Finite frames and Sigma-Delta quantization

John J. Benedetto
Norbert Wiener Center, Department of Mathematics
University of Maryland, College Park

http://www.norbertwiener.umd.edu

Norbert Wiener Center
Outline and collaborators

1. Finite frames
2. Sigma-Delta quantization – theory and implementation
3. Sigma-Delta quantization – number theoretic estimates

Collaborators: Matt Fickus (frame force); Alex Powell and Ö zgür Yilmaz ($\sum - \Delta$ quantization); Alex Powell, Aram Tangboondouangjít, and Ö zgür Yilmaz ($\sum - \Delta$ quantization and number theory).
Finite Frames

Frames

Frames $F = \{e_n\}_{n=1}^{N}$ for $d$-dimensional Hilbert space $H$, e.g., $H = \mathbb{K}^d$, where $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$.

Any spanning set of vectors in $\mathbb{K}^d$ is a frame for $\mathbb{K}^d$. 

Finite frames and Sigma-Delta quantization – p.3/64
Frames $F = \{e_n\}_{n=1}^{N}$ for $d$-dimensional Hilbert space $H$, e.g., $H = \mathbb{K}^d$, where $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$.

- Any spanning set of vectors in $\mathbb{K}^d$ is a frame for $\mathbb{K}^d$.
- $F \subseteq \mathbb{K}^d$ is $A$-tight if

$$\forall x \in \mathbb{K}^d, A\|x\|^2 = \sum_{n=1}^{N} |\langle x, e_n \rangle|^2$$
Finite Frames

Frames

Frames $F = \{e_n\}_{n=1}^{N}$ for $d$-dimensional Hilbert space $H$, e.g., $H = \mathbb{K}^d$, where $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$.

- Any spanning set of vectors in $\mathbb{K}^d$ is a frame for $\mathbb{K}^d$.
- $F \subseteq \mathbb{K}^d$ is $A$-tight if

$$\forall x \in \mathbb{K}^d, \ A\|x\|^2 = \sum_{n=1}^{N} |\langle x, e_n \rangle|^2$$

- If $\{e_n\}_{n=1}^{N}$ is a finite unit norm tight frame (FUN-TF) for $\mathbb{K}^d$, with frame constant $A$, then $A = N/d$. 

Finite frames and Sigma-Delta quantization – p.3/64
Finite Frames

Frames

Frames $F = \{e_n\}_{n=1}^N$ for $d$-dimensional Hilbert space $H$, e.g., $H = \mathbb{K}^d$, where $\mathbb{K} = \mathbb{C}$ or $\mathbb{K} = \mathbb{R}$.

- Any spanning set of vectors in $\mathbb{K}^d$ is a frame for $\mathbb{K}^d$.

- $F \subseteq \mathbb{K}^d$ is $A$-tight if

  $$\forall x \in \mathbb{K}^d, \ A\|x\|^2 = \sum_{n=1}^{N} |\langle x, e_n \rangle|^2$$

- If $\{e_n\}_{n=1}^N$ is a finite unit norm tight frame (FUN-TF) for $\mathbb{K}^d$, with frame constant $A$, then $A = \frac{N}{d}$.

- Let $\{e_n\}$ be an $A$-unit norm TF for any separable Hilbert space $H$. $A \geq 1$, and $A = 1 \iff \{e_n\}$ is an ONB for $H$ (Vitali).
The geometry of finite tight frames

The vertices of platonic solids are FUN-TFs.
The geometry of finite tight frames

- The vertices of platonic solids are FUN-TFs.
- Points that constitute FUN-TFs do not have to be equidistributed, e.g., ONBs and Grassmanian frames.
The vertices of platonic solids are FUN-TFs.

Points that constitute FUN-TFs do not have to be equidistributed, e.g., ONBs and Grassmanian frames.

FUN-TFs can be characterized as minimizers of a “frame potential function” (with Fickus) analogous to
The geometry of finite tight frames

- The vertices of platonic solids are FUN-TFs.
- Points that constitute FUN-TFs do not have to be equidistributed, e.g., ONBs and Grassmanian frames.
- FUN-TFs can be characterized as minimizers of a “frame potential function” (with Fickus) analogous to Coulomb’s Law.
Frame force and potential energy

\[ F : S^{d-1} \times S^{d-1} \setminus D \rightarrow \mathbb{R}^d \]

\[ P : S^{d-1} \times S^{d-1} \setminus D \rightarrow \mathbb{R}, \]

where \( P(a, b) = p(\|a - b\|) \), \( p'(x) = -xf(x) \)

- Coulomb force

\[ CF(a, b) = (a - b)/\|a - b\|^3, \quad f(x) = 1/x^3 \]
Frame force and potential energy

\[
F : S^{d-1} \times S^{d-1} \setminus D \rightarrow \mathbb{R}^d
\]

\[
P : S^{d-1} \times S^{d-1} \setminus D \rightarrow \mathbb{R},
\]

where \( P(a, b) = p(\|a - b\|) \), \( p'(x) = -xf(x) \)

dots Coulomb force

\[
CF(a, b) = \frac{(a - b)}{\|a - b\|^3}, \quad f(x) = \frac{1}{x^3}
\]

dots Frame force

\[
FF(a, b) = \langle a, b \rangle (a - b), \quad f(x) = 1 - \frac{x^2}{2}
\]
Frame force and potential energy

\[ F : S^{d-1} \times S^{d-1} \setminus D \rightarrow \mathbb{R}^d \]

\[ P : S^{d-1} \times S^{d-1} \setminus D \rightarrow \mathbb{R}, \]

where \( P(a, b) = p(\|a - b\|), \quad p'(x) = -xf(x) \)

- Coulomb force

\[ CF(a, b) = (a - b)/\|a - b\|^3, \quad f(x) = 1/x^3 \]

- Frame force

\[ FF(a, b) = <a, b>(a - b), \quad f(x) = 1 - x^2/2 \]

- Total potential energy for the frame force

\[ TFP(\{x_n\}) = \sum_{m=1}^{N} \sum_{n=1}^{N} |<x_m, x_n>|^2 \]
Characterization of FUN-TFs

For the Hilbert space $H = \mathbb{R}^d$ and $N$, consider
\[ \{x_n\}_{1}^{N} \in S^{d-1} \times \ldots \times S^{d-1} \] and

\[
T F P(\{x_n\}) = \sum_{m=1}^{N} \sum_{n=1}^{N} | \langle x_m, x_n \rangle |^2.
\]

**Theorem** Let $N \leq d$. The minimum value of $T F P$, for the frame force and $N$ variables, is $N$; and the *minimizers* are precisely the orthonormal sets of $N$ elements for $\mathbb{R}^d$. 
Characterization of FUN-TFs

For the Hilbert space $H = \mathbb{R}^d$ and $N$, consider
\[ \{x_n\}_1^N \in S^{d-1} \times \ldots \times S^{d-1} \] and
\[ TFP(\{x_n\}) = \sum_{m=1}^N \sum_{n=1}^N |<x_m, x_n>|^2. \]

**Theorem** Let $N \leq d$. The minimum value of $TFP$, for the frame force and $N$ variables, is $N$; and the minimizers are precisely the orthonormal sets of $N$ elements for $\mathbb{R}^d$.

**Theorem** Let $N \geq d$. The minimum value of $TFP$, for the frame force and $N$ variables, is $N^2/d$; and the minimizers are precisely the FUN-TFs of $N$ elements for $\mathbb{R}^d$. 
Characterization of FUN-TFs

For the Hilbert space $H = \mathbb{R}^d$ and $N$, consider
\[ \{x_n\}_{n=1}^N \in S^{d-1} \times \ldots \times S^{d-1} \]
and
\[ TFP(\{x_n\}) = \sum_{m=1}^N \sum_{n=1}^N |<x_m, x_n>|^2. \]

**Theorem** Let $N \leq d$. The minimum value of $TFP$, for the frame force and $N$ variables, is $N$; and the minimizers are precisely the orthonormal sets of $N$ elements for $\mathbb{R}^d$.

**Theorem** Let $N \geq d$. The minimum value of $TFP$, for the frame force and $N$ variables, is $N^2/d$; and the minimizers are precisely the FUN-TFs of $N$ elements for $\mathbb{R}^d$.

**Problem** Find FUN-TFs analytically, effectively, computationally.
Given $u_0$ and $\{x_n\}_{n=1}^{\infty}$

$$u_n = u_{n-1} + x_n - q_n$$

$$q_n = Q(u_{n-1} + x_n)$$

First Order $\Sigma\Delta$
A quantization problem

**Qualitative Problem** Obtain *digital* representations for class \( X \), suitable for storage, transmission, recovery.

**Quantitative Problem** Find dictionary \( \{e_n\} \subseteq X \):

1. Sampling [continuous range \( K \) is not digital]

\[ \forall x \in X, \ x = \sum x_ne_n, \ x_n \in K (\mathbb{R} \text{ or } \mathbb{C}). \]

2. Quantization. Construct finite alphabet \( A \) and

\[ Q : X \to \{ \sum q_ne_n : q_n \in A \subseteq K \} \]

such that \( |x_n - q_n| \) and/or \( \|x - Qx\| \) small.

**Methods** Fine quantization, e.g., PCM. Take \( q_n \in A \) close to given \( x_n \). Reasonable in 16-bit (65,536 levels) digital audio.

Coarse quantization, e.g., \( \Sigma\Delta \). Use fewer bits to exploit redundancy.
\[ \mathcal{A}_{K}^{\delta} = \{(-K + 1/2)\delta, (-K + 3/2)\delta, \ldots, (-1/2)\delta, (1/2)\delta, \ldots, (K - 1/2)\delta\} \]

\[
Q(u) = \arg \min \{|u - q| : q \in \mathcal{A}_{K}^{\delta}\} = q_u
\]
Setting

Let $x \in \mathbb{R}^d$, $\|x\| \leq 1$. Suppose $F = \{e_n\}_{n=1}^{N}$ is a FUN-TF for $\mathbb{R}^d$. Thus, we have

$$x = \frac{d}{N} \sum_{n=1}^{N} x_n e_n$$

with $x_n = \langle x, e_n \rangle$. Note: $A = N/d$, and $|x_n| \leq 1$.

**Goal** Find a “good” quantizer, given

$$A^\delta_{K} = \{(-K + \frac{1}{2})\delta, (-K + \frac{3}{2})\delta, \ldots, (K - \frac{1}{2})\delta\}.$$ 

**Example** Consider the alphabet $\mathcal{A}_1^2 = \{-1, 1\}$, and $E_7 = \{e_n\}_{n=1}^{7}$, with $e_n = (\cos(\frac{2n\pi}{7}), \sin(\frac{2n\pi}{7})).$
$\mathcal{A}^2_1 = \{ -1, 1 \}$ and $E_7$

$\Gamma_{\mathcal{A}^2_1}(E_7) = \left\{ \frac{2}{7} \sum_{n=1}^{7} q_n e_n : q_n \in \mathcal{A}^2_1 \right\}$
Replace $x_n \leftrightarrow q_n = \arg\{\min |x_n - q| : q \in A^\delta_K\}$. Then

$$\tilde{x} = \frac{d}{N} \sum_{n=1}^{N} q_n e_n$$

satisfies

$$\|x - \tilde{x}\| \leq \frac{d}{N} \left\| \sum_{n=1}^{N} (x_n - q_n) e_n \right\| \leq \frac{d}{N} \frac{\delta}{2} \sum_{n=1}^{N} \|e_n\| = \frac{d}{2} \delta.$$ 

Not good!

**Bennett’s “white noise assumption”**

Assume that $(\eta_n) = (x_n - q_n)$ is a sequence of independent, identically distributed random variables with mean 0 and variance $\frac{\delta^2}{12}$. Then the mean square error (MSE) satisfies

$$\text{MSE} = E \|x - \tilde{x}\|^2 \leq \frac{d}{12A} \delta^2 = \frac{(d\delta)^2}{12N}.$$
Remarks

1. Bennett’s “white noise assumption” is not rigorous, and not true in certain cases.

2. The MSE behaves like $C/A$. In the case of $\Sigma \Delta$ quantization of bandlimited functions, the MSE is $O(A^{-3})$ (Gray, Güntürk and Thao, Bin Han and Chen). PCM does not utilize redundancy efficiently.

3. The MSE only tells us about the average performance of a quantizer.
$A_1^2 = \{-1, 1\}$ and $E_7$

Let $x = (\frac{1}{3}, \frac{1}{2})$, $E_7 = \{(\cos(\frac{2\pi n}{7}), \sin(\frac{2\pi n}{7}))\}_{n=1}^7$. Consider quantizers with $A = \{-1, 1\}$.
\[ A_1^2 = \{-1, 1\} \text{ and } E_7 \]

Let \( x = \left( \frac{1}{3}, \frac{1}{2} \right) \), \( E_7 = \{(\cos(\frac{2n\pi}{7}), \sin(\frac{2n\pi}{7}))\}_{n=1}^{7} \). Consider quantizers with \( A = \{-1, 1\} \).
\[ \mathcal{A}_1^2 = \{-1, 1\} \] and \( E_7 \)

Let \( x = \left( \frac{1}{3}, \frac{1}{2} \right) \), \( E_7 = \{(\cos(\frac{2n\pi}{7}), \sin(\frac{2n\pi}{7}))\}_{n=1}^7 \). Consider quantizers with \( \mathcal{A} = \{-1, 1\} \).
$\mathcal{A}_{1}^{2} = \{-1, 1\}$ and $E_{7}$

Let $x = (\frac{1}{3}, \frac{1}{2})$, $E_{7} = \{(\cos(\frac{2n\pi}{7}), \sin(\frac{2n\pi}{7}))\}_{n=1}^{7}$. Consider quantizers with $\mathcal{A} = \{-1, 1\}$. 

![Diagram showing a scatter plot with points representing $x_{\Sigma\Delta}$ and $x_{PCM}$ with quantization levels indicated by dots.]}
Let $F = \{e_n\}_{n=1}^{N}$ be a frame for $\mathbb{R}^d$, $x \in \mathbb{R}^d$.

Define $x_n = \langle x, e_n \rangle$.

Fix the ordering $p$, a permutation of $\{1, 2, \ldots, N\}$.

Quantizer alphabet $A^\delta_K$

Quantizer function $Q(u) = \text{arg}\{\min |u - q| : q \in A^\delta_K\}$

Define the first-order $\Sigma\Delta$ quantizer with ordering $p$ and with the quantizer alphabet $A^\delta_K$ by means of the following recursion.

\[
\begin{align*}
    u_n - u_{n-1} &= x_{p(n)} - q_n \\
    q_n &= Q(u_{n-1} + x_{p(n)})
\end{align*}
\]

where $u_0 = 0$ and $n = 1, 2, \ldots, N$. 
Stability

The following stability result is used to prove error estimates.

**Proposition** If the frame coefficients \( \{x_n\}_{n=1}^N \) satisfy

\[
|x_n| \leq (K - 1/2)\delta, \quad n = 1, \ldots, N,
\]

then the state sequence \( \{u_n\}_{n=0}^N \) generated by the first-order \( \Sigma\Delta \) quantizer with alphabet \( \mathcal{A}^{\delta}_K \) satisfies \( |u_n| \leq \delta/2, n = 1, \ldots, N. \)

The first-order \( \Sigma\Delta \) scheme is equivalent to

\[
u_n = \sum_{j=1}^{n} x_{p(j)} - \sum_{j=1}^{n} q_j, \quad n = 1, \ldots, N.
\]
The following stability result is used to prove error estimates.

**Proposition** If the frame coefficients $\{x_n\}_{n=1}^N$ satisfy

$$|x_n| \leq (K - 1/2)\delta, \quad n = 1, \cdots, N,$$

then the state sequence $\{u_n\}_{n=0}^N$ generated by the first-order $\Sigma\Delta$ quantizer with alphabet $A_{K}^{\delta}$ satisfies $|u_n| \leq \delta/2, n = 1, \cdots, N$.

- The first-order $\Sigma\Delta$ scheme is equivalent to

$$u_n = \sum_{j=1}^{n} x_{p(j)} - \sum_{j=1}^{n} q_j, \quad n = 1, \cdots, N.$$

- Stability results lead to **tiling problems** for higher order schemes.
**Definition** Let $F = \{e_n\}_{n=1}^{N}$ be a frame for $\mathbb{R}^d$, and let $p$ be a permutation of $\{1, 2, \ldots, N\}$. The *variation* $\sigma(F, p)$ is

$$\sigma(F, p) = \sum_{n=1}^{N-1} \|e_p(n) - e_p(n+1)\|.$$
Error estimate

**Definition** Let $F = \{e_n\}_{n=1}^N$ be a frame for $\mathbb{R}^d$, and let $p$ be a permutation of $\{1, 2, \ldots, N\}$. The *variation* $\sigma(F, p)$ is

$$\sigma(F, p) = \sum_{n=1}^{N-1} \|e_{p(n)} - e_{p(n+1)}\|.$$ 

**Theorem** Let $F = \{e_n\}_{n=1}^N$ be an $A$-FUN-TF for $\mathbb{R}^d$. The approximation

$$\tilde{x} = \frac{d}{N} \sum_{n=1}^N q_n e_{p(n)}$$

generated by the first-order $\Sigma \Delta$ quantizer with ordering $p$ and with the quantizer alphabet $A^\delta_K$ satisfies

$$\|x - \tilde{x}\| \leq \frac{(\sigma(F, p) + 1)d}{N} \frac{\delta}{2}.$$
Let $E_7$ be the FUN-TF for $\mathbb{R}^2$ given by the 7th roots of unity. Randomly select 10,000 points in the unit ball of $\mathbb{R}^2$. Quantize each point using the $\Sigma\Delta$ scheme with alphabet $A^{1/4}$. The figures show histograms for $||x - \tilde{x}||$ when the frame coefficients are quantized in their natural order $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ (left) and order $x_1, x_4, x_7, x_3, x_6, x_2, x_5$ (right).
$E_N = \{e^N_n\}_{n=1}^{N}, e^N_n = (\cos(2\pi n/N), \sin(2\pi n/N))$. Let $x = \left(\frac{1}{\pi}, \sqrt{\frac{3}{17}}\right)$.

$$x = \frac{d}{N} \sum_{n=1}^{N} x^N_n e^N_n, \quad x^N_n = \langle x, e^N_n \rangle.$$ 

Let $\tilde{x}_N$ be the approximation given by the 1st order $\Sigma\Delta$ quantizer with alphabet $\{-1, 1\}$ and natural ordering. log-log plot of $||x - \tilde{x}_N||$. 
Improved estimates

\[ E_N = \{e_n^N\}_{n=1}^N, \text{Nth roots of unity FUN-TFs for } \mathbb{R}^2, \, x \in \mathbb{R}^2, \]
\[ ||x|| \leq (K - 1/2)\delta. \]

Quantize

\[ x = \frac{d}{N} \sum_{n=1}^{N} x_n^N e_n^N, \quad x_n^N = \langle x, e_n^N \rangle \]

using 1st order \(\Sigma\Delta\) scheme with alphabet \(A_K^\delta\).

**Theorem** If \(N\) is even and large then \(||x - \tilde{x}|| \lesssim \frac{\delta \log N}{N^{5/4}}\).

If \(N\) is odd and large then \(\frac{\delta}{N} \lesssim ||x - \tilde{x}|| \leq \frac{(2\pi + 1)d}{N} \frac{\delta}{2}\).

**Remark** The proof uses the analytic number theory approach developed by Sinan Güntürk, and the theorem is true more generally.
Harmonic frames

Zimmermann and Goyal, Kelner, Kovačević, Thao, Vetterli.

\( H = \mathbb{C}^d \). An harmonic frame \( \{ e_n \}_{n=1}^N \) for \( H \) is defined by the rows of the Bessel map \( L \) which is the complex \( N \)-DFT \( N \times d \) matrix with \( N - d \) columns removed.
Harmonic frames

Zimmermann and Goyal, Kelner, Kovačević, Thao, Vetterli.

- $H = \mathbb{C}^d$. An harmonic frame $\{e_n\}_{n=1}^{N}$ for $H$ is defined by the rows of the Bessel map $L$ which is the complex $N$-DFT $N \times d$ matrix with $N - d$ columns removed.

- $H = \mathbb{R}^d$, $d$ even. The harmonic frame $\{e_n\}_{n=1}^{N}$ is defined by the Bessel map $L$ which is the $N \times d$ matrix whose $n$th row is

$$e_n^N = \sqrt{\frac{2}{d}} \left( \cos\left(\frac{2\pi n}{N}\right), \sin\left(\frac{2\pi n}{N}\right), \ldots, \cos\left(\frac{2\pi (d/2)n}{N}\right), \sin\left(\frac{2\pi (d/2)n}{N}\right) \right).$$
Harmonic frames

Zimmermann and Goyal, Kelner, Kovačević, Thao, Vetterli.

- $H = \mathbb{C}^d$. An harmonic frame $\{e_n\}_{n=1}^{N}$ for $H$ is defined by the rows of the Bessel map $L$ which is the complex $N$-DFT $N \times d$ matrix with $N - d$ columns removed.

- $H = \mathbb{R}^d$, $d$ even. The harmonic frame $\{e_n\}_{n=1}^{N}$ is defined by the Bessel map $L$ which is the $N \times d$ matrix whose $n$th row is

$$e_n^N = \sqrt{\frac{2}{d}} \left( \cos \left( \frac{2\pi n}{N} \right), \sin \left( \frac{2\pi n}{N} \right), \ldots, \cos \left( \frac{2\pi (d/2)n}{N} \right), \sin \left( \frac{2\pi (d/2)n}{N} \right) \right).$$

- Harmonic frames are FUN-TFs.
Harmonic frames

Zimmermann and Goyal, Kelner, Kovačević, Thao, Vetterli.

- $H = \mathbb{C}^d$. An harmonic frame $\{e_n\}_{n=1}^{N}$ for $H$ is defined by the rows of the Bessel map $L$ which is the complex $N$-DFT $N \times d$ matrix with $N - d$ columns removed.

- $H = \mathbb{R}^d$, $d$ even. The harmonic frame $\{e_n\}_{n=1}^{N}$ is defined by the Bessel map $L$ which is the $N \times d$ matrix whose $n$th row is

$$e_n^N = \sqrt{\frac{2}{d}} \left( \cos\left(\frac{2\pi n}{N}\right), \sin\left(\frac{2\pi n}{N}\right), \ldots, \cos\left(\frac{2\pi (d/2) n}{N}\right), \sin\left(\frac{2\pi (d/2) n}{N}\right) \right).$$

- Harmonic frames are FUN-TFs.

- Let $E_N$ be the harmonic frame for $\mathbb{R}^d$ and let $p_N$ be the identity permutation. Then

$$\forall N, \sigma(E_N, p_N) \leq \pi d(d + 1).$$
Theorem Let $E_N$ be the harmonic frame for $\mathbb{R}^d$ with frame bound $N/d$. Consider $x \in \mathbb{R}^d$, $\|x\| \leq 1$, and suppose the approximation $\tilde{x}$ of $x$ is generated by a first-order $\Sigma\Delta$ quantizer as before. Then

$$\|x - \tilde{x}\| \leq \frac{d^2(d + 1) + d}{N} \frac{\delta}{2}.$$ 

Hence, for harmonic frames (and all those with bounded variation),

$$\text{MSE}_{\Sigma\Delta} \leq \frac{C_d}{N^2} \delta^2.$$
Theorem  Let $E_N$ be the harmonic frame for $\mathbb{R}^d$ with frame bound $N/d$. Consider $x \in \mathbb{R}^d$, $\|x\| \leq 1$, and suppose the approximation $\tilde{x}$ of $x$ is generated by a first-order $\Sigma\Delta$ quantizer as before. Then

$$\|x - \tilde{x}\| \leq \frac{d^2(d + 1) + d}{N} \delta \frac{\delta}{2}.$$ 

Hence, for harmonic frames (and all those with bounded variation),

$$\text{MSE}_{\Sigma\Delta} \leq \frac{C_d}{N^2} \delta^2.$$ 

This bound is clearly superior asymptotically to

$$\text{MSE}_{\text{PCM}} = \frac{(d\delta)^2}{12N}.$$ 

Finite frames and Sigma-Delta quantization – p.25/64
The digital encoding

\[ \text{MSE}_{\text{PCM}} = \frac{(d\delta)^2}{12N} \]

in PCM format leaves open the possibility that decoding (reconstruction) could lead to

\[ \text{“MSE}_{\text{PCM}}^{\text{opt}} \ll O\left(\frac{1}{N}\right). \]

Goyal, Vetterli, Thao (1998) proved

\[ \text{“MSE}_{\text{PCM}}^{\text{opt}} \sim \frac{\tilde{C}_d}{N^2} \delta^2. \]

**Theorem** The first order \( \Sigma \Delta \) scheme achieves the asymptotically optimal MSE\(_{\text{PCM}}\), for harmonic frames.
Proof of Improved Estimates theorem

If $N$ is even and large then $\|x - \tilde{x}\| \lesssim \frac{\delta \log N}{N^{5/4}}$.

If $N$ is odd and large then $\frac{\delta}{N} \lesssim \|x - \tilde{x}\| \leq \frac{(2\pi + 1)d}{N} \frac{\delta}{2}$.
Sigma-Delta quantization–number theoretic estimates

Proof of Improved Estimates theorem

- If $N$ is even and large then $\|x - \tilde{x}\| \lesssim \frac{\delta \log N}{N^{3/4}}$.
- If $N$ is odd and large then $\frac{\delta}{N} \lesssim \|x - \tilde{x}\| \leq \frac{(2\pi + 1)d\delta}{N}$.

∀$N$, $\{e_n^N\}_{n=1}^N$ is a FUN-TF.
Proof of Improved Estimates theorem

If $N$ is even and large then $||x - \tilde{x}|| \lesssim \frac{\delta \log N}{N^{5/4}}$.

If $N$ is odd and large then $\frac{\delta}{N} \lesssim ||x - \tilde{x}|| \leq \frac{(2\pi + 1)d \delta}{2}$.

$\forall N, \{e_n^N\}_{n=1}^N$ is a FUN-TF.

$$x - \tilde{x}_N = \frac{d}{N} \left( \sum_{n=1}^{N-2} v_n^N (f_n^N - f_{n+1}^N) + v_{N-1}^N f_{N-1}^N + u_N^N e_N^N \right)$$

$$f_n^N = e_n^N - e_{n+1}^N, \quad v_n^N = \sum_{j=1}^n u_j^N, \quad \tilde{u}_n^N = \frac{u_n^N}{\delta}$$
Sigma-Delta quantization–number theoretic estimates

Proof of Improved Estimates theorem

- If \( N \) is even and large then \( ||x - \tilde{x}|| \lesssim \frac{\delta \log N}{N^{5/4}} \).
- If \( N \) is odd and large then \( \frac{\delta}{N} \lesssim ||x - \tilde{x}|| \leq \frac{(2\pi + 1)d}{N} \delta \).

\( \forall N, \{e_n^N\}_{n=1}^N \) is a FUN-TF.

\[
x - \tilde{x}_N = \frac{d}{N} \left( \sum_{n=1}^{N-2} v_n^N (f_n^N - f_{n+1}^N) + v_{N-1}^N f_{N-1}^N + u_N^N e_N^N \right)
\]

\[
f_n^N = e_n^N - e_{n+1}^N, \quad v_n^N = \sum_{j=1}^n u_j^N, \quad \tilde{u}_n^N = \frac{u_n^N}{\delta}
\]

To bound \( v_n^N \).
Koksma Inequality

Discrepancy

The discrepancy $D_N$ of a finite sequence $x_1, \ldots, x_N$ of real numbers is

$$D_N = D_N(x_1, \ldots, x_N) = \sup_{0 \leq \alpha < \beta \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\alpha, \beta]}(\{x_n\}) - (\beta - \alpha) \right|,$$

where $\{x\} = x - \lfloor x \rfloor$. 
**Koksma Inequality**

- **Discrepancy**
  The discrepancy $D_N$ of a finite sequence $x_1, \ldots, x_N$ of real numbers is
  \[
  D_N = D_N(x_1, \ldots, x_N) = \sup_{0 \leq \alpha < \beta \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\alpha, \beta)}(\{x_n\}) - (\beta - \alpha) \right|,
  \]
  where $\{x\} = x - \lfloor x \rfloor$.

- **Koksma Inequality**
  \[
  g : [-1/2, 1/2) \to \mathbb{R} \text{ of bounded variation and } \{\omega_j\}_{j=1}^{n} \subset [-1/2, 1/2) \implies
  \left| \frac{1}{n} \sum_{j=1}^{n} g(\omega_j) - \int_{-\frac{1}{2}}^{\frac{1}{2}} g(t)dt \right| \leq \text{Var}(g) \text{Disc} \left( \{\omega_j\}_{j=1}^{n} \right).
  \]
Koksma Inequality

**Discrepancy**
The discrepancy $D_N$ of a finite sequence $x_1, \ldots, x_N$ of real numbers is

$$D_N = D_N(x_1, \ldots, x_N) = \sup_{0 \leq \alpha < \beta \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\alpha, \beta)}(\{x_n\}) - (\beta - \alpha) \right|,$$

where $\{x\} = x - \lfloor x \rfloor$.

**Koksma Inequality**

$g : [-1/2, 1/2) \to \mathbb{R}$ of bounded variation and

$\{\omega_j\}_{j=1}^{n} \subset [-1/2, 1/2) \implies$

$$\left| \frac{1}{n} \sum_{j=1}^{n} g(\omega_j) - \int_{-1/2}^{1/2} g(t) dt \right| \leq \text{Var}(g) \text{Disc} \left( \{\omega_j\}_{j=1}^{n} \right).$$

With $g(t) = t$ and $\omega_j = \tilde{u}_j^N$,

$$|v_n^N| \leq n\delta \text{Disc} \left( \{\tilde{u}_j^N\}_{j=1}^{n} \right).$$
Erdös-Turán Inequality

\[ \exists C > 0, \forall K, \text{Disc}\left(\{\tilde{u}_n^N\}_{n=1}^j\right) \leq C\left(\frac{1}{K} + \frac{1}{j} \sum_{k=1}^K \frac{1}{k} \left| \sum_{n=1}^j e^{2\pi i k \tilde{u}_n^N} \right|\right). \]
Erdös-Turán Inequality

\[ \exists C' > 0, \forall K, \text{Disc}\left(\{\tilde{u}_n^N\}_{n=1}^j\right) \leq C\left(\frac{1}{K} + \frac{1}{j} \sum_{k=1}^K \frac{1}{k} \left| \sum_{n=1}^j e^{2\pi ik\tilde{u}_n^N} \right| \right). \]

To approximate the exponential sum.
Approximation of Exponential Sum

(1) Güntürk’s Proposition
\[ \forall N, \exists X_N \in \mathcal{B}_{\Omega/N} \]
such that \( \forall n = 0, \ldots, N, \)
\[ X_N(n) = u_n^N + c_n \frac{\delta}{2}, \quad c_n \in \mathbb{Z} \]
and \( \forall t, \left| X_N'(t) - h \left( \frac{t}{N} \right) \right| \lesssim \frac{1}{N} \)

(2) Bernstein’s Inequality
If \( x \in \mathcal{B}_{\Omega} \), then \( \| x^{(r)} \|_\infty \leq \Omega^r \| x \|_\infty \)

\[ \widehat{\mathcal{B}}_{\Omega} = \{ T \in A' (\mathbb{R}) : \operatorname{supp} T \subseteq [-\Omega, \Omega] \} \]

\[ \mathcal{M}_{\Omega} = \{ h \in \mathcal{B}_{\Omega} : h' \in L^\infty (\mathbb{R}) \text{ and all zeros of } h' \text{ on } [0, 1] \text{ are simple} \} \]

We assume \( \exists h \in \mathcal{M}_{\Omega} \) such that \( \forall N \) and \( \forall 1 \leq n \leq N, \ h(n/N) = x_n^N \).
Approximation of Exponential Sum

(1) Güntürk’s Proposition

∀N, ∃X_N ∈ B_{Ω/N} such that ∀n = 0, ..., N,

X_N(n) = u_n^N + c_n \frac{\delta}{2}, c_n ∈ \mathbb{Z}

and ∀t, \left| X'_N(t) - h\left(\frac{t}{N}\right) \right| \lesssim \frac{1}{N}

(2) Bernstein’s Inequality

If x ∈ B_Ω, then \|x^{(r)}\|_∞ \leq Ω^r \|x\|_∞

∀t, \left| X''_N(t) - \frac{1}{N} h'\left(\frac{t}{N}\right) \right| \lesssim \frac{1}{N^2}

• \hat{B}_Ω = \{T ∈ A'(\mathbb{R}) : \text{supp}T ⊆ [-Ω, Ω]\}

• \mathcal{M}_Ω = \{h ∈ B_Ω : h' ∈ L^∞(\mathbb{R}) and all zeros of h' on [0, 1] are simple\}

• We assume ∃h ∈ \mathcal{M}_Ω such that ∀N and ∀1 ≤ n ≤ N, h(n/N) = x_n^N.
Van der Corput Lemma

Let $a, b$ be integers with $a < b$, and let $f \in C^2([a, b])$ with $f''(x) \geq \rho > 0$ for all $x \in [a, b]$ or $f''(x) \leq -\rho < 0$ for all $x \in [a, b]$ then

$$\left| \sum_{n=a}^{b} e^{2\pi i f(n)} \right| \leq \left( |f'(b) - f'(a)| + 2 \right) \left( \frac{4}{\sqrt{\rho}} + 3 \right).$$
Van der Corput Lemma

Let $a, b$ be integers with $a < b$, and let $f \in C^2([a, b])$ with $f''(x) \geq \rho > 0$ for all $x \in [a, b]$ or $f''(x) \leq -\rho < 0$ for all $x \in [a, b]$ then

$$\left| \sum_{n=a}^{b} e^{2\pi if(n)} \right| \leq \left( |f'(b) - f'(a)| + 2 \right) \left( \frac{4}{\sqrt{\rho}} + 3 \right).$$

∀ $0 < \alpha < 1$, ∃ $N_\alpha$ such that ∀ $N \geq N_\alpha$,

$$\left| \sum_{n=1}^{j} e^{2\pi ik\tilde{u}_n^N} \right| \leq N^\alpha + \frac{\sqrt{k}N^{1-\frac{\alpha}{2}}}{\sqrt{\delta}} + \frac{k}{\delta}.$$
Choosing appropriate $\alpha$ and $K$

Putting $\alpha = 3/4$, $K = N^{1/4}$ yields

$$\exists \tilde{N} \text{ such that } \forall N \geq \tilde{N}, \text{Disc}\left(\left\{\tilde{u}^N_n\right\}_{n=1}^j\right) \lesssim \frac{1}{N^{1/4}} + \frac{N^{3/4} \log(N)}{j}$$
Choosing appropriate $\alpha$ and $K$

Putting $\alpha = 3/4$, $K = N^{1/4}$ yields

$$\exists \tilde{N} \text{ such that } \forall N \geq \tilde{N}, \text{Disc} \left( \left\{ \tilde{u}_n^N \right\}_{n=1}^j \right) \lesssim \frac{1}{N^{1/4}} + \frac{N^{3/4} \log(N)}{j}$$

Conclusion

$$\forall n = 1, \ldots, N, \ |v^N_n| \lesssim \delta N^{3/4} \log N$$
A sequence \((x_n)\) of real numbers is said to be uniformly distributed modulo 1 (u.d. mod 1) if

\[
\forall 0 \leq \alpha < \beta \leq 1, \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} 1_{[\alpha, \beta)}(\{x_n\}) = \beta - \alpha,
\]

where \(\{a\} = a - [a]\) denotes the fractional part of a real number \(a\).

**Examples** The following sequences are u.d. mod 1:

\[
\begin{align*}
0 & \quad 0 & \quad 1 & \quad 0 & \quad 1 & \quad 2 & \quad 0 & \quad 1 & \quad k - 1 \\
\overline{1} & \quad \overline{2} & \quad \overline{2} & \quad \overline{3} & \quad \overline{3} & \quad \overline{3} & \quad \overline{k} & \quad \overline{k} & \quad \overline{k} & \quad \overline{k}
\end{align*}
\]
Uniform distribution and discrepancy

A sequence \((x_n)\) of real numbers is said to be uniformly distributed modulo 1 (u.d. mod 1) if

\[
\forall 0 \leq \alpha < \beta \leq 1, \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} 1_{[\alpha,\beta)}(\{x_n\}) = \beta - \alpha,
\]

where \(\{a\} = a - [a]\) denotes the fractional part of a real number \(a\).

**Examples** The following sequences are u.d. mod 1:

1. \(0\ 0\ 1\ 0\ 1\ 2\ 0\ 1\ \frac{k - 1}{k}, \ldots\)
2. The sequence \((n\alpha)\) where \(\alpha\) is an irrational number.
Uniform distribution and discrepancy

A sequence \((x_n)\) of real numbers is said to be uniformly distributed modulo 1 (u.d. mod 1) if

\[
\forall 0 \leq \alpha < \beta \leq 1, \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} 1_{[\alpha, \beta)}(\{x_n\}) = \beta - \alpha,
\]

where \(\{a\} = a - [a]\) denotes the fractional part of a real number \(a\).

**Examples** The following sequences are u.d. mod 1:

- \(0, 0, 1, 0, 1, 2, 0, 1, \ldots, k - 1, \ldots\)
- The sequence \((n\alpha)\) where \(\alpha\) is an irrational number.
- The sequence \((\log F_n)\) where \(F_n\) denotes the Fibonacci sequence \(F_0 = 0, F_1 = 1, F_n = F_{n-1} + F_{n-2}\) for \(n \geq 2\).
Criteria for uniform distribution

**Weyl Criterion:** The sequence \( (x_n) \) is u.d. mod 1 if and only if for all integers \( h \neq 0 \)

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = 0.
\]
Criteria for uniform distribution

- **Weyl Criterion:** The sequence \((x_n)\) is u.d. mod 1 if and only if for all integers \(h \neq 0\)

  \[
  \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} = 0.
  \]

- The sequence \((x_n)\) is u.d. mod 1 if and only if for every Riemann-integrable function \(f\) on \([0, 1]\), the following equation holds:

  \[
  \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(\{x_n\}) = \int_{0}^{1} f(x) \, dx.
  \]
Discrepancy

The discrepancy $D_N$ of a finite sequence $x_1, \ldots, x_N$ of real numbers is defined by

$$D_N = D_N(x_1, \ldots, x_N) = \sup_{0 \leq \alpha < \beta \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} 1_{[\alpha, \beta]}(\{x_n\}) - (\beta - \alpha) \right|.$$  

We define $D_N^*$ by

$$D_N^* = D_N^*(x_1, \ldots, x_N) = \sup_{0 < \alpha \leq 1} \left| \frac{1}{N} \sum_{n=1}^{N} 1_{[0, \alpha]}(\{x_n\}) - \alpha \right|.$$  

The discrepancy of a sequence measures how well the sequence is distributed over a unit interval.
Properties of discrepancy

The sequence $\omega$ is u.d. mod 1 if and only if $\lim_{N \to \infty} D_N(\omega) = 0$. 
The sequence $\omega$ is u.d. mod 1 if and only if $\lim_{N \to \infty} D_N(\omega) = 0$.

For any sequence of $N$ numbers we have $1/N \leq D_N \leq 1$. 
Properties of discrepancy

- The sequence $\omega$ is u.d. mod 1 if and only if $\lim_{N \to \infty} D_N(\omega) = 0$.
- For any sequence of $N$ numbers we have $1/N \leq D_N \leq 1$.
- The discrepancy $D_N$ and $D^*_N$ are related by the inequality

$$D^*_N \leq D_N \leq 2D^*_N.$$
Properties of discrepancy

- The sequence $\omega$ is u.d. mod 1 if and only if $\lim_{N \to \infty} D_N(\omega) = 0$.
- For any sequence of $N$ numbers we have $1/N \leq D_N \leq 1$.
- The discrepancy $D_N$ and $D_N^*$ are related by the inequality

\[ D_N^* \leq D_N \leq 2D_N^*. \]

- Explicit formula for $D_N^*$: If $x_1 \leq x_2 \leq \cdots \leq x_N$ are $N$ numbers in $[0, 1)$, $x_0 = 0$, and $x_{N+1} = 1$, then

\[ D_N^* = \max_{j=0,\ldots,N} \max(|x_j - j/N|, |x_{j+1} - j/N|). \]
LeVeque inequality Let $x_1, \ldots, x_N$ be a finite sequence of real numbers. Then

$$D_N \leq \left( \frac{6}{\pi^2} \sum_{h=1}^{\infty} \frac{1}{h^2} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right|^2 \right)^{1/3}.$$
LeVeque inequality Let $x_1, \ldots, x_N$ be a finite sequence of real numbers. Then

$$D_N \leq \left( \frac{6}{\pi^2} \sum_{h=1}^{\infty} \frac{1}{h^2} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \right|^2 \right)^{1/3}.$$

Erdős-Turán inequality Let $m$ be a positive integer and $x_1, \ldots, x_N$ a finite sequence of real numbers, then

$$D_N \leq \frac{6}{m + 1} + \frac{4}{\pi} \sum_{k=1}^{m} \left( \frac{1}{k} - \frac{1}{m + 1} \right) \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i k x_n} \right|. $$
Estimation of discrepancy

- **LeVeque inequality** Let \( x_1, \ldots, x_N \) be a finite sequence of real numbers. Then

\[
D_N \leq \left( \frac{6}{\pi^2} \sum_{h=1}^{\infty} \frac{1}{h^2} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi ihx_n} \right|^2 \right)^{1/3}.
\]

- **Erdös-Turán inequality** Let \( m \) be a positive integer and \( x_1, \ldots, x_N \) a finite sequence of real numbers, then

\[
D_N \leq \frac{6}{m + 1} + \frac{4}{\pi} \sum_{k=1}^{m} \left( \frac{1}{k} - \frac{1}{m + 1} \right) \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi ikx_n} \right|.
\]

- **Corollary of Erdös-Turán inequality** There exists \( C \) such that for all \( m > 0 \) and for any \( x_1, \ldots, x_N \in \mathbb{R} \),

\[
D_N \leq C \left( \frac{1}{m} + \sum_{h=1}^{m} \frac{1}{h} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi ihx_n} \right| \right).
\]
Assume WLOG that $x_1 \leq x_2 \leq \cdots \leq x_N$ are numbers in $[0, 1)$. Set

$$\Delta_N(x) = \frac{1}{N} \sum_{n=1}^{N} 1_{[0,x)}(x_n) - x$$

for $0 \leq x \leq 1$ and extend this function with period 1 to $\mathbb{R}$. Note a typical graph of $\Delta_N(x)$, with $x_1 = 1/8, x_2 = 2.3/8, x_3 = 5.4/8, x_4 = 7.2/8$. 

![Graph of Δ₄(x)](image-url)
Assume that

\[ \int_{0}^{1} \Delta_N(x) = 0. \]
Proof of E-T inequality (continued)

Assume that
\[ \int_0^1 \Delta_N(x) = 0. \]

Set \( S_h = \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i h x_n} \) for \( h \in \mathbb{Z} \). Then by computing the Fourier coefficients of \( \Delta_N(x) \) we have for integer \( h \neq 0 \)
\[ \frac{S_h}{-2\pi i h} = \int_0^1 \Delta_N(x) e^{2\pi i h x} \, dx. \]
Let $m$ be a positive integer and $a$ a real number. Then

$$\sum_{h=-m}^{m}^* (m + 1 - |h|) e^{-2\pi i h a} \frac{S_h}{-2\pi i h}$$

$$= \int_{-a}^{1-a} \Delta_N (x + a) \left( \sum_{h=-m}^{m} (m + 1 - |h|) e^{2\pi i h x} \right) dx,$$

where $\sum^*$ means 0 is not included in the range of the sum.
Proof of E-T inequality (continued)

Let \( m \) be a positive integer and \( a \) a real number. Then

\[
\sum_{h=-m}^{m}^{\ast} (m + 1 - |h|) e^{-2\pi i a} \frac{S_h}{-2\pi i h}
\]

\[
= \int_{-a}^{1-a} \Delta_N(x + a) \left( \sum_{h=-m}^{m} (m + 1 - |h|) e^{2\pi i h x} \right) dx,
\]

where \( \sum^{\ast} \) means 0 is not included in the range of the sum.

Because of the periodicity of the integrand, the last integral can be taken over \([−1/2, 1/2]\).
Proof of E-T inequality (continued)

Note the following discrete Fejér kernel:

\[ \sum_{h=-m}^{m} (m + 1 - |h|) e^{2\pi i h x} = \frac{\sin^2 (m + 1) \pi x}{\sin^2 \pi x}. \]
Proof of E-T inequality (continued)

Note the following discrete Fejér kernel:

\[ \sum_{h=-m}^{m} (m + 1 - |h|) e^{2\pi i h x} = \frac{\sin^2(m + 1) \pi x}{\sin^2 \pi x}. \]

We have

\[ \left| \int_{-1/2}^{1/2} \Delta_N(x + a) \frac{\sin^2(m + 1) \pi x}{\sin^2 \pi x} \, dx \right| \]

\[ \leq \frac{1}{2\pi} \sum_{h=-m}^{m} (m + 1 - |h|) \frac{|S_h|}{|h|} \]

\[ = \frac{1}{\pi} \sum_{h=1}^{m} (m + 1 - h) \frac{|S_h|}{h}. \]
Proof of E-T inequality (continued)

From the nature of the graph of $\Delta_N(x)$ we either have $\Delta_N(b) = -D_N^*$ or $\Delta_N(b + 0) = \lim_{x \to b^+} \Delta_N(x) = D_N^*$ for some $b \in [0, 1]$. We will deal with the second case.
Proof of E-T inequality (continued)

From the nature of the graph of $\Delta_N(x)$ we either have
$\Delta_N(b) = -D_N^*$ or $\Delta_N(b + 0) = \lim_{x \to b^+} \Delta_N(x) = D_N^*$ for some $b \in [0, 1]$. We will deal with the second case.

For $b < t \leq b + D_N^*$, we have

$$\Delta_N(t) = D_N^* + \Delta_N(t) - \Delta_N(b + 0) \geq D_N^* + b - t.$$
Proof of E-T inequality (continued)

From the nature of the graph of $\Delta_N(x)$ we either have
$\Delta_N(b) = -D_N^*$ or $\Delta_N(b + 0) = \lim_{x \to b^+} \Delta_N(x) = D_N^*$ for some
$b \in [0, 1]$. We will deal with the second case.

For $b < t \leq b + D_N^*$, we have

$$\Delta_N(t) = D_N^* + \Delta_N(t) - \Delta_N(b + 0) \geq D_N^* + b - t.$$  

Choose $a = b + \frac{1}{2} D_N^*$. Then for $|x| < \frac{1}{2} D_N^*$,

$$\Delta_N(x + a) \geq D_N^* + b - x - a = \frac{1}{2} D_N^* - x.$$
Proof of E-T inequality (continued)

\[ \Delta_4(x) \]

\[ D_4^* \rightarrow \]

slope \( = \frac{\Delta_4(t) - \Delta_4(b+0)}{t-b} \geq -1 \)

slope \( = -1 \)
Proof of E-T inequality (continued)

Breaking the integral into $\left[ -\frac{1}{2}, -\frac{D_N^*}{2} \right], \left[ -\frac{D_N^*}{2}, \frac{D_N^*}{2} \right], \left[ \frac{D_N^*}{2}, \frac{1}{2} \right]$ and using the symmetry argument and the inequality

$$\sin x \geq \frac{2}{\pi} x,$$

for $x \in [0, \pi/2]$, we obtain –
Proof of E-T inequality (continued)

\[
\left( \int_{-D_N/2}^{D_N/2} + \int_{-D_N/2}^{1/2} + \int_{D_N/2}^{1/2} \right) \Delta_N(x + a) \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx \\
\geq \int_{-D_N/2}^{D_N/2} \left( \frac{D_N}{2} - x \right) \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx - D_N^* \int_{-1/2}^{D_N/2} \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx \\
- D_N^* \int_{D_N}^{1/2} \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx \\
= D_N^* \int_{0}^{D_N/2} \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx - 2D_N^* \int_{D_N/2}^{1/2} \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx \\
= D_N^* \int_{0}^{1/2} \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx - 3D_N^* \int_{D_N/2}^{1/2} \frac{\sin^2(m + 1)\pi x}{\sin^2 \pi x} \, dx \\
\geq \frac{m + 1}{2} D_N^* - 3D_N^* \int_{D_N/2}^{1/2} \frac{dx}{4x^2} > \frac{m + 1}{2} D_N^* - \frac{3}{2}.
\]
Proof of E-T inequality (continued)

\[ D_N^* \leq \frac{3}{m + 1} + \frac{2}{\pi} \sum_{h=1}^{m} \left( \frac{1}{h} - \frac{1}{m + 1} \right) |S_h|. \]

Since \( D_N \leq 2D_N^* \), we have

\[ D_N \leq \frac{6}{m + 1} + \frac{4}{\pi} \sum_{h=1}^{m} \left( \frac{1}{h} - \frac{1}{m + 1} \right) |S_h|. \]
\[ D_N^* \leq \frac{3}{m + 1} + \frac{2}{\pi} \sum_{h=1}^{m} \left( \frac{1}{h} - \frac{1}{m + 1} \right) |S_h|. \]

Since \( D_N \leq 2D_N^* \), we have
\[ D_N \leq \frac{6}{m + 1} + \frac{4}{\pi} \sum_{h=1}^{m} \left( \frac{1}{h} - \frac{1}{m + 1} \right) |S_h|. \]

Show that for any finite sequence \( x_1, \ldots, x_N \) in \([0, 1)\), there exists \( c \in [0, 1) \) such that \( \{x_1 + c\}, \ldots, \{x_N + c\} \) satisfies
\[ \int_0^1 \Delta_N(x)dx = 0. \]

This completes the proof of the Erdös-Turán inequality.
Koksma inequality

**Theorem** Let $f$ be a function of bounded variation on $[0, 1]$ with total variation $\text{Var}(f)$ and let $x_1, \ldots, x_N$ be given points in $[0, 1)$ with discrepancy $D_N$. Then

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(x_n) - \int_{0}^{1} f(t) dt \right| \leq \text{Var}(f) \cdot D_N.$$ 

**Proof**

- Assume that $x_1 \leq x_2 \leq \cdots \leq x_N$ are points in $[0, 1)$. Set $x_0 = 0$ and $x_{N+1} = 1$.

- Riemann-Stieltjes integration by parts gives –

$$\frac{1}{N} \sum_{n=1}^{N} f(x_n) - \int_{0}^{1} f(t) dt = \sum_{n=0}^{N} \int_{x_n}^{x_{n+1}} (t - \frac{n}{N}) df(t).$$
Proof of K inequality (continued)

For a fixed $n$ with $0 \leq n \leq N$ and for $x_n \leq t \leq x_{n+1}$, one can prove

$$\left| t - \frac{n}{N} \right| \leq \max \left( \left| x_n - \frac{n}{N} \right|, \left| x_{n+1} - \frac{n}{N} \right| \right) \leq D_N^*.$$ 

Since $D_N^* \leq D_N$, the Koksma inequality follows by elementary properties of the Riemann-Stieltjes integral.
Additional results and definitions

- **Exponential sum theorem**  Let $a$ and $b$ be integers with $a < b$, and let $f$ be twice differentiable on $[a, b]$ with $|f''(x)| \geq \rho > 0$ for $x \in [a, b]$. Then

  \[
  \left| \sum_{n=a}^{b} e^{2\pi if(n)} \right| \leq (|f'(b) - f'(a)| + 2) \left( \frac{4}{\sqrt{\rho}} + 3 \right).
  \]

- The class of **bandlimited functions** $\mathcal{B}_\Omega$ consists of all real valued functions $f \in L^\infty(\mathbb{R})$ whose $\widehat{f}$ (as a distribution) is a finite Borel measure supported on $[-\Omega, \Omega]$.

- By the Paley-Wiener theorem, each element of $\mathcal{B}_\Omega$ is the restriction of an entire function to the real line.

- A function $h \in \mathcal{B}_\Omega$ is said to belong to the class $\mathcal{M}_\Omega$ if $h' \in L^\infty(\mathbb{R})$ and all zeros of $h'$ on $[0, 1]$ are simple.
Bernstein inequality If $f \in \mathcal{B}_\Omega$ then for all integers $r \geq 0$,

$$\|f^{(r)}\|_\infty \leq \Omega^r \|f\|_\infty.$$
**Bernstein inequality** If $f \in B_\Omega$ then for all integers $r \geq 0$,

$$\|f^{(r)}\|_\infty \leq \Omega^r \|f\|_\infty.$$  

**Theorem (Güntürk)** Let $f \in B_\Omega$. Then for each $\lambda > 1$, there exists an analytic function $X_\lambda \in B_\Omega$ such that for all $t \in \mathbb{R}$,

$$X_\lambda(t) - X_\lambda(t - 1) = f\left(\frac{t}{\lambda}\right)$$

and there exists a constant $C_\lambda(f) > 0$ such that

$$\|X_\lambda' - f\left(\frac{.}{\lambda}\right)\|_\infty \leq C_\lambda(f) \frac{1}{\lambda}.$$
Theorem

Let $\{F_N\}_N^\infty$ be a family of unit norm tight frames for $\mathbb{R}^d$, with $F_N = \{e_n^N\}_{n=1}^N$. Suppose $x \in \mathbb{R}^d$ satisfies $\|x\| \leq (K - 1/2)\delta$ for some positive integer $K$ and $\delta > 0$ in the $\Sigma\Delta$ scheme. Let $\{x_n^N\}_{n=1}^N$ be the sequence of frame coefficients of $x$ with respect to $F_N$. If, for some $\Omega > 0$, there exists $h \in \mathcal{M}_\Omega$ such that

$$\forall N \text{ and } 1 \leq n \leq N, \quad x_n^N = h(n/N),$$

and if $N$ is sufficiently large, then

$$|v_n^N| \leq C(x)\delta N^{3/4} \log N,$$

where $v_n^N = \sum_{j=1}^n u_j^N$ and $C(x)$ is a constant that depends on $x$. 
Proof of theorem

For each $N \geq d$, let $u^N_n$, for $n = 1, \ldots, N$, be the state variables obtained from the $\Sigma\Delta$ scheme and define

$$\tilde{u}^N_n = \frac{u^N_n}{\delta}.$$

Set $u^N_0 = 0$. 
Proof of theorem

For each $N \geq d$, let $u_n^N$, for $n = 1, \ldots, N$, be the state variables obtained from the $\Sigma\Delta$ scheme and define

$$\tilde{u}_n^N = \frac{u_n^N}{\delta}.$$  

Set $u_0^N = 0$.

Then, by the stability result,

$$|\tilde{u}_n^N| \leq \frac{1}{2}.$$  

Finite frames and Sigma-Delta quantization – p.52/64
Proof of theorem

For each $N \geq d$, let $u^N_n$, for $n = 1, \ldots, N$, be the state variables obtained from the $\Sigma\Delta$ scheme and define

$$\tilde{u}^N_n = \frac{u^N_n}{\delta}.$$  

Set $u^N_0 = 0$.

Then, by the stability result,

$$|\tilde{u}^N_n| \leq \frac{1}{2}.$$  

Identify the interval $[-1/2, 1/2]$ with $[0, 1]$ and apply Koksma’s inequality:
Proof of theorem (continued)

\[ |v_n^N| = \delta \left| \sum_{j=1}^{n} \tilde{u}_j^N \right| = n\delta \left| \frac{1}{n} \sum_{j=1}^{n} \tilde{u}_j^N - \int_{-1/2}^{1/2} y \, dy \right| \]
\[ \leq n\delta \text{Var}(x) D_n(\tilde{u}_1^N, \ldots, \tilde{u}_n^N) = n\delta D_n(\tilde{u}_1^N, \ldots, \tilde{u}_n^N) \]

Next apply the Erdös-Turán inequality: There exists a constant \( C > 0 \) such that for all \( K \in \mathbb{N} \),

\[ D_n(\tilde{u}_1^N, \ldots, \tilde{u}_n^N) \leq C \left( \frac{1}{K} + \frac{1}{n} \sum_{k=1}^{K} \frac{1}{k} \left| \sum_{j=1}^{n} e^{2\pi ik\tilde{u}_j^N} \right| \right). \]
Using the Güntürk’s theorem, we have that for each integer $N \geq d$ there exists an analytic function $X_N \in B_\Omega$ such that for all $1 \leq n \leq N$,

$$X_N(n) = u_n^N + c_n \delta/2,$$

for some $c_n \in \mathbb{Z}$. Each $c_n$ has the same parity as $n$. Moreover, for all $t \in \mathbb{R}$,

$$|X'_N(t) - h\left(\frac{t}{N}\right)| \leq \frac{C_N(x)}{N},$$

for some constant $C_N(x) > 0$.

The Bernstein inequality yields

$$|X''_N(t) - \frac{1}{N} h'\left(\frac{t}{N}\right)| \leq \frac{C_N(x)}{N^2},$$

for some constant $C_N(x) > 0$. 
Proof of theorem (continued)

Let $z_1 < z_2 < \cdots < z_{n^*}$ be zeros of $h'$ in $[0, 1]$. Let $0 < \alpha < 1$ be a fixed constant to be determined later. Define intervals $I_j$ and $J_j$ by

$$\forall j = 1, \ldots, n^*, I_j = [Nz_j - N^\alpha, Nz_j + N^\alpha]$$

$$\forall j = 1, \ldots, n^* - 1, J_j = [Nz_j + N^\alpha, Nz_{j+1} - N^\alpha]$$

$$J_0 = [1, Nz_1 - N^\alpha] \text{ and } J_{n^*} = [Nz_{n^*} + N^\alpha, N].$$

Adjust intervals appropriately when $z_1 = 0$ or $z_{n^*} = 1$. 
Proof of theorem (continued)

Let \( z_1 < z_2 < \cdots < z_{n^*} \) be zeros of \( h' \) in \([0, 1] \). Let \( 0 < \alpha < 1 \) be a fixed constant to be determined later. Define intervals \( I_j \) and \( J_j \) by

\[
\forall j = 1, \ldots, n^*, I_j = [Nz_j - N^\alpha, Nz_j + N^\alpha]
\]

\[
\forall j = 1, \ldots, n^* - 1, J_j = [Nz_j + N^\alpha, Nz_{j+1} - N^\alpha]
\]

\[
J_0 = [1, Nz_1 - N^\alpha] \text{ and } J_{n^*} = [Nz_{n^*} + N^\alpha, N].
\]

Adjust intervals appropriately when \( z_1 = 0 \) or \( z_{n^*} = 1 \).

When \( N \) is sufficiently large then

\[
J_0 \cup I_1 \cup J_1 \cup \cdots \cup I_{n^*} \cup J_{n^*} = [1, N].
\]
Proof of theorem (continued)

\begin{align*}
J_0 & \quad I_1 & \quad J_1 & \quad \ldots & \quad I_{n^*} & \quad J_{n^*} \\
Nz_1 + N^\alpha & \quad Nz_{n^*} - N^\alpha \\
1 & \quad \uparrow & \quad \uparrow & \quad \uparrow & \quad \uparrow & \quad N \\
Nz_1 - N^\alpha & \quad Nz_2 - N^\alpha & \quad Nz_{n^*} + N^\alpha
\end{align*}
Proof of theorem (continued)

It follows from the properties of $h \in \mathcal{M}_\Omega$ that there exist constants $C_1(x), C_2(x)$ and $C_3(x)$ such that for sufficiently large $N$ and for all $t \in J_0 \cup \cdots \cup J_{n^*}$,

$$|h'(\frac{t}{N})| \geq C_1(x) \frac{1}{N^{1-\alpha}},$$

$$|X''(t)| \geq C_2(x) \frac{1}{N^{2-\alpha}}, \quad |X_N'(t)| \leq C_3(x).$$

By the exponential sum theorem, we have that there exists a constant $C'(x) > 0$ and an index $N_0$ such that for all $N \geq N_0$ and $k \geq 1$

$$\sum_{j=0}^{n^*} \left| \sum_{n \in \mathbb{N} \cap J_j} e^{2\pi ik\mathcal{u}_n^N} \right| \leq C'(x) \left( \frac{\sqrt{k}}{\sqrt{\delta}} N^{1-\alpha/2} + \frac{k}{\delta} \right).$$
Proof of theorem (continued)

- For all $N \geq d$ and $k \geq 1$,

$$\sum_{j=1}^{n^*} \left| \sum_{n \in \mathbb{N} \cap I_j} e^{2\pi i k \tilde{u}_n^N} \right| \leq 2n^* N^\alpha.$$

- There exists a constant $C(x) > 0$ such that for all $k \geq 1$ and $N \geq N_0$,

$$\left| \sum_{j=1}^{n} e^{2\pi i k \tilde{u}_j^N} \right| \leq C(x) (N^\alpha + \frac{\sqrt{k} N^{1-\alpha/2}}{\sqrt{\delta}} + \frac{k}{\delta}).$$

- Now choose $\alpha = 3/4$ and $K$, an integer larger than $N_0^{1/4}$ and note that

$$\sum_{k=1}^{n} \frac{1}{k} \leq 1 + \log n, \quad \sum_{k=1}^{n} \frac{1}{\sqrt{k}} \leq 2\sqrt{n}.$$
Proof of theorem (continued)

Then there exist constants $C_1(x), C_2(x), C_3(x) > 0$ such that for sufficiently large $N$

\[ |v_n^N| \leq n\delta C_1(x) \left( \frac{1}{K} + \frac{1}{n} \sum_{k=1}^{K} \left( \frac{N^{3/4}}{k} + \frac{N^{1-3/8}}{\sqrt{\delta}} \cdot \frac{1}{\sqrt{k}} + \frac{1}{\delta} \right) \right) \]

\[ \leq n\delta C_2(x) \left( \frac{1}{N^{1/4}} + \frac{N^{3/4}}{n} \log N + \frac{2N^{3/4}}{n\sqrt{\delta}} + \frac{N^{1/4}}{n\delta} \right) \]

\[ \leq C_2(x) \left( \frac{\delta n}{N^{1/4}} + \delta N^{3/4} \log N + \sqrt{\delta} N^{3/4} + N^{1/4} \right) \]

\[ \leq C_3(x) \delta N^{3/4} \log N. \]

This completes the proof of the theorem.
\{F_N\}_{N=d}^{\infty} be a family of unit norm tight frames for \(\mathbb{R}^d\), for which each \(F_N = \{e_n^N\}_{n=1}^{N}\) satisfies the zero sum condition. Suppose \(x \in \mathbb{R}^d\) satisfies \(\|x\| \leq (K - 1/2)\delta\) for some positive integer \(K\) and \(\delta > 0\) in the \(\Sigma\Delta\) scheme. Let \(\{x_n^N\}_{n=1}^{N}\) be the sequence of frame coefficients of \(x\) with respect to \(F_N\), and suppose there exists \(h \in \mathcal{M}_\Omega, \Omega > 0\), such that

\[
\forall N \text{ and } 1 \leq n \leq N, \quad x_n^N = h(n/N).
\]

Additionally, suppose that \(f_n^N = e_n^N - e_{n+1}^N, n = 1, \ldots, N\), satisfies \(\|f_n^N\| \leq \frac{c_1}{N}\) for some constant \(c_1 > 0\) and \(\|f_n^N - f_{n+1}^N\| \leq \frac{c_2}{N^2}\), for some constant \(c_2 > 0\), and set \(u_0^N = 0\) in the \(\Sigma\Delta\) scheme. Then –
Corollary – conclusion

If \( N \) is even and sufficiently large, then

\[
\|x - \tilde{x}_N\| \leq C_1(x) \frac{\delta \log N}{N^{5/4}},
\]

for some constant \( C_1(x) > 0 \).

If \( N \) is odd and sufficiently large, then

\[
C_2(x) \frac{\delta}{N} \leq \|x - \tilde{x}_N\| \leq \frac{\delta d}{2N} (\sigma(F_N, p_N) + 1),
\]

for some constant \( C_2(x) > 0 \).
Proof of corollary

\[ x - \tilde{x}_N = \frac{d}{N} \left( \sum_{n=1}^{N-1} u_n^N (e_n^N - e_{n+1}^N) + u_N^N e_N^N \right) \]

\[ = \frac{d}{N} \left( \sum_{n=1}^{N-2} v_n^N (f_n^N - f_{n+1}^N) + v_{N-1}^N f_{N-1}^N + u_N^N e_N^N \right), \]

where \( f_n^N = e_n^N - e_{n+1}^N, v_n^N = \sum_{j=1}^{n} e_j^N, v_0^N = 0. \)
Proof of corollary

\[ x - \tilde{x}_N = \frac{d}{N} \left( \sum_{n=1}^{N-1} u_n^N (e_n^N - e_{n+1}^N) + u_N^N e_N^N \right) \]

\[ = \frac{d}{N} \left( \sum_{n=1}^{N-2} v_n^N (f_n^N - f_{n+1}^N) + v_{N-1}^N f_{N-1}^N + u_N^N e_N^N \right), \]

where \( f_n^N = e_n^N - e_{n+1}^N, v_n^N = \sum_{j=1}^{n} e_j^N, v_0^N = 0. \)

By the Theorem, we have

\[ \left\| \frac{d}{N} \left( \sum_{n=1}^{N-2} v_n^N (f_n^N - f_{n+1}^N) + v_{N-1}^N f_{N-1}^N \right) \right\| \leq C_1(x) \frac{\delta \log N}{N^{5/4}}, \]

for some constant \( C_1(x) > 0. \)
Proof of corollary (continued)

If \( N \) is even then \( u_N^N = 0 \) and so

\[
\|x - \tilde{x}_N\| \leq C_1(x) \frac{\delta \log N}{N^{5/4}},
\]

for some constant \( C_1(x) > 0 \) and for sufficiently large \( N \).

If \( N \) is odd then \( |u_N^N| = \delta/2 \) and we have, combining with previous results,

\[
C_2(x) \frac{\delta}{N} \leq \|x - \tilde{x}_N\| \leq \frac{\delta d}{2N} (\sigma(F_N, p_N) + 1),
\]

for some constant \( C_2(x) > 0 \) and for sufficiently large \( N \).

This completes the proof of the Corollary.
That's all folks!

Norbert Wiener Center