Discrete Denoising for Channels with Memory

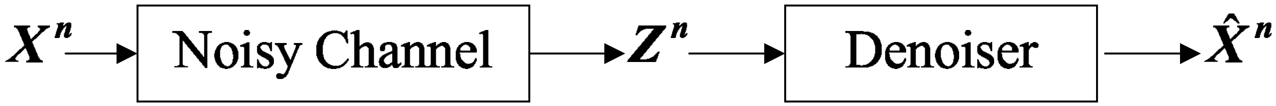
Tsachy Weissman

Joint work with
Rui Zhang

Department of Electrical Engineering

Stanford University

Denoising



- $X^n = (X_1, \ldots, X_n)$ is a noise-free signal of interest corrupted by a channel
- Observe a noise-corrupted sequence $Z^n = (Z_1, \ldots, Z_n)$
- Objective is to estimate X^n from Z^n
- Λ measures goodness of reconstruction $\hat{X}^n$: $\frac{1}{n}\sum_{t=1}^{n} \Lambda(X_t, \hat{X}_t)$

Discrete Denoising

- X_i, Z_i, $\hat{X}_i$ take values in the finite alphabet $\mathcal{A} = \{0, \ldots, M-1\}$

For concreteness, we focus on:

- Modulo-M additive noise $Z_i = X_i \oplus N_i$
- Channel characterized by noise process $\{N_i\}$

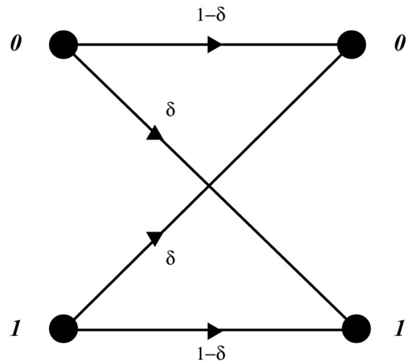
Applications

- Text Correction
- Image Denoising
- Reception of Uncoded Data
- DNA Sequence Analysis and Processing
- Systematic Source/Channel Decoding
- Pattern Recognition
- Separation of Superimposed Signals
- Computer Memory with Defects

⋮

Example

- Binary source sequence: 000111110000111100
- Channel: BSC, i.e., $\{N_i\} \sim$



$$000100000100000101 0$$

- Corrupted Sequence: $\Rightarrow 000011110100111011 0$
- Loss Function: Λ is Hamming loss
- Objective: Minimize Bit Error Rate given the observation of n-block.

Universality Setting

- Noiseless Source unknown
- Channel known

Previous Approaches to Universal Discrete Denoising (for memoryless channels)

- Hidden Markov process modelling: EM + Forward-Backward (aka BCJR, Baum-Welch, Discrete-time Wonham Filtering)

- Compression-based denoising:
  - Occam Filters [Natarajan '95]
  - ⋮
  - * Kolmogorov Sampler [Donoho '02]
    * Empirical Dist. of R-D Codes [Weissman & Ordentlich '04]

Question:

In this setting, can optimum source-dependent performance be attained?

DUDE Algorithm: General Idea

(Weissman, Ordentlich, Seroussi, Verdú & Weinberger '05)

- Fix context length k. For each letter to be denoised, do:

- Find left k-context $(\ell_1, \ldots, \ell_k)$ and right k-context $(r_1, \ldots, r_k)$

ℓ_1	ℓ_2	$\cdots$	ℓ_k	$\bullet$	r_1	r_2	$\cdots$	r_k

- Count all occurrences of letters with left k-context $(\ell_1, \ldots, \ell_k)$ and right k-context $(r_1, \ldots, r_k)$.

- Make decision using
  - the loss function
  - the channel transition matrix
  - the count vector
  - the observed letter to be denoised.

- Note: Decision rule does not depend on n, k, nor on $(\ell_1, \ldots, \ell_k)$ and $(r_1, \ldots, r_k)$

Example: DUDE for BSC + Hamming loss (error rate)

For each bit b , count how many bits that have the same left and right k-contexts are equal to b and how many are equal to $\bar{b}$. If the ratio of these counts is below

$$\frac{2\delta(1-\delta)}{(1-\delta)^2+\delta^2}$$

then b is deemed to be an error introduced by the BSC.

Properties of the DUDE

- Universally achieves asymptotically optimum performance
- Linear complexity in time and space
- Does well in practice

Some follow-up work on DUDE

- (Dembo & Weissman '04)
  "Universal Denoising for the Finite-Input-General-Output Channel"
- (Gemelos, Sigurjónsson & Weissman '04)
  "Universal Minimax Discrete Denoising Under Channel Uncertainty"
- (Ordentlich, Weissman, Weinberger, Somekh-Baruch & Merhav '04)
  "Discrete Universal Filtering Through Incremental Parsing"
- (Ordentlich, Weinberger & Weissman '04)
  "Efficient pruning of multi-directional context trees with applications to universal denoising and compression"
- (Ordentlich, Seroussi, Verdú, Viswanathan, Weinberger & Weissman '04)
  "Channel Decoding of Systematically Encoded Unknown Redundant Sources,"
- (Chen, Diggavi, Dusad & Muthukrishnan '05)
  "Efficient String Matching Algorithms for Combinatorial Universal Denoising,"
- (Yu & Verdú '05)
  "Schemes for Bi-Directional Modeling of Discrete Stationary Sources,"

All consider memoryless channels

Our Focus: Channels with Memory

Assumption 1. *Noise process,* $\{N_i\}$ *, is*

- *stationary*
- α*-mixing with* $\sum_{t=1}^{\infty} \alpha_t < \infty$

where the α-mixing coefficients are defined as:

$$\alpha_t = \sup_{\{k \leq l \leq m \leq n:\ m-l \geq t\}} \max_{u_k^l, u_m^n} \left| P(N_k^l = u_k^l, N_m^n = u_m^n) - P(N_k^l = u_k^l) P(N_m^n = u_m^n) \right|$$

Channels with Memory (cont.)

Define the $M^{2k+1} \times M^{2k+1}$ channel transition matrix as

$$\Pi^k_{-k}(x^k_{-k}, z^k_{-k}) = P(N^k_{-k} = z^k_{-k} \ominus x^k_{-k})$$

Assumption 2. Π^k_{-k} *is non-singular for every* k*.*

Note the relation

$$P^T_{Z^k_{-k}} = P^T_{X^k_{-k}} \cdot \Pi^k_{-k}, \tag{1}$$

which implies, with Assumption 2,

$$P^T_{X^k_{-k}} = P^T_{Z^k_{-k}} \cdot \left(\Pi^k_{-k}\right)^{-1}$$

Optimum Denoising: Known Source

$$\hat{X}_i^{opt}(z^n) = \text{arg}\min_{\hat{x}\in\mathcal{A}} E[\Lambda(X_i, \hat{x})|Z^n = z^n]$$

Motivating Derivation

- Optimum kth order sliding-window denoiser (source distribution known)

$$\hat{X}_0^{opt}(z_{-k}^k) = \arg\min_{\hat{x}} E[\Lambda(X_0,\hat{x})|Z_{-k}^k = z_{-k}^k]$$

$$= \arg\min_{\hat{x}} \sum_a \Lambda(a,\hat{x}) \left[\sum_{x_{-k}^k : x_0 = a} P_{X_{-k}^k}(x_{-k}^k) P_{N_{-k}^k}(z_{-k}^k \ominus x_{-k}^k) \right]$$

- Motivated by $P^T_{X_{-k}^k} = P^T_{Z_{-k}^k} \cdot \left(\Pi_{-k}^k\right)^{-1}$, we take

$$\hat{X}_i(z_{i-k}^{i+k}) = \arg\min_{\hat{x}} \sum_a \Lambda(a,\hat{x}) \left[\sum_{x_{-k}^k : x_0 = a} \left[\hat{P}^T_{Z_{-k}^k} \cdot \left(\Pi_{-k}^k\right)^{-1} \right] (x_{-k}^k) P_{N_{-k}^k}(z_{i-k}^{i+k} \ominus x_{-k}^k) \right]$$

Let $\hat{X}^{n,k}$ denote the overall denoiser obtained.

Computation of $(\Pi^k_{-k})^{-1}$

Note that $\Pi^k_{-k}\left(x^k_{-k}, z^k_{-k}\right) = \Pi^k_{-k}\left(\tilde{x}^k_{-k}, \tilde{z}^k_{-k}\right)$ whenever $z^k_{-k} \ominus x^k_{-k} = \tilde{z}^k_{-k} \ominus \tilde{x}^k_{-k}$.

Theorem 1. *Let $\mathcal{F}_M$ denote the $M \times M$ Fourier matrix*

$$\mathcal{F}_M(l,m) = \frac{1}{\sqrt{M}} \exp\left\{-j\frac{2\pi}{M} lm\right\}$$

and

$$\mathcal{H}_n = \mathcal{F}_M^{\otimes n}.$$

(a) *$\mathcal{H}_{2k+1}$ diagonalizes Π^k_{-k}, i.e., $\Pi^k_{-k} = \mathcal{H}^H_{2k+1}\Gamma\mathcal{H}_{2k+1}$, where Γ is diagonal $M^{2k+1} \times M^{2k+1}$.*
(b) *diag$(\Gamma) = \mathcal{H}_{2k+1} \cdot P_{N^k_{-k}}$.*

Thus we get

- $\hat{P}_{X^k_{-k}} = \left(\Pi^k_{-k}\right)^{-T} \cdot \hat{P}_{Z^k_{-k}} = \mathcal{H}_{2k+1} \cdot \left[\left(\mathcal{H}^*_{2k+1} \cdot \hat{P}_{Z^k_{-k}}\right) \oslash \left(\mathcal{H}_{2k+1} \cdot P_{N^k_{-k}}\right)\right]$,
- Computation is $O(kM^{2k})$, compared with $O(M^{6k})$ of direct computation.

Example: in case $M = 2$, each Π^k_{-k} is diagonalized by the Hadamard transform (Iordache, Tăbus & Astola '02) and (Giurcăneanu & Yu '05)

Complexity

M = Alphabet size; k = Order of sliding-window denoiser; n =Data block length

- **Pre-processing.**
  $(\Pi^k_{-k})^{-1}$: $O(kM^{2k})$
- **Computation of counts.**
  $\hat{P}_{Z^k_{-k}}[z^n]$: $O(kn)$
- **Computation of decoding rule.**
  $\hat{P}_{X^k_{-k}}$: $O(M^{4k})$
  $\left\{\hat{X}_0(z^k_{-k})\right\}_{z^k_{-k}}$: $O(M^{4k})$
- **Denoising.**
  $\hat{X}_i(z^{i+k}_{i-k})$: $O(kn)$

Total number of operations: $O(kn + M^{4k}) = O(n \log n)$ provided $k_n = c \log n$

Total space: $O(n)$

Selection of k

- Complexity:
  - We have seen that $k_n = c \log n$ gives $O(n \log n)$ complexity
- More basic tradeoff:
  - k too short $\mapsto$ suboptimum performance
  - k too long ($\Leftrightarrow$ too short n) $\mapsto$ counts are unreliable

Performance Criterion

Formally, a denoiser $\hat{X}^n$ is a mapping $\mathcal{A}^n \to \mathcal{A}^n$. For any x^n, z^n let

$$L_{\hat{X}^n}(x^n, z^n) = \frac{1}{n} \sum_{t=1}^{n} \Lambda(x_t, \hat{X}_t(z^n)),$$

where Λ is the given loss function.

Universal Asymptotic Optimality

Theorem 2. *Let* $\hat{X}^n_{univ} \triangleq \hat{X}^{n,k_n}$ *where* $k_n \to \infty$ *and satisfies*

$$\frac{1}{n} k_n M^{12k_n} \left\| \left(\Pi^{k_n}_{-k_n} \right)^{-1} \right\|^2 \longrightarrow 0 \quad \textit{as} \quad n \to \infty.$$

1. *Stochastic Setting : For any stationary process* $\mathbf{X} = (X_1, X_2 \ldots)$

$$\lim_{n\to\infty} EL_{\hat{X}^n_{univ}}(X^n, Z^n) = \lim_{n\to\infty} \min_{\hat{X}^n} EL_{\hat{X}^n}(X^n, Z^n)$$

where the minimization on the right side is over all denoisers.

2. *Semi-Stochastic Setting : For all* $\{x_n\}_{n\geq 1}$, $x^n \in \mathcal{A}^n$,

$$L_{\hat{X}^n_{univ}}(x^n, Z^n) - D_{k_n}(x^n, Z^n) \longrightarrow 0 \quad \textit{in probability}$$

where $D_k(x^n, z^n) = \min_{f:\mathcal{A}^{2k+1}\to\mathcal{A}} \left[\frac{1}{n-2k} \sum_{i=k+1}^{n-k} \Lambda\left(x_i, f(z_{i-k}^{i+k})\right) \right]$.

Example 1: Memoryless Noise

- For a memoryless channel, $\Pi_{-k}^{k} = (\Pi_{-0}^{0})^{\otimes(2k+1)}$

- Therefore, $\left\| \left(\Pi_{-k}^{k}\right)^{-1} \right\| = \left\| \left(\Pi_{-0}^{0}\right)^{-1} \right\|^{2k+1}$

- $k_n = c \log n$ suffices for

$$\frac{1}{n} k_n M^{12k_n} \left\| \left(\Pi_{-k_n}^{k_n}\right)^{-1} \right\|^2 \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

Remarks:

1. Can be shown using ideas similar to those in [Dembo & Weissman '04] that our scheme coincides with the DUDE in this case

2. Bounds in DUDE paper allow $k_n = C \log n$, for $C > c$

Example 2: Binary Noise Modulated by An Arbitrarily Distributed State Process

- Let $\{S_i\}$ be an arbitrarily distributed state process and $\{N_i\}$ be a binary process whose components are independent when conditioned on $\{S_i\}$, where $N_i | S_i = s \sim \text{Bernoulli}(\delta_s)$ for every $s \in \mathcal{S}$

- Let $\delta = \sup_{s \in \mathcal{S}} \delta_s$ and assume $\delta < 1/2$

- It can be shown that $\left\| \left(\Pi_{-k}^{k}\right)^{-1} \right\| \leq 1/(1-2\delta)^{2k+1}$

- $k_n = c \log n$ suffices for

$$\frac{1}{n} k_n M^{12k_n} \left\| \left(\Pi_{-k_n}^{k_n}\right)^{-1} \right\|^2 \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

Example 3: Contagion Channels

- Contagion channels (F. Alajaji & T. Fuja '94) are binary additive noise channels, where the noise process is an M-th order Markov process with transition probabilities characterized by $P(N_t = 1 | N_{t-M}^{t-1} = n_{t-M}^{t-1}) = \frac{\varepsilon + w(n_{t-M}^{t-1})\delta}{1+M\delta}$, where w denotes Hamming weight, $\varepsilon = P(N_t = 1)$

- Can show

$$\left\| \left(\Pi_{-k}^{k}\right)^{-1} \right\| \leq \left(\frac{1-2\varepsilon}{1+M\delta} \right)^{-(2k+1)}$$

- $k_n = c \log n$ suffices for

$$\frac{1}{n} k_n M^{12k_n} \left\| \left(\Pi_{-k_n}^{k_n}\right)^{-1} \right\|^2 \longrightarrow 0 \quad \text{as} \quad n \to \infty$$

Proof Sketch: Semi-stochastic Setting

- To show

$$L_{\hat{X}^n_{univ}}(x^n, Z^n) - D_{k_n}(x^n, Z^n) \longrightarrow 0 \quad \text{in probability}$$

- Define

$$q_k(z^n, x^n)[a, u_{-k}^k] = \frac{1}{n-2k} |\{k+1 \le i \le n-k : x_i = a, z_{i-k}^{i+k} = u_{-k}^k\}|$$

$$\hat{q}_k(z^n)[a, u_{-k}^k] = \sum_{x_{-k}^k : x_0 = a} \left[\hat{P}_{Z_{-k}^k}[z^n]^T \cdot \left(\Pi_{-k}^k\right)^{-1} \right] (x_{-k}^k) P_{N_{-k}^k}(u_{-k}^k \ominus x_{-k}^k)$$

- With the following fact

$$|L_{\hat{X}^{n,k}}(x^n, z^n) - D_k(x^n, z^n)| \le \Lambda_{max} M^{2k+2} \|q_k(z^n, x^n) - \hat{q}_k(z^n)\|$$

- It is sufficient to show

$$P\left(\|\hat{q}_k(Z^n) - q_k(Z^n, x^n)\| \ge \epsilon\right) \le M^{8k+2} \frac{\left(4k + 1 + 2\sum_{t=1}^{\infty} \alpha_t^{(N)}\right) \left\| \left(\Pi_{-k}^k\right)^{-1} \right\|^2}{\epsilon^2 (n-2k)}$$

Proof Sketch: Stochastic Setting

- To show

$$\lim_{n\to\infty} EL_{\hat{X}^n_{univ}}(X^n, Z^n) = \lim_{n\geq\infty} \min_{\hat{X}^n} EL_{\hat{X}^n}(X^n, Z^n)$$

- It follows from the proof for semi-stochastic setting:

$$P\left(L_{\hat{X}^n_{univ}}(X^n, Z^n) \geq D_k(X^n, Z^n) + \varepsilon\right) \leq \varepsilon$$

$$EL_{\hat{X}^n_{univ}}(X^n, Z^n) \leq ED_k(X^n, Z^n) + \varepsilon + \varepsilon\Lambda_{max} = ED_k(X^n, Z^n) + \varepsilon(1 + \Lambda_{max})$$

- Together with

$$ED_k(X^n, Z^n) \leq E\left[\min_{\hat{x}\in\mathcal{A}} E\left[\Lambda(X_0, \hat{x})|Z_{-k}^k\right]\right]$$

$$\lim_{k\to\infty} E\left[\min_{\hat{x}\in\mathcal{A}} E\left[\Lambda(X_0, \hat{x})|Z_{-k}^k\right]\right] = \inf_{n\geq 1} \min_{\hat{X}^n} EL_{\hat{X}^n}(X^n, Z^n)$$

- It follows that

$$\limsup_{n\to\infty} EL_{\hat{X}^n_{univ}}(X^n, Z^n) \leq \inf_{n\geq 1} \min_{\hat{X}^n} EL_{\hat{X}^n}(X^n, Z^n)$$

Extension 1: General Stationary Channels

Consider a general channel characterized by $\{P(\cdot|x_{-\infty}^{\infty})\}_{x_{-\infty}^{\infty}}$ and satisfying

1. stationarity

2. $P(z^k_{-k}|x_{-\infty}^{\infty}) = P(z^k_{-k}|\tilde{x}_{-\infty}^{\infty})$ whenever $x^k_{-k} = \tilde{x}^k_{-k}$, $\forall k$

   [then take $\Pi^k_{-k}(x^k_{-k}, z^k_{-k}) = P(z^k_{-k}|x^k_{-k})$, and right side will make sense]

3. $\sum_{t=1}^{\infty} \alpha_t < \infty$,

where α-mixing coefficients are now defined as:

$$\alpha_t = \sup_{x_{-\infty}^{\infty}} \sup_{\{k \le l \le m \le n:\ m-l \ge t\}} \max_{u^l_k, u^n_m} \left| P(z^l_k, z^n_m | x^l_k, x^n_m) - P(z^l_k|x^l_k) P(z^n_m|x^n_m) \right|$$

Scheme and its performance guarantees carry over verbatim
[replacing $P_{N^k_{-k}}(z^{i+k}_{i-k} \ominus x^k_{-k})$ by $P(Z^k_{-k} = z^{i+k}_{i-k}|x^k_{-k})$]

Example of Family of Non-Additive Stationary Channels Satisfying Assumptions

- $\{N_t\}$ is stationary noise process as before
- Channel input-output relationship is

$$Z_i = f(x_i, N_{i-l}^{i+l})$$

Extension 2: Non-Stationary Channels

In stationary case $\hat{P}^T_{Z^k_{-k}} \cdot \left(\Pi^k_{-k}\right)^{-1}$ was good estimate of $(2k+1)$ th-order empirical distribution of input sequence. So now replace

$\hat{P}^T_{Z^k_{-k}} \cdot \left(\Pi^k_{-k}\right)^{-1} = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} \mathbf{1}^T_{\{Z^{i+k}_{i-k} = \cdot\}} \cdot \left(\Pi^k_{-k}\right)^{-1}$

by the more general form

$\frac{1}{n-2k} \sum_{i=k+1}^{n-k} \mathbf{1}^T_{\{Z^{i+k}_{i-k} = \cdot\}} \cdot \left(\Pi^{i+k}_{i-k}\right)^{-1}$, leading to $\hat{X}_i(z^{i+k}_{i-k}) =$

$$\arg\min_{\hat{x}} \sum_a \Lambda(a, \hat{x}) \left\{ \sum_{x^k_{-k}: x_0 = a} \left[\sum_{j=k+1}^{n-k} \mathbf{1}^T_{\left\{Z^{j+k}_{j-k} = \cdot\right\}} \cdot \left(\Pi^{j+k}_{j-k}\right)^{-1} \right] (x^k_{-k}) \Pi^{i+k}_{i-k}(x^k_{-k}, z^{i+k}_{i-k}) \right\}$$

Semi-stochastic performance guarantees carry over: For all $\{x_n\}_{n \geq 1}$, $x^n \in \mathcal{A}^n$,

$$\limsup_{n \to \infty} \text{ in probability} \left[L_{\hat{X}^n_{univ}}(x^n, Z^n) - D_{k_n}(x^n, Z^n) \right] \leq 0.$$

growth condition for k_n now being

$$\frac{1}{n} k_n M^{12 k_n} \sup_i \left\| \left(\Pi^{i+k_n}_{i-k_n}\right)^{-1} \right\|^2 \longrightarrow 0 \quad \text{as} \quad n \to \infty.$$

Extension 3: Multi-dimensional Index

- Replace contexts by neighborhoods
- Analogous assumptions on α-mixing and non-singularity of channel
- Analogous performance guarantees in both stochastic and semi-stochastic setting

A Variation

- Considered also the following modified denoiser:
  1. Set k , the order of the sliding-window denoiser to be used
  2. Select a value k' , $k' < k$, and obtain $\hat{P}_{X_{-k'}^{k'}} = (\Pi_{-k'}^{k'})^{-T} \cdot \hat{P}_{Z_{-k'}^{k'}}$
  3. Obtain $\hat{P}_{X_{-k}^{k}}$ through left- and right-extension of $\hat{P}_{X_{-k'}^{k'}}$ by assuming $\mathbf{X}$ is a Markov process of order no greater than $2k'$, i.e.,

$$\hat{P}_{X_{-k}^{k}}(x_{-k}^{k}) = \hat{P}_{X_{-k'}^{k'}}\left(x_{-k'}^{k'}\right) \prod_{i=1}^{k-k'} \left[\hat{P}_{X_{-k'}^{k'}}\left(x_{k'+i} | x_{-k'+i}^{k'+i-1}\right) \hat{P}_{X_{-k'}^{k'}}\left(x_{-k'-i} | x_{-k'-i+1}^{k'-i}\right) \right],$$

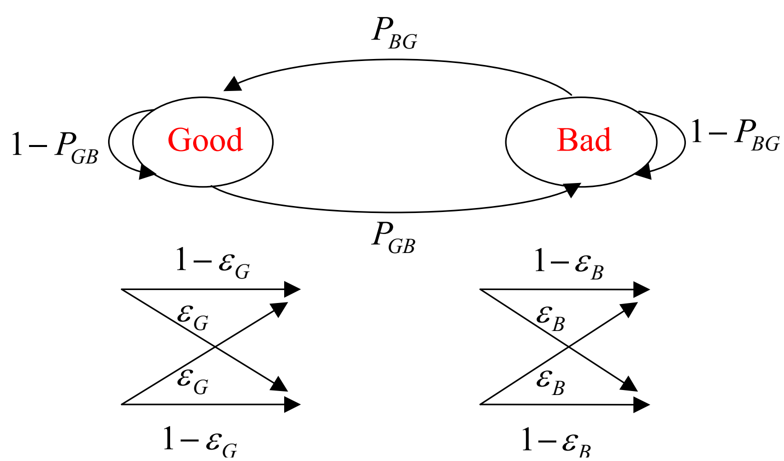
  4. Do the denoising assuming $\hat{P}_{X_{-k}^{k}}$
- The modified denoiser attains:
  - Observed that $\hat{P}_{X_{-k}^{k}}$ thus obtained is closer to $P_{X_{-k}^{k}}$ than $(\Pi_{-k}^{k})^{-T} \cdot \hat{P}_{Z_{-k}^{k}}$
  - Need compute $(\Pi_{-k'}^{k'})^{-1}$ rather than $(\Pi_{-k}^{k})^{-1}$

Similar idea applicable for multi-dimensional data.

Initial justification in (Moon & Weissman'05): "Universal Filtering via Hidden Markov Modelling"

Experiment 1: Burst-Noise Channel Corrupting a 1st-order Markov Chain

- The source sequence is a first-order symmetric binary Markov process with the transition probability, p
- The noise sequence is a binary two-state hidden Markov process with parameters $[\epsilon_G, \epsilon_B, P_{GB}, P_{BG}]$

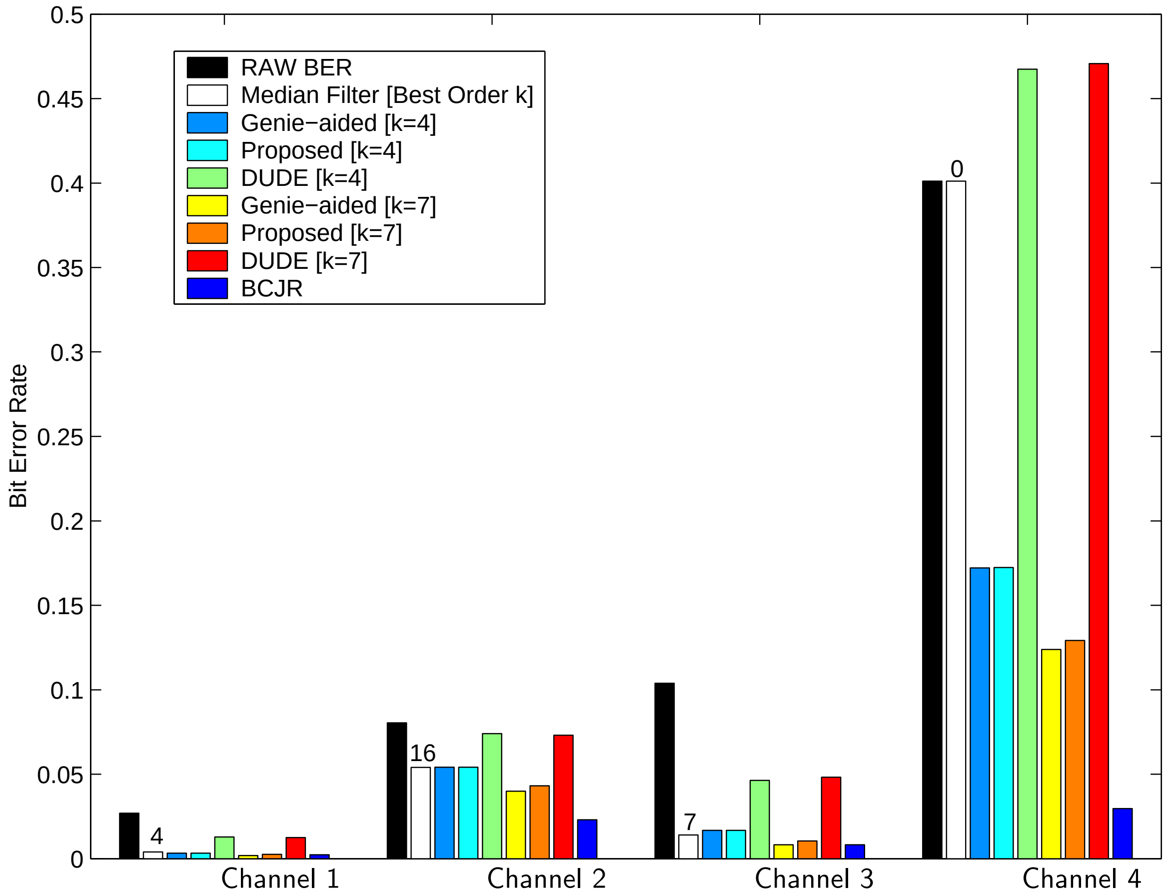


Reference Schemes

- **Median Filter**$[k]$
  The $2k+1$ sliding-window median filter by "majority-vote" decoding
- **Genie-aided**$[k]$
  $\arg\min_{f:\mathcal{A}^{2k+1}\to\mathcal{A}}\left[\frac{1}{n-2k}\sum_{i=k+1}^{n-k}\Lambda\left(x_i, f(z_{i-k}^{i+k})\right)\right]$
- **Proposed**$[k]$
  The proposed universal $2k+1$ sliding-window denoiser. The modified denoiser is used for $k=7$ with $k'=2$.
- **DUDE**$[k]$
  The DUDE for DMC by taking the channel as an equivalent DMC with the cross-over probability, $p_e = \frac{\epsilon_B P_{GB}+\epsilon_G P_{BG}}{P_{GB}+P_{BG}}$
- **BCJR**
  The optimum denoiser with known source statistics, implemented by the BCJR algorithm (the "forward-backward" recursions)

1D Denoising Results: Bit Error Rate

Source transition probability, $p = 0.01$, $n = 10^6$



	Channel 1	Channel 2	Channel 3	Channel 4
$[\varepsilon_G \; \varepsilon_B \; P_{GB} \; P_{BG}]$	[0.01 0.2 0.01 0.1]	[0.01 0.8 0.01 0.1]	[0.01 0.2 0.01 0.01]	[0.01 0.8 0.01 0.01]

Image Denoising: Setup

- The source signals are three binary images: (1) Text Image (Shannon's paper): $10^3 \times 10^3$ (2) Half-toned Image (Einstein's Portrait): 900×900 (3) Black-and-white Image (Lena): 256×256
- The noise sequence is a binary two-state hidden Markov random field with parameters $[\varepsilon_G, \varepsilon_B, \alpha_G, \alpha_B]$
- States are 8-nearest-neighbor Gibbs field characterized by

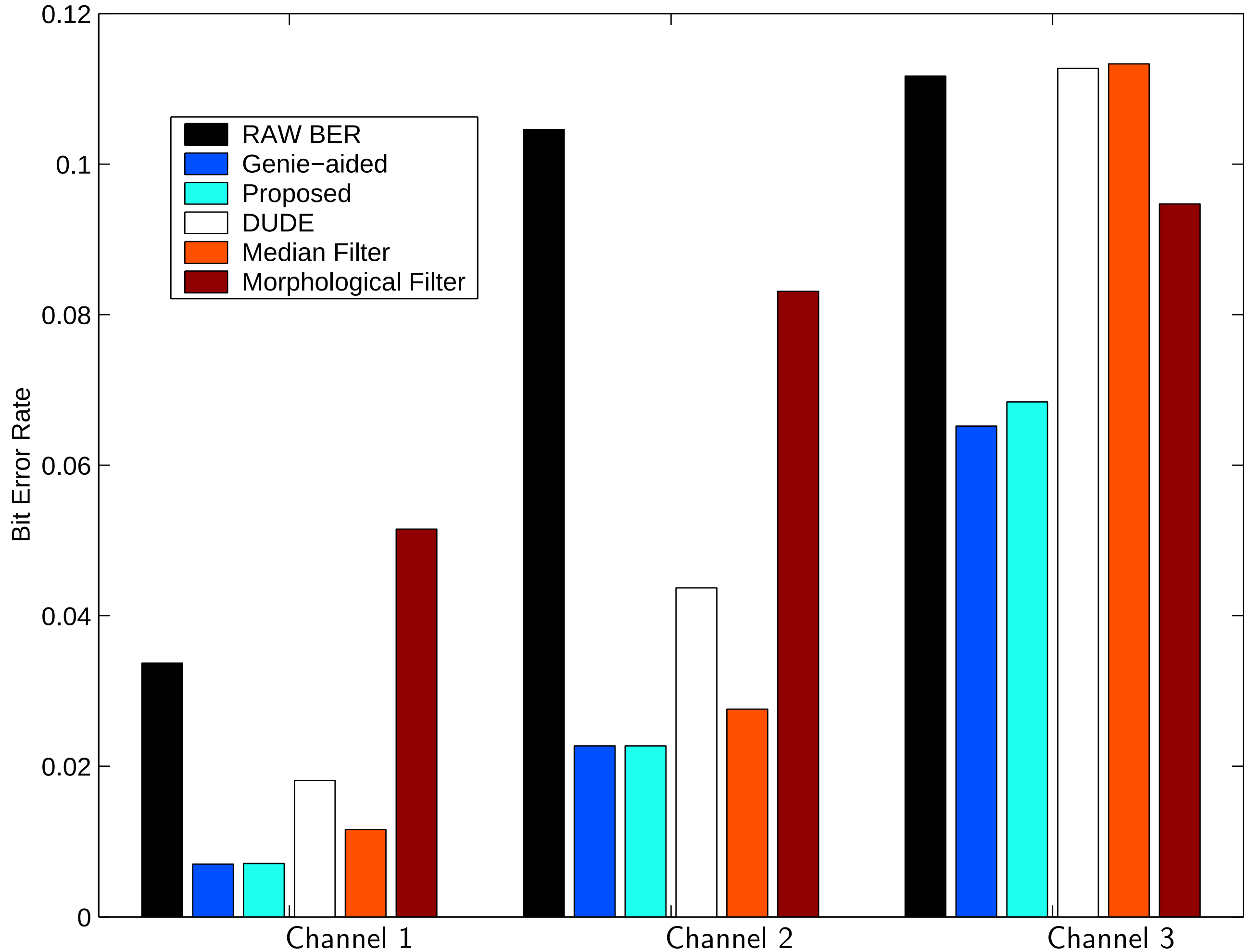
$$P(S_{i,j} = s_{i,j} | S_{\mathcal{N}_{i,j}} = s_{\mathcal{N}_{i,j}}) = \frac{\exp^{-\left[V_1(s_{i,j}) + \sum_{(i,j)} \sum_{(k,l) \in \mathcal{N}_{i,j}} V_2(s_{i,j}, s_{k,l})\right]}}{\sum_{s_{i,j}} \exp^{-\left[V_1(s_{i,j}) + \sum_{(i,j)} \sum_{(k,l) \in \mathcal{N}_{i,j}} V_2(s_{i,j}, s_{k,l})\right]}}$$

  where $s_{i,j} \in \{G, B\}$, $V_1(s_{i,j}) = \alpha_{s_{i,j}}$, $V_2(s_{i,j}, s_{k,l}) = 2\delta(s_{i,j}, s_{k,l}) - 1$
- Two-state Markov random field with 50 Gibbs sampling iterations

Image Denoising: Reference Schemes

- **Genie-aided**
  The best 3×3 sliding-window denoiser
- **Proposed**
  The proposed universal 3×3 sliding-window denoiser
- **DUDE**
  The 3×3 sliding-window DUDE for DMC
- **Median Filter**
  The 3×3 sliding-window median filter
- **Morphological Filter**
  The morphological filter uses a 3×3 structure element and implements the CLOSE and then the OPEN operation to the noise corrupted image

Image Denoising Results: Text Image



	Channel 1	Channel 2	Channel 3
$[\varepsilon_G\ \varepsilon_B\ \alpha_G\ \alpha_B]$	[0.01 0.2 0.2 0]	[0.01 0.2 0 0]	[0.01 0.8 0.2 0]

Image Denoising Results: Text Image Corrupted by Channel 1

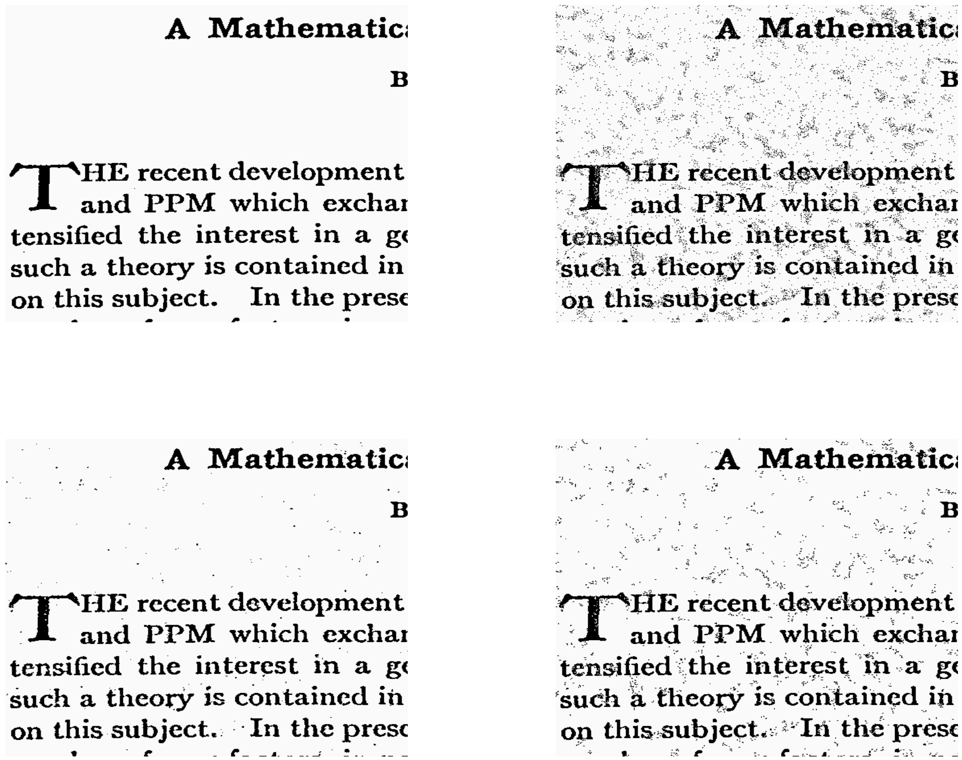
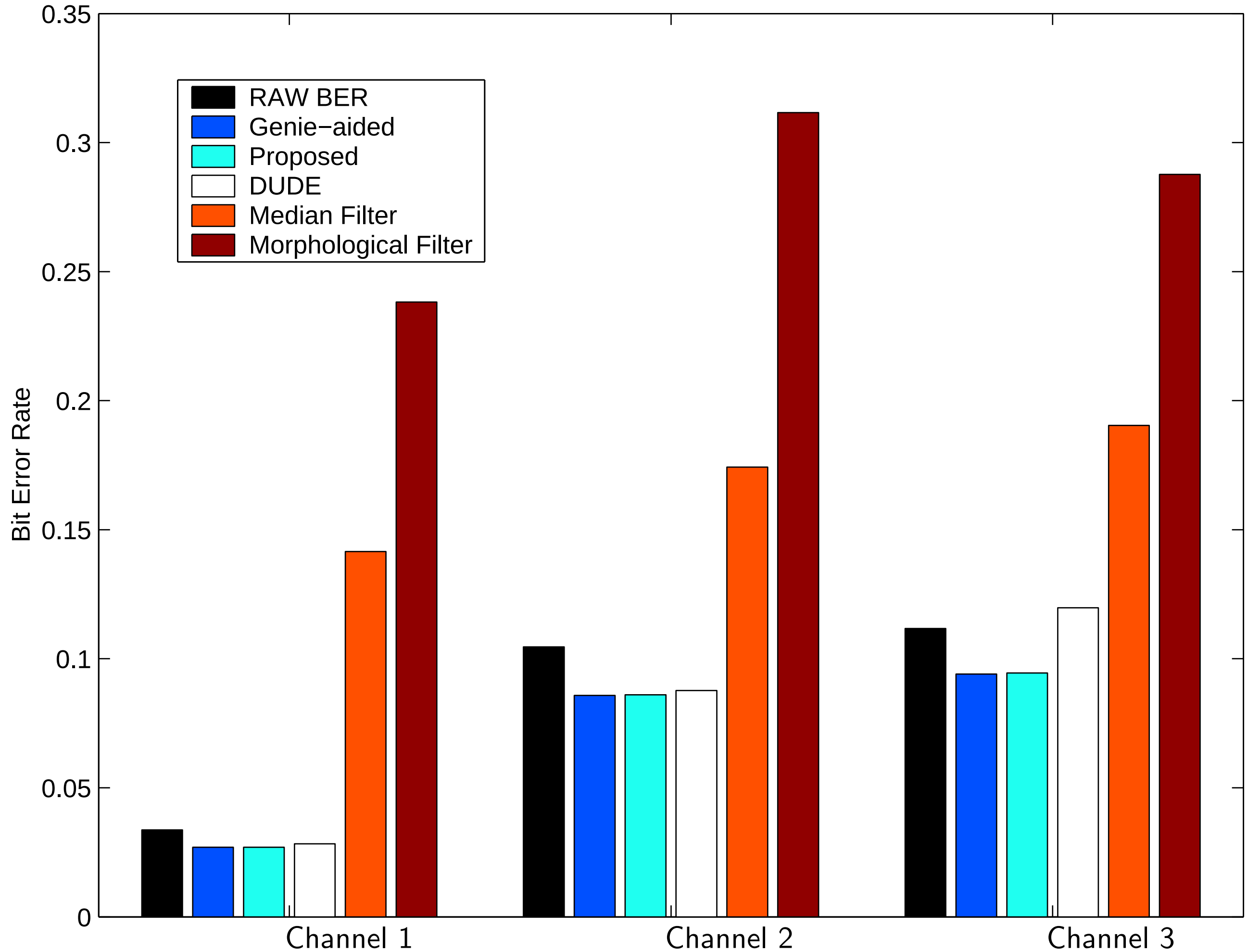


Figure 1: top-left : noiseless image; top-right : noisy image; bottom-left : denoised image by the proposed denoiser; bottom-right: denoised image by DUDE

Image Denoising Results: Half-toned Image



	Channel 1	Channel 2	Channel 3
$[\varepsilon_G\ \varepsilon_B\ \alpha_G\ \alpha_B]$	[0.01 0.2 0.2 0]	[0.01 0.2 0 0]	[0.01 0.8 0.2 0]

Image Denoising Results: Half-toned Image Corrupted by Channel 3

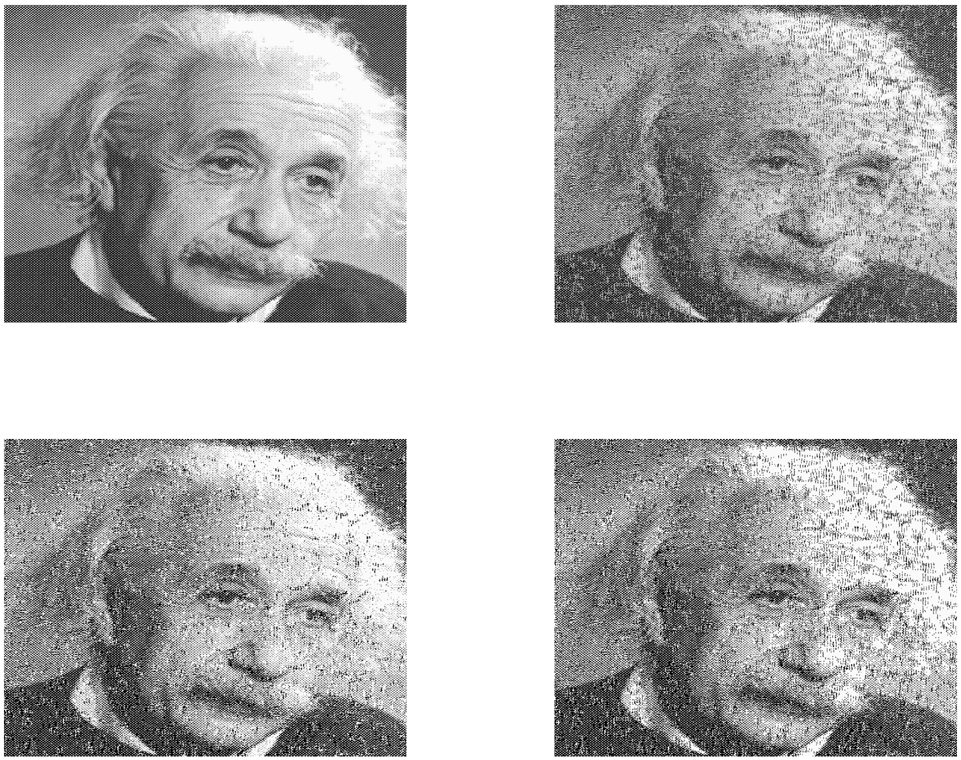
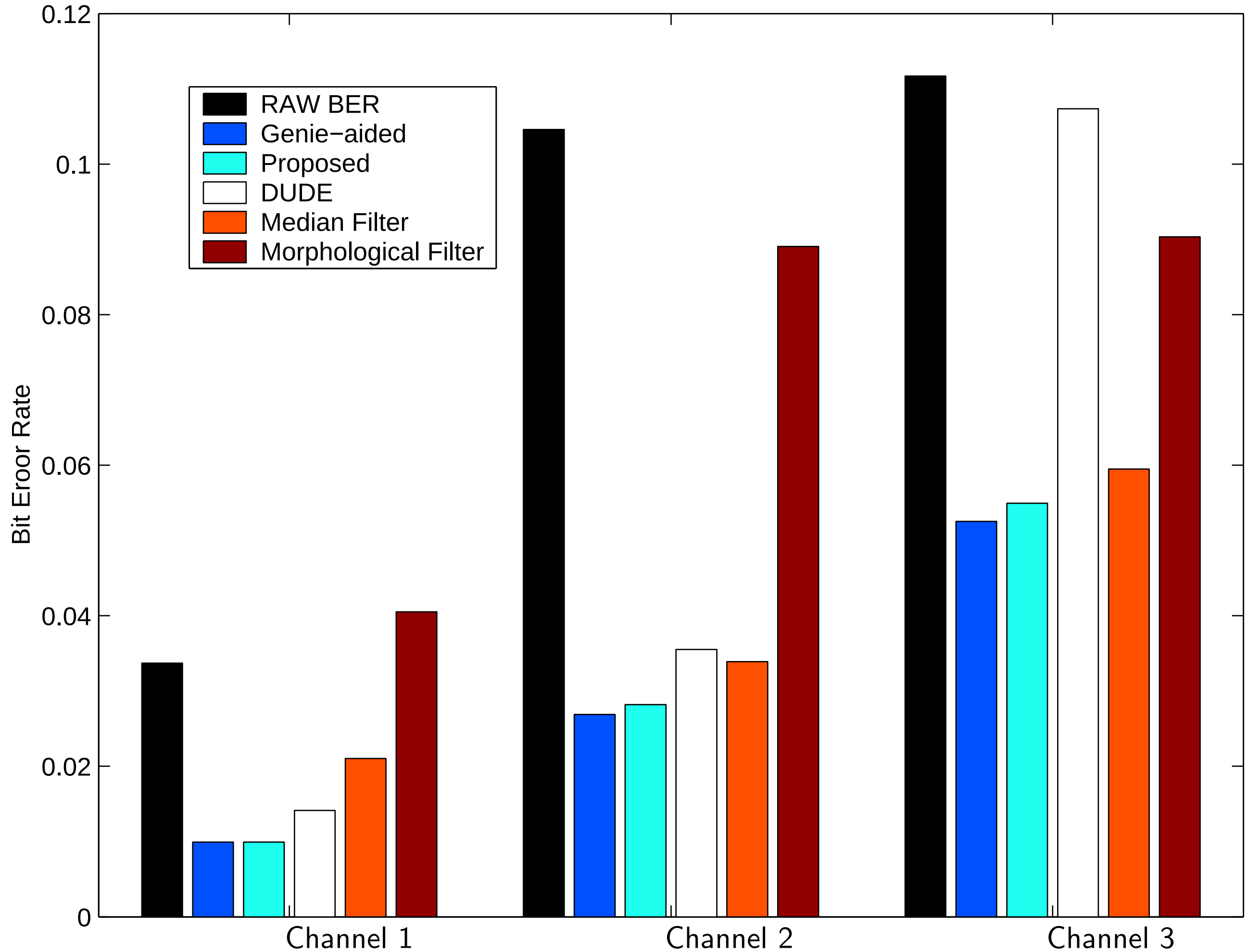


Figure 2: top-left : noiseless image; top-right : noisy image; bottom-left : denoised image by the proposed denoiser; bottom-right: denoised image by DUDE

Image Denoising Results: Black-and-white Image



	Channel 1	Channel 2	Channel 3
$[\varepsilon_G\ \varepsilon_B\ \alpha_G\ \alpha_B]$	[0.01 0.2 0.2 0]	[0.01 0.2 0 0]	[0.01 0.8 0.2 0]

Image Denoising Results: Black-and-white Image Corrupted by Channel 3

Figure 3: top-left : noiseless image; top-right : noisy image; bottom-left : denoised image by the proposed denoiser; bottom-right: denoised image by DUDE

Conclusions

- Considered discrete denoising of an unknown source corrupted by a known channel with memory

- Presented a practical denoiser that is universally asymptotically optimal

- Experimental results indicate that much is to be gained in practice by taking the channel memory into account