Data Compression at High Sampling Rates

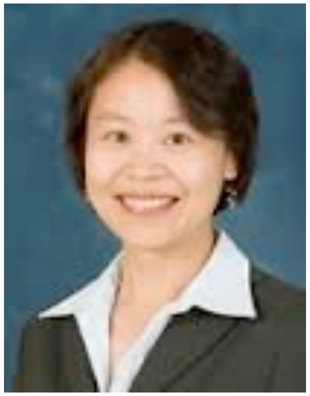
David L. Neuhoff

CSP Seminar

ECE, University of Michigan

Nov. 7, 2013

Collaborators

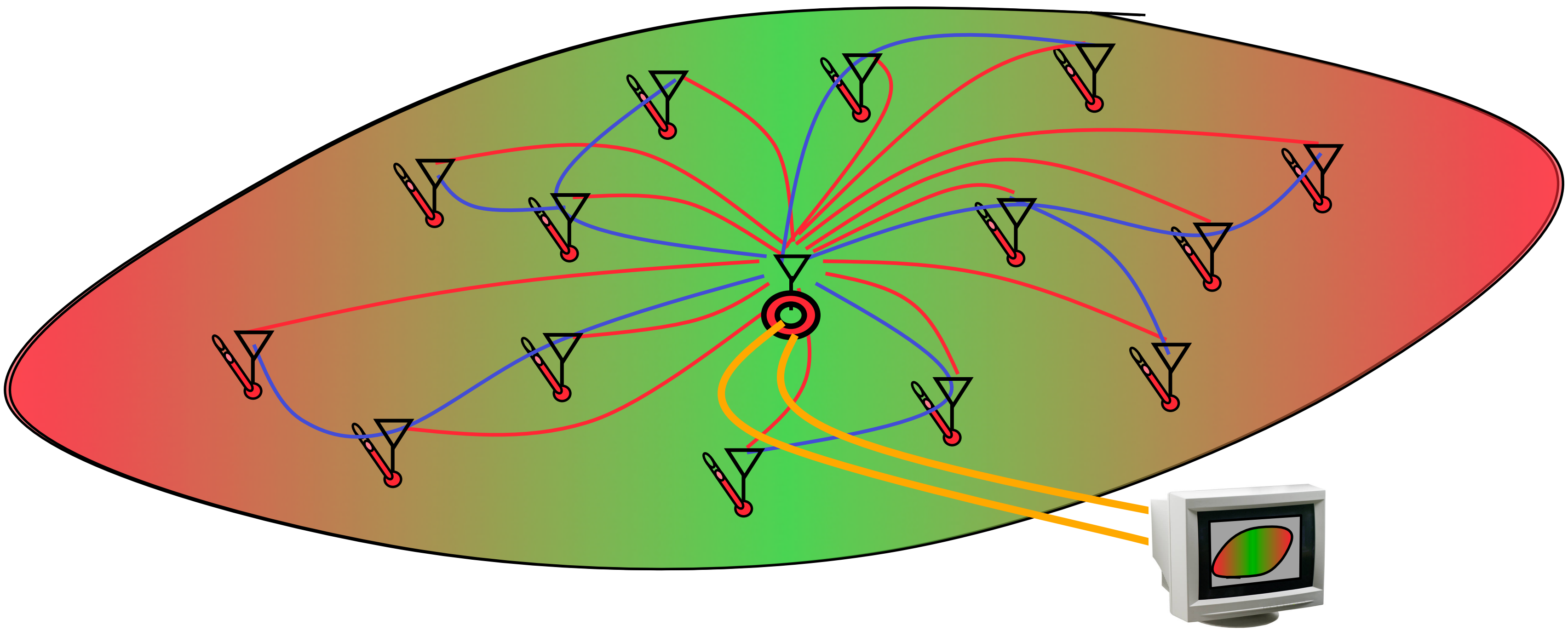


Mingyan Liu

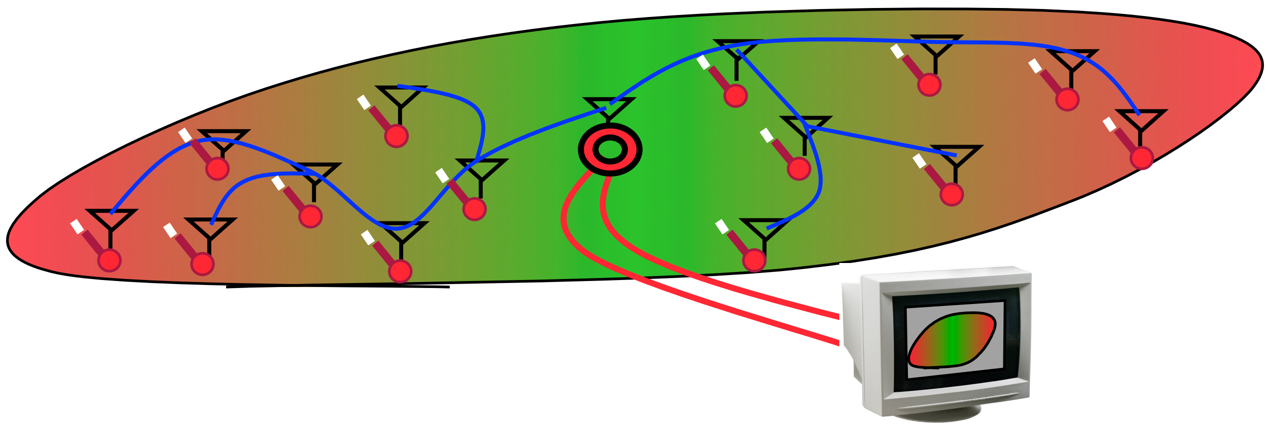
Daniel Marco

Sandeep Pradhan

Field gathering:
Sampling, Encoding, Transporting, Reconstructing

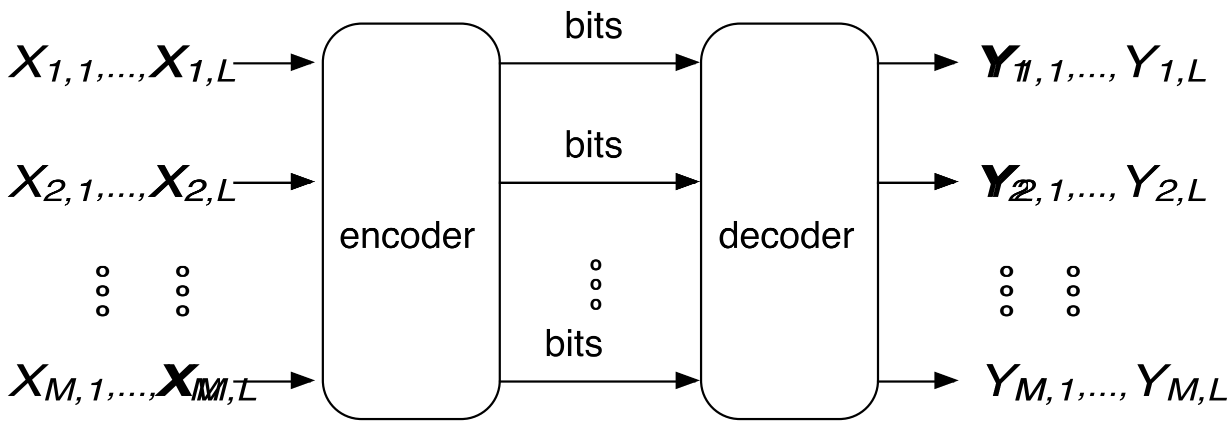


Field-Gathering Wireless Sensor Network

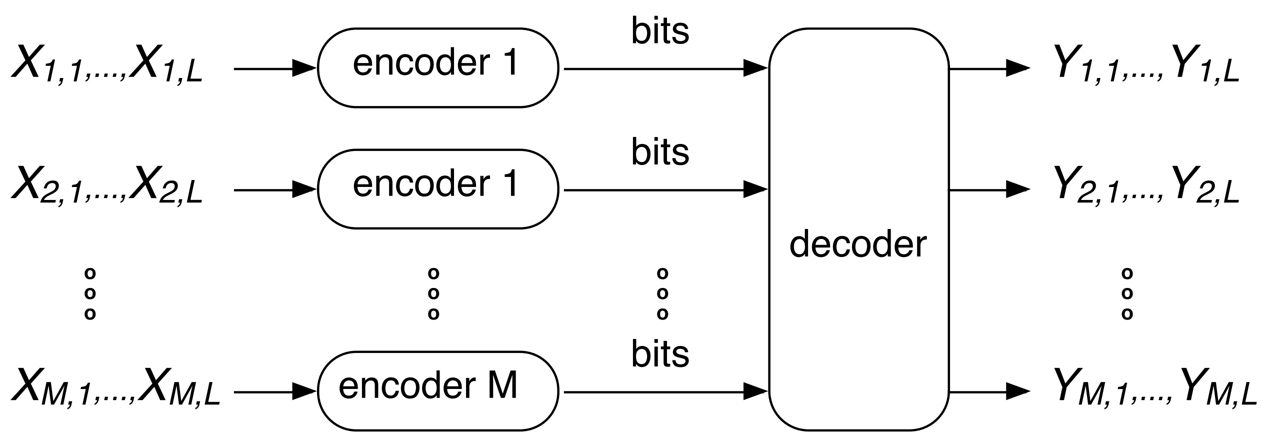


- Sensors sample a field in two-dim'l region at discrete sequence of times.
- Each sensor source encodes its time-sequence of samples.
  This requires distributed lossy source coding.
- Communication network conveys bits to collector.
- Decoder at collector reconstructs snapshots of field (not just at sensor locations).
- We focus here on performance of source coding, not communication network.
- Competing Goals: minimize
  - rate = avg. number of bits per unit area per snapshot
  - MSE distortion, integrated over entire region

Centralized Coding



Distributed Coding



Principal Goals and Questions

- Goal: design encoders and decoder to minimize coding rate subject to MSE distortion being at most target value *d.*

  coding rate = bits/unit-area/time step

  = sampling rate x coding rate per sample

  (sensors/unit area) x (bits/sensor/time step)

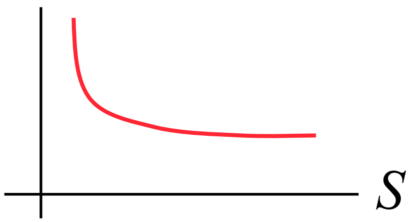
- For a given random field model *X*, class of coding schemes *C*, sampling rate *S*, and target distortion *d*, the coding rate per sample can be as small as the *operational rate-distortion function* $R_{X,C,S}(d)$.

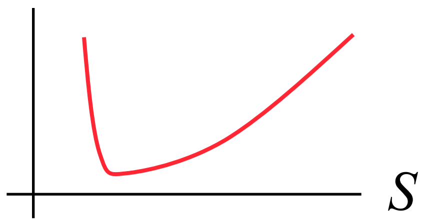
- Hence, given *X*, *C*, *S*, *d,* coding rate can be as small as

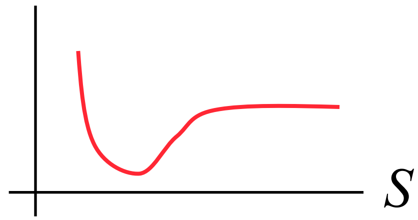
  $$S \times R_{X,C,S}(d)$$

- Goal: For different code classes *C,* find limit of $S \times R_{X,C,S}(d)$ for large *S.*

- Question: Like which of the following does $S \times R_{X,C,S}(d)$ behave?







Simplify to 1-dimensional, continuous-time signals

- Not so much theory is known for source coding for continuous-time sources, even in 1-dimension

Four Classes of Lossy Source Codes to be Used with High-Rate Sampling

We'll analyze the following classes for stationary, Gaussian, continuous-time sources

- Transform + VQ
  (centralized & optimal)
- Scalar quantization + entropy-rate coding:
  (centralized or distributed, no transform, suboptimal)
- Distributed VQ:
  (no transform, suboptimal)
- Transform + scalar quantization:
  (centralized, suboptimal)

Ideas Leading to a Conjecture

- For continuous-time source
  - Best performance of lossy source codes is given by Shannon rate-distortion function $R_{sh}(d)$.
  - For Gaussian source, a parametric expression for $R_{sh}(d)$ is known (Kolmogorov).
  - Performance as good as $R_{sh}(d)$ can be attained (to within ε) by sampling at very high rate, coding samples, and reconstructing cont.-time signal.
- For discrete-time (sampled) source
  - Best rate-distortion performance with distributed coding is not known, except for two Gaussian sources.
  - Uniform scalar quantization plus entropy-rate coding has performance close to $R_{sh}(d)$ for any discrete-time source.
  - Entropy-rate coding can be done in distributed fashion with as small rate as centralized coding. (Slepian-Wolf coding)
- Conjecture: uniform scalar quantization + distributed entropy-rate coding is distributed coding system that works well for cont-time source with high sampling rate (performance might be close to optimal).

  (Idea: entropy-rate coding will exploit strong sample dependences to mitigate large sampling rate.)

Ideas Leading to a Conjecture

- For continuous-time source
  - Best performance of lossy source codes is given by Shannon rate-distortion function $R_{sh}(d)$.
  - For Gaussian source, a parametric expression for $R_{sh}(d)$ is known (Kolmogorov)
  - Performance as good as $R_{sh}(d)$ can be attained (to within ε) by sampling at very high rate, coding samples, and reconstructing cont.-time signal.
- For discrete-time (sampled) source
  - Best rate-distortion performance with distributed coding is not known, except for two Gaussian sources.
  - Uniform scalar quantization plus entropy-rate coding has performance close to $R_{sh}(d)$ for any discrete-time source.
  - Entropy-rate coding can be done in distributed fashion with as small rate as centralized coding. (Slepian-Wolf coding)
- Conjecture: uniform scalar quantization + distributed entropy-rate coding is distributed coding system that works well for cont-time source with high sampling rate (performance might be close to optimal).

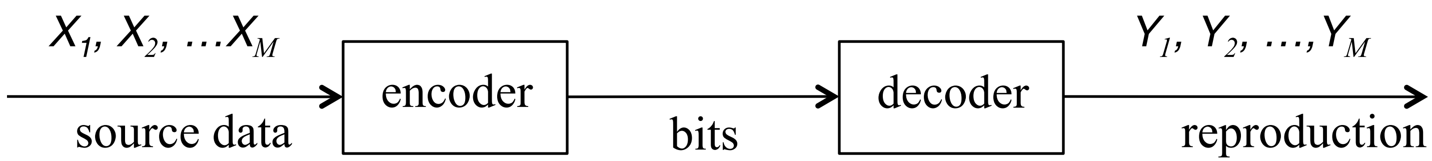
  (Idea: entropy-rate coding will exploit strong sample dependences to mitigate large sampling rate.)

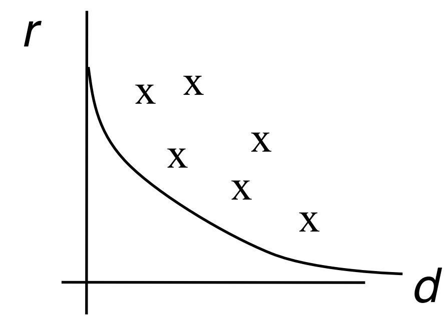
Four Strategies For Lossy Source Coding Based on High-Rate Sampling

We'll analyze the following for stationary, Gaussian, continuous-time sources

- Transform + VQ
  (centralized & optimal)
- Scalar quantization + entropy-rate coding:
  (centralized or distributed, no transform, suboptimal)
- Distributed VQ:
  (no transform, suboptimal)
- Transform + scalar quantization:
  (centralized, suboptimal)

Review: Lossy Source Coding in Discrete-Time





- Code: c = encoder & decoder (e.g. block code)
- Performance = rate & MSE distortion
  - $R(c)$ = # bits/sample
  - $D(c) = \frac{1}{M}\sum_{i=1}^{M} E(X_i - Y_i)^2$
- Operational rate-distortion function ORDF for class of codes C
  - $R_C(d)$ = least rate of any c in C with $D(c) \le d$
- Shannon rate-distortion theory: if C includes all block codes with all blocklengths

$$R_C(d) \;=\; R_{sh}(d) \;\triangleq\; \lim_{N\to\infty} \inf_{p(\underline{y}|\underline{x})} \frac{1}{N} I(\underline{X};\underline{Y}) \;\triangleq\; \text{Shannon rate-dist'n func.}$$

- Example: IID Gaussian $R_{sh}(d) = \max\left\{\frac{1}{2}\log\frac{\sigma^2}{d}, 0\right\}$
- High resolution theory: when d small $R_C(d) \approx \frac{1}{2}\log\frac{\sigma^2}{d} + \eta_{C,X}$

where $\eta_{C,X}$ depends on class of codes and source statistics [Bennett, Zador].

ORDF: Block Codes & DT Gaussian Source

- With orthogonal transform

$$R(c) = \frac{1}{M}\sum_{i=1}^{M} R(c_i), \quad D(c) = \frac{1}{M}\sum_{i=1}^{M} D(c_i)$$

- Using optimal codes with distortions d_i

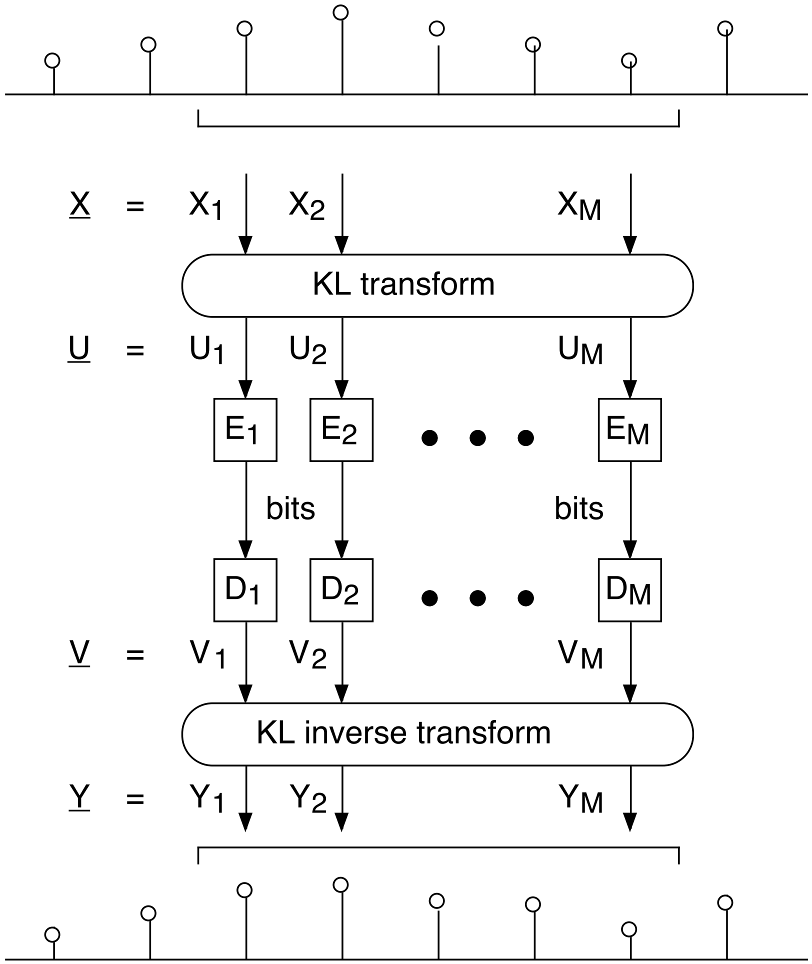
$$R(c) = \frac{1}{M}\sum_{i=1}^{M} \max\left\{\frac{1}{2}\log\frac{\lambda_i}{d_i}, 0\right\}$$

where λ_i = e. val. of cov. matrix K_M of $\underline{X}$

- Minimize over $d_1,\ldots,d_M \geq 0$ s.t. $\frac{1}{M}\sum_{i=1}^{M} d_i \leq d$ to find there exists d' s.t.

$$d_i = \min\{d', \lambda_i\}, \text{ all } i,$$

$$R(d) = \frac{1}{M}\sum_{i=1}^{M} \max\left\{\frac{1}{2}\log\frac{\lambda_i}{d'}, 0\right\}$$



ORDF: Block Codes & DT Gaussian Source

- With orthogonal transform

$$R(c) = \frac{1}{M}\sum_{i=1}^{M} R(c_i), \quad D(c) = \frac{1}{M}\sum_{i=1}^{M} D(c_i)$$

- Using optimal codes with distortions d_i

$$R(c) = \frac{1}{M}\sum_{i=1}^{M} \max\left\{\frac{1}{2}\log\frac{\lambda_i}{d_i}, 0\right\}$$

where λ_i = e. val. of cov. matrix K_M of $\underline{X}$

- Minimize over $d_1,\ldots,d_M \geq 0$ s.t. $\frac{1}{M}\sum_{i=1}^{M} d_i \leq d$ to find there exists d' *s.t.*

$$d_i = \min\{d', \lambda_i\}, \text{ all } i,$$

$$R(d) = \frac{1}{M}\sum_{i=1}^{M} \max\left\{\frac{1}{2}\log\frac{\lambda_i}{d'}, 0\right\}$$

- Convenient parametric form: $\theta \geq 0$

$$\hat{d}_M(\theta) = \frac{1}{M}\sum_{i=1}^{M} \min\{\lambda_i, \theta\}$$

$$\hat{R}_M(\theta) = \frac{1}{M}\sum_{i=1}^{M} \max\left\{\frac{1}{2}\log\frac{\lambda_i}{\theta}, 0\right\}$$

- Take $M \to \infty$, using asympt. e. val. dist'n thm:

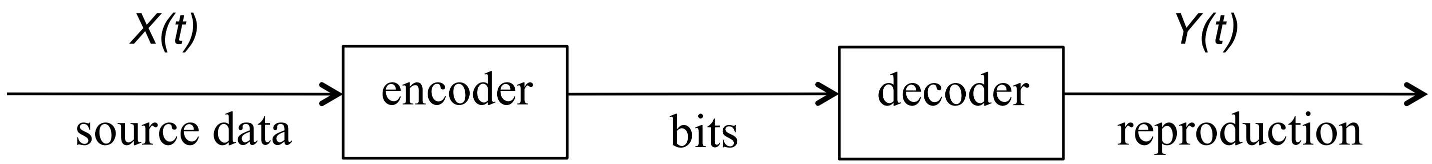
$$\hat{d}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \min\{\Phi(\Omega), \theta\}\, d\Omega$$

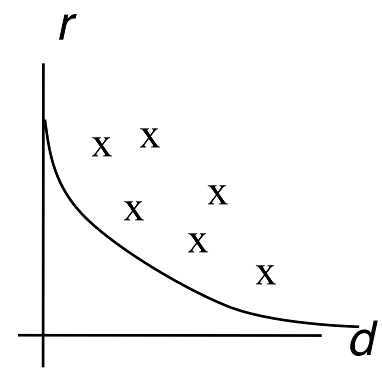
$$\hat{r}(\theta) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \max\left\{\frac{1}{2}\log\frac{\Phi(\Omega)}{\theta}, 0\right\} d\Omega$$

where $\Phi(\Omega)$ is power spect. density of discrete-time process X.

[Kolmogorov `56]

Review: Lossy Source Coding in Continuous-Time





- Code: c = encoder & decoder (e.g. block code)
- Performance = rate & MSE distortion
  - $R(c)$ = # bits/second
  - $D(c) = \frac{1}{T}\int_0^T E(X(t)-Y(t))^2$
- Operational rate-distortion function ORDF for class of codes C
  - $R_C(d)$ = least rate of any c in C with $D(c) \le d$

ORDF: CT Gaussian Source

- Sample at N samples/sec
- When N large,

  $D \approx$ distortion in decoded samples

  $R = R(c) \times N$ bits/sec
- Take limit as $N \rightarrow \infty$ of

$$\hat{d}(\theta, N) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \min \{\Phi_N(\Omega), \theta\} \, d\Omega$$

$$\hat{r}(\theta, N) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \max \left\{ \frac{1}{2} \log \frac{\Phi_N(\Omega)}{\theta}, 0 \right\} d\Omega$$

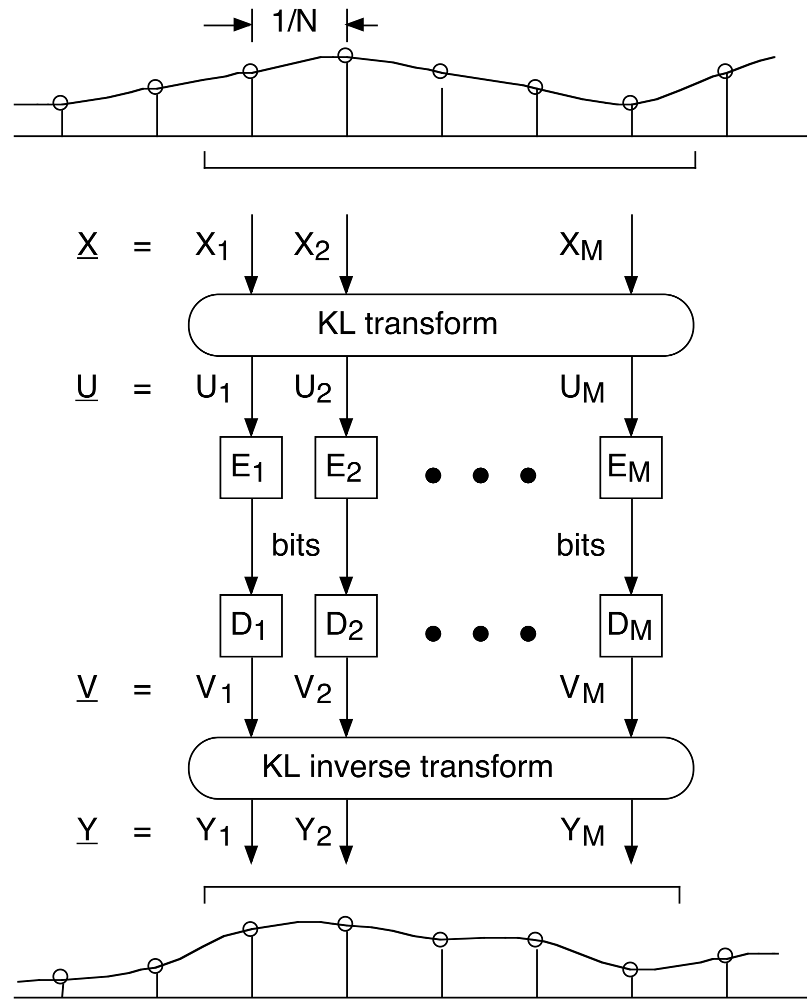
- Change variables and use

  $$N\Phi_N(N\omega) \to S(\omega) \text{ as } N \to \infty$$

  where $S(\omega)$ is power spectral density of $X(t)$.
- To get *inverse water pouring formulas:*

$$\mathcal{D}(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \min \{S(\omega), \theta\} \, d\omega$$

$$\mathcal{R}(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \max \left\{ \frac{1}{2} \log \frac{S(\omega)}{\theta}, 0 \right\} d\omega$$

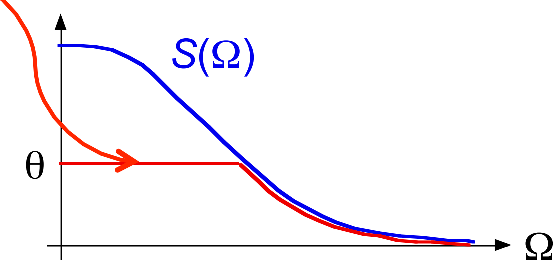


[Kolmogorov `56, Berger `71]

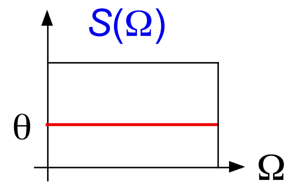
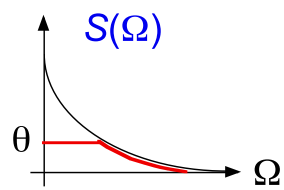
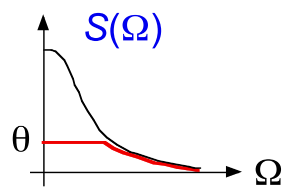
Summary and Interpretaton

$$\mathcal{D}(\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\left\{S(\omega), \theta\right\} d\omega \qquad \mathcal{R}(\theta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \max\left\{\frac{1}{2}\log\frac{S(\omega)}{\theta}, 0\right\} d\omega$$

distortion profile



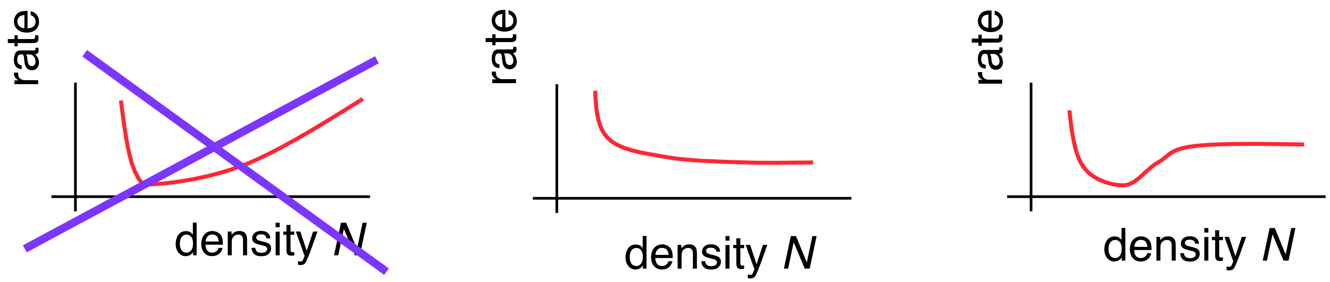
Examples

	$S(\omega)$		$\mathcal{R}(d)$
bandlimited			$\frac{1}{2}\log_2 \frac{1}{d} + c$
exponential spectrum	$e^{-\lvert\omega\rvert}$		$c\left(c \ln \frac{1}{d} + \ln \ln \frac{c}{d} + c + o(1)\right)^2$
Gauss-Markov	$\frac{2}{\omega^2 + 1}$		$\frac{c}{d}$

- Heavier tailed spectrum $\Longrightarrow$ larger $\mathcal{R}(d)$ at small d.

Summary

- For coding with transform and vector quantization.



- Probably the middle one.

Ideas Leading to a Conjecture

- For continuous-time source
  - Best performance of lossy source codes is given by Shannon rate-distortion function $R_{sh}(d)$.
  - For Gaussian source, a parametric expression for $R_{sh}(d)$ is known (Kolmogorov).
  - Performance as good as $R_{sh}(d)$ can be attained (to within ε) by sampling at very high rate, coding samples, and reconstructing cont.-time signal.
- For discrete-time (sampled) source
  - Best rate-distortion performance with distributed coding is not known, except for two Gaussian sources.
  - Uniform scalar quantization plus entropy-rate coding has performance close to $R_{sh}(d)$ for any discrete-time source.
  - Entropy-rate coding can be done in distributed fashion with as small rate as centralized coding. (Slepian-Wolf coding)
- Conjecture: uniform scalar quantization + distributed entropy-rate coding is distributed coding system that works well for cont-time source with large sampling rate (performance might be close to optimal).

  (Idea: entropy-rate coding will exploit strong sample dependence to mitigate large sampling rate.)

Four Strategies For Lossy Source Coding Based on High-Rate Sampling

We'll analyze the following for stationary, Gaussian, continuous-time sources

- Transform + VQ
  (centralized & optimal)
- Scalar quantization + entropy-rate coding:
  (centralized or distributed, no transform, suboptimal)
- Distributed VQ:
  (no transform, suboptimal)
- Transform + scalar quantization:
  (centralized, suboptimal)

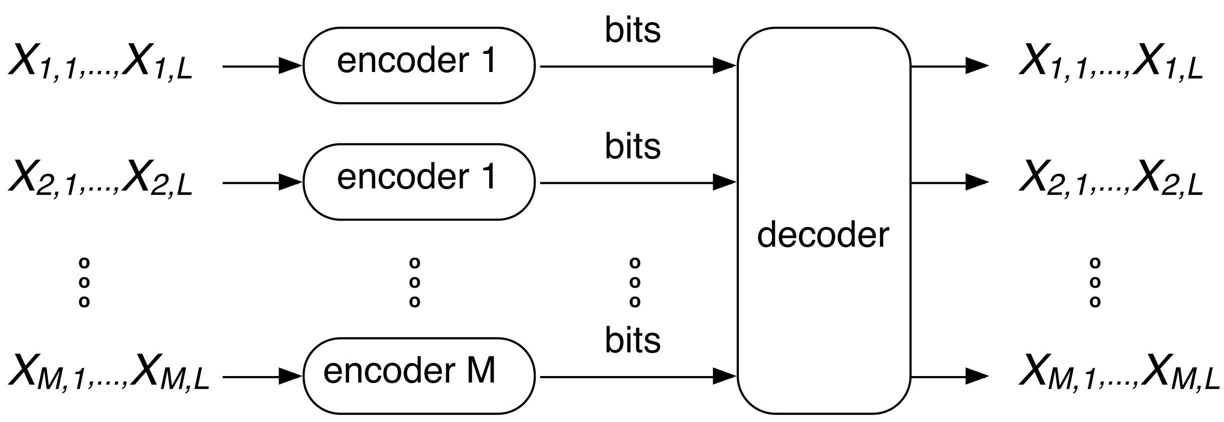
Review: Centralized and Distributed Entropy-Rate Coding for Discrete-Time Source

- Entropy-Rate Coding (ERC)
  - The lowest rate with which a discrete-stationary source can be lossless encoded (for example with block-to-variable-length codes or conditional codes) its entropy-rate

$$H_\infty \triangleq \lim_{L \to \infty} \frac{1}{L} H(X_1,\ldots,X_L) = \lim_{L \to \infty} H(X_L \,/\, X_1,\ldots,X_{L-1}) \text{ bits/sample}$$

- Example: stationary Markov source $H_\infty = H(X_2 / X_1)$

- ERC can be done in a distributed fashion at the same rate! [Slepian-Wolf `73]

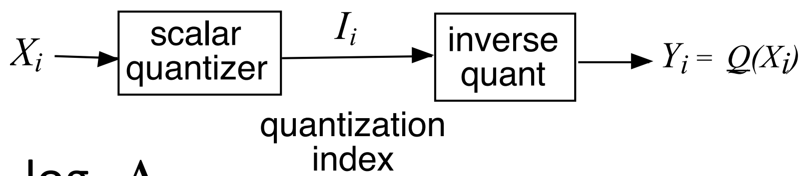


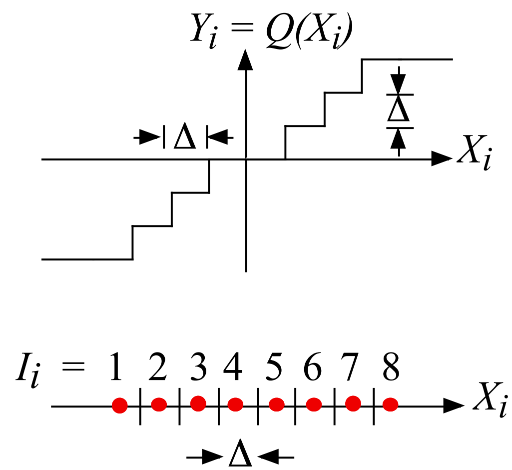
Uniform Scalar Quantization (USQ) with ERC

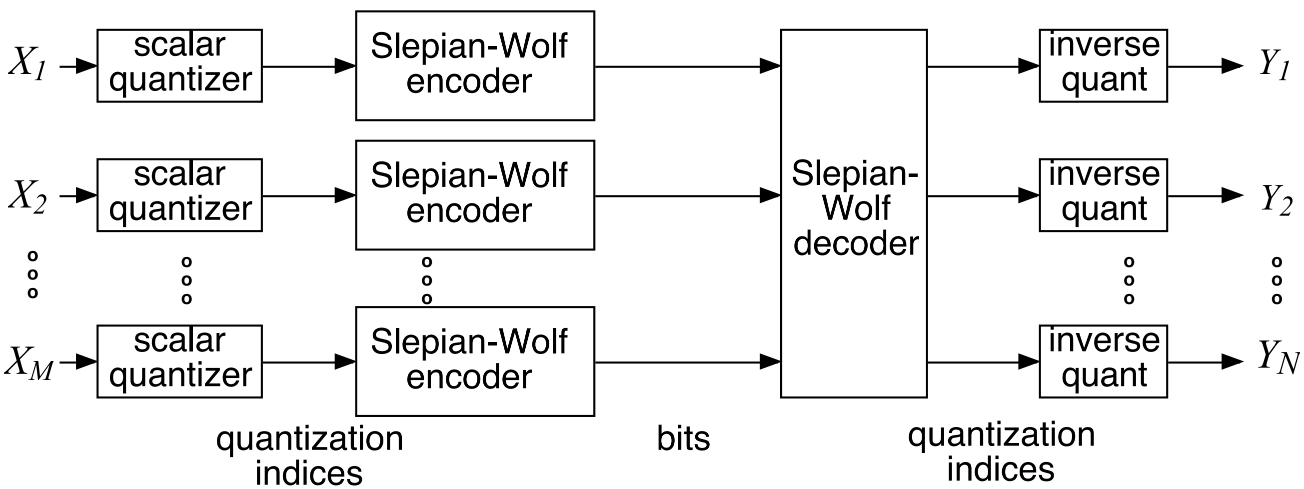
- Assume small step size Δ
- Distortion: $D(c) \cong \frac{\Delta^2}{12}$
- Rate: $R(c) = H_\infty(Y) = H_\infty(I) \cong H_\infty(X) - \log_2 \Delta$
- For a stationary, discrete-time source, the ORDF of USQ with ERC is

$$R_C(d) \cong R_{sh}(d) + 0.255$$

- With distributed ERC, USQ + ERC becomes distributed coding system.







Ideas Leading to a Conjecture

- For continuous-time source
  - Best performance of lossy source codes is given by Shannon rate-distortion function $R_{sh}(d)$.
  - For Gaussian source, a parametric expression for $R_{sh}(d)$ is known (Kolmogorov).
  - Performance as good as $R_{sh}(d)$ can be attained (to within ε) by sampling at very high rate, coding samples, and reconstructing cont.-time signal.
- For discrete-time (sampled) source
  - Best rate-distortion performance with distributed coding is not known, except for two Gaussian sources.
  - Uniform scalar quantization plus entropy-rate coding has performance close to $R_{sh}(d)$ for any discrete-time source.
  - Entropy-rate coding can be done in distributed fashion with as small rate as centralized coding. (Slepian-Wolf coding)

- Conjecture: uniform scalar quantization + distributed entropy-rate coding is distributed coding system that works well for cont-time source with large sampling rate (performance might be close to optimal).

  (Idea: entropy-rate coding will exploit strong sample dependence to mitigate large sampling rate.)

The Good News

- Suppose we sample with rate N, quantize, entropy-rate code, and reconstruct cont-time signal.
  - $D \cong$ MSE of quantizer on samples; not affected by sampling rate
  - $R = N \times R(c)$ bits/sec

    $\cong N \times H_\infty(N)$ (entropy-rate is a function of sampling rate)

- Theorem 1: [Marco-DN 2009]

  For any stationary source and quantizer

$$H_\infty(N) \to 0 \text{ as } N \to \infty$$

- Proof sketch: (recall I_i is index produced by quantizer in response to X_i)
  for any L,

$$H_\infty(N) \le L\,H(I_1,\ldots,I_L) = \frac{1}{L}\sum_{n=1}^{L} H(I_1 \mid I_1,\ldots,I_{n-1})$$

$$\le \frac{1}{L}H(I_1) + \frac{L-1}{L}H(I_2 \mid I_1) \quad \text{by stationarity}$$

$$\to 0 \text{ as } N \to \infty \quad \text{because } \Pr(I_2 = I_1) \to 1$$

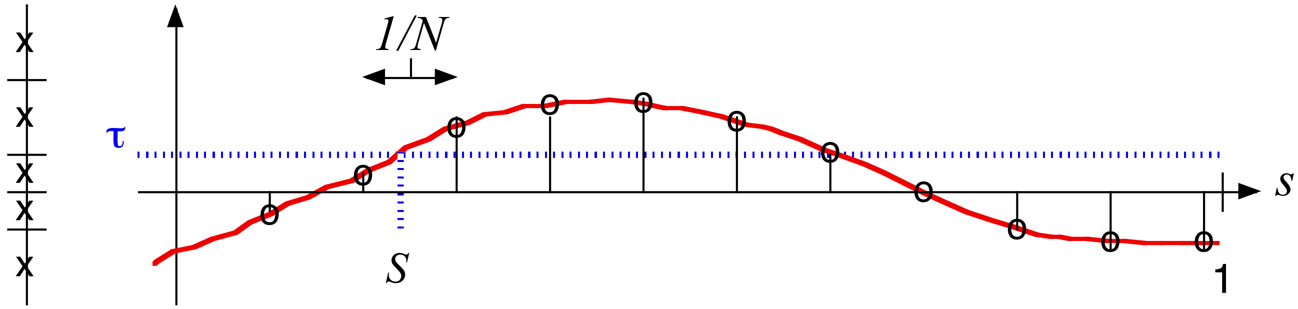
The Bad News … Conjecture is False

- Theorem 2: [Marco-DN 2009]

  For virtually any stationary source and quantizer

$$N H_{\infty}(N) \rightarrow \infty \text{ as } N \rightarrow \infty$$

Key Observation[1]



- S = location of 1st threshold crossing in $[0,1]$; $S = 2$ if no crossing.
- $H(S) = \infty$, since S is rand. variable with a continuous component.
- From quantizer indices $I_1, \ldots, I_N$, can make estimate S_N of S s.t.

$$|S - S_N| \leq \frac{1}{N} \quad \text{with high probability}$$

- This implies $H(S_N) \to \infty$.
- Also $H(I_1,\ldots,I_N) \geq H(S_N)$, since S_N is a function of $I_1,\ldots,I_N$
- Thus $\lim_{N\to\infty} N H_\infty(N) = \lim_{N\to\infty} N \frac{H(I_1,\ldots,I_N)}{N} = \lim_{N\to\infty} H(I_1,\ldots,I_N) = \infty$

---

[1] Courtesy of Bruce Hajek

Another Explanation

- $R = N \times R(c)$

  $\cong N \times (R_{sh,N}(d) + 0.255)$

  $= N \times R_{sh,N}(d) + N \times 0.255$

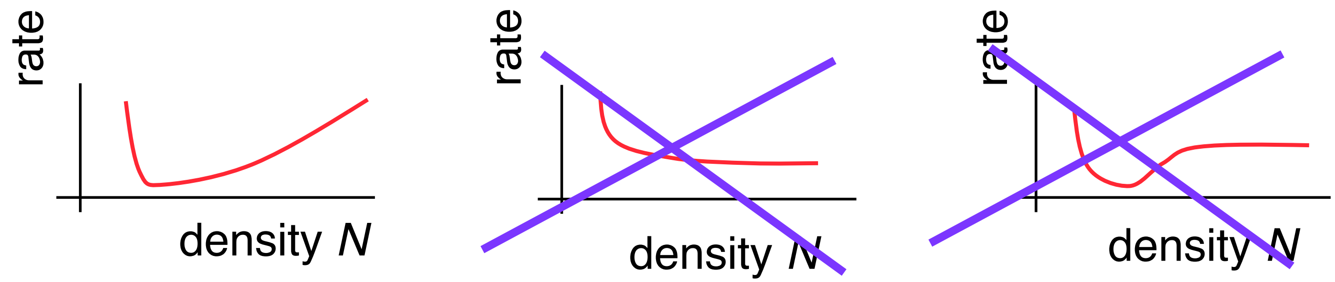
  $\rightarrow R_{sh}(d) + \infty$

- The weakness of this argument is that $R(c)$ is small when N is large, whereas the high resolution approximation used is generally valid only when rate is large.

Summary

- *Bad news* from Theorem 2: For scalar quantization and S-W distributed lossless coding,

$$N\,R_N(d) \;\rightarrow\; \infty$$



- With identical scalar quantizers, when sensors are dense, entropy coding cannot sufficiently exploit increased correlation to mitigate increased number of sensors
- N.B.: It is not that field gathering with identical scalar quantizers is infeasible. But with such, there is a finite best sampling density.

At What Rate Does $NH_\infty(N) \to \infty$?

- Theorem: [Marco-DN 2010] For unif. scalar quant. with step size Δ, infinitely many levels and stat'ry Gaussian $\mathcal{N}(0,\sigma^2)$ source with autocorr. func. $\rho(\tau)$,

$$N H_\infty(N) \le H(I_1,\ldots,I_N) \le NH(I_2 \mid I_1) \cong -Nm\sqrt{1-\rho(1/N)}\,\log_2\sqrt{1-\rho(1/N)}$$

where

$$m = -\frac{2\sqrt{2}}{\pi}\sum_{k=0}^{\infty} e^{-\frac{(k+1/2)^2\Delta^2}{2\sigma^2}}$$

- Examples: When N is large,

$$\rho(\tau) = e^{-|\tau|} \quad\Rightarrow\quad NH(I_2 \mid I_1) \cong \frac{m}{2}\sqrt{N}\log_2 N$$

$$\rho(\tau) = e^{-\tau^2} \quad\Rightarrow\quad NH(I_2 \mid I_1) \cong \frac{m}{2}\log_2 N$$

Why Does Scalar Quantization Perform So Poorly With Dense Sensors?

- Is it a flaw of all distributed source coding schemes?
- Or just a flaw of scalar quantization based schemes? ⇐ *this one*

- Consider Distributed Vector Quantization (VQ)

- Kashyap et al. [2005]
  - Showed that for a stationary, Gaussian source and ideal distributed lossy coding, $N\,R_N(d)$ remains finite as N increases.

- Pradhan & DN [2006,2013]
  - Made a similar analysis.

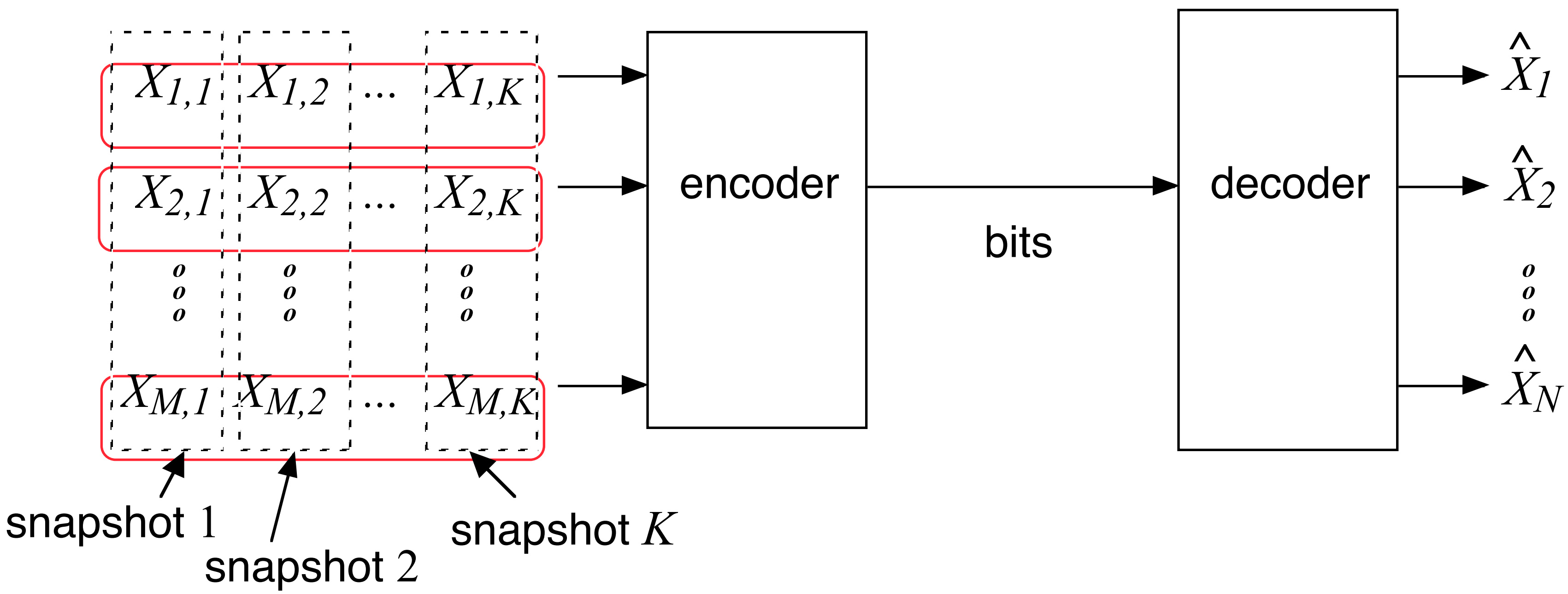
- Note: VQ dimension must increase with sampling rate N.

Four Strategies For Lossy Source Coding Based on High-Rate Sampling

We'll analyze the following for stationary, Gaussian, continuous-time sources

- Transform + VQ
  (centralized & optimal)
- Scalar quantization + entropy-rate coding:
  (centralized or distributed, no transform, suboptimal)
- Distributed VQ:
  (no transform, suboptimal)
- Transform + scalar quantization:
  (centralized, suboptimal)

Distributed Vector Quantization



- M sensors, $1/N$ apart in spatial interval $[0, M/N]$
- Spatial sampling rate = N
- For sampled source, $R_{DVQ}(M,d)$ known only for $M=2$ Gaussian sources [*Wagner, et al. 2007*]
- Kashyap et al. [05] and Pradhan-DN [06,10] applied Berger-Tung bound [77] to obtain upper bounds to $R_{DVQ}(M,d)$.

Berger-Tung Bound for Distributed VQ

- Lower bound to least rate of distributed encoding of M sources with MSE d :

$$R_{DVQ}(M,d) \geq R_{BT}(M,d) = \inf_{p} \frac{1}{M} I_p(X_1 \ldots X_M ; Y_1 \ldots Y_M)$$

where “inf” is over test channels with MSE ≤ D.

Has same form as M-th-order Shannon rate-distortion function, except

– Components of test channel are conditionally independent given source inputs

$$p(y_1,\ldots,y_M \mid x_1,\ldots x_M) = \prod_{i=1}^{M} p(y_i \mid x_i)$$

– In determining MSE, the test channel output $Y_1 \ldots Y_M$ is followed by an optimal estimator for inputs $X_1 \ldots X_M$ from source.

Applying Berger-Tung and Kuhn-Tucker

- Choose test channel:

$$Y_i = \frac{1}{1+\theta}\left(X_i + Z_i\right), \quad i = 1,\ldots,M$$

with Z_i's IID, $\mathcal{N}(0,\theta)$

- Use Kuhn-Tucker:

$$R_{BT,\theta} = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}\log_2\left(\frac{\lambda_i}{\theta}+1\right)$$

$$D_{BT,\theta} = \frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_i\theta}{\lambda_i+\theta}$$

where $\lambda_1,\ldots,\lambda_M$ are the eigenvalues of covariance matrix of $X_1 \ldots X_M$

in comparison

for centralized VQ

$$R_{Sh,\theta} = \frac{1}{M}\sum_{i=1}^{M}\max\left\{\frac{1}{2}\log_2\frac{\lambda_i}{\theta},0\right\}$$

$$D_{Sh,\theta} = \frac{1}{M}\sum_{i=1}^{M}\min\{\lambda_i,\theta\}$$

Take limit as M → ∞

- Begin with

$$R_{BT,\theta} = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{2}\log_2\left(\frac{\lambda_i}{\theta}+1\right)$$

$$D_{BT,\theta} = \frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_i\theta}{\lambda_i+\theta}$$

- Let $M \to \infty$

$$R_{BT,\theta} \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\log_2\left(\frac{\Phi(\Omega)}{\theta}+1\right)d\Omega$$

$$D_{BT,\theta} \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\Phi(\Omega)\theta}{\Phi(\Omega)+\theta}d\Omega$$

for centralized VQ

$$R_{Sh,\theta} = \frac{1}{M}\sum_{i=1}^{M}\max\left\{\frac{1}{2}\log_2\frac{\lambda_i}{\theta},0\right\}$$

$$D_{Sh,\theta} = \frac{1}{M}\sum_{i=1}^{M}\min\{\lambda_i,\theta\}$$

$$R_{sh,\theta} \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\max\left\{\frac{1}{2}\log_2\frac{\Phi(\Omega)}{\theta},0\right\}d\Omega$$

$$D_{Sh,\theta} \to \frac{1}{2\pi}\int_{-\pi}^{\pi}\min\{\Phi(\Omega),\theta\}\,d\Omega$$

Let Sampling Rate $N \to \infty$

- Change variables -- let $\omega = \Omega N$

$$R_{BT,\theta} \to \frac{1}{2\pi}\int_{-\pi N}^{\pi N} \frac{1}{2}\log_2\left(\frac{\Phi(\omega/N)/N}{\theta/N}+1\right)\frac{1}{N}d\omega$$

$$D_{BT,\theta} \to \frac{1}{2\pi}\int_{-\pi N}^{\pi N} \frac{\Phi(\omega/N)\theta}{\Phi(\omega/N)+\theta}\frac{1}{N}d\omega$$

- Let sampling rate $N \to \infty$; let $\phi = \theta N$; then $\Phi(\omega/N)/N \to S(\omega)$ as $N \to \infty$

$$N\,R_{BT,\theta} \to \mathcal{R}_{BT,\theta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{2}\log_2\left(\frac{S(\omega)}{\theta}+1\right)d\omega$$

$$D_{BT,\theta} \to \mathcal{D}_{BT,\theta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S(\omega)\theta}{S(\omega)+\theta}d\omega$$

- This upper bound to optimal performance of distributed coding coding might be tight.

Comparison

Distributed Coding
(attainable rate)

$$R_{BT,\theta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{1}{2}\log_2\left(\frac{S(\omega)}{\theta}+1\right) d\omega$$

$$D_{BT,\theta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{S(\omega)\theta}{S(\omega)+\theta}\, d\omega$$

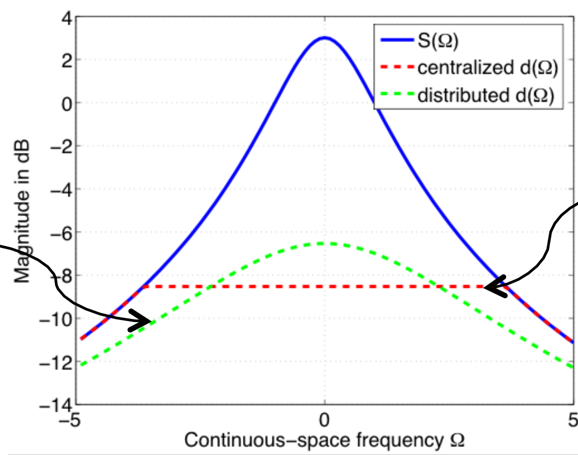
Centralized Coding
(optimal rate)

$$R_{Sh,\theta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \max\left\{\frac{1}{2}\log_2\frac{S(\omega)}{\theta}, 0\right\} d\omega$$

$$D_{Sh,\theta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \min\{S(\omega),\phi\}\, d\omega$$

Distortion Profiles:

$$\frac{S(\omega)\theta}{S(\omega)+\theta}$$



$$\min\{S(\omega),\phi\}$$

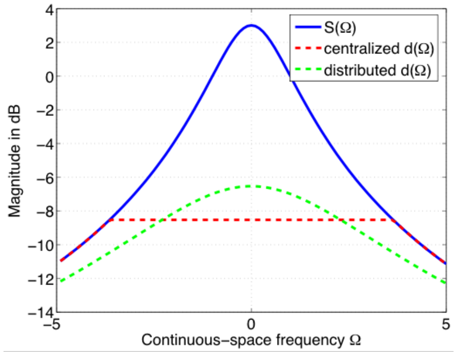
distributed coding cannot use transform, and so cannot have sharp cutoff bandlimiting.

Example

- Source -- stationary, Gauss-Markov,

$$\rho(\tau) = e^{-|\tau|}, \quad S(\omega) = \frac{2}{1+\omega^2}$$

$S(\omega)$



Distributed Coding
(attainable rate)

$$R_{BT}(d) = \frac{1}{2\ln 2}\left(\frac{1}{d} - 1\right)$$

$= 6.5$ bits/m

Centralized Coding
(optimal rate)

$$R_{Sh}(d) \cong \frac{1}{2\ln 2}\left(\frac{0.81}{d} - 1\right) \text{ for small } d$$

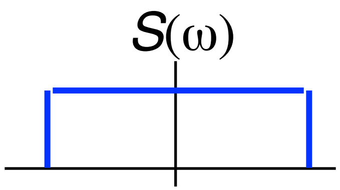
$= 5.1$ bits/m

$d = 0.1$

Example

- Source -- stationary, flat bandlimited

$$S(\omega) = \begin{cases} \pi/\omega_o, & |\omega| \le \omega_o \\ 0, & \text{else} \end{cases}$$



Distributed Coding
(attainable rate)

$$R_{BT}(d) = \frac{\omega_o}{2\pi} \log_2 \frac{1}{d}$$

= 1.7 bits/m

Centralized Coding
(optimal rate)

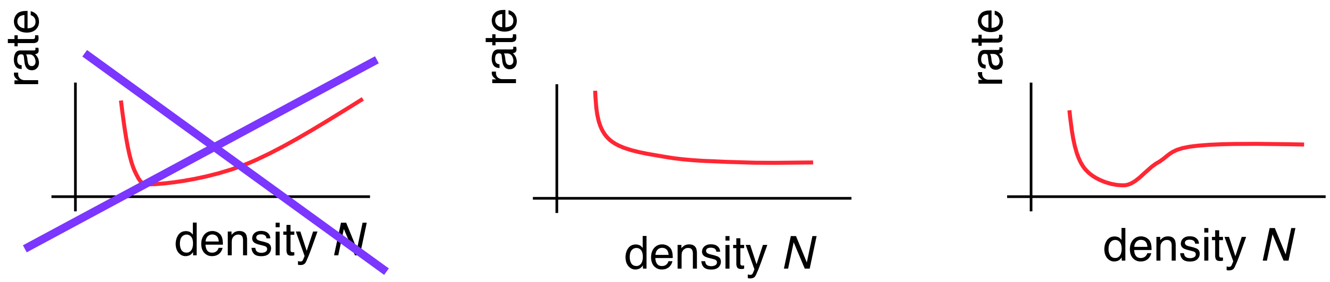
$$R_{Sh}(d) = \frac{\omega_o}{2\pi} \log_2 \frac{1}{d}$$

= 1.7 bits/m

d = 0.1

Summary

- For coding with distributed vector quantization.



- Probably the middle one.

Are Scalar Quantizers Always Bad with Dense Samples?

- Not always!

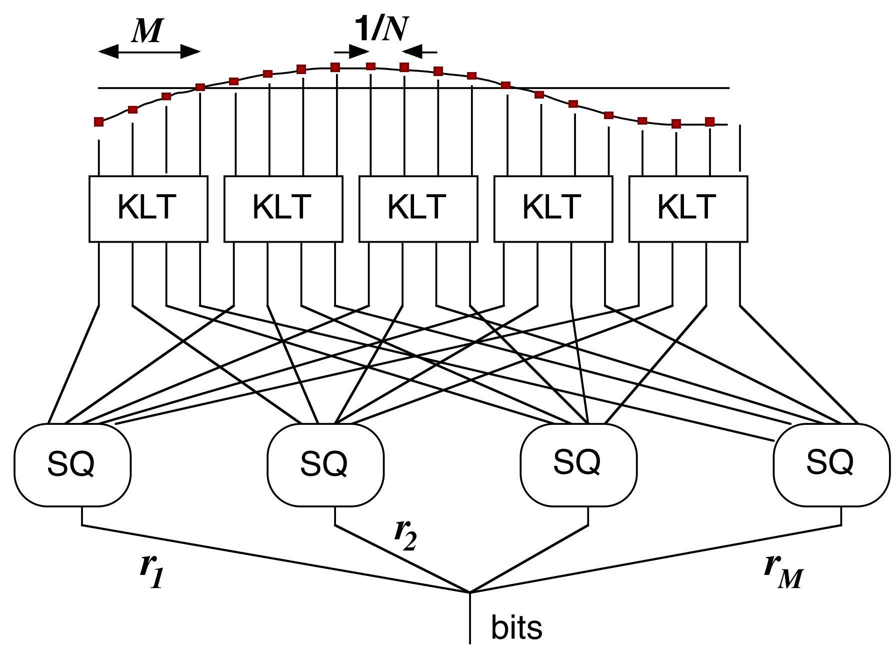
Four Strategies For Lossy Source Coding Based on High-Rate Sampling

We'll analyze the following for stationary, Gaussian, continuous-time sources

- Transform + VQ
  (centralized & optimal)
- Scalar quantization + entropy-rate coding:
  (centralized or distributed, no transform, suboptimal)
- Distributed VQ:
  (no transform, suboptimal)
- Transform + scalar quantization:
  (centralized, suboptimal)

Transform, scalar quantization, entropy coding

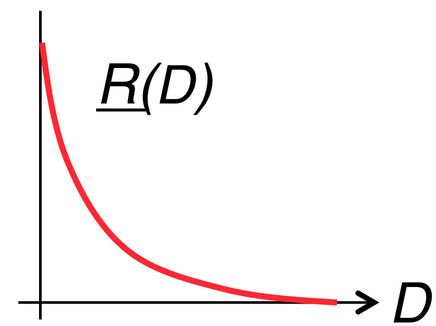
- Proceed as before ...
- Sampling rate N rate over $[0,\infty)$
- M-dimensional KLT produces M indep. Gaussian coef's with variances equal to eigenval's of covar. matrix of $X_1, X_2, \ldots, X_M$: $\lambda_1^{(M)}, \ldots, \lambda_M^{(M)}$
- Independently scalar quantize and entropy code each type of transform coefficient, instead of optimally VQ encoding.
- Optimize the rate allocation for coefficients $r_1, r_2, \ldots, r_M$
- Take M to infinity.
- Take N to infinity.



Transform, scalar quantization, entropy coding

- Rate: $R = \frac{1}{M}\sum_{i=1}^{M} r_i$
- Distortion: $D = \frac{1}{M}\sum_{i=1}^{M} d_i$
- Let $\underline{R}(d)$ denote ORDF for scalar quantizing with entropy coding a unit variance Gaussian variable,

  Assume $\underline{R}(d)$ is convex.



- Then for i th coef. $r_i = \underline{R}\left(\frac{d_i}{\lambda_i^{(M)}}\right)$
- Use Kuhn-Tucker theory to optimize d_i's.

Transform, scalar quantization, entropy coding

- Given $\phi < 0$, Kuhn-Tucker gives

$$d_i = \lambda_i \underline{D}'(\phi\lambda_i)$$

where $\underline{D}'(\cdot)$ is the inverse of the derivative of $\underline{R}(\cdot)$

- Substituting this gives optimal rate-distortion pairs, parameterized by ϕ

$$R_{Tr,\phi} = \frac{1}{M}\sum_{i=1}^{M} \underline{R}(\underline{D}'(\phi\lambda_i^{(M)}))$$

$$D_{Tr,\phi} = \frac{1}{M}\sum_{i=1}^{M} \lambda_i^{(M)} \min\{1, \underline{D}'(\phi\lambda_i^{(M)})\}$$

- Dimension $M \to \infty$

$$R_{Tr,\phi} \to \frac{1}{2\pi}\int_{-\pi}^{\pi} \underline{R}(\underline{D}'(\phi\Phi(\Omega)))\, d\Omega$$

$$D_{Tr,\phi} \to \frac{1}{2\pi}\int_{-\pi}^{\pi} \min\{1, \underline{D}'(\phi\Phi(\Omega))\}\, d\Omega$$

Transform, scalar quantization, entropy coding

- Change variables -- let $\omega = \Omega N$

$$R_{Tr,\phi} \to \frac{1}{2\pi} \int_{-\pi N}^{\pi N} \underline{R}(\underline{D}'(\phi\Phi(\omega/N))) \frac{1}{N} d\omega$$

$$D_{Tr,\phi} \to \frac{1}{2\pi} \int_{-\pi N}^{\pi N} \min\{1, \underline{D}'(\phi\Phi(\omega/N))\} \frac{1}{N} d\omega$$

- Sampling rate $N \to \infty$, let $\phi = \theta N$, $\Phi(\omega/N)/N \to S(\omega)$ as $N \to \infty$

$$NR \to R_{Tr,\theta} = \frac{1}{2\pi} \int_{-\infty}^{-\infty} \underline{R}\left(\underline{D}'\left(\frac{S(\omega)}{\theta}\right)\right) d\omega$$

$$D \to D_{Tr,\theta} = \frac{1}{2\pi} \int_{-\infty}^{-\infty} S(\omega) \min\left\{1, \underline{D}'\left(\frac{S(\omega)}{\theta}\right)\right\} d\omega$$

Transform, scalar quantization, entropy coding

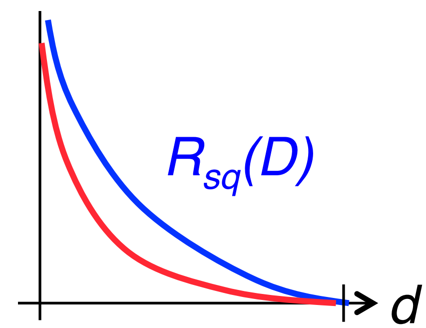
- To repeat

$$R_{\mathrm{Tr},\theta} = \frac{1}{2\pi} \int_{-\infty}^{-\infty} \underline{R}(\underline{D}'\left(\frac{S(\omega)}{\theta}\right) d\omega$$

$$D_{\mathrm{Tr},\theta} = \frac{1}{2\pi} \int_{-\infty}^{-\infty} S(\omega) \min\left\{1, \underline{D}'\left(\frac{S(\omega)}{\theta}\right)\right\} d\omega$$

- Since these are finite, scalar quantization does not lead to catastrophic performance, provided it is preceded by a transform.
- Note: If *$\underline{R}$(.)* is replaced by Shannon rate-distortion function for Gaussian samples, the above reduces to Shannon rate-distortion function for continuous-time Gaussian source.
- [Pradhan-DN, 2007,13]

Why Does Transform Coding With Scalar Quantization Not Suffer Catastrophically Bad Performance?



- Without transform, scalar quant. + ent. coding has rate

$$R_N(d) \approx R_{sh,N}(d) + O(1)$$

$$R \approx N\, R_N(d) + N O(1) \rightarrow R_{Sh,N}(d) + \infty$$

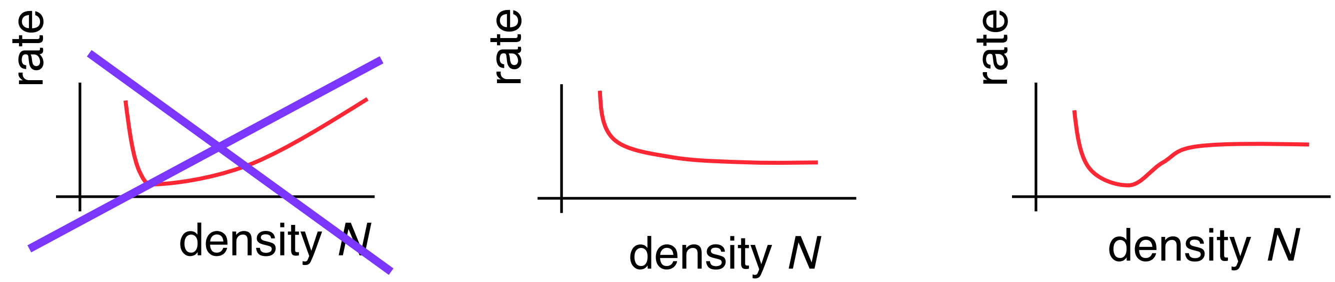
- However: $O(1)$ "loss" goes to zero as d approaches variance.
- With KLT, variances are eigenvalues.
- **Lemma:** For any $\delta > 0$, fraction of eigenvalues $> \delta$ goes to zero.
- With transform coding

$$R \approx N \frac{1}{M} \sum_{i=1}^{M} r_i = N \frac{1}{M} \sum_{i=1}^{M} \underline{R}\left(\frac{d_i}{\lambda_i^{(M)}} \right)$$

- For most i, $d_i \approx \lambda_i$, so there is virtually no loss $\Rightarrow$ overall loss is small.

Summary

- For coding with transform and quantization.



- Probably the middle one.

Overall Summary

- Can attain optimal rate-distortion performance with high-rate sampling and transform coding
- Can attain good rate-distortion performance with high-rate sampling and
  - Transform coding with scalar quantization
  - Distributed coding
- Cannot attain good rate-distortion performance with high-rate sampling and direct scalar quantization, even with entropy-rate coding.
- To attain good performance, the dimension of the quantizer (in time) must grow as sampling rate grows.
- If one wishes to use scalar quantization plus ERC, one should not use too large a sampling rate, because entropy-rate does not decrease fast enough to mitigate the effect of high sampling rate.
- In centralized transform coding, scalar quantization does not cause a problem because most coefficients are scalar quantized at very low rates at which there is very little loss relative to high-dimensional VQ.

Ongoing Work

- High-resolution, high-sampling-rate analysis:
  - We are finding closed form expressions for ORDF $R_C(d)$ for distributed and transform coding when sampling rate is large and distortion d is constrained to be small.
- Convergence of discrete-time power spectral density to continuous-time power spectral density:
  - We are identifying conditions under which one can rigorously prove

    $$N\Phi_N(N\omega) \to S(\omega) \text{ as } N \to \infty$$

    and finding counterexamples, where conditions do not hold.