 = 1$.
$a \triangleq (1, a_0, \cdots, a_m)$

LP problem $\alpha_m = \min_{a:a_0=1} E(\epsilon_m^2)$; $\quad a \Leftrightarrow A(f) = \sum_{n=0}^{m} a_n e^{-i2\pi nf}$

Suggests LP model for synthesizing speech — yields roughly same spectra as input



LPC model: a, α_m, voicing decision, pitch (if voiced)

but no *coding* yet, need to produce bits $\cdots$

Select best **approximation** $\hat{\sigma}/\hat{A}$ to model from a finite set. Traditionally done by **scalar quantization** of parameter vector from [Saito and Itakura (1966)]:

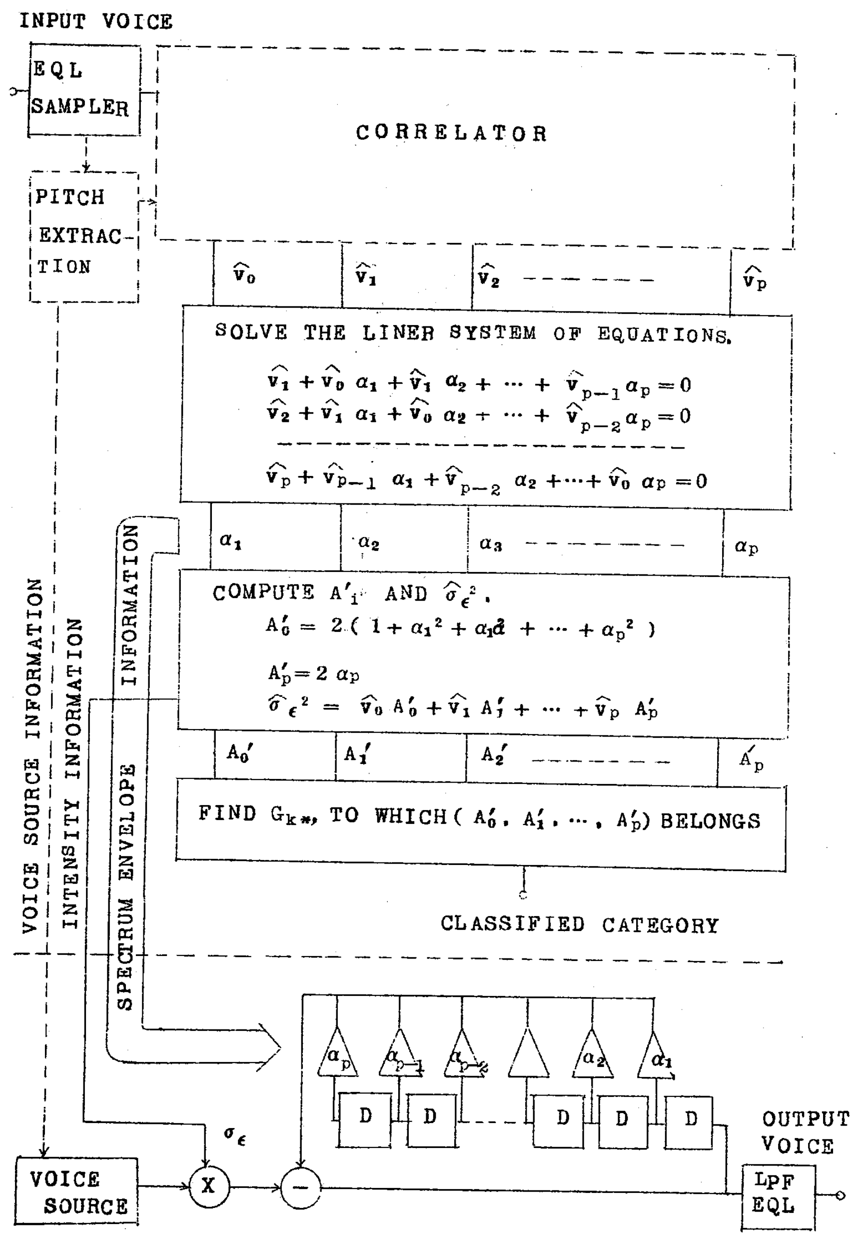


図 5.　新しいパラメータ伝送方式

LPC and VoIP

LPC was combined with original Network Communications Protocol (NCP) by Danny Cohen and the Network Speech Communications Group to obtain first understandable realtime two-way LPC packet speech communication. 3.5 kbs over ARPAnet between CHI and MIT-LL (1974) . Ancestor of VoIP and all real-time signal communication and streaming on the Internet — combine compression with suitable network protocol.

Very Low Rate Speech Coding

Early LPC produced understandable speech at low rates (e.g., 2400 bps), but quality not high

Could get comparable quality at lower rates using VQ on the parameters [Chaffee & Omura (1974–1975), Buzo, Matsuyama,Gray^2, Markel (1978–1980)].

800 bps Speech Coder [Dave Wong, Fred Juang, A.H. (Steen) Gray (1982)] very low rate, good quality for secure and military applications.

To get high quality with more bits needed something better.

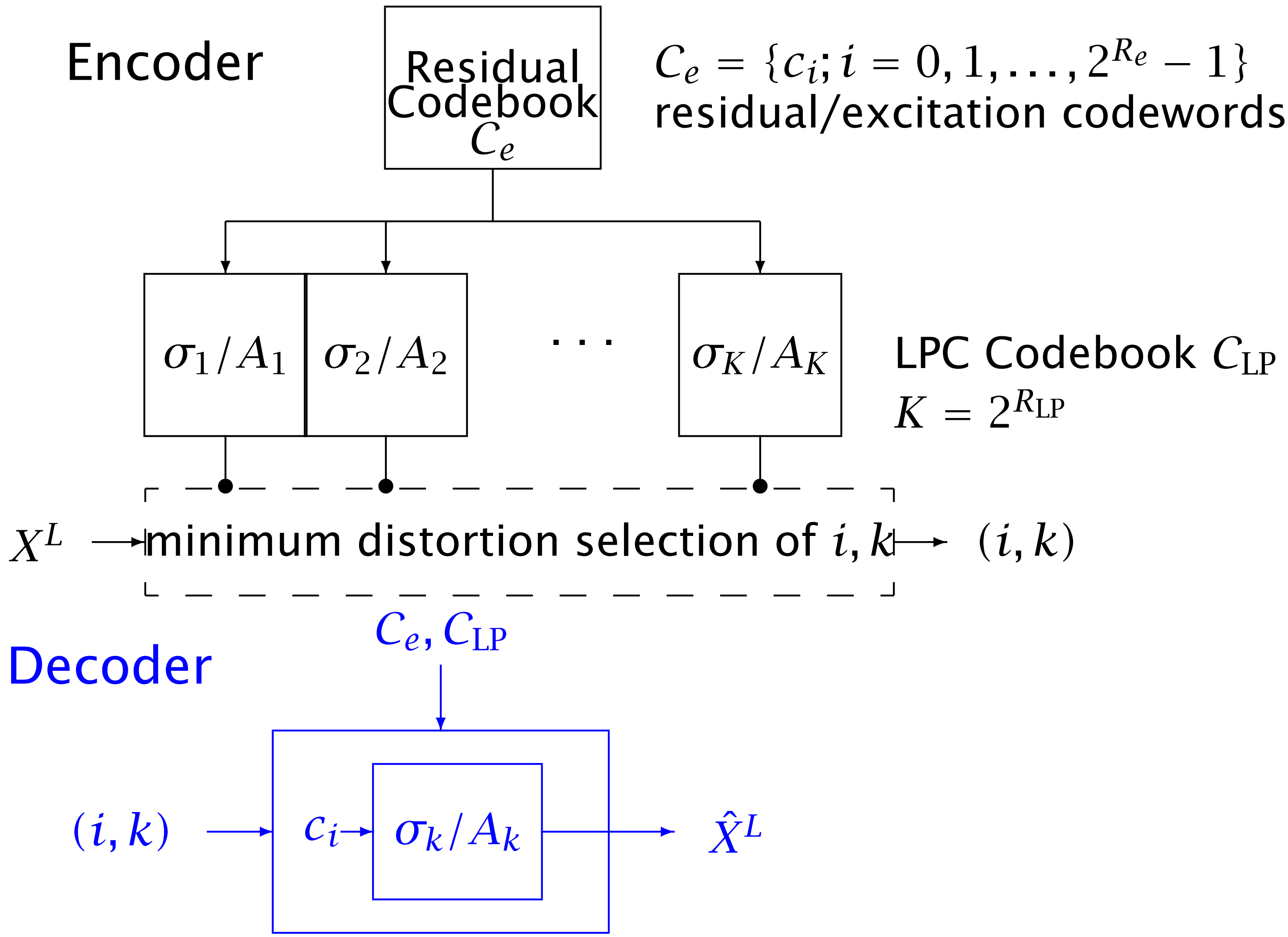
CELP/VXC/AbS Waveform Coding

Incorporate **compressed model** into closed-loop waveform coder for original signal.

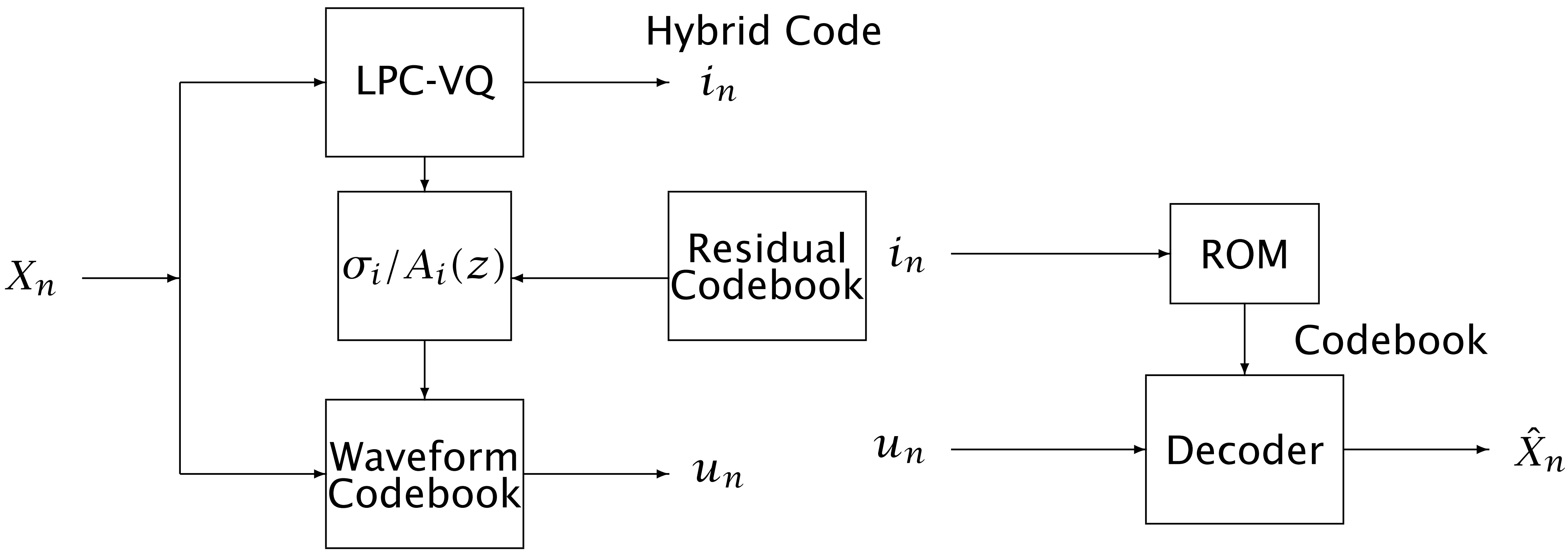
Idea: Instead of receiver generating locally the excitation to the model filter for synthesis, send best digital input sequence to filter in sense of matching model output to original signal — Choose VQ codeword yielding excitation with minimum *overall distortion* from input vector to final reproduction vector.

Y. Matsuyama (1978, 1981,1982) incorporated LPC VQ model into an LP-based waveform coder with minimum perceptually weighted squared error distortion match of input speech to output of model filter driven by residual/excitation codebook (using tree waveform encoding)

First idea: universal code — design LP models by Lloyd clustering using spectral distortion [Buzo (1978)], pick best model/excitation combination by parallel trees search:



Second idea: Pick LP model first using spectral distortion, then code tree colored by that model:



Both systems are LP models excited by an excitation codebook. In hindsight can also interpret as early example of *forward adaptive VQ*, *classified or switched codebook VQ*, *analysis-by-synthesis VQ*, and *code excited linear prediction*.

Adoul, Debray, and Dalle (1980) proposed a similar system which selected one of a collection of predictors for a DPCM coder designed by Lloyd clustering with respect to a spectral distance. *VQ on predictor selection, not on residuals*

Stewart (1981, 1982) extended (and credited) Matsuyama to large model codebooks, trellis waveform codes (VA) designed based on training data using a simple perceptual distortion weighting.

Schroeder and Atal (1981) introduced an AbS coder under the assumption of *perfect (no quantization)* LP coefficients and tree encoding of residual.

Atal and Remde (1982) introduced a multipulse driven LP model for AbS coder. Also assumed *perfect* LP coefficients.

Atal and Schroeder (1984) incorporated quantization of the model and introduced the name *code excited linear prediction (CELP)*.

Most histories of voice coding attribute invention of CELP and AbS LP coding to these three Atal et al. papers. Stewart is rarely (and Matsuyama almost never) mentioned.

Adoul et al. introduced CELP with low complexity algebraic residual codes (ACELP) (1987), ancestor of compression in most cell phones

nearly 6.8 billion, almost 1/person on earth

With time CELP incorporated many bells and whistles, including long and short prediction, fixed and adaptive codebooks, and post-filtering. Led to Federal Standard CELP (1991) and later cell phones and VoIP.

A Few Lowlights ☠

For some reason the field of data compression seems to attract more than its share of scammers, and in a few cases they have managed to collect millions of dollars in investment money before either being exposed, indicted, or disappearing. Mark Nelson, Dr Dobb's Bloggers

From *The data compression news blog*, www.c10n.info/archives/415: The standard compression scam is executed with the following steps:

- Visionary develops an astonishing breakthrough in compression.
- Company announces this amazing breakthrough without any validation, to promote himself and his company (the company has already issued several statements and releases on this technology).
- Visionary surrounds himself with key names, hires an ex-Qualcomm division president
- Makes sure that all management in the company are “in” through issuing stock (all upper management, the BOD and audit committee members have received stock or options from IPEX, or associated companies Digicorp (NASDAQ:DCGO) and/or Patient Safety Technologies (AMEX:PST)).
- Files patents with major patent firm (though they will never be issued).
- Throughout the process, insiders sell off millions of shares pocketing a fortune.
- Investors spend millions on a triumph of “hope over reality”. The company eventually folds leaving investors holding the bag.

Some examples:

Zeosync (2002) Guaranteed lossless 100:1 compression of any data. $50 M loss.

Zeosync explicitly claimed that they *superseded Claude Shannon's work*.

Jan Sloot Claimed amazing new compression album that could *compress a 1 hour HD move down to 8 Kbytes*. Mysteriously died in 1999 the day before being funded by venture capitalists. http://en.wikipedia.org/wiki/Jan_Sloot, www.endlesscompression.com.

Web Technologies (1995) Datafiles/16 Claimed could compress files larger than 64kB to about 1/16 original size. DataFiles/16 compressed files, but when decompressed, *those files bore no resemblance to their originals*.

June 1992 *Byte*, Vol 17 No 6:

According to Earl Bradley, WEB Technologies' vice president of sales and marketing, the compression algorithm used by DataFiles/16 *is not subject to the laws of information theory*.

Repeated or iterated or recursive compression Jules Gilbert's 1992 patent. Patent claimed to compress any digital file losslessly by at least one bit.

Fractal Codes Barnsely and Sloane (1987, 1988 *Byte* (10,000:1),1988 DCC (1,000:1)). Chaos and iterated function systems (IFS), Iterated Systems, Inc. (incorporated 1988).

And then there are the simply silly:

Confusion of compression with dimension reduction

Projection/downsampling/compressive sampling is not compression in Shannon sense until consider conversion into bits.

Equal Area/ Maximum Entropy Quantization (1971, 1973)

Use in LPC speech quoted led to Markel's Theorem: *Any system is optimal according to some criterion.*

Compression Ratio Confusion If original codebook size $N_1 = 2^{R_1}$ and reproduction codebook size $N_2 = 2^{R_2}$, is compression ratio $N_1/N_2 = 2^{R_1}/2^{R_2}$ or $\log_2 N_1/\log_2 N_2 = R_1/R_2$?

E.g., if VQ with original codebook (vector alphabet) of size $1024 = 2^{10}$ and reproduction codebook of size $2 = 2^1$, is the compression ratio $1024 : 1$ or only $10 : 1$?

Back to the Future: Task-oriented Quantization

Combining compression with signal processing: (classification, estimation, information extraction, learning), *task driven quantization* or *functional quantization*. Many examples show better to design compression in context of application, e.g., quantizing sensor readings for eventual use in classifiers. [Itakura (1966), Hilbert (1977), Kassam (1977), Poor and Thomas (1977), Picimbono and Duvaut (1988), Gersho (1996), Misra, Goyal, and Varshney (2008)]

A few other omitted important topics: Effects of channel noise, distributed quantization, universal lossy compression, multiple description , noisy source quantization, successive refinement/multiresolution/layered/progressive.

— $\mathcal{FIN}$ —

Primary Sources

Note: Most of the details of the citations of this presentation can be found in these papers and books. If some are obscure, contact me at rmgray@stanford.edu.

L. D. Davisson and R. M. Gray, eds., *Data Compression*, Dowden, Hutchinson and Ross, Benchmark Papers in Electrical Engineering and Computer Sciences, Vol. 14, 1976.

R. M. Gray,"Spectral analysis of quantization noise," *IEEE Transactions on Information Theory*, pp. 1220–1244, November 1990.

A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Kluwer (now Springer), 1992.

S.P. Lipshitz, R. A. Wannamaker, and J. Vanderkooy, "Quantization and dither: a theoretical survey," *J. Audio Eng. Soc.*, May 1992, pp. 355–375.

R. M. Gray and T. G. Stockham, Jr. "Dithered quantizers,"*IEEE Transactions on Information Theory*, May 1993, pp. 805–812.

R.M. Gray "Quantization noise in $\Delta\Sigma$ A/D converters," Chapter 2 of *Delta-Sigma Data Converters*, edited by S. Norsworthy, R. Schreier, and G. Temes, IEEE Press, 1997, pp. 44–74.

R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, pp. 2325–2384, October 1998. (Commemorative Issue, 1948–1998).

T. Berger and J.D. Gibson, "Lossy Source Coding," *IEEE Transactions on Information Theory*, pp. 2693–2723, Oct. 1998.

R.M. Gray, *Linear Predictive Coding and the Internet Protocol: A survey of LPC and a History of Realtime Digital Speech on Packet Networks*, Now Publishers, 2010.

W. A. Pearlman, “Milestones and Trends in Image Compression,” VCIP 2010.

http://www.faqs.org/faqs/compression-faq/

John C. Kieffer, “Ten milestones in the history of source coding theory,”
http://backup.itsoc.org/review/kieffer.html

D.L. Neuhoff, private communications.