# Beta Encoders

Ronald DeVore

University of South Carolina

# Mathematician's Perplexity

- Ingrid Daubechies, Sinan Gunturk, Vinay Vaishampayan

# Mathematician's Perplexity

- Ingrid Daubechies, Sinan Gunturk, Vinay Vaishampayan
- Analog/Digital conversion

# Mathematician's Perplexity

- Ingrid Daubechies, Sinan Gunturk, Vinay Vaishampayan
- Analog/Digital conversion
- Why Sigma-Delta?

# BANDLIMITED SIGNALS

Assume:

1. Class $\mathcal{B}$

a. $f$ is bandlimited $\hat{f}(\omega) = 0$, $|\omega| > \pi$

b. $f \in L_2 \cap L_\infty$ and $\|f\|_{L_\infty} < 1$

Shannon-Whitaker Formula

$$f(t) = \sum_{n \in \mathbb{Z}} f(n) \frac{\sin(t-n)}{(t-n)} = \sum_{n \in \mathbb{Z}} f(n) \text{sinc}(t-n)$$

Nyquist Sample Rate Is One

# Most Natural Encoding: PCM1

1. $m \geq 1$, $0 < x < 1$
2. $B_m(x) = b_1(x)2^{-1} + \cdots + b_m(x)2^{-m}$ first $m$-terms of binary expansion of $x$.
3. Encode: $f \longrightarrow \{(b_1(f(n)), \ldots, b_m(f(n)\}_{n \in \mathbb{Z}}$
4. Decode: $\bar{f}_n := B_m(f(n))$

$$\bar{f} = \sum_{n \in \mathbb{Z}} \bar{f}_n \text{sinc}(t - n)$$

# Optimal Bit Performance

- Fix $[0, T]$ on which we want to recover $f$

# Optimal Bit Performance

- Fix $[0, T]$ on which we want to recover $f$
- For bit budget of $m$ bits per Nyquist sample it seems that PCM1 has minimal distortion:

$$\|f - \bar{f}\|_{L_2[0,T]} \ll 2^{-m}$$

for any $T > 0$

# Optimal Bit Performance

- Fix $[0, T]$ on which we want to recover $f$
- For bit budget of $m$ bits per Nyquist sample it seems that PCM1 has minimal distortion:

$$\|f - \bar{f}\|_{L_2[0,T]} \ll 2^{-m}$$

for any $T > 0$

- Why is PCM1 not preferred in practice?

# Whoops: Some Problems

- Need all samples of $f$

# Whoops: Some Problems

- Need all samples of $f$
- PCM not stable:

$$\sum_{n\in\mathbb{Z}} |\text{sinc}(t-n)| = \infty$$

# Whoops: Some Problems

- Need all samples of $f$
- PCM not stable:

$$\sum_{n \in \mathbb{Z}} |\text{sinc}(t-n)| = \infty$$

- Can correct both of these by slight oversampling

# Whoops: Some Problems

- Need all samples of $f$
- PCM not stable:

$$\sum_{n\in\mathbb{Z}} |\text{sinc}(t-n)| = \infty$$

- Can correct both of these by slight oversampling
- Let $\lambda > 1$ Take $g$ with $\hat{g}$ smooth so that

$$\hat{g}_\lambda(\omega) = 1, \quad |\omega| \le \pi$$

$$\hat{g}_\lambda(\omega) = 0, \quad |\omega| > \lambda\pi$$

# Oversampling Continued

- For any bandlimited $f \in \mathcal{B}$:

$$f = \frac{1}{\lambda} \sum_{n \in \mathbb{Z}} f(\frac{n}{\lambda}) g_\lambda (t - \frac{n}{\lambda})$$

# Oversampling Continued

- For any bandlimited $f \in \mathcal{B}$:

$$f = \frac{1}{\lambda} \sum_{n \in \mathbb{Z}} f(\frac{n}{\lambda}) g_\lambda(t - \frac{n}{\lambda})$$

- $\hat{g}_\lambda$ smooth implies $g_\lambda$ decays exponentially

$$\frac{1}{\lambda} \sum_{n \in \mathbb{Z}} |g_\lambda(t - \frac{n}{\lambda})| \leq M, \quad \text{for all } t$$

# Oversampling Continued

- For any bandlimited $f \in \mathcal{B}$:

$$f = \frac{1}{\lambda} \sum_{n \in \mathbb{Z}} f(\frac{n}{\lambda}) g_\lambda(t - \frac{n}{\lambda})$$

- $\hat{g}_\lambda$ smooth implies $g_\lambda$ decays exponentially

$$\frac{1}{\lambda} \sum_{n \in \mathbb{Z}} |g_\lambda(t - \frac{n}{\lambda})| \leq M, \quad \text{for all } t$$

- PCM Encoding: $f \longrightarrow \{(b_1(f(n)), \ldots, b_m(f(n)\}_{n \in [-a, T+a]}$

# Oversampling Continued

- For any bandlimited $f \in \mathcal{B}$:

$$f = \frac{1}{\lambda} \sum_{n \in \mathbb{Z}} f(\frac{n}{\lambda}) g_\lambda(t - \frac{n}{\lambda})$$

- $\hat{g}_\lambda$ smooth implies $g_\lambda$ decays exponentially

$$\frac{1}{\lambda} \sum_{n \in \mathbb{Z}} |g_\lambda(t - \frac{n}{\lambda})| \leq M, \quad \text{for all } t$$

- PCM Encoding: $f \longrightarrow \{(b_1(f(n)), \ldots, b_m(f(n)\}_{n \in [-a, T+a]}$
- Still not the answer: Sigma-Delta preferred over PCM in practice

# Sigma-Delta Modulation: First order

1. Let $\lambda >> 1$
2. We want to assign one bit to each sample $f(\frac{n}{\lambda})$
3. Set $u_0 = 0$ and define recursively

$$\begin{cases} u_n = u_{n-1} + f(\frac{n}{\lambda}) - q_n^\lambda \\ q_n^\lambda = \text{ sign } \left(u_{n-1} + f(\frac{n}{\lambda})\right) \end{cases} ,$$

4. Read $f(\frac{1}{\lambda})$ assign $q_1^\lambda$, Read $f(\frac{2}{\lambda})$ assign $q_2^\lambda$, etc.

# Sigma-Delta continued

4'. Can define a similar reccursion running backwards

5. Decode:

$$f_\lambda(t) := \frac{1}{\lambda} \sum_{n\in\mathbb{Z}} q_n^\lambda g_\lambda(t - \frac{n}{\lambda})$$

6. $u_n$ state variable tracks differences in running sums:

$$u_n = \sum_{k=1}^{n} [f(\frac{k}{\lambda}) - q_k^\lambda]$$

# What is rate distortion for Sigma-Delta?

- Summation by parts

$$
\begin{aligned}
f(t) - f_\lambda(t) &= \frac{1}{\lambda}\sum_{n\in\mathbb{Z}}[f(\frac{n}{\lambda}) - q_n^\lambda]g_\lambda(t-\frac{n}{\lambda}) \\
&= \frac{1}{\lambda}\sum_{n\in\mathbb{Z}}[u_n - u_{n-1}]g_\lambda(t-\frac{n}{\lambda}) \\
&= \frac{1}{\lambda}\sum_{n\in\mathbb{Z}}u_n[g_\lambda(t-\frac{n}{\lambda}) - g_\lambda(t-\frac{n+1}{\lambda})]
\end{aligned}
$$

# Rate Distortion continued:

- If state variable $|u_n|$ bounded by $M$, then

$$|f(t) - f_\lambda(t)| \leq \frac{M}{\lambda} \sum_{n \in \mathbb{Z}} \int_{\frac{n}{\lambda}}^{\frac{n+1}{\lambda}} |g'_\lambda(s)|\, ds \leq C\frac{M}{\lambda}$$

# Rate Distortion continued:

- If state variable $|u_n|$ bounded by $M$, then

$$|f(t) - f_\lambda(t)| \leq \frac{M}{\lambda} \sum_{n \in \mathbb{Z}} \int_{\frac{n}{\lambda}}^{\frac{n+1}{\lambda}} |g'_\lambda(s)|\, ds \leq C \frac{M}{\lambda}$$

- Prove $|u_n| \leq 1$ by induction

$$u_n = \underbrace{u_{n-1} + f(\frac{n}{\lambda})}_{\in[-2,2]} - \text{sign}(u_{n-1} + f(\frac{n}{\lambda}))$$

# Compare:

- $m$ number of bits per Nyquist sample

# Compare:

- $m$ number of bits per Nyquist sample
- PCM has distortion $O(2^{-m})$

# Compare:

- $m$ number of bits per Nyquist sample
- PCM has distortion $O(2^{-m})$
- Sigma-Delta has distortion $O(1/m)$

# Compare:

- $m$ number of bits per Nyquist sample
- PCM has distortion $O(2^{-m})$
- Sigma-Delta has distortion $O(1/m)$
- Why use Sigma-Delta?

# Where are we?

- We still have no explanation of why engineers prefer Sigma-Delta modulation

# Where are we?

- We still have no explanation of why engineers prefer Sigma-Delta modulation
- From the viewpoint of rate distortion PCM is better

# Where are we?

- We still have no explanation of why engineers prefer Sigma-Delta modulation
- From the viewpoint of rate distortion PCM is better
- The answer must lie elsewhere

# Where are we?

- We still have no explanation of why engineers prefer Sigma-Delta modulation
- From the viewpoint of rate distortion PCM is better
- The answer must lie elsewhere
- Error in computation: circuit implementation

# Imperfect Quantizers

- In circuit implementation Quantizers will not be perfect

# Imperfect Quantizers

- In circuit implementation Quantizers will not be perfect
- Consider the imperfect implementation of $Q(x) := \text{sign } x$

$$
\begin{aligned}
Q_n(x) &= \text{sign}(x) \quad \text{for } |x| \geq \tau \\
|Q_n(x)| &= 1 \quad \text{for } |x| \leq \tau
\end{aligned}
$$

# Imperfect Quantizers

- In circuit implementation Quantizers will not be perfect
- Consider the imperfect implementation of $Q(x) := \text{sign } x$

$$
\begin{aligned}
Q_n(x) &= \text{sign}(x) \quad \text{for } |x| \geq \tau \\
|Q_n(x)| &= 1 \quad \text{for } |x| \leq \tau
\end{aligned}
$$

- Here $\tau$ can vary at each implementation but $|\tau| \leq \mu$ with $\mu$ fixed

# Imperfect quantizer in PCM

- Suppose $x = 1/2 + \delta$ with $0 < \delta < \tau$. Then first bit $b_1(x)$ may be wrong

# Imperfect quantizer in PCM

- Suppose $x = 1/2 + \delta$ with $0 < \delta < \tau$. Then first bit $b_1(x)$ may be wrong
- $b_1(x) \neq Q(x)$

# Imperfect quantizer in PCM

- Suppose $x = 1/2 + \delta$ with $0 < \delta < \tau$. Then first bit $b_1(x)$ may be wrong
- $b_1(x) \neq Q(x)$
- $|x - \bar{x}| \geq \delta$

# Imperfect quantizer in PCM

- Suppose $x = 1/2 + \delta$ with $0 < \delta < \tau$. Then first bit $b_1(x)$ may be wrong
- $b_1(x) \neq Q(x)$
- $|x - \bar{x}| \geq \delta$
- $|f(t) - \bar{f}(t)| \geq c\delta$

# Imperfect quantization in Sigma-Delta Modulation

New Dynamical System

$$\begin{cases} \bar{u}_n = u_{n-1} + f(\frac{n}{\lambda}) - \bar{q}_n^\lambda \\ \bar{q}_n^\lambda = Q_n\left(\bar{u}_{n-1} + f(\frac{n}{\lambda})\right) \end{cases} ,$$

CLAIM $|\bar{u}_n| \leq 1 + \delta$

$$\underbrace{u_{n-1} + f(\frac{n}{\lambda})}_{\in[-2-\delta,2+\delta]} - \bar{q}_n^\lambda$$

# Error Analysis

1. $\bar{f}(t) = \frac{1}{\lambda} \sum_{n\in\mathbb{Z}} \bar{q}_n^{\lambda} g_{\lambda}(t - \frac{n}{\lambda})$
2. $|f(t) - \bar{f}(t)| \leq C/\lambda$

- We get same error bounds with quantization error

# Error Analysis

1. $\bar{f}(t) = \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} \bar{q}_n^\lambda g_\lambda(t - \frac{n}{\lambda})$
2. $|f(t) - \bar{f}(t)| \leq C/\lambda$

- We get same error bounds with quantization error
- PCM gives error $\delta$, Sigma-Delta gives error $C/\lambda$

# Error Analysis

1. $\bar{f}(t) = \frac{1}{\lambda} \sum_{n \in \mathbb{Z}} \bar{q}_n^\lambda g_\lambda(t - \frac{n}{\lambda})$
2. $|f(t) - \bar{f}(t)| \leq C/\lambda$

- We get same error bounds with quantization error
- PCM gives error $\delta$, Sigma-Delta gives error $C/\lambda$
- Self correction is due to the feedback loop

# Error Analysis

1. $\bar{f}(t) = \frac{1}{\lambda} \sum_{n \in \mathbb{Z}} \bar{q}_n^\lambda g_\lambda(t - \frac{n}{\lambda})$
2. $|f(t) - \bar{f}(t)| \leq C/\lambda$

- We get same error bounds with quantization error
- PCM gives error $\delta$, Sigma-Delta gives error $C/\lambda$
- Self correction is due to the feedback loop
- Same analysis works for higher order Sigma-Delta

# Can we have best of both worlds?

- PCM offers exponential decay in distortion but no quantization error correcting

# Can we have best of both worlds?

- PCM offers exponential decay in distortion but no quantization error correcting
- Sigma-Delta offers quantization error correction but not exponential decay in distortion

# Can we have best of both worlds?

- PCM offers exponential decay in distortion but no quantization error correcting
- Sigma-Delta offers quantization error correction but not exponential decay in distortion
- Can we have both exponential rate distortion and quantization error correction

# Return to binary encoding

Quantizer $Q(y) = 1, \quad y \geq 1, \; Q(y) = 0, \quad y < 1$

$$\begin{aligned} u_1 &= 2x \\ b_1 &= Q(u_1) \\ u_{n+1} &= 2(u_n - b_n) \\ b_{n+1} &= Q(u_{n+1}) \end{aligned}$$

$$x = \sum_{n=1}^{\infty} b_n 2^{-n}$$

# EncoderDaubechies-DeVore-Günturk-Vas

- Let $1 < \beta < 2$

# EncoderDaubechies-DeVore-Günturk-Vas

- Let $1 < \beta < 2$
- If $x \in [0, 1]$, then $x = \sum_k b_k \beta^{-k}$

# EncoderDaubechies-DeVore-Güntürk-Vas

- Let $1 < \beta < 2$
- If $x \in [0, 1]$, then $x = \sum_k b_k \beta^{-k}$
- This decomposition is not unique

# EncoderDaubechies-DeVore-Güntürk-Vas

- Let $1 < \beta < 2$
- If $x \in [0,1]$, then $x = \sum_k b_k \beta^{-k}$
- This decomposition is not unique
- Can use redundancy to have quantization error correcting

# Delay Buffer:

- Idea is to delay assigning bit of one till sure. Can do this because we can always catch up

# Delay Buffer:

- Idea is to delay assigning bit of one till sure. Can do this because we can always catch up
- To ensure representation exists after assigning $b_1, \ldots, b_n$, we need

$$0 \leq x - \sum_{k=1}^{n} b_k \beta^{-k} \leq \sum_{k=n+1}^{\infty} \beta^{-k} = \frac{\beta^{-n}}{\beta - 1}$$

# Implement Delay

- We shall use a delay $\delta > 0$

# Implement Delay

- We shall use a delay $\delta > 0$
- Ideal quantization

$$Q(x) = 1, \quad x \geq 1 + \delta$$

$$Q(x) = 0, \quad x < 1 + \delta$$

# Determine bits

Encoding a number $x \in [0,1)$

$$u_1 = \beta x$$

$$b_1 = Q(u_1)$$

$$u_{n+1} = \beta(u_n - b_n)$$

$$b_{n+1} = Q(u_{n+1})$$

$$x = \sum_{n=1}^{\infty} b_n \beta^{-n}$$

# Imperfect quanizer

$$\bar{Q}(x) = 1, \quad x \geq 1 + \delta + \tau$$

$$\bar{Q}(x) = 0, \quad x < 1 + \delta - \tau$$

$$\bar{Q}(x) \in \{0, 1\}$$

# Encoding a number

$$\bar{u}_1 = \beta x$$

$$\bar{b}_1 = \bar{Q}(u_1)$$

$$\bar{u}_{n+1} = \beta(\bar{u}_n - \bar{b}_n)$$

$$\bar{b}_{n+1} = \bar{Q}(u_{n+1})$$

# ance of beta encoder under imperfect qua

Theorem Given $\mu$ and suppose each $\tau$ in the imperfect quantizer satisfies $|\tau| \leq \mu$. If delay $\delta$ satisfies
(i) $\mu \leq \delta$
(ii) $1 < \beta < \frac{2+\mu+\delta}{1+\mu+\delta}$
Then, for each $x \in [0, 1)$, we have

$$|x - \sum_{k=1}^{n} \bar{b}_k \beta^{-k}| \leq C\beta^{-n}$$

# Encoding Signals

- The beta encoder can be used to build an encoder for signals with the same exponential decay in the face of imprecise quantizers

# Encoding Signals

- The beta encoder can be used to build an encoder for signals with the same exponential decay in the face of imprecise quantizers
- Sample slightly above Nyquist rate

# Encoding Signals

- The beta encoder can be used to build an encoder for signals with the same exponential decay in the face of imprecise quantizers
- Sample slightly above Nyquist rate
- Quantize sample $f(\frac{n}{\lambda})$ using $m$ bits from Beta encoder

# Encoding Signals

- The beta encoder can be used to build an encoder for signals with the same exponential decay in the face of imprecise quantizers
- Sample slightly above Nyquist rate
- Quantize sample $f(\frac{n}{\lambda})$ using $m$ bits from Beta encoder
- This gives quantized $\bar{f}_n$

# Encoding Signals continued

- $\bar{f}_n$ decoding of these bits

$$|f(t) - \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} \bar{f}_n g_\lambda(t - \frac{n}{\lambda})| \leq C\beta^{-m}$$

# Encoding Signals continued

- $\bar{f}_n$ decoding of these bits

$$|f(t) - \frac{1}{\lambda}\sum_{n\in\mathbb{Z}} \bar{f}_n g_\lambda(t - \frac{n}{\lambda})| \leq C\beta^{-m}$$

- This encoder is impervious to quantization error