System-Aware Compression

Optimizing Systems from the Compression Standpoint

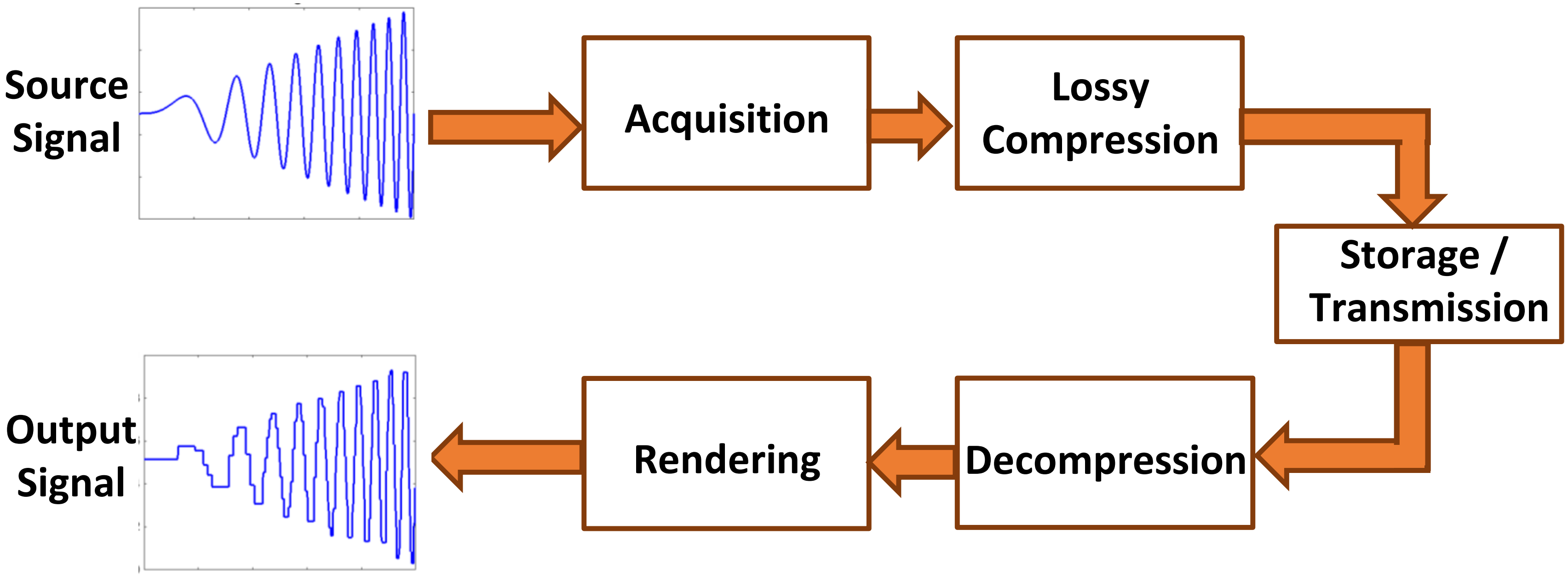
Yehuda Dar

Joint work with **Alfred Bruckstein** and **Michael Elad**

CS Department, Technion – Israel Institute of Technology

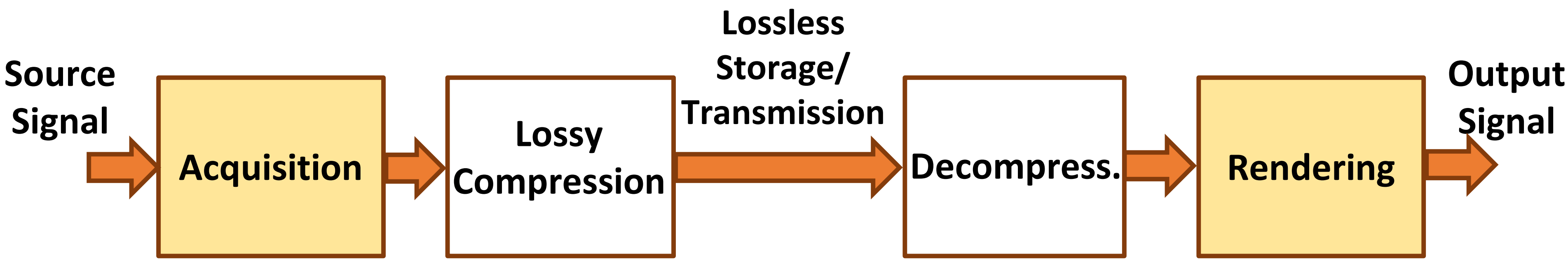
IEEE International Symposium on Information Theory 2018

Motivation: Acquisition-Rendering Systems



Standard compression techniques ignore the system structure!

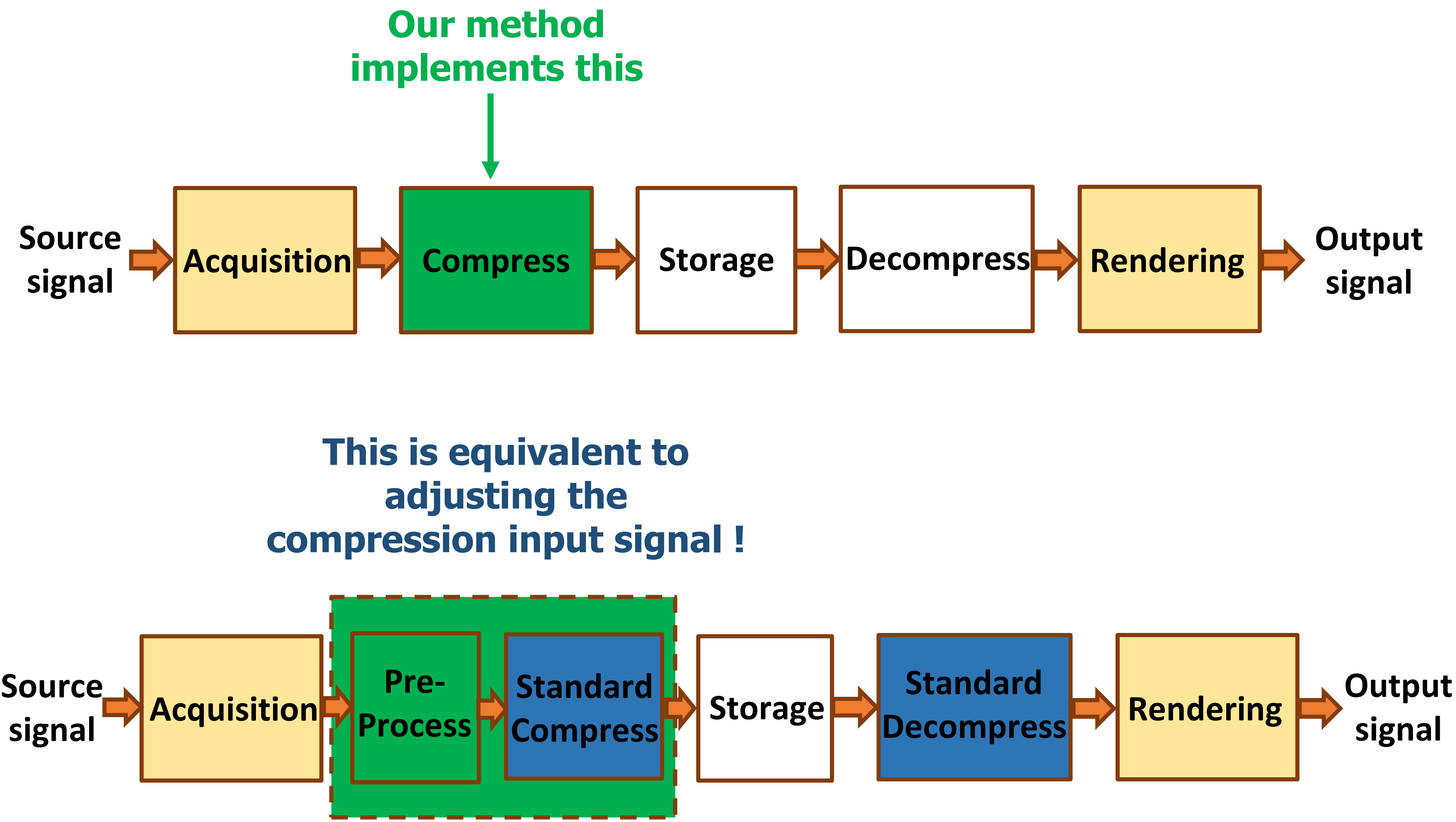
The Perspective of Remote Source Coding



An instance of the remote source coding problem [Dobrushin & Tsybakov '62, Wolf & Ziv '70]

- **Previous works** studied the **statistical perspective** using an **explicit source model**.
- **Here** we focus on a **deterministic setting**.

The Proposed Compression Method



The Proposed Compression Method:
The Challenge

Requirement #1:

Using **standard** compression techniques.

Problem

Ignore the system structure → **suboptimal !**

Requirement #2:

Having **system-optimized** compression.

Problem

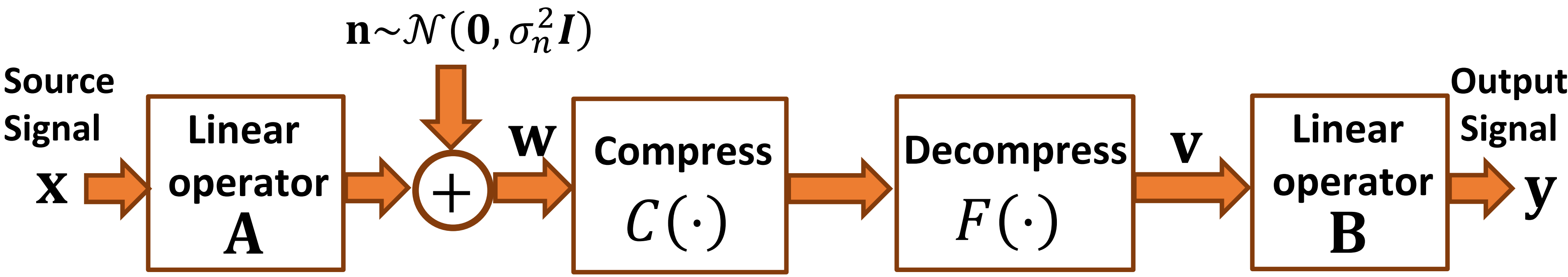
Needs a **system-specific** design.

The proposed framework:

Extends **standard compression** techniques to **system-optimized** compression procedures.

Moreover, we do **not rely on any post filtering!**

Model and Problem Formulation:
System's End-to-End Distortion Metric



The distortion metric should

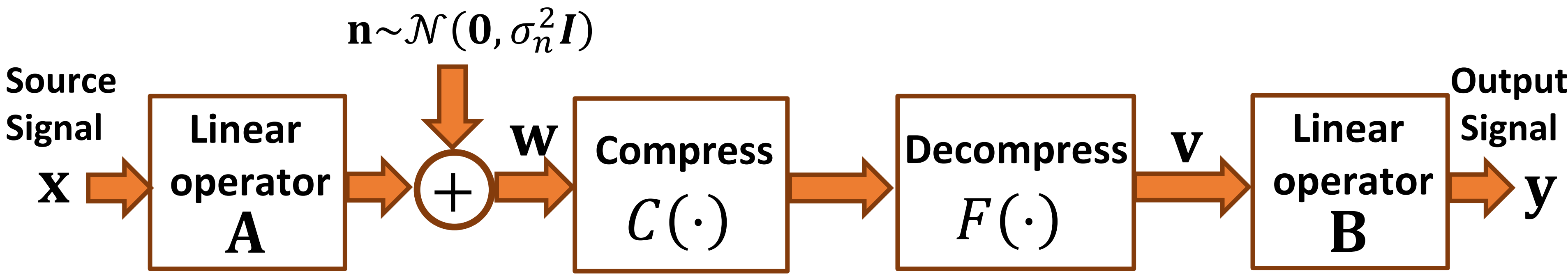
- Reflect the **distance** of $\mathbf{y}$ from the **unknown source** $\mathbf{x}$.
- Rely on the **degraded observation** $\mathbf{w} = \mathbf{A}\mathbf{x} + \mathbf{n}$.

Motivating metric:

$$d_s(\mathbf{w}, \mathbf{y}) = \frac{1}{M} \|\mathbf{w} - \mathbf{A}\mathbf{y}\|_2^2$$

The **ideal result** of $\mathbf{y}_{ideal} = \mathbf{x}$ corresponds to $d_s(\mathbf{w}, \mathbf{y}_{ideal}) = \frac{1}{M} \|\mathbf{n}\|_2^2 \approx \sigma_n^2$

Model and Problem Formulation:
System's End-to-End Distortion Metric



The distortion metric should also

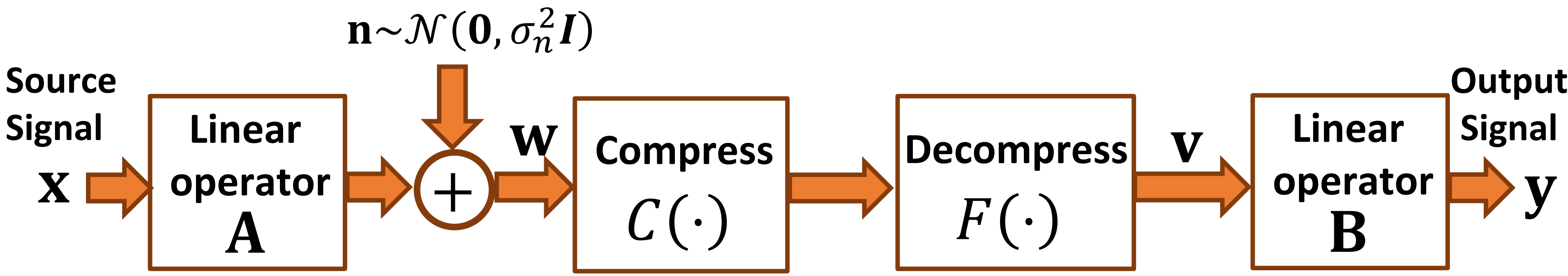
- Consider the **post-decompression** operator $\mathbf{B}$.
- Be evaluated for the **decompressed signal** $\mathbf{v}$.

Proposed metric:

$$d_c(\mathbf{w}, \mathbf{v}) \triangleq d_s(\mathbf{w}, \mathbf{B}\mathbf{v}) = \frac{1}{M} \|\mathbf{w} - \mathbf{A}\mathbf{B}\mathbf{v}\|_2^2$$

The **ideal result** of $\mathbf{y}_{ideal} = \mathbf{P}_{\mathbf{B}}\mathbf{x}$ corresponds to $D_0 \triangleq \frac{1}{M} \|\mathbf{A}(\mathbf{I} - \mathbf{P}_{\mathbf{B}})\mathbf{x} + \mathbf{n}\|_2^2$.

Model and Problem Formulation:
System-Aware Compression Optimization



Our compression goal:

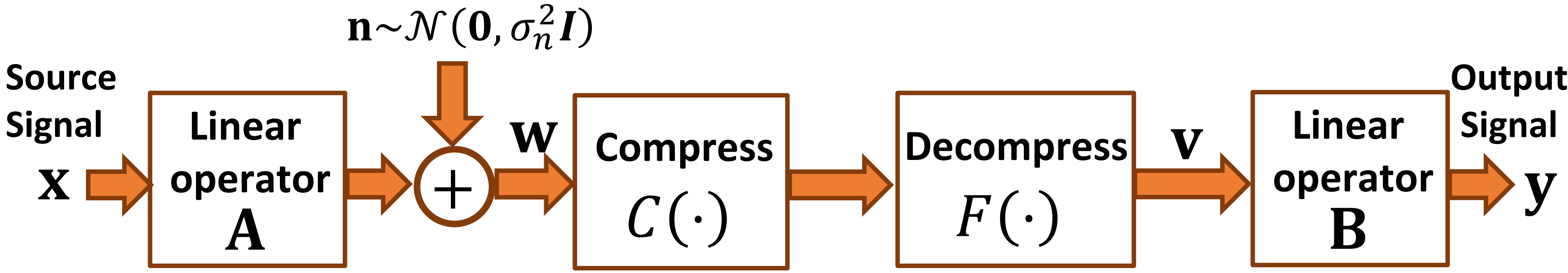
Minimize the **bit-cost** under a **system distortion constraint**

$$\hat{\mathbf{v}} = \underset{\mathbf{v} \in \mathcal{S}}{\operatorname{argmin}} \, R(\mathbf{v})$$

$$\text{s.t.} \quad D_0 \leq \frac{1}{M} \|\mathbf{w} - \mathbf{A}\mathbf{B}\mathbf{v}\|_2^2 \leq D_0 + D$$

An operational rate-distortion optimization for the input signal $\mathbf{w}$ with respect to the compression architecture $\{\mathcal{S}, R\}$.

Model and Problem Formulation:
System-Aware Compression Optimization

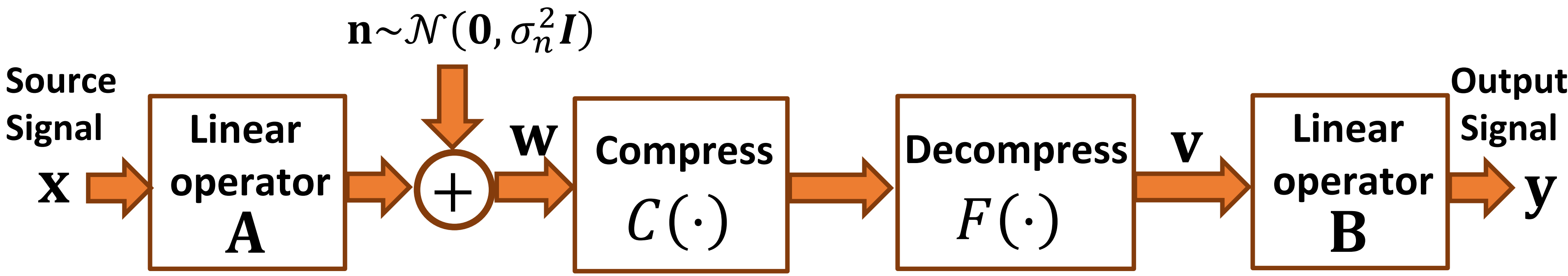


The unconstrained **Lagrangian** optimization:

$$\hat{\mathbf{v}} = \underset{\mathbf{v} \in \mathcal{S}}{\operatorname{argmin}} \; R(\mathbf{v}) + \lambda \frac{1}{M} \|\mathbf{w} - \mathbf{ABv}\|_2^2$$

where λ is a Lagrange multiplier corresponding to some distortion level D_λ

The Practical Challenge



The unconstrained **Lagrangian** optimization:

$$\hat{\mathbf{v}} = \underset{\mathbf{v} \in \mathcal{S}}{\operatorname{argmin}} \; R(\mathbf{v}) + \lambda \frac{1}{M} \|\mathbf{w} - \mathbf{A}\mathbf{B}\mathbf{v}\|_2^2$$

Computationally hard due to $\mathbf{A}$ and $\mathbf{B}$!

Design Overview

The Variable Splitting Trick

The hard optimization:

$$\hat{\mathbf{v}} = \underset{\mathbf{v}\in\mathcal{S}}{\operatorname{argmin}}\ R(\mathbf{v}) + \frac{\lambda}{M}\|\mathbf{w} - \mathbf{A}\mathbf{B}\mathbf{v}\|_2^2$$

Two complicated parts considering $\mathbf{v}$

Variable splitting:

$$(\hat{\mathbf{v}}, \hat{\mathbf{z}}) = \underset{\mathbf{v}\in\mathcal{S}\ ,\ \mathbf{z}\in\mathbb{R}^M}{\operatorname{argmin}}\ R(\mathbf{v}) + \frac{\lambda}{M}\|\mathbf{w} - \mathbf{A}\mathbf{B}\mathbf{z}\|_2^2$$

subject to $\mathbf{v} = \mathbf{z}$

where $\mathbf{z}$ is an additional vector due to the split.

The ADMM Optimization Form

Augmented Lagrangian (scaled version) and the **Method of Multipliers**:

$$\left(\hat{\mathbf{v}}^{(t)}, \hat{\mathbf{z}}^{(t)}\right) = \underset{\mathbf{v}\in\mathcal{S}\ ,\ \mathbf{z}\in\mathbb{R}^M}{\operatorname{argmin}} R(\mathbf{v}) + \frac{\lambda}{M}\|\mathbf{w} - \mathbf{ABz}\|_2^2 + \frac{\beta}{2}\left\|\mathbf{v} - \mathbf{z} + \mathbf{u}^{(t)}\right\|_2^2$$

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + \left(\hat{\mathbf{v}}^{(t)} - \hat{\mathbf{z}}^{(t)}\right)$$

where $\mathbf{u}^{(t)} \in \mathbb{R}^M$ is the scaled dual-variable and β is a parameter introduced in the augmented Lagrangian.

Alternating minimization provides the **ADMM form**:

$$\hat{\mathbf{v}}^{(t)} = \underset{\mathbf{v}\in\mathcal{S}}{\operatorname{argmin}} R(\mathbf{v}) + \frac{\beta}{2}\left\|\mathbf{v} - \tilde{\mathbf{z}}^{(t)}\right\|_2^2$$

$$\hat{\mathbf{z}}^{(t)} = \underset{\mathbf{z}\in\mathbb{R}^M}{\operatorname{argmin}} \frac{\lambda}{M}\|\mathbf{w} - \mathbf{ABz}\|_2^2 + \frac{\beta}{2}\left\|\mathbf{z} - \tilde{\mathbf{v}}^{(t)}\right\|_2^2$$

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + \left(\hat{\mathbf{v}}^{(t)} - \hat{\mathbf{z}}^{(t)}\right)$$

where

$$\tilde{\mathbf{z}}^{(t)} = \hat{\mathbf{z}}^{(t-1)} - \mathbf{u}^{(t)}$$

$$\tilde{\mathbf{v}}^{(t)} = \hat{\mathbf{v}}^{(t)} + \mathbf{u}^{(t)}$$

Design Overview

Leveraging a Standard Compression

$$\hat{\mathbf{v}}^{(t)} = \underset{\mathbf{v}\in\mathcal{S}}{\operatorname{argmin}}\, R(\mathbf{v}) + \frac{\beta}{2}\left\|\mathbf{v} - \tilde{\mathbf{z}}^{(t)}\right\|_2^2$$

$$\hat{\mathbf{z}}^{(t)} = \underset{\mathbf{z}\in\mathbb{R}^M}{\operatorname{argmin}}\, \frac{\lambda}{M}\|\mathbf{w} - \mathbf{ABz}\|_2^2 + \frac{\beta}{2}\left\|\mathbf{z} - \tilde{\mathbf{v}}^{(t)}\right\|_2^2$$

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + \left(\hat{\mathbf{v}}^{(t)} - \hat{\mathbf{z}}^{(t)}\right).$$

Compression optimization considering **regular squared error**

We suggest to

replace this step with an application of a **standard compression method**

Leveraging a Standard Compression

$$b^{(t)} = StandardCompress\left(\tilde{\mathbf{z}}^{(t)}, \rho\right)$$

$$\hat{\mathbf{v}}^{(t)} = StandardDecompress\left(b^{(t)}\right)$$

$$\hat{\mathbf{z}}^{(t)} = \underset{\mathbf{z}\in\mathbb{R}^M}{\operatorname{argmin}} \frac{\lambda}{M}\|\mathbf{w}-\mathbf{ABz}\|_2^2 + \frac{\beta}{2}\left\|\mathbf{z}-\tilde{\mathbf{v}}^{(t)}\right\|_2^2$$

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + \left(\hat{\mathbf{v}}^{(t)} - \hat{\mathbf{z}}^{(t)}\right).$$

Compression optimization considering **regular squared error**

We suggest to

replace this step with an application of a **standard compression method**

ℓ_2-Constrained Deconvolution

$$b^{(t)} = StandardCompress\left(\tilde{\mathbf{z}}^{(t)}, \theta\right)$$

$$\hat{\mathbf{v}}^{(t)} = StandardDecompress\left(b^{(t)}\right)$$

$$\hat{\mathbf{z}}^{(t)} = \underset{\mathbf{z}\in\mathbb{R}^M}{\operatorname{argmin}} \frac{\lambda}{M}\|\mathbf{w} - \mathbf{ABz}\|_2^2 + \frac{\beta}{2}\left\|\mathbf{z} - \tilde{\mathbf{v}}^{(t)}\right\|_2^2$$

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + \left(\hat{\mathbf{v}}^{(t)} - \hat{\mathbf{z}}^{(t)}\right).$$

ℓ_2-Constrained Deconvolution
Optimization of simple quadratic terms

This can be easily solved!

The Proposed Approach: Concept

The complicated optimization: $\hat{\mathbf{v}} = \underset{\mathbf{v}\in\mathcal{S}}{\operatorname{argmin}}\ R(\mathbf{v}) + \frac{\lambda}{M}\|\mathbf{w} - \mathbf{A}\mathbf{B}\mathbf{v}\|_2^2$

is addressed via **repeated** application of **3 simple procedures**:

associated with

$$\hat{\mathbf{v}} = \underset{\mathbf{v}\in\mathcal{S}}{\operatorname{argmin}}\ R(\mathbf{v}) + \frac{\lambda}{M}\|\mathbf{w} - \mathbf{v}\|_2^2$$

conceptually employed in modern compression techniques

Standard Compression

Standard Decompression

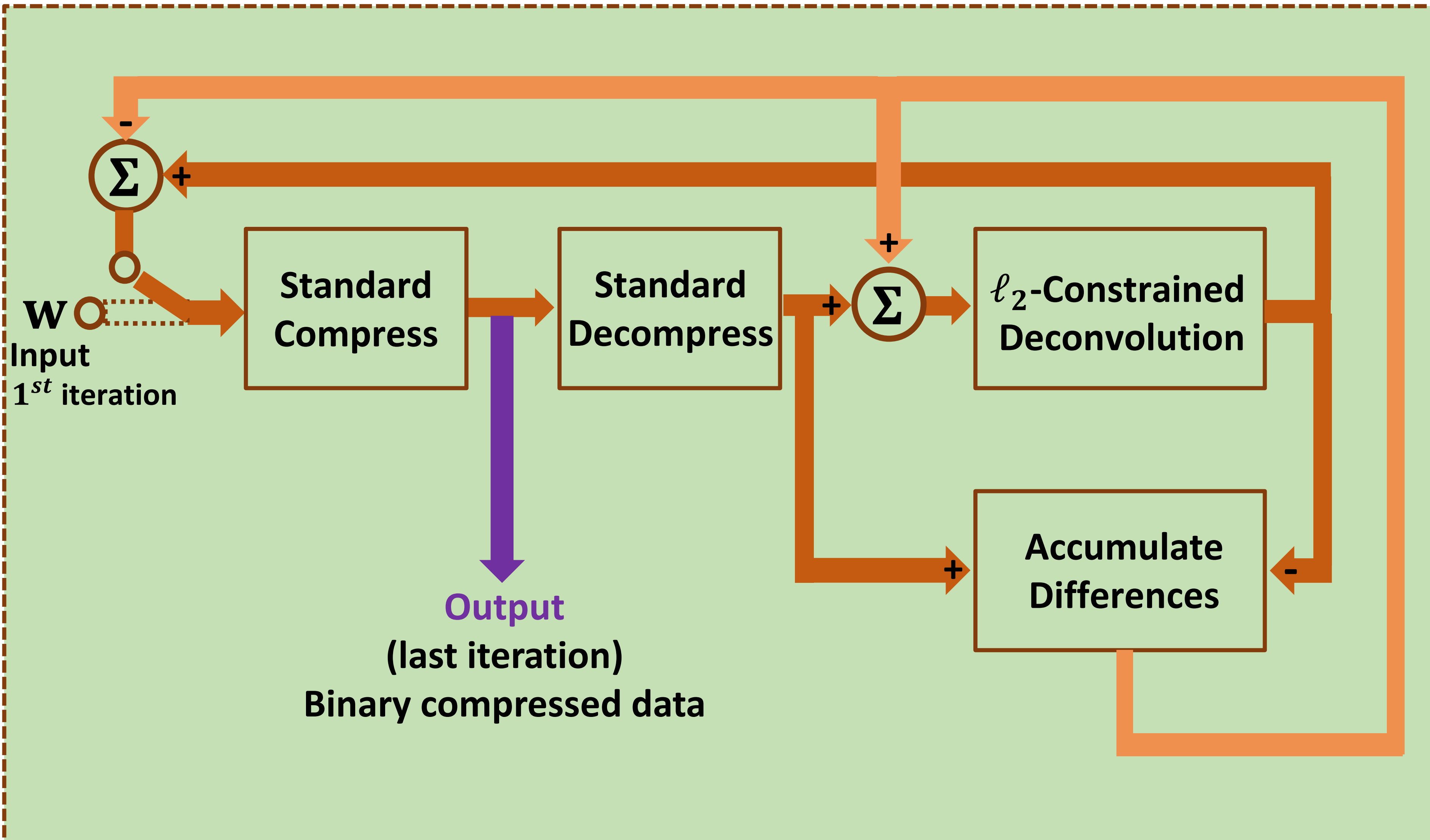
Examples: HEVC/H.265, JPEG2000, …

Simple Deconvolution

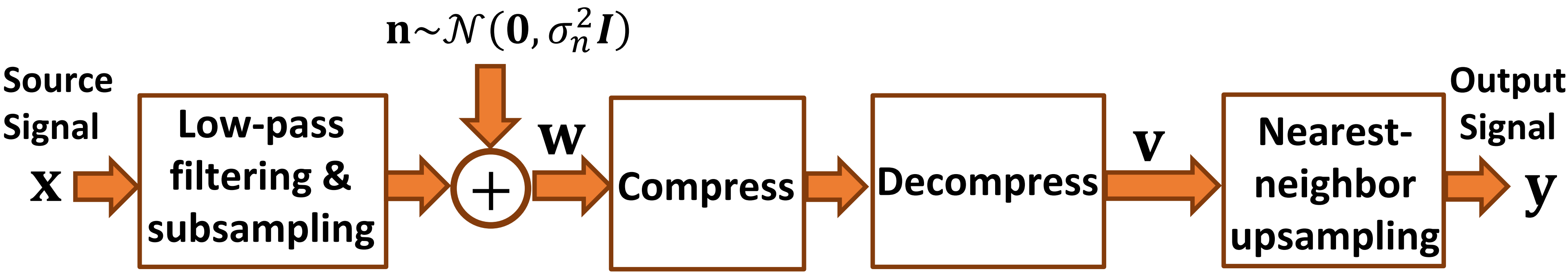
If $\mathbf{AB}$ blurs, this stage sharpens

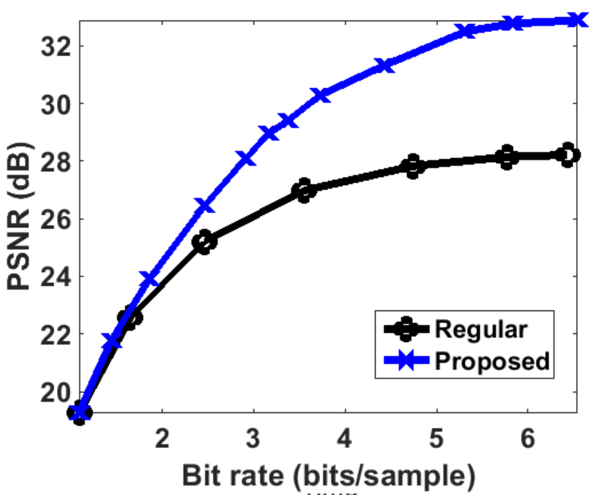
Design Overview

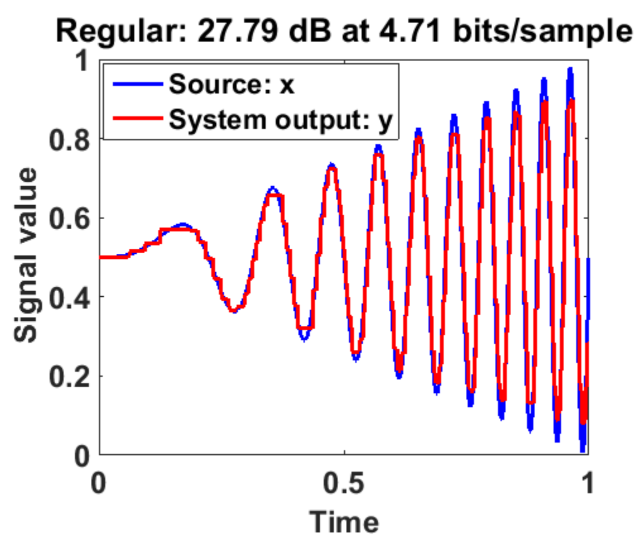
The Proposed Compression Method

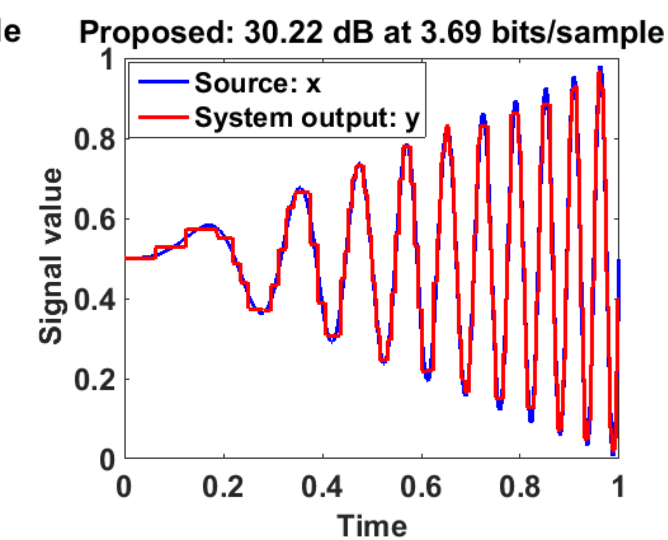


Experiments: Acquisition-Rendering Systems



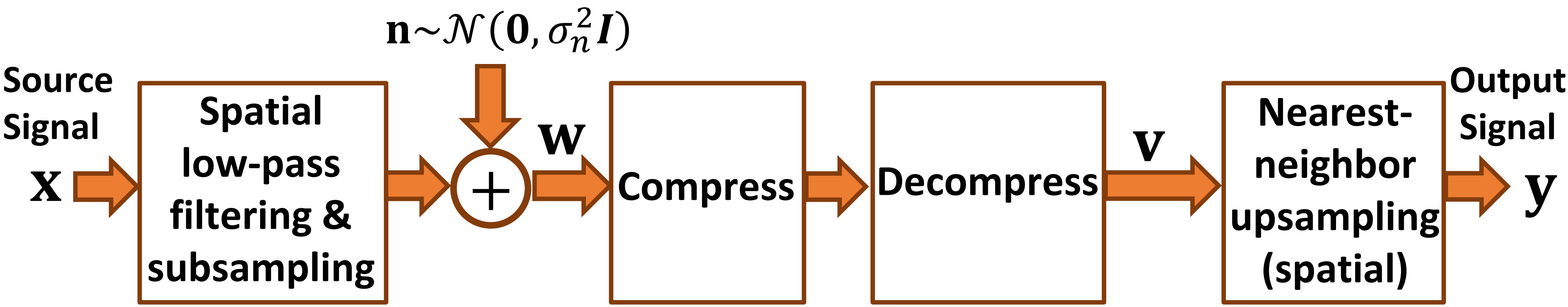


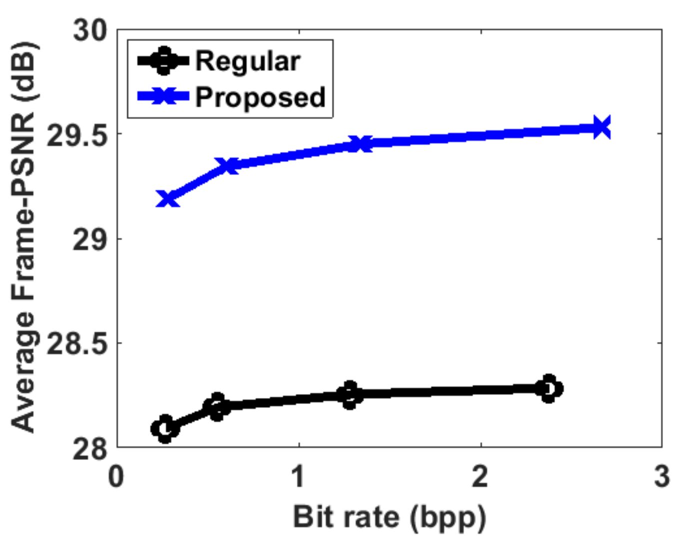


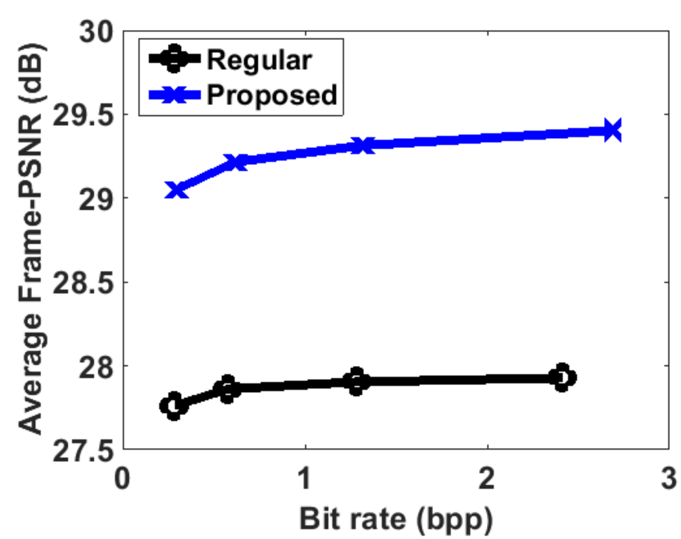


Results for 1D signals based on an adaptive tree-based coding technique.

Experiments: Acquisition-Rendering Systems







Results for video coding based on the HEVC standard.

Experiments: Acquisition-Rendering Systems

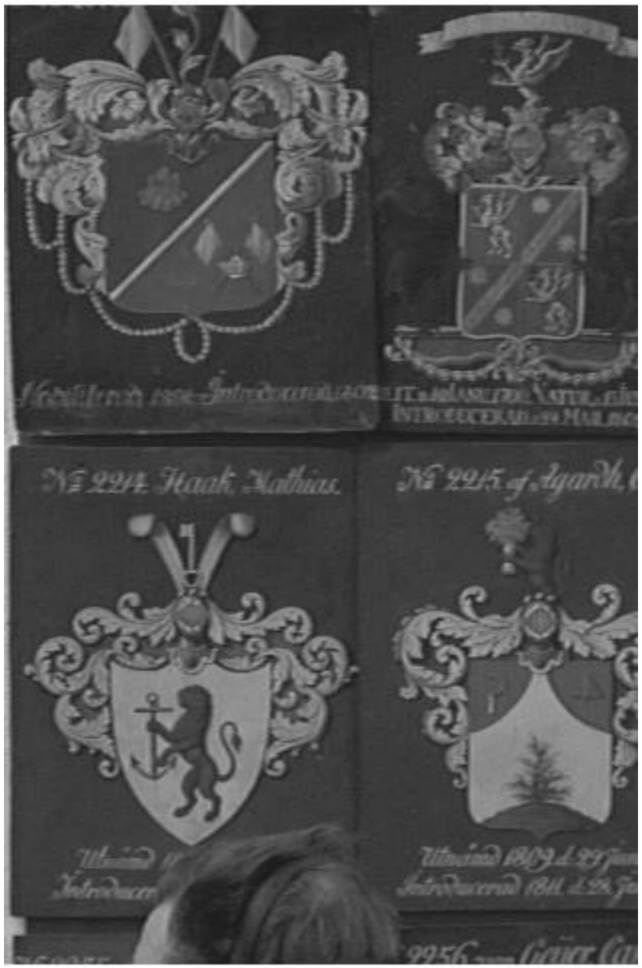
Source frame

Rendered frame: HEVC
28.29 dB at 2.37 bpp

Rendered frame: Proposed
29.45 dB at 1.34 bpp

Experiments: Acquisition-Rendering Systems

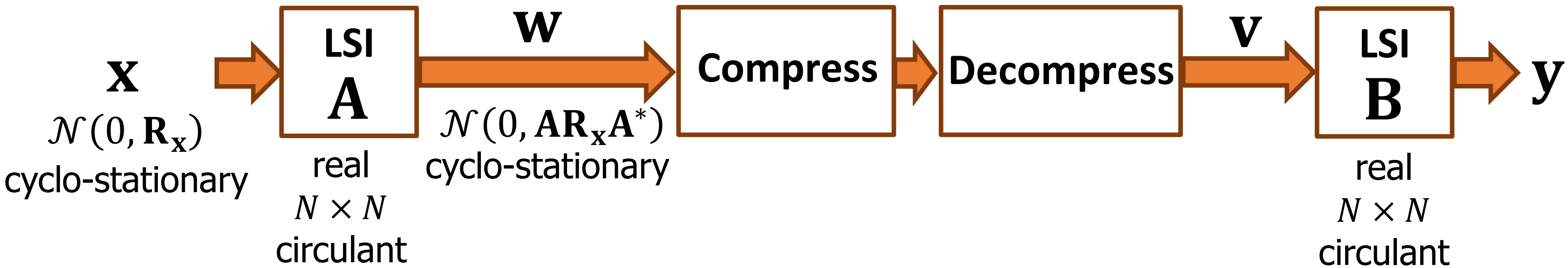
Source frame



Rendered frame: HEVC
27.93 dB at 2.41 bpp

Rendered frame: Proposed
29.31 dB at 1.31 bpp

Rate-Distortion Theoretic Analysis



$\left\{\lambda_k^{(\mathbf{x})}\right\}_{k=0}^{N-1}$ $\left\{a_k^F\right\}_{k=0}^{N-1}$ eigenvalues obtained via DFT-based diagonalization $\left\{b_k^F\right\}_{k=0}^{N-1}$

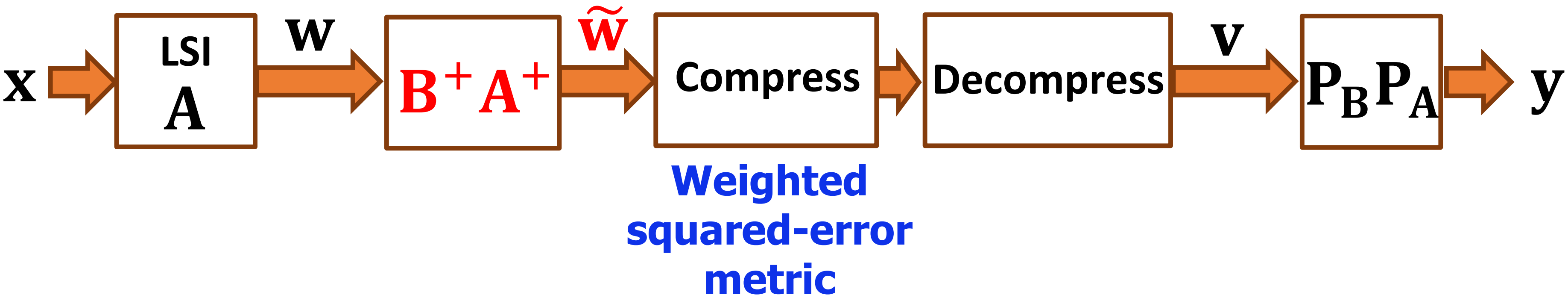
Rate-Distortion Problem #1: **Basic Form**

$$\min_{p_{\mathbf{v}|\mathbf{w}}} \; I(\mathbf{w}; \mathbf{Bv})$$

$$\text{s.t.} \quad N E\{D_0\} \le E\{\|\mathbf{w} - \mathbf{ABv}\|_2^2\} \le N(E\{D_0\} + D)$$

Minimal expected distortion: $E\{D_0\} = \frac{1}{N} \sum_{k:\, a_k^F \neq 0\,,\, b_k^F = 0} \left|a_k^F\right|^2 \lambda_k^{(\mathbf{x})}$ → defined by the intersection of $\mathbf{B}$'s nullspace and $\mathbf{A}$'s range.

Rate-Distortion Theoretic Analysis



Rate-Distortion Problem #2:

Pseudoinverse filtering of compression input

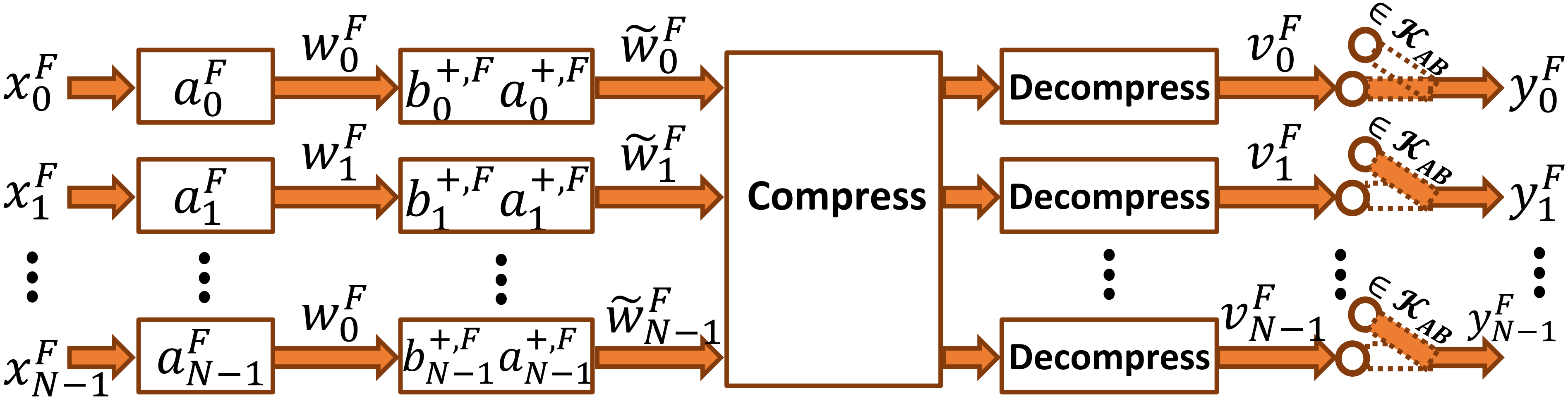
$$\min_{p_{\mathbf{v}|\tilde{\mathbf{w}}}} \; I(\tilde{\mathbf{w}}; \mathbf{P_B P_A v})$$

$$\text{s.t.} \quad E\{\|\mathbf{AB}(\tilde{\mathbf{w}} - \mathbf{v})\|_2^2\} \le ND$$

where $\tilde{\mathbf{w}} = \mathbf{B}^+\mathbf{A}^+\mathbf{w}$

$\mathbf{P_A}$ and $\mathbf{P_B}$ project onto $\mathbf{A}$ and $\mathbf{B}$ ranges, respectively.

Rate-Distortion Theoretic Analysis



Rate-Distortion Problem #3:
DFT-domain distortion allocation for independent Gaussian variables

$$\min_{\{D_k\}_{k\in\mathcal{K}_{AB}}} \sum_{k\in\mathcal{K}_{AB}} \frac{1}{2}\log\left(\frac{\lambda_k^{(\widetilde{\mathbf{w}})}}{D_k}\right)$$

$$\text{s.t.} \sum_{k\in\mathcal{K}_{AB}} \left|a_k^F b_k^F\right| D_k \leq ND$$

$$0 \leq D_k \leq \lambda_k^{(\widetilde{\mathbf{w}})} \quad , k \in \mathcal{K}_{AB}$$

DFT-domain range of AB: $\mathcal{K}_{AB} = \{k \mid a_k^F \neq 0 \text{ and } b_k^F \neq 0\}$

Rate-Distortion Theoretic Analysis

System & component specific water levels: $\theta_k^{(AB)} \triangleq \dfrac{\theta}{\left|a_k^F b_k^F\right|^2}$

Optimal distortion allocation $(k \in \mathcal{K}_{AB})$

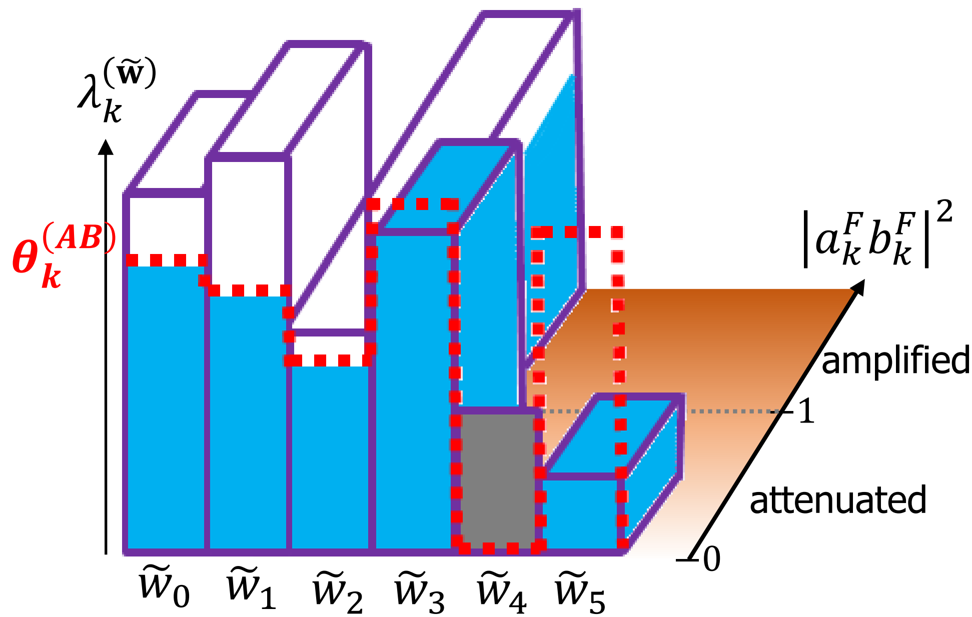
$$D_k = \begin{cases} \theta_k^{(AB)} & , 0 \le \theta_k^{(AB)} < \lambda_k^{(\tilde{\mathbf{w}})} \\ \lambda_k^{(\tilde{\mathbf{w}})} & , \theta_k^{(AB)} \ge \lambda_k^{(\tilde{\mathbf{w}})} \end{cases}$$

where θ is set such that $\sum_{k \in \mathcal{K}_{AB}} \left|a_k^F b_k^F\right| D_k = ND$

Optimal rate allocation

$$R_k = \begin{cases} \frac{1}{2} \log\left(\frac{\lambda_k^{(\tilde{\mathbf{w}})}}{\theta_k^{(AB)}}\right) & , k \in \mathcal{K}_{AB} \text{ and } 0 \le \theta_k^{(AB)} < \lambda_k^{(\tilde{\mathbf{w}})} \\ 0 & , \text{otherwise} \end{cases}$$

Reverse water-filling of **volumes** imposed by $\tilde{\mathbf{w}}$ components' variances and the **system operators**.



System nullspace components are ignored and not coded.

Conceptual Resemblances Between Theory and Practice

Analysis Gaussian signals and LSI operators	**Proposed Method** Non-Gaussian signals and linear operators
Extension of the **standard** distortion allocation	Repeated application of a **standard** compression technique
Pseudoinverse filtering of compression input	The ℓ_2-constrained deconvolution is a **softened pseudoinverse filter**: $\hat{\mathbf{z}}^{(t)} = \left(\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{B} + \tilde{\beta}\mathbf{I}\right)^{-1}\left(\mathbf{B}^*\mathbf{A}^*\mathbf{A}\mathbf{B}\tilde{\mathbf{w}} + \tilde{\beta}\tilde{\mathbf{v}}^{(t)}\right)$
$\tilde{\mathbf{w}}$ components **importance** determined by **$\mathbf{AB}$ spectrum**	Also applied, e.g., for LSI operators: $\hat{z}_k^{F,(t)} = \frac{\left\vert a_k^F b_k^F \right\vert^2 \tilde{w}_k^F + \tilde{\beta}\tilde{v}_k^{F,(t)}}{\left\vert a_k^F b_k^F \right\vert^2 + \tilde{\beta}}$

Conclusions

- **Practical** framework for addressing **complicated rate-distortion optimizations**.
  - Demonstrated for optimizing **acquisition-rendering** systems
  - Analyzed for Gaussian signals and LSI operators.

Our contributions with respect to existing works:

- **Operational** rate-distortion optimization for a **given signal**.
- Distortion metric imposed by the **lack of a statistical source model**.
- **Practical** compression methodology relying on **ADMM** and **standard compression** techniques.
- Applications to **image and video** compression.

Conclusions

For additional details and related studies see:

Y. Dar, M. Elad, A. M. Bruckstein, **"Optimized Pre-Compensating Compression"**, **IEEE Trans. on Image Processing**, 2018.

Y. Dar, M. Elad, A. M. Bruckstein, **"Restoration by Compression"**, ***Submitted*** *to* **IEEE Trans. on Signal Processing**, arXiv preprint:1711.05147.

Y. Dar, A. M. Bruckstein, M. Elad, R. Giryes, **"Postprocessing of Compressed Images via Sequential Denoising"**, **IEEE Trans. on Image Processing**, 2016.

Y. Dar, M. Elad, A. M. Bruckstein, **"System-Aware Compression"**, **IEEE International Symposium on Information Theory** (ISIT) 2018.

Y. Dar, M. Elad, A. M. Bruckstein, **"Compression for Multiple Reconstructions"**, **IEEE International Conference on Image Processing** (ICIP) 2018.

Y. Dar, A. M. Bruckstein, M. Elad, **"Image Restoration via Successive Compression"**, in **Picture Coding Symposium** 2016.