MIMO Channel Estimation with Score-Based Generative Priors learned from Noisy Data

Asad Aali[1], Marius Arvinte[1,2], Sidharth Kumar[1], Jonathan I. Tamir[1]

[1]Chandra Family Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, TX, USA

[2]Intel Corporation, Hillsboro, OR, USA

TEXAS The University of Texas at Austin

ArXiv Link

UT CSI Lab

Introduction

- Generative models trained on clean data distribution have shown to outperform end-to-end supervised deep learning.
- A large collection of clean training data is prohibitively expensive to acquire.
- Our method approximately learns a generative model of the clean distribution from noisy data.
- We present SURE-Score: a novel loss function that leverages Stein's unbiased risk estimate (SURE) to jointly denoise the data and learn a score function

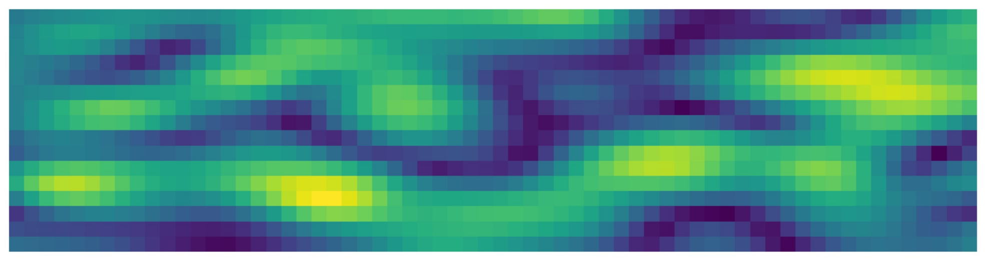
Wireless System Theory

- MIMO forward model: $\mathbf{Y} = \mathbf{HP} + \mathbf{N}$.

$\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$	Channel state information matrix
$\mathbf{p}_i \in \mathbb{C}^{N_t}$	Pilot symbol, P = $(p_0, p_1, \ldots, p_b)$, where $b = \alpha_{pilot} * N_t$
$\sigma^2_{\text{pilot}}\mathbf{I}$	Complex Additive White Gaussian Noise

- Narrowband, point-to-point MIMO communication scenario
- Channel estimation requires estimating H, using the received pilot matrix Y, while having knowledge of the transmitted pilot matrix P

$$\widetilde{H} = H + w, w \sim N\left(0, \sigma_w^2 I\right)$$



Example Clustered Delay Line (CDL-C) channel (magnitude)

Training SURE-Score

$$\mathcal{L}(\theta) = \alpha \underbrace{\left(\mathbb{E}_{\widetilde{H}, n_i}\left[\sigma_i^2 \left\| s_\theta\big(\overbrace{\widetilde{H} + \sigma_w^2 s_\theta(\widetilde{H})}^{f_\theta(\widetilde{H})} + n_i\big) + \frac{n_i}{\sigma_i^2}\right\|_2^2\right]\right)}_{\text{Score Loss}} + \underbrace{\left(\mathbb{E}_{\widetilde{H}, w}\left[\left\|\overbrace{\sigma_w^2 s_\theta(\widetilde{H})}^{f_\theta(\widetilde{H}) - \widetilde{H}}\right\|_2^2 + 2\sigma_w^2 \operatorname{div}_{\widetilde{H}}\big(\overbrace{\widetilde{H} + \sigma_w^2 s_\theta(\widetilde{H})}^{f_\theta(\widetilde{H})}\big)\right]\right)}_{\text{Modified SURE Loss}}$$

Where $\operatorname{div}_{\widetilde{H}}\left(\widetilde{H} + \sigma_w^2 s_\theta(\widetilde{H})\right) = tr\left(J_{\widetilde{H} + \sigma_w^2 s_\theta(\widetilde{H})}\right)$

Where α is appropriate scaling applied to score model

$$\underbrace{\mathbb{E}_{H,w}\left[\left\|f_\theta(\widetilde{H}) - \widetilde{H}\right\|_2^2 + 2\sigma_w^2 \operatorname{div}_{\widetilde{H}}\left(f_\theta(\widetilde{H})\right)\right]}_{\text{Regular SURE Loss}}$$

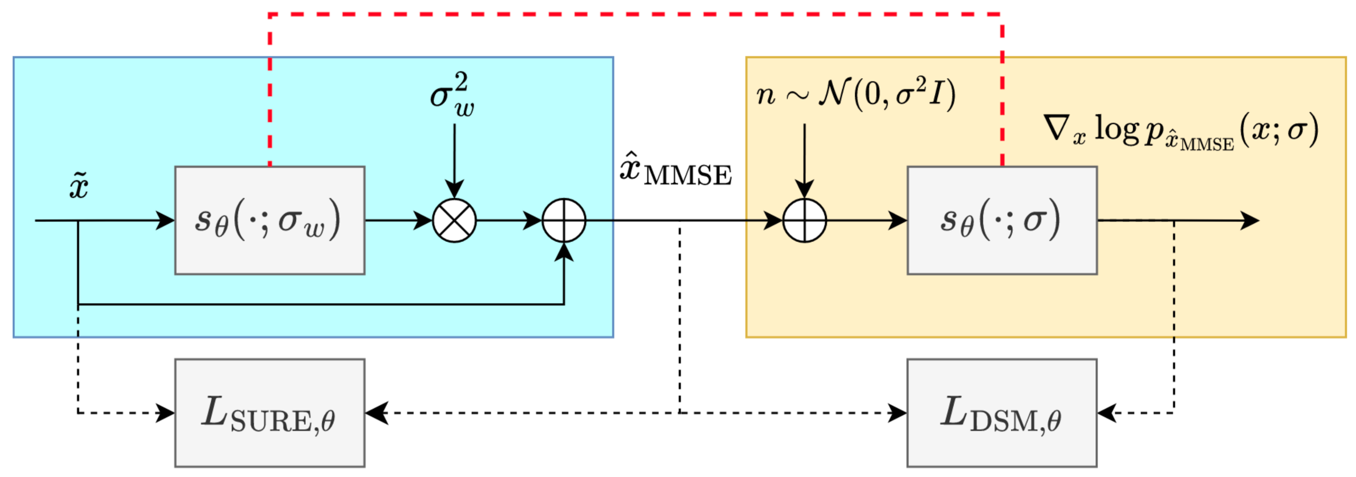


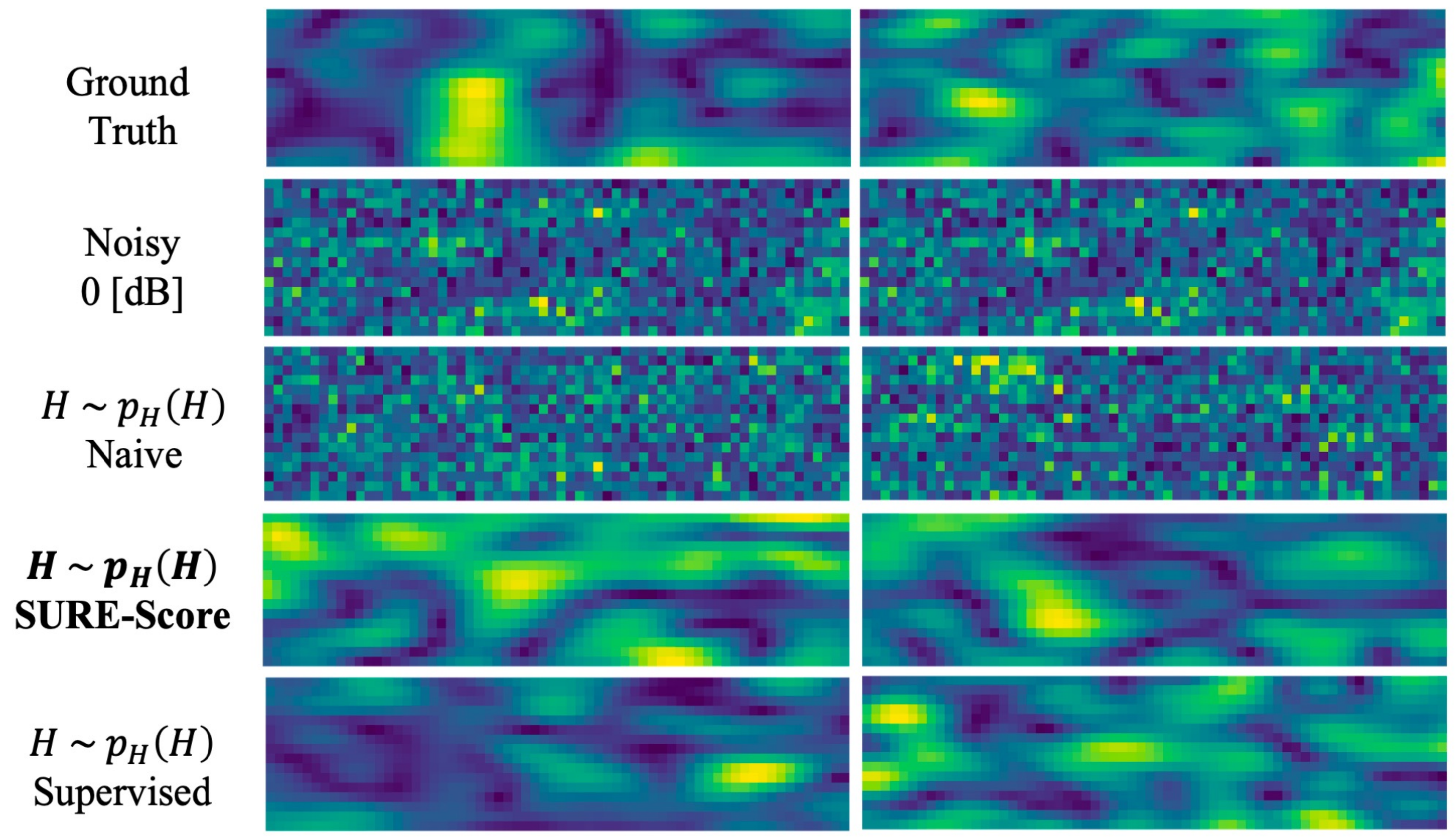
Fig. 1. Flow of SURE-Score during training. The same deep neural network s_θ is used first for denoising and subsequently for denoising score matching.

Generic PyTorch Training Pipeline

Sampling - Annealed Langevin Dynamics

$$\mathbf{H}_{t+1} \leftarrow \mathbf{H}_t + \alpha_t \cdot \nabla \log p_H(\mathbf{H}_t) + \beta_t \cdot \zeta_t.$$

- $\alpha_t \cdot \psi_H(\mathbf{H}_t)$ increases the likelihood of the current sample.
- $\beta_t \cdot \zeta_t$ represents a perturbation to the above process.



Let p_H denote the distribution of MIMO (CDL-C) channels for a stochastic environment.

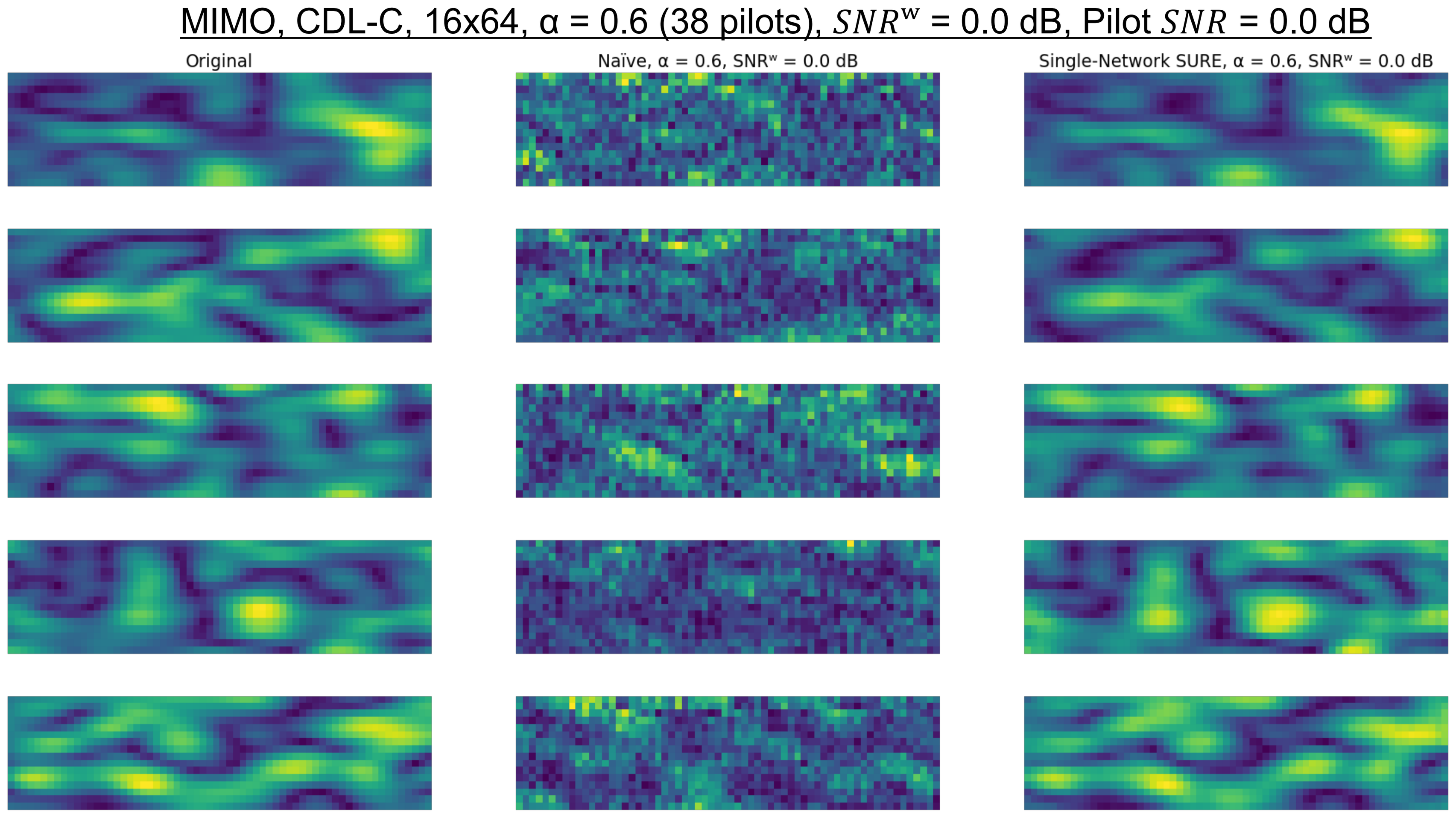
$$\psi_H(\mathbf{H}) = \nabla \log p_H(\mathbf{H}),$$

Fig. 2. Prior sampling for three methods: Naive, SURE-Score at SNR^w 0 dB, and Supervised. Each column is different realization of a CDL-C channel.

Posterior Reconstruction

$$\mathbf{H}_{\text{est},i+1} = \mathbf{H}_{\text{est},i} + \alpha_i \cdot (\nabla \log p_{Y|H}(\mathbf{Y}|\mathbf{H}_{\text{est},i}) + \nabla \log p_H(\mathbf{H}_{\text{est},i})) + \sqrt{2\beta \cdot \alpha_i} \cdot \sigma_{z_i} \cdot \zeta,$$

MIMO, CDL-C, 16x64, α = 0.6 (38 pilots), SNR^w = 0.0 dB, Pilot SNR = 0.0 dB



$$H \sim p(H|Y)$$

Fig. 3.
Naive: Sampling using score model trained directly on noisy channels

Single-Network SURE: Sampling using score model trained on channels denoised using a single network

Posterior Reconstruction – Benchmarking

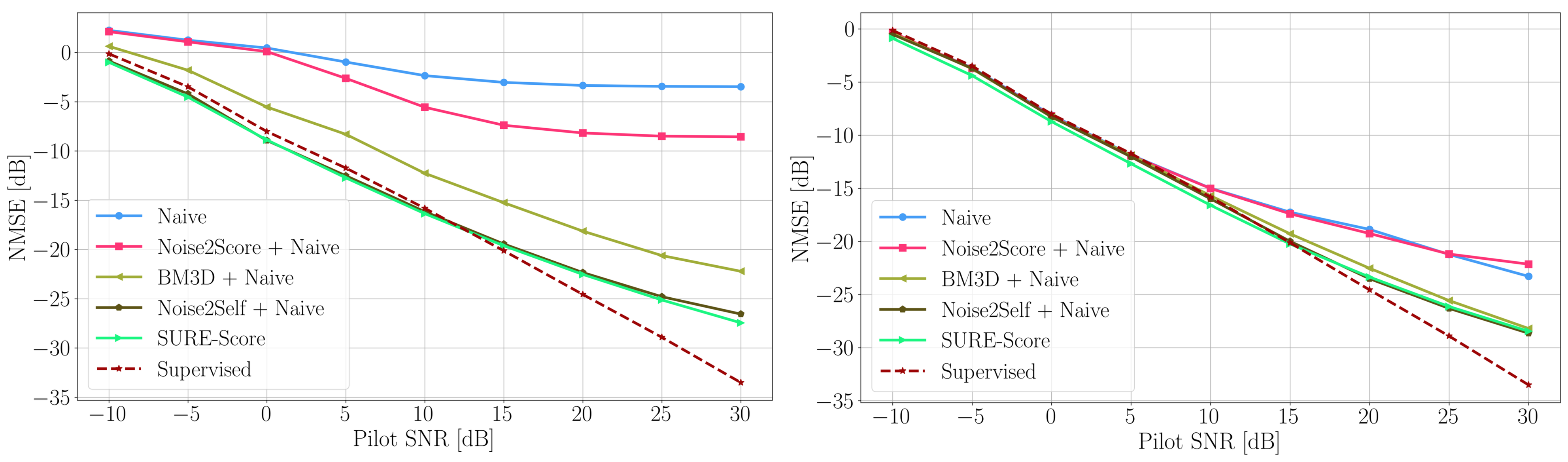


Fig. 3. Channel estimation performance at α = 0.6 (38 pilots) using score models trained on CDL-C channels at SNR^w: 0 dB (left) and 10 dB (right).

Key Takeaways:

- **SURE-Score** performs close to **optimal** with respect to supervised DSM except at higher pilot SNR
- Naive training plateaus in estimation performance because of ***overfitting***
- Noise2Score and BM3D suffer at lower SNR^w and improve at higher SNR
- Performance gap at high pilot SNR likely due to performance limits of MMSE denoiser and finite training data

Generalized SURE-Score

- Goal: Learn the score directly from noisy measurements y

$$y = Ax + n$$

- Where n is a zero-mean Gaussian random vector
- A is full-rank

Methodology:

- Utilize extended SURE principle to obtain unbiased MSE estimate for exponential family noise

$$s(h) = \|x\|^2 + \|h(u)\|^2 +$$

$$2\left(Tr\left(\frac{\partial h(u)}{\partial u}\right) + h^T(u)\frac{\partial \ln q(u)}{\partial u}\right)$$

- Use a single-network to jointly denoise the data and learn score-function

Discussion and Conclusion

- Self-supervised techniques can **match** supervised techniques in denoising and inverse problem performance
- Reconstruction performance with and without access to ground truth measurements is equivalent at low SNRs and comparable at high SNRs
- **Next Steps:** Our work currently assumes white Gaussian noise corruption but could be extended to arbitrary exponential families

Selected References

1. Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," Advances in neural information processing systems, vol. 32, 2019.
2. A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, "Robust compressed sensing mri with deep generative priors," Advances in Neural Information Processing Systems, vol. 34, pp. 14938–14954, 2021.
3. M.Arvinte and J.I.Tamir,"Mimo channel estimation using score-based generative models," IEEE Transactions on Wireless Communications, 2022.
4. C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, "Unsu- pervised learning with stein's unbiased risk estimator," arXiv preprint arXiv:1805.10531, 2018.