



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

# **Generative Priors and Sampling Optimization for Compressed Sensing in 2D NMR Spectroscopy**

Semester Project

Nikolai Argatoff

October 3. 2025

Advisors: PhD Candidate Nicolas Schmid, Dr. Mojmir Mutny  
Supervisor: Prof. Andreas Krause

Master's programme in Applied Mathematics, ETH Zürich

---

## Abstract

Nuclear magnetic resonance (NMR) spectroscopy is one of the most important methods for studying molecular structure in chemistry and biology. However, the acquisition of multi-dimensional NMR spectra can demand several days of spectrometer time. Non-uniform sampling (NUS) reduces acquisition time by measuring only a fraction of the signal. In NUS, the main challenge is accurately reconstructing the full spectrum from incomplete data.

Classical reconstruction methods like compressed sensing typically rely on a priori assumptions, while supervised deep learning approaches are limited by fixed undersampling schedules. With most prior work focused on discriminative settings, the use of generative priors for robust NUS reconstruction has remained unexplored.

This work presents the first successful application of diffusion models to two-dimensional NUS reconstruction. We utilize diffusion posterior sampling (DPS), a generative framework fully independent of the undersampling schedule. The performance of DPS is evaluated against a supervised deep neural network and a compressed sensing method on experimental  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra. We show that DPS consistently outperforms the supervised model and achieves the highest reconstruction quality in the most challenging case, the dense spectrum of the large protein MALT1.

Further, a new data-driven method is proposed to optimize undersampling patterns for arbitrary reconstruction algorithms, using an energy-based Ising model trained with policy gradients. The resulting masks consistently outperform standard Poisson-gap schedules and enable successful reconstruction of MALT1 with only 9.4 % of the data.

---

## Acknowledgements

I would like to express my sincere gratitude to my advisors, Nicolas Schmid and Mojmír Mutný, for their support and patience during countless discussions, whether over coffee in Winterthur or online, and for involving me in such an engaging and fun project. From them, I have learned not only theoretical and practical approaches but also the essence of the scientific method. I also thank Prof. Andreas Krause for his formal supervision and for his inspiring PAI lectures, which introduced me to machine learning.

---

# Contents

---

<b>Contents</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Diffusion models for Inverse Problems . . . . .	1
1.2 NMR Spectroscopy and Non-Uniform-Sampling . . . . .	1
1.3 Undersampling schedules . . . . .	2
1.4 Contributions . . . . .	2
<b>2 Background</b>	<b>4</b>
2.1 Diffusion models . . . . .	4
2.1.1 Score matching . . . . .	4
2.1.2 Score-based generative modeling and DDPMs . . . . .	5
2.1.3 Diffusion models through Stochastic Differential Equations . . . . .	6
2.1.4 Diffusion Models for Inverse Problems . . . . .	7
2.2 NMR Spectroscopy . . . . .	8
2.2.1 FID and NMR spectrum . . . . .	9
2.2.2 Two-dimensional NMR . . . . .	9
2.2.3 NUS reconstruction problem . . . . .	10
2.2.4 Virtual-Echo representation . . . . .	11
2.3 Ising Model . . . . .	12
<b>3 Methods</b>	<b>13</b>
3.1 The reconstruction problem . . . . .	13
3.2 Optimization of sampling patterns . . . . .	14
3.2.1 Undersampling schedules as Ising spins . . . . .	14
<b>4 Experiments</b>	<b>17</b>
4.1 Reconstruction problem . . . . .	17
4.1.1 Training . . . . .	17
4.1.2 Reconstruction performance . . . . .	17
4.2 Performance of optimized schedules . . . . .	18
4.2.1 DPS as the reconstruction algorithm . . . . .	21
4.2.2 IST as the reconstruction algorithm . . . . .	22
<b>5 Discussion</b>	<b>26</b>
5.1 Reconstruction with Diffusion Posterior Sampling . . . . .	26
5.2 Optimizing undersampling patterns with the Ising model . . . . .	27

<b>Bibliography</b>	<b>29</b>
<b>A Appendix</b>	<b>34</b>
A.1 Generating synthetic training data . . . . .	34
A.2 Iterative Soft Thresholding (IST) . . . . .	34
A.3 Reconstruction figures . . . . .	35

## Chapter 1

---

# Introduction

---

### 1.1 Diffusion models for Inverse Problems

Diffusion models are celebrated for making significant advances in photorealistic image generation [1–6]. Recently, a rich literature has been developed around the use of diffusion models to solve inverse problems [7–17]. An inverse problem refers to the estimation of an unknown quantity  $\mathbf{x}^*$  from an incomplete and noisy measurement

$$\mathbf{y} = \mathcal{A}(\mathbf{x}^*) + \boldsymbol{\eta},$$

where a measurement is represented by a forward operator  $\mathcal{A}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $\boldsymbol{\eta}$  is noise. In many cases, inverse problems are ill-defined with a large family of suitable solutions. Diffusion models address this issue by providing a prior over the solution space. Their generative process can then be guided towards a solution consistent with the measurement [18, 19].

Such priors have been widely applied to solve image restoration tasks like inpainting, super-resolution and colorization [7, 9–11, 13, 16, 17, 20, 21]. Aside from a few notable applications, such as in medical imaging [22–25], the broader potential of diffusion priors for inverse problems in science and engineering remains largely unexplored and is still an active area of research [26].

### 1.2 NMR Spectroscopy and Non-Uniform-Sampling

Nuclear Magnetic Resonance (NMR) spectroscopy is one of the most valuable techniques for identifying molecular structure, dynamics and interactions. Its applications range from protein structure determination and metabolomics to pharmaceutical development. As noted by Keeler in his textbook, '*That NMR is useful for chemists will be taken as self-evident.*' [27].

While indispensable, the acquisition of full multidimensional NMR spectra is often prohibitively time-consuming. Three-dimensional NMR experiments often require several days of costly instrument time, while fully sampled four-dimensional experiments are practically intractable. In a method called Non-Uniform Sampling (NUS) the acquisition time is greatly reduced by measuring only a small subset of the full signal [28–30]. The associated inverse

problem of recovering a high-quality spectrum from incomplete NMR data is called NUS reconstruction problem.

NMR spectra obtained from NUS measurements must be recovered through specialized reconstruction algorithms. Given that NMR spectra are approximately sparse, a wide range of methods have been developed within the framework of compressed sensing (CS) [31, 32]. Among these, iterative soft thresholding (IST) [33–36] is widely employed as a practical surrogate for  $l_1$ -norm minimization [37]. Beyond CS-based approaches, traditional reconstruction techniques such as SMILE [38] and maximum entropy methods [39, 40] have also been used.

Deep neural networks have been proposed as promising alternatives to classical methods for NUS reconstruction [41–44]. Among them, the WaveNet-inspired WNN architecture by Jahangiri et al. [43] achieved improvements over IST and SMILE. However, the majority of existing approaches are discriminative and trained in a supervised fashion, requiring paired data and fixed sampling schemes, while generative strategies remain largely unexplored. A recent study published during the course of this work, DiffNMR, employed diffusion models for time-domain inpainting along the indirect dimension, but reported only limited reconstruction performance [45].

### 1.3 Undersampling schedules

For many reconstruction algorithms, the reconstruction fidelity is greatly influenced by the choice of undersampling schedule. Hyberts et al. [46] observed that schedules with large gaps at the edges of the acquisition grid generally perform worse than those with central gaps. Moreover, sufficient randomness is required to avoid aliasing artifacts and violations of the Nyquist theorem. To address these issues, they introduced sinusoidally weighted Poisson-gap sampling (PGS), in which the gap size  $k$  is drawn from a Poisson distribution  $f(k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$ , with  $\lambda$  modulated as  $\lambda = \Lambda \sin \theta$  to bias sampling density toward the edges. PGS was shown to substantially outperform uniform random and standard Poisson-gap sampling when combined with FM reconstruction [46].

More recently, a systematic study [47] investigated seven subsampling strategies using the hmsIST reconstruction method [36]. They found that although no single sampling strategy will always yield the absolute best result, some schedules like PGS generally produced high-fidelity reconstructions more reliably. By contrast, uniform random sampling was shown to frequently generate poor schedules when used with IST [47].

It is appealing to replace hand-crafted priors with learned undersampling schedules, leveraging techniques from machine learning. Significant progress in this direction has been made in the MRI literature, where one popular strategy is to optimize undersampling masks by backpropagating through a deep reconstruction network [25, 48–50], often using the Gumbel–Softmax trick [51]. However, these approaches are inherently tied to the reconstruction network. Developing methods that learn optimal undersampling schedules for any particular reconstruction algorithm could be highly valuable for the NMR community.

### 1.4 Contributions

The contributions of this work are twofold. First, we present the first successful application of unconditional diffusion models to NUS reconstruction in 2D NMR. We use a modification of Diffusion Posterior Sampling (DPS) [9] as the reconstruction algorithm and evaluate its

performance on  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra of three proteins: Ubiquitin (8.6 kDa), Azurin (14 kDa), and MALT1 (44 kDa). DPS is benchmarked against two baselines: the conditional deep learning model WNN of Jahangiri et al. [43] and our own implementation of IST based on [52, 53]. Our results show that DPS outperforms the WNN in almost all cases and achieves competitive performance with IST for smaller proteins Ubiquitin and Azurin. Remarkably, DPS surpasses both IST and WNN for the most challenging protein MALT1 both in terms of performance metrics, but also visually.

Our second contribution is a simple method for optimizing undersampling schedules for arbitrary reconstruction algorithms. We use an energy-based model inspired by the Ising model from statistical physics [54] and represent the sampling patterns as spin configurations. The model parameters are learned using the REINFORCE algorithm [55], following the approach of [56], in which the partition function term is eliminated with a Leave-One-Out baseline [57]. Sampled Ising schedules are learned for both DPS- and IST-based reconstructions, and they are shown to consistently outperform standard Poisson-gap schedules on synthetic datasets as well as on the experimental MALT1 spectrum. Moreover, we use integer optimization to solve the global optimum of the Ising energy function, producing a deterministic mask with high reconstruction fidelity even at 9.4 % undersampling.

## Chapter 2

---

# Background

---

## 2.1 Diffusion models

As is often the case in research, breakthroughs come about from the integration of insights developed across multiple studies. In this section, we cover some of the important concepts that led up to the development of diffusion models as we conceptually know them today.

### 2.1.1 Score matching

Suppose that we want to approximate an unknown probability density function  $p_{\text{data}}(\mathbf{x})$  from which we only have access to an independent and identically distributed dataset  $\{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^N$ . Using a parameterization  $q(\mathbf{x}; \boldsymbol{\theta})$ , we can construct a normalized density

$$p(\mathbf{x}; \boldsymbol{\theta}) = \frac{q(\mathbf{x}; \boldsymbol{\theta})}{Z(\boldsymbol{\theta})}, \quad (2.1)$$

where  $\boldsymbol{\theta}$  are the parameters and  $Z(\boldsymbol{\theta})$  is an untractable normalization constant. Such  $q(\mathbf{x}; \boldsymbol{\theta})$  is also referred to as a non-normalized probabilistic model.

We define the *score* of the density  $p_{\text{data}}(\mathbf{x})$  to be  $\nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})$ . Similarly, we define the *score function*  $\mathbf{s}(\mathbf{x}; \boldsymbol{\theta}) = \nabla_{\mathbf{x}} \log p(\mathbf{x}; \boldsymbol{\theta})$  which can be checked to not depend on  $Z(\boldsymbol{\theta})$ . Leveraging this property, Hyvärinen [58] introduced the method of *score matching*, which estimates parameters by solving

$$\min_{\boldsymbol{\theta}} \frac{1}{2} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} \|\nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x}) - \mathbf{s}(\mathbf{x}; \boldsymbol{\theta})\|^2. \quad (2.2)$$

The obvious problem of not knowing the true score  $\nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})$  can be avoided by a simple integration by parts, which gives up to a constant

$$\min_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} \left[ \frac{1}{2} \|\mathbf{s}(\mathbf{x}; \boldsymbol{\theta})\|^2 + \text{Tr}(\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \boldsymbol{\theta})) \right]. \quad (2.3)$$

This objective can then be optimized using the empirical mean over the available samples. In practice, the score function  $\mathbf{s}(\mathbf{x}; \boldsymbol{\theta})$  is parameterized by a neural network, commonly

referred to as a *score network*. However, computing  $\text{Tr}(\nabla_{\mathbf{x}} \mathbf{s}(\mathbf{x}; \theta))$  becomes intractable for high-dimensional data  $\mathbf{x}$ , such as images.

The computation can be avoided using yet another mathematical trick proposed by Vincent [59] known as *denoising score matching*. By convolving the data with some noise kernel  $q_\sigma(\tilde{\mathbf{x}} | \mathbf{x})$  we get a perturbed data density

$$p_\sigma(\tilde{\mathbf{x}}) = \int p_{\text{data}}(\mathbf{x}) q_\sigma(\tilde{\mathbf{x}} | \mathbf{x}) d\mathbf{x}, \quad (2.4)$$

for some small noise level  $\sigma$ . In the case of a Gaussian kernel this amounts to  $\tilde{\mathbf{x}} = \mathbf{x} + \sigma \mathbf{z}$  with  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$ . Estimating the score of the *perturbed* density then leads to a computationally efficient training objective:

$$\min_{\theta} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x}), \tilde{\mathbf{x}} \sim q_\sigma(\tilde{\mathbf{x}} | \mathbf{x})} \| \mathbf{s}_\theta(\mathbf{x}) + \nabla_{\mathbf{x}} \log q_\sigma(\tilde{\mathbf{x}} | \mathbf{x}) \|^2. \quad (2.5)$$

Again, for a Gaussian kernel above denoising score matching objective simplifies to

$$\min_{\theta} \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x}), \mathbf{z} \sim \mathcal{N}(0, I)} \left\| \mathbf{s}(\mathbf{x} + \sigma \mathbf{z}; \theta) + \frac{\mathbf{z}}{\sigma} \right\|^2, \quad (2.6)$$

which can be seen as predicting the added noise to the sample. The diffusion models are trained with a series of denoising score matching objectives like above at multiple noise scales.

### 2.1.2 Score-based generative modeling and DDPMs

The aim of generative modeling is to learn a model that can produce samples following the same distribution as the dataset  $\{\mathbf{x}_i \in \mathbb{R}^d\}_{i=1}^N$ . Song and Ermon introduced Score Matching with Langevin Dynamics (SMDL) [5], demonstrating that trained score networks enable generative modeling via Langevin dynamics:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \alpha \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x}) + \sqrt{2\alpha} \mathbf{z}_t, \quad (2.7)$$

where  $\mathbf{z}_t \sim \mathcal{N}(\mathbf{0}, I)$  and  $\alpha > 0$  is the step size. However, directly applying Eq. (2.7) often produces low-quality samples due to poor state-space exploration and inaccuracies in the score approximation in low-density regions. To address this, Song and Ermon [5] proposed perturbing the data at multiple noise scales  $\{\sigma_i\}_{i=1}^L$ , which substantially improves sample quality. They introduced noise-conditional score networks  $s(\mathbf{x}, \sigma; \theta)$  that estimate the score of perturbed densities at each noise level. Generation is then performed with annealed Langevin dynamics, beginning at the highest noise scale  $\sigma_L$  and gradually reducing noise to produce clean samples.

Another major breakthrough was the work of Ho et al. [2] on denoising diffusion probabilistic models (DDPMs). Diffusion models are inspired by non-equilibrium thermodynamics and

slowly corrupt data samples through a forward Markov chain by iteratively adding Gaussian noise [1]:

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}\left(\sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t I\right), \quad t = 1, \dots, T. \quad (2.8)$$

The noise schedule  $\{\beta_t\}_{t=1}^T$  is chosen such that  $\mathbf{x}_0 \sim p_{\text{data}}$  is mapped to pure Gaussian noise  $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, I)$ . The noising process can be efficiently calculated in one step

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \mathbf{z}, \quad (2.9)$$

where  $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$  and  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$ . For small enough  $\beta_t$  the reverse process can be parametrized as a normal distribution:

$$p(\mathbf{x}_{t-1} \mid \mathbf{x}_t; \theta) = \mathcal{N}(\mu(\mathbf{x}_t, t; \theta), \Sigma(\mathbf{x}_t, t; \theta)), \quad (2.10)$$

where  $\mu(\mathbf{x}_t, t; \theta)$  and  $\Sigma(\mathbf{x}_t, t; \theta)$  are neural networks for learning the mean and variance. Drawing a random sample  $\mathbf{x}_T$  and iteratively running the reverse chain  $p(\mathbf{x}_{t-1} \mid \mathbf{x}_t; \theta)$  for  $t = T, T-1, \dots, 1$  transforms the reverse process into a generative model.

Among other practical improvements to the diffusion models, Ho et al. [2] showed that parameterizing the mean as

$$\mu(\mathbf{x}_t, t; \theta) = \frac{1}{\sqrt{1 - \beta_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \varepsilon(\mathbf{x}_t, t; \theta) \right), \quad (2.11)$$

with  $\varepsilon(\mathbf{x}_t, t; \theta)$  being a neural network yields a training objective equivalent to denoising score matching.

### 2.1.3 Diffusion models through Stochastic Differential Equations

The score based generative modeling and diffusion probabilistic modeling are unified by Song et al. [4] into a single framework through continuous time stochastic differential equations (SDEs). For times  $t \in [0, T]$ , the sample  $\mathbf{x}(t)$  evolves through a forward diffusion process defined by an Ito SDE

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t) dt + g(t) d\mathbf{w}, \quad (2.12)$$

where  $\mathbf{f}(\cdot, t) : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is the drift,  $g(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$  is the diffusion coefficient, and  $\mathbf{w}$  is a standard  $n$ -dimensional Brownian motion. The forward SDE converts the samples of the data distribution  $\mathbf{x}(0) \sim p_{\text{data}}$  into pure noise samples from a simple prior distribution  $\mathbf{x}(T) \sim p_T$ .

Anderson [60] showed that the reverse process of Eq. (2.12) is a diffusion process running backwards in time

$$d\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \nabla_{\mathbf{x}} \log p_t(\mathbf{x})] dt + g(t) d\mathbf{w}, \quad (2.13)$$

where the infinitesimals above describe the reversed time flow from  $T$  to 0 and  $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$  is the score of each marginal distribution.

Training a time-dependent score network  $\mathbf{s}(\mathbf{x}, t; \theta)$  with denoising score matching gives an approximation of the score  $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ . Solving the reverse SDE with numerical methods generates a sample from  $p_{\text{data}}$ .

Song et al. [4] showed that the forward process of DDPM is the discretization of Eq. (2.12) with the drift  $\mathbf{f}(\mathbf{x}, t) = -\beta(t)\mathbf{x}/2$  and diffusion coefficient  $g(t) = \sqrt{\beta(t)}$ , where  $\beta(t)$  is the noise schedule. When  $t \rightarrow \infty$ , this process has fixed variance and is referred to as Variance Preserving (VP) SDE<sup>1</sup>.

#### 2.1.4 Diffusion Models for Inverse Problems

Given the forward operator  $\mathcal{A}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the inverse problem with additive noise is given by

$$\mathbf{y} = \mathcal{A}(\mathbf{x}^*) + \boldsymbol{\eta}, \quad (2.14)$$

where  $\boldsymbol{\eta} \in \mathbb{R}^m$  is noise. In ill-posed inverse problems recovery is impossible without a priori assumptions about the properties of  $\mathbf{x}^*$ . In the probabilistic approach  $\mathbf{x}^*$  is assumed to be sample from a prior distribution  $p(\mathbf{x})$  and the forward model Eq. (2.14) gives rise to the likelihood  $p(\mathbf{y}|\mathbf{x})$ . Solution to the inverse problem is distributed according to the Bayes rule:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}. \quad (2.15)$$

Imposing the diffusion model as the prior, the samples from the posterior  $p(\mathbf{x}|\mathbf{y})$  follow the reverse SDE

$$d\mathbf{x} = \{\mathbf{f}(\mathbf{x}, t) - g(t)^2 [\nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \nabla_{\mathbf{x}} \log p_t(\mathbf{y}|\mathbf{x})]\} dt + g(t) d\mathbf{w}. \quad (2.16)$$

The score of the prior  $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$  can be approximated with a pretrained score network  $\mathbf{s}(\mathbf{x}, t; \theta)$ . However, the likelihood term  $\nabla_{\mathbf{x}} \log p_t(\mathbf{y}|\mathbf{x}(t))$  is easy to compute only at  $t = 0$ . To illustrate this, consider an inverse problem with the Gaussian noise model  $\boldsymbol{\eta} \sim \mathcal{N}(0; \sigma^2 I)$ . In this case, the dependency between  $\mathbf{y}$  and  $\mathbf{x}(0) = \mathbf{x}^*$  is given by

$$\nabla_{\mathbf{x}} \log p_0(\mathbf{y} | \mathbf{x}(0)) = -\frac{1}{2\sigma^2} \nabla_{\mathbf{x}} \|\mathbf{y} - \mathcal{A}(\mathbf{x})\|_2^2 \Big|_{\mathbf{x}=\mathbf{x}(0)}. \quad (2.17)$$

---

<sup>1</sup>The SDE equivalent to SMLD has exploding variance and is called Variance Exploding (VE) SDE. In this work, we focus on VP SDE framework.

For a general noise level  $t$ , the noisy samples  $\mathbf{x}(t)$  and  $\mathbf{y}$  have no such closed-form dependency. Hence, the likelihood must be either estimated numerically or the measurement consistency is indirectly enforced by other means like projections to the measurement subspace, see [18, 19] for a thorough review.

Chung et al. [9] introduced Diffusion Posterior Sampling (DPS), which provides an approximation of the likelihood term  $\nabla_{\mathbf{x}} \log p_t(\mathbf{y} | \mathbf{x}(t))$ . Analogously to Eq. (2.9), in the VP-SDE framework the forward diffusion process is given by

$$\mathbf{x}(t) = \sqrt{\bar{\alpha}(t)} \mathbf{x}(0) + \sqrt{1 - \bar{\alpha}(t)} \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, I). \quad (2.18)$$

In order to reverse the direction, DPS utilizes an estimator of  $\mathbf{x}(0)$  given  $\mathbf{x}(t)$  using a remarkable result known as Tweedie's formula [61]:

$$\hat{\mathbf{x}}_0(\mathbf{x}(t)) := \mathbb{E}[\mathbf{x}(0) | \mathbf{x}(t)] = \frac{1}{\sqrt{\bar{\alpha}(t)}} [\mathbf{x}(t) + (1 - \bar{\alpha}(t)) \nabla_{\mathbf{x}} \log p_t(\mathbf{x}(t))]. \quad (2.19)$$

DPS approximates the likelihood  $p_t(\mathbf{y} | \mathbf{x}(t))$  by plugging in  $\hat{\mathbf{x}}_0$  into the closed form likelihood formula  $p_0(\mathbf{y} | \mathbf{x}(0))$ . For the case of Gaussian noise considered in Eq. (2.20), the DPS approximation gives

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{y} | \mathbf{x}(t)) \approx -\frac{1}{2\sigma^2} \nabla_{\mathbf{x}} \|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0(\mathbf{x}))\|_2^2 \Big|_{\mathbf{x}=\mathbf{x}(t)}. \quad (2.20)$$

Finally, the score of the posterior can be approximated by

$$\nabla_{\mathbf{x}} \log p_t(\mathbf{x} | \mathbf{y}) = \nabla_{\mathbf{x}} \log p_t(\mathbf{x}) + \nabla_{\mathbf{x}} \log p_t(\mathbf{y} | \mathbf{x}) \quad (2.21)$$

$$\approx \mathbf{s}(\mathbf{x}, t; \boldsymbol{\theta}) - \frac{1}{2\sigma^2} \nabla_{\mathbf{x}} \|\mathbf{y} - \mathcal{A}(\hat{\mathbf{x}}_0(\mathbf{x}))\|_2^2 \Big|_{\mathbf{x}=\mathbf{x}(t)}, \quad (2.22)$$

where  $\hat{\mathbf{x}}_0(\mathbf{x}) = [\mathbf{x} + (1 - \bar{\alpha}(t)) \mathbf{s}(\mathbf{x}, t; \boldsymbol{\theta})] / \sqrt{\bar{\alpha}(t)}$  is the approximation of Tweedie's formula with the score network. For numerical implementation with  $i \in \{1, 2, \dots, N\}$  points, we discretize  $\mathbf{x}_i = \mathbf{x}(iT/N)$ , and put  $\beta_i = \beta(iT/N)$ .

## 2.2 NMR Spectroscopy

NMR spectroscopy is based on the fact that atomic nuclei possess a magnetic quantum number  $I$ , called nuclear spin. Biochemists typically focus on nuclei with spin quantum number  $I = \frac{1}{2}$ , such as  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$ , which have a twofold degenerate ground state [27]. The application of a large external magnetic field  $B_0$  breaks the symmetry, creating an energy difference  $\Delta E$  proportional to the applied field. In equilibrium, a large spin population in a bulk material tends to slightly favor the lower energy state following the Boltzmann statistics.

The spin population can be excited to higher-energy states by applying a secondary oscillating magnetic field  $B_1$ . Resonance occurs when the frequency of this field matches the energy gap  $\Delta E$ , also referred to as the Larmor frequency. The resulting macroscopic magnetization of the nuclei precesses around the main magnetic field, inducing an alternating voltage in the surrounding detection coil [27]. This voltage, known as the free induction decay (FID), is sampled at regular intervals and digitally stored. By applying a discrete Fourier transform to the FID, one obtains the NMR spectrum, which provides the foundation for analyzing the molecular structure of the sample.

### 2.2.1 FID and NMR spectrum

To a good approximation, the FID can be modeled as a superposition of  $N$  decaying complex exponentials:

$$X_{\text{FID}}(t) = \sum_{k=1}^N a_k e^{i(2\pi\omega_k t + \varphi_k)} e^{-t/\tau_k}, \quad (2.23)$$

where  $\omega_k$  is referred to as the chemical shift,  $\varphi_k$  is the phase,  $\tau_k$  is the relaxation time, and  $A_k$  is the amplitude of  $k$ :th wave. The real and imaginary components of the FID are recorded separately with quadrature detection.

The Fourier transform gives a complex spectrum

$$S(\omega) = \mathcal{F}[X](\omega) = \sum_{k=1}^N a_k (A_k(\omega) + iD_k(\omega)), \quad (2.24)$$

where  $\mathcal{F}$  is the Fourier transform and  $A_k(\omega)$  and  $D_k(\omega)$  are known as the absorption and dispersion Lorentzian spectral lines:

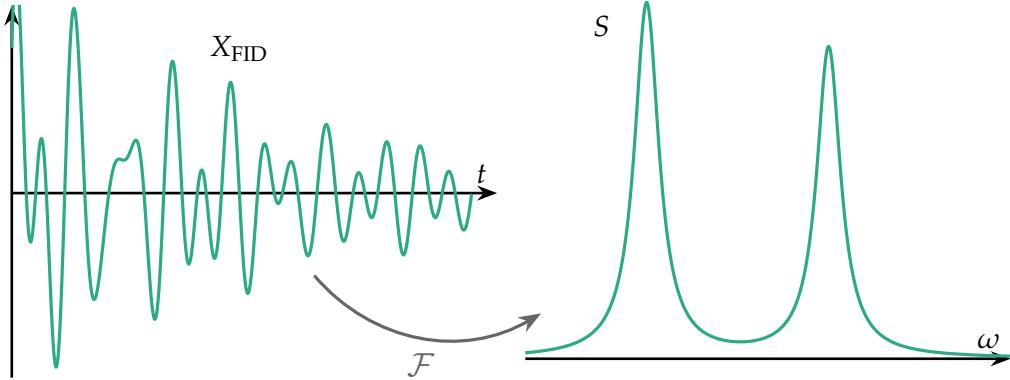
$$A_k(\omega) = \frac{1}{1 + \tau_k^2(\omega - \omega_k)^2}, \quad D_k(\omega) = \frac{\tau_k(\omega - \omega_k)}{1 + \tau_k^2(\omega - \omega_k)^2}. \quad (2.25)$$

An illustration of the FID and its spectrum is shown in Fig. 2.1. The dispersive part of the spectrum is typically discarded. Multiple post-processing steps are typically carried out both on the FID and the spectrum, such as applying weighting functions, zero-filling and phase correction. Although we expressed the above theory in continuous form, both the FID and the spectrum are in practice matrices of uniformly spaced samples.

### 2.2.2 Two-dimensional NMR

The FID in Eq. (2.23) is a one-dimensional time-domain signal. In more sophisticated experiments, the FID can be measured as a function of multiple time variables, giving rise to multi-dimensional NMR.

In the case of two-dimensional NMR (2D-NMR), the sample is first excited by an initial pulse, after which the magnetization evolves during a period  $[0, t_1]$ . A second pulse is then applied, and the resulting free induction decay is recorded as a function of a second time



**Figure 2.1:** Schematic of the free induction decay (FID) signal and its spectrum, showing only the real components.

variable  $t_2$ . Repeating this procedure for different values of  $t_1$  produces a set of time-domain signals measured in terms of  $t_2$ , which together form a two-dimensional time signal [27]. Similarly to Eq. (2.23), we can represent this 2D-FID by

$$X_{\text{FID}}(t_1, t_2) = \sum_{k=1}^N a_k e^{i(2\pi\omega_{k,1}t_1 + \varphi_{k,1})} e^{-t_1/\tau_{k,1}} e^{i(2\pi\omega_{k,2}t_2 + \varphi_{k,2})} e^{-t_2/\tau_{k,2}}. \quad (2.26)$$

Subsequently, the NMR spectrum  $S(\omega_1, \omega_2)$  is a function of two frequency variables. The first dimension is referred to as *indirect* dimension and the second as *direct* dimension, since only  $t_2$  is measured directly.

### 2.2.3 NUS reconstruction problem

As can be seen from the above procedure, acquiring a multi-dimensional NMR spectrum at by measuring uniformly at high resolution is experimentally expensive. Acquisition times can be significantly reduced with *non-uniform sampling* (NUS). In 2D NMR with non-uniform sampling, the direct dimension  $t_2$  is typically measured in full, while only a subset of points in the indirect dimension  $t_1$  are acquired, with the remaining entries treated as missing and set to zero.

Let  $U$  be a binary diagonal matrix representing some undersampling sequence and denote the uniformly sampled FID as a matrix  $X_{\text{FID}}$  with respective spectrum  $S = \mathcal{F}[X_{\text{FID}}]$  where  $\mathcal{F}$  is the discrete Fourier transform (DFT). The undersampled time signal is given by

$$X_{\text{NUS}} = UX_{\text{FID}}. \quad (2.27)$$

The NUS spectrum is obtained from

$$S_{\text{NUS}} = \mathcal{F}[U] * S, \quad (2.28)$$

where  $*$  is convolution with the kernel  $\mathcal{F}[U]$ . The kernel is referred to as the point spread function (PSF). Convolution with the PSF introduces substantial distortions, making the NUS spectrum unsuitable for extraction of chemical information.

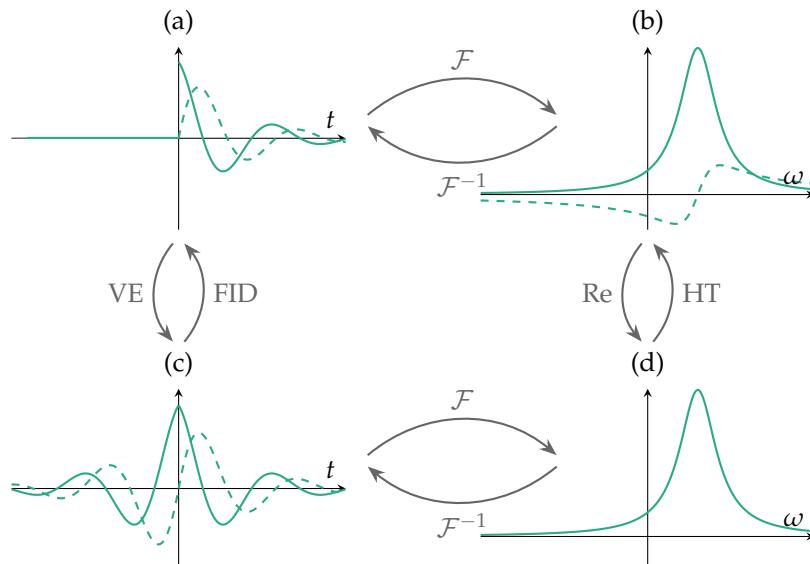
### 2.2.4 Virtual-Echo representation

Although spectroscopists ultimately care about the absorptive spectrum, transitioning between the time and frequency domains requires both components. For causal time signals such as the FID, the real and imaginary parts of the spectrum are not independent, but they are linked by the Kramers–Kronig relations [62, 63].

The causality of the FID is utilized in NMR to represent the time domain signal in *virtual-echo* form [64]. Since the time signal is known to start at  $t = 0$ , we may extend the FID for  $t < 0$  by padding with zeros. Virtual echo signal is obtained by the conversion of the FID to a Hermitian form:

$$X_{VE}(t) = \begin{cases} X_{FID}(t), & t \geq 0; \\ X_{FID}^\dagger(-t), & t < 0. \end{cases} \quad (2.29)$$

Here  $\dagger$  denotes the conjugation. This procedure is visualized in (a) to (c) direction in Fig. 2.2.



**Figure 2.2: Virtual echo representation.** (a) FID (b) A spectrum with real and imaginary components. (c) The virtual echo (VE). (d) Real component of the original spectrum. Real and imaginary parts are represented by solid and dashed lines. Transitions between (a) and (b), and between (c) and (d), are given by the Fourier transform  $\mathcal{F}$ . The direction (b) to (d) corresponds to taking the real part, while the reverse operation is obtained via the Hilbert transform (HT).

Lastly, we note that the virtual echo properties can be seen from the elementary relations of the Fourier transform. Let  $x(t)$  be a time-domain signal, and denote its Fourier transform by  $X(\omega)$ . Then, recalling that  $\text{Re}(X(\omega)) = (X(\omega) + X^\dagger(\omega))/2$ , we have

$$\mathcal{F}^{-1}[\text{Re}(X)][t] = \frac{1}{2}(x(t) + x^\dagger(-t)). \quad (2.30)$$

When we have a causal signal with  $x(t) = 0$  for  $t < 0$ , then  $x(t)$  and  $x^\dagger(-t)$  do not overlap, and form (Hermitian) mirror-images of each other about the origin.

## 2.3 Ising Model

The Ising model is one of the most studied models in statistical physics that was introduced to describe ferromagnetism [54]. It falls into the class of non-normalized models introduced in Section 2.1.1, more specifically energy-based models (EBMs) with the density

$$p(\sigma; \theta) = \frac{e^{-E(\sigma; \theta)}}{Z(\theta)}, \quad (2.31)$$

where  $E(\sigma; \theta)$  is an arbitrary function, called the *energy*. The Ising model uses  $N$  binary spin variables  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$  where each spin  $\sigma_i \in \{-1, +1\}$ . The model's energy function is defined as

$$E(\sigma; \theta) = - \sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j - \sum_{i=1}^N h_i \sigma_i, \quad (2.32)$$

where  $\theta = \{J, \mathbf{h}\}$  represents the model parameters. The first term captures pairwise interactions between connected spins indicated by  $\langle i, j \rangle$ , while the second term introduces a bias on individual spins.

The normalization constant, also known as the partition function, is a sum over  $2^N$  terms:

$$Z(\theta) = \sum_{\sigma \in \{\pm 1\}^N} e^{-E(\sigma; \theta)}. \quad (2.33)$$

Finding the spin configuration corresponding to the lowest energy, also known as the ground state, requires solving an integer optimization problem

$$\sigma^* = \operatorname{argmin}_{\sigma \in \{\pm 1\}^N} - \sum_{\langle i,j \rangle} J_{ij} \sigma_i \sigma_j - \sum_{i=1}^N h_i \sigma_i, \quad (2.34)$$

which is NP-complete in the case of non-planar connectivity [65]. The gradient of the log-probability for an EBM is given by

$$\nabla_{\theta} \log p(\sigma; \theta) = -\nabla_{\theta} E(\sigma; \theta) - \nabla_{\theta} \log Z(\theta) \quad (2.35)$$

$$= -\nabla_{\theta} E(\sigma; \theta) + \mathbb{E}_{\sigma \sim p(\sigma; \theta)} [\nabla_{\theta} E(\sigma; \theta)], \quad (2.36)$$

where the main challenge is approximating the second term  $\nabla_{\theta} \log Z(\theta)$ .

## Chapter 3

---

# Methods

---

The quality of NUS reconstructions depends critically on both the choice of reconstruction algorithm and the undersampling mask  $U$ . In this chapter, we introduce our approach for performing reconstruction using diffusion posterior sampling, followed by a framework for optimizing sampling patterns.

### 3.1 The reconstruction problem

Recovering the NMR spectrum from NUS measurements is a linear inverse problem in the frequency space with a forward operator  $\mathcal{A}(\mathbf{x}) = \mathcal{F}[U] * \mathbf{x}$ , where  $\mathbf{x} \in \mathbb{C}^{n \times n'}$  is the NMR spectrum,  $\mathcal{F} \in \mathbb{C}^{n \times n}$  is the DFT, and  $U \in \mathbb{C}^{n \times n}$  is an undersampling schedule. We focus on 2D NMR where the indirect dimension is undersampled according to an undersampling ratio  $m/n$  where  $m < n$  denotes the number of measured points.

To reconstruct NUS spectra with unconditional diffusion models, we employ Diffusion Posterior Sampling (DPS) [9], as introduced in Section 2.1.4. Following previous work on NUS reconstruction [41, 43], we incorporate an additional data consistency step at the end of the DPS denoising loop. As shown in line 9 of Algorithm 1, this step restores the acquired experimental data in the time domain.

---

**Algorithm 1** DPS for NUS reconstruction

---

- 1: **Input:** Number of steps  $N$ , measurement  $\mathbf{y}$ , undersampling mask  $U$ , step size  $\{\zeta_i\}_{i=1}^N$ , variance schedule  $\{\tilde{\sigma}_i\}_{i=1}^N$
  - 2:  $\mathbf{x}_N \sim \mathcal{N}(\mathbf{0}, I)$
  - 3: **for**  $i = N - 1$  **to** 0 **do**
  - 4:    $\hat{\mathbf{s}} \leftarrow \mathbf{s}(\mathbf{x}_i, i; \theta)$
  - 5:    $\hat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\mathbf{x}_i + (1 - \bar{\alpha}_i)\hat{\mathbf{s}})$
  - 6:    $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$
  - 7:    $\mathbf{x}_{i-1}'' \leftarrow \frac{\sqrt{\alpha_i(1 - \bar{\alpha}_{i-1})}}{1 - \bar{\alpha}_i} \mathbf{x}_i + \frac{\sqrt{\bar{\alpha}_{i-1}\beta_i}}{1 - \bar{\alpha}_i} \hat{\mathbf{x}}_0 + \tilde{\sigma}_i \mathbf{z}$  ▷ DDPM sampling step
  - 8:    $\mathbf{x}_{i-1}' \leftarrow \mathbf{x}_{i-1}'' - \zeta_i \nabla_{\mathbf{x}_i} \|\mathbf{y} - \mathcal{F}[U] * \hat{\mathbf{x}}_0\|_2^2$  ▷ DPS likelihood step
  - 9:    $\mathbf{x}_{i-1} \leftarrow \mathbf{x}_{i-1}' - \mathcal{F}[U] * \mathbf{x}_{i-1}' + \mathbf{y}$  ▷ NUS data consistency
  - 10: **end for**
  - 11: **return**  $\hat{\mathbf{x}}_0$
-

## 3.2 Optimization of sampling patterns

We aim to learn a distribution  $p(U; \theta)$  over a family of optimal undersampling schedules  $U$  for reconstructing NUS spectra. For full generality, we denote an arbitrary reconstruction algorithm by  $\text{rec}[\cdot]$ . Given a noisy NUS measurement  $\mathbf{y} = \mathcal{F}[U] * \mathbf{x} + \eta$ , we denote the reconstruction by  $\hat{\mathbf{x}} = \text{rec}[\mathbf{y}]$ . The fully sampled spectra  $\mathbf{x}$  are assumed to follow a distribution  $p_{\text{data}}(\mathbf{x})$ . Our training objective can be stated as

$$\theta^* = \operatorname{argmax}_{\theta} \mathbb{E}_{U \sim p(U; \theta), \mathbf{x} \sim p_{\text{data}}(\mathbf{x}), \mathbf{y} \sim p(\mathbf{y} | \mathbf{x}, U)} [r(\mathbf{x}, \hat{\mathbf{x}})], \quad (3.1)$$

where  $r(\cdot, \cdot)$  is a reward function that quantifies reconstruction fidelity. As we may not differentiate through an arbitrary reconstruction process  $\text{rec}[\cdot]$ , finding an optimal  $\theta^*$  is challenging.

Since  $U$  corresponds to binary sampling patterns, in our approach we parametrize  $p(U; \theta)$  using one of the simplest distributions capable of capturing correlations between binary variables.

### 3.2.1 Undersampling schedules as Ising spins

For compactness, the optimization problem in Eq. (3.1) can be expressed as a loss given by negative expected reward

$$\mathcal{L}(\theta) = \mathbb{E}_{U \sim p(U; \theta)} [-r(U)], \quad (3.2)$$

where the reward is defined as  $r(U) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x}), \mathbf{y} \sim p(\mathbf{y} | \mathbf{x}, U)} [r(\mathbf{x}, \hat{\mathbf{x}})]$ .

The REINFORCE estimator [55] provides an unbiased estimate of the gradient of this expectation:

$$\nabla_{\theta} \mathcal{L}(\theta) = \mathbb{E}_{U \sim p(U; \theta)} [-\nabla_{\theta} \log p(U; \theta) r(U)]. \quad (3.3)$$

We represent each undersampling schedule  $U$  by an Ising spin configuration  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$ , where  $\sigma_i = +1$  denotes a measured point and  $\sigma_i = -1$  denotes an unmeasured point. With  $p(U; \theta)$  parametrized as an Ising energy-based model (EBM), cf. Eq. (2.31), the REINFORCE gradient becomes

$$\nabla_{\theta} \mathcal{L}(\theta) = \mathbb{E}_{\sigma \sim p(\sigma; \theta)} [-\nabla_{\theta} \log p(\sigma; \theta) r(\sigma)], \quad (3.4)$$

$$= \mathbb{E}_{\sigma \sim p(\sigma; \theta)} [(\nabla_{\theta} E(\sigma; \theta) + \nabla_{\theta} \log Z(\theta)) r(\sigma)], \quad (3.5)$$

where we have used Eq. (2.35). The partition function term in Eq. (3.4) can be estimated using MCMC methods, with especially efficient variants developed for Ising models [66]. However, we follow the approach of [56], which uses a clever baseline to cancel the partition function term entirely.

The leave-one-out baseline was proposed by Kool et al. [57] to reduce the variance of the REINFORCE estimator for structured prediction tasks. Drawing  $k$  samples  $\sigma^{(1)}, \dots, \sigma^{(k)}$ , a baseline for sample  $i$  can be constructed from other samples  $j \neq i$ . This gives a REINFORCE Leave-One-Out (RLOO) estimator:

$$\nabla_{\theta} \mathcal{L}(\theta) \approx \frac{1}{k} \sum_{i=1}^k -\nabla_{\theta} \log p(\sigma^{(i)}; \theta) \left( r(\sigma^{(i)}) - \frac{1}{k-1} \sum_{j \neq i} r(\sigma^{(j)}) \right). \quad (3.6)$$

B  nk  estad et al. [56] utilized above estimator for  $k = 2$  which cancels the spin independent term  $\nabla_{\theta} \log Z(\theta)$ . Eq. (3.6) simplifies to

$$\nabla_{\theta} \mathcal{L}(\theta) \approx \frac{1}{2} \left( r(\sigma^{(1)}) - r(\sigma^{(2)}) \right) \cdot \nabla_{\theta} \left( E(\sigma^{(1)}; \theta) - E(\sigma^{(2)}; \theta) \right). \quad (3.7)$$

For many reward functions, the loss can be trivially minimized by aligning all spins upward, which corresponds to measuring everything in NUS. To enforce a fixed sampling budget, B  nk  estad et al. [56] introduced an additional penalty term in the loss.

We propose an alternative approach where sampling is restricted to the submanifold where the total magnetization<sup>1</sup> is conserved. Sampling within this constrained manifold is carried out using Kawasaki dynamics see Algorithm 2). We put everything together in Algorithm 3 which shows the full training framework.

---

**Algorithm 2** Kawasaki MCMC

- 1: **Input:** Ising weights  $\theta = \{J, \mathbf{h}\}$ , Number of spins  $N$ , total magnetization  $m$ , number of steps  $M$ , temperature  $T$
  - 2: Initialize  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)$  with  $m$  spins up and  $N - m$  spins down
  - 3: **for**  $t = 1$  to  $M$  **do**
  - 4:     Select at random  $i, j$  with  $\sigma_i = -\sigma_j$
  - 5:     Compute:
$$\Delta E = 2\sigma_i \left( \sum_{k=1}^N J_{ik} \sigma_k + h_i \right) + 2\sigma_j \left( \sum_{k=1}^N J_{jk} \sigma_k + h_j \right) - 4J_{ij} \sigma_i \sigma_j$$
  - 6:     Sample  $u \sim \mathcal{U}([0, 1])$
  - 7:     **if**  $\Delta E < 0$  or  $\exp(-\Delta E / T) > u$  **then**
  - 8:         Swap  $\sigma_i, \sigma_j$
  - 9:     **end if**
  - 10: **end for**
  - 11: **return**  $\sigma$
- 

<sup>1</sup>that is the fraction of spins pointing up

**Algorithm 3** REINFORCE algorithm for learning Ising-based NUS schedules

- 1: **Input:** NMR data  $p_{\text{data}}$ , reconstructor  $\text{rec}[\cdot]$ , reward function  $r(\cdot, \cdot)$ , batch size  $B$ , number of steps  $S$ , Kawasaki-MCMC parameters
- 2: Initialize  $\boldsymbol{\theta} = \{J, \mathbf{h}\}$
- 3: **for**  $s = 1, \dots, S$  **do**
- 4:     Sample  $\sigma^{(1)}, \sigma^{(2)}$  with Kawasaki-MCMC (Algorithm 2)
- 5:     Obtain  $U^{(1)}, U^{(2)}$  from  $\sigma^{(1)}, \sigma^{(2)}$
- 6:     Draw  $\mathbf{x}^{(k)} \sim p_{\text{data}} \forall k \in \{1, \dots, B\}$
- 7:     **for**  $i \in \{1, 2\}$  **do**
- 8:         Undersample  $\mathbf{y}^{(k,i)} = \mathcal{F}[U^{(i)}] * \mathbf{x}^{(k)} \forall k \in \{1, \dots, B\}$
- 9:         Reconstruct  $\hat{\mathbf{x}}^{(k,i)} = \text{rec}[\mathbf{y}^{(k,i)}] \forall k \in \{1, \dots, B\}$
- 10:          $r^{(i)} = \frac{1}{B} \sum_{k=1}^B r(\mathbf{x}^{(k)}, \hat{\mathbf{x}}^{(k,i)})$
- 11:     **end for**
- 12:     Compute RLOO estimator:

$$\mathbf{g} = \frac{1}{2}(r^{(1)} - r^{(2)}) \nabla_{\boldsymbol{\theta}} (E(\sigma^{(1)}; \boldsymbol{\theta}) - E(\sigma^{(2)}; \boldsymbol{\theta}))$$

- 13:     Update  $\boldsymbol{\theta} \leftarrow \text{Adam}(\boldsymbol{\theta}, \mathbf{g})$
- 14: **end for**

## Chapter 4

---

# Experiments

---

### 4.1 Reconstruction problem

#### 4.1.1 Training

We use the ADM repository and the associated UNet-based architecture by Dhariwal & Nichol [3]. Since each decaying plane wave in the FID can be described by only a few parameters, a relatively small network is sufficient to capture the relevant features. Accordingly, we choose a conservative configuration shown in Table 4.1. We use the default training settings, but replace the learned variance with a fixed cosine noise schedule to further reduce the number of trainable parameters. To construct the training data, we generate 50k synthetic spectra adapting the data creation process of Jahangiri et al. [43, 67] (see Appendix A.1 for details). For DPS, all experiments were conducted on an NVIDIA GeForce RTX 4090 GPU.

**Table 4.1:** Model card for pretrained NMR diffusion model

Component	Specification
Input resolution	$256 \times 256$
Base channels	64
Residual blocks per level	2
Attention	$8 \times 8$
Parameters	29727873
Training steps	12 500

#### 4.1.2 Reconstruction performance

**Dataset.** We evaluate DPS (see Algorithm 1) on  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra of three proteins that represent increasing levels of reconstruction difficulty: Ubiquitin (8.6 kDa), Azurin (14 kDa), and MALT1 (44 kDa). These spectra were chosen to enable direct comparison with the WNN model of Jahangiri et al. [43, 67], a conditional deep neural network evaluated on the same datasets, which makes it a suitable baseline. In addition, we benchmark against the traditional compressed sensing algorithm IST (see Algorithm 4), using implementation based on [52, 53].

**Undersampling schedules.** In previous work on WNN [43], the best performance was achieved when training with a single Poisson-gap sampling (PGS) schedule. We make use of the pretrained WNN models available in the MR-Ai repository [43, 67], where ten models

are provided for each of three undersampling ratios, with each model trained on a distinct PGS schedule. These ratios correspond to 11, 21, and 31 sampled points out of 128 in total, i.e. 8.6 %, 16.4 %, and 24.2 %, respectively. For our initial comparison, we evaluate DPS, WNN, and IST on these same schedules.

**Metrics.** Reconstruction quality is evaluated using three standard performance metrics: the coefficient of determination ( $R^2$ ), root mean squared error (RMSE) and Pearson correlation (Corr). We primarily report  $R^2$ , as we found it to be the most informative and sensitive metric, and provide RMSE and Corr in Table 4.2 for completeness.

**Algorithm details.** Compared to the original DPS framework, our implementation in Algorithm 1 includes an additional step to enforce data consistency in the time domain. Reconstruction is performed using a sliding window with stride 128 along the direct dimension, processing the spectrum piecewise. To account for the stochasticity of the diffusion process, DPS reconstructions are averaged over 10 independent runs with different noise realizations. Similar averaging over multiple initial guesses did not affect IST performance in our experiments. For DPS we use a fixed learning rate  $\zeta_i = 2.5$  and the default 1000 steps. All experiments were conducted on an NVIDIA GeForce RTX 4090 GPU. All methods operate in the virtual echo representation, where spectra are zerofilled and undersampling masks are symmetrized to 256 points.

**Results.** Figure 4.1 summarizes the reconstruction performance of DPS, IST, and WNN across the three proteins and four undersampling ratios. For Ubiquitin at 8.6 % undersampling, DPS and IST perform similarly, both clearly outperforming WNN. For Azurin at 16.4 %, the three methods yield nearly identical results, with IST achieving a higher  $R^2$  in the third decimal place. For MALT1 at 16.4 % and 24.2 %, DPS shows a clear advantage over both IST and WNN, and is the only method to exceed  $R^2 > 0.99$  at 24.2 % and  $R^2 > 0.98$  at 16.4 % undersampling.

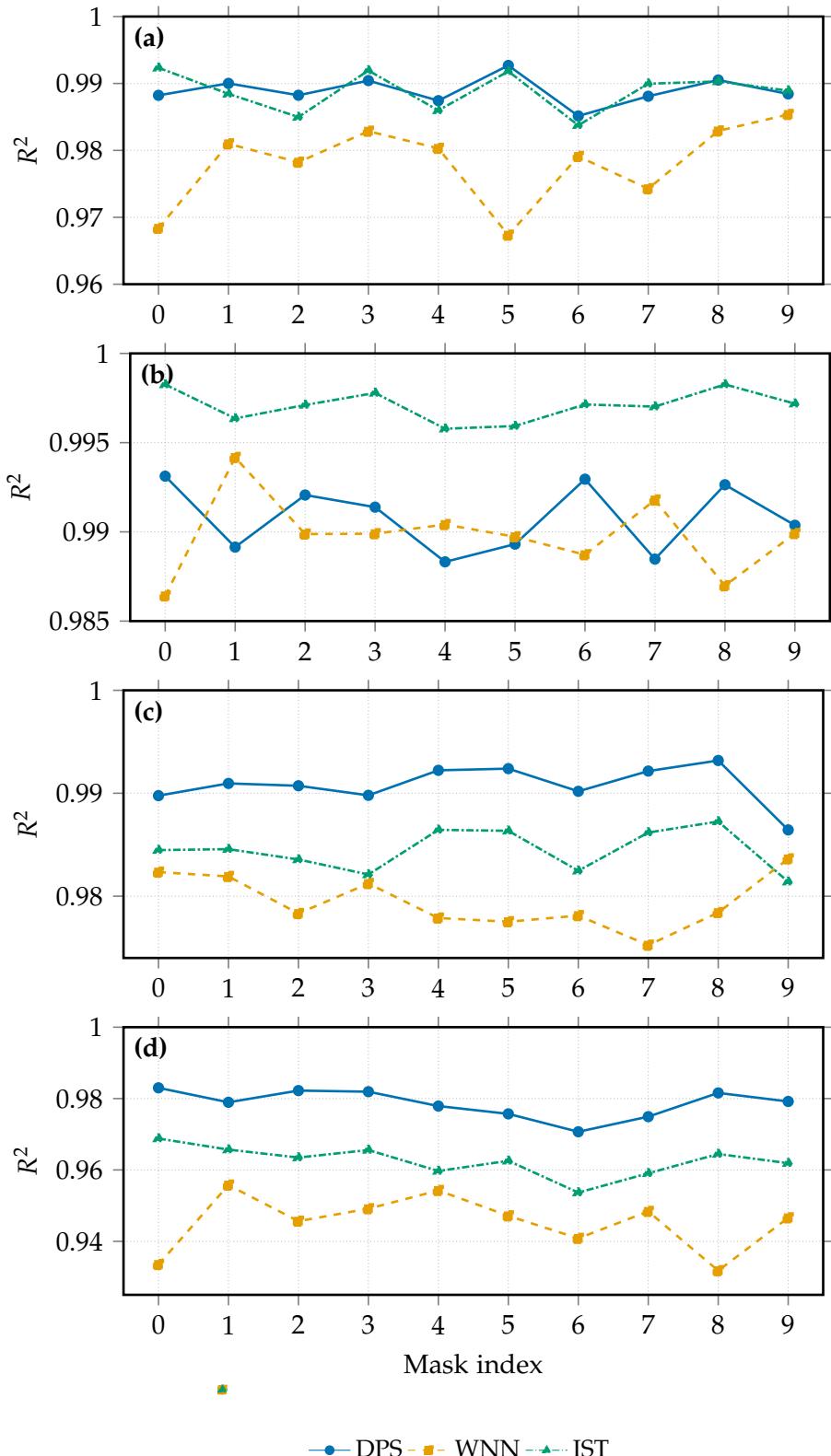
Reconstructions of MALT1 at 16.4 % undersampling are shown in Fig. A.3 where recovers the cleanest spectrum, while IST and WNN both exhibit distortions and artefacts. Reconstructions of Ubiquitin and Azurin and MALT1 at 24.2 % undersampling are shown in Appendix A.3.

The average performance across proteins and undersampling levels is summarized in Table 4.2. DPS and IST perform nearly identically on Ubiquitin and on Azurin at 11 sampled points, with IST achieving the best results for Azurin at 21 points. For MALT1, DPS clearly outperforms both IST and WNN at 21 and 31 points. With the exception of Azurin at 21 points, the fidelity of WNN reconstructions is lower than the other two methods.

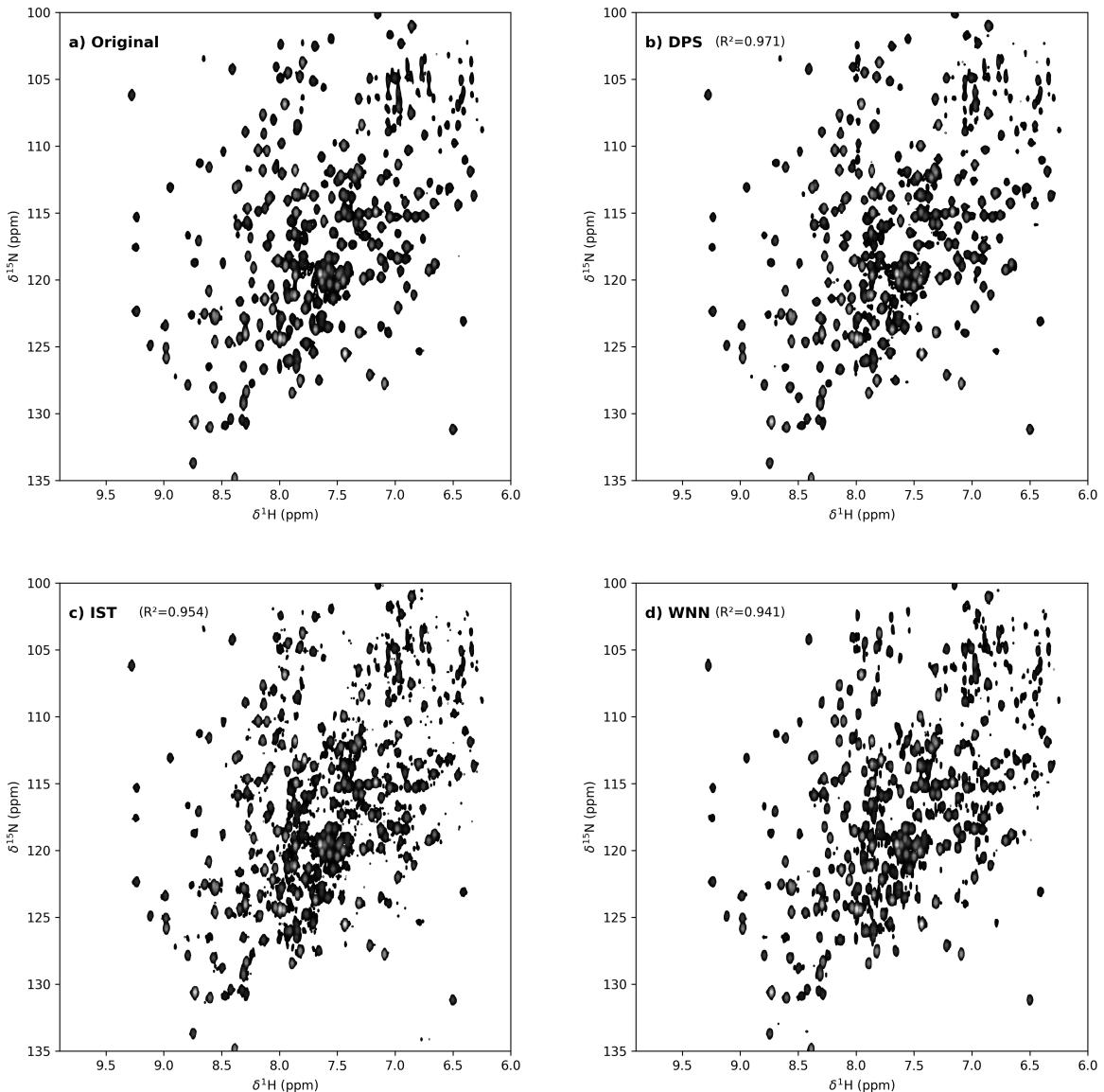
**Robustness to undersampling.** To assess performance across undersampling ratios, we use five sinusoidally weighted Poisson-gap (PGS) masks at each sparsity level, generated with the tool provided by Harvard Medical School (HMS) [36]. Figure 4.3 compares reconstructions of MALT1 obtained with DPS and IST under these schedules. Starting from a sampling density of 12/128 (9.35 %), DPS consistently surpasses IST. As can be seen from Fig. 4.7, DPS and IST perform similarly at this extreme undersampling level if only judged by  $R^2$ , but the visual quality show that DPS is already preferable.

## 4.2 Performance of optimized schedules

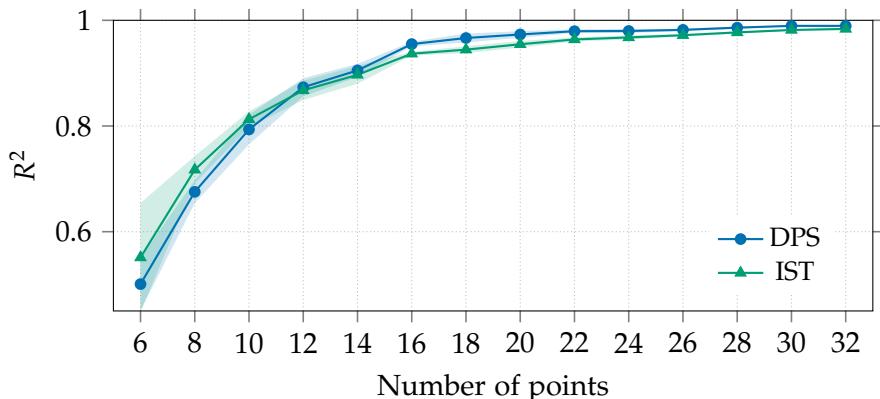
**Training.** We train an Ising model corresponding to a sampling density of 12/128, using both DPS and IST reconstruction algorithms with the training specifications summarized in Table 4.3. Training is performed using REINFORCE with the RLOO gradient estimator as



**Figure 4.1:** Coefficient of determination ( $R^2$ ) obtained across 10 distinct Poisson-gap sampling (PGS) masks at fixed undersampling ratios. (a) Ubiquitin with  $r = 8.6\%$  (11/128 points), (b) Azurin with  $r = 16.4\%$  (21/128 points), and (c) MALT1 with  $r = 24.2\%$  (31/128 points). (d) MALT1 with  $r = 16.4\%$  (21/128 points). Results are shown for DPS, WNN, and IST reconstructions under identical PG schedules. Note the different spacing on the vertical axis.



**Figure 4.2:** Reconstructed spectra for MALT1 undersampled at 21/128 (16.4 %, mask 5): (a) fully sampled reference, (b) DPS reconstruction, (c) IST reconstruction, (d) WNN reconstruction.



**Figure 4.3:** Mean reconstruction accuracy ( $R^2$ ) as a function of the number of sampled points out of 128. Shaded bands show  $\pm$  one standard deviation across five masks given by Poisson-Gap sampling.

**Table 4.2:** Mean reconstruction performance (RMS ↓, Corr ↑,  $R^2$  ↑) averaged over ten masks for each spectrum and undersampling levels. Each column indicates the spectrum and undersampling level out of 128 points.

Method	Ubiquitin-11	Azurin-11	Azurin-21	MALT1-21	MALT1-31	Metric
WNN	0.015	0.017	0.006	0.017	0.010	RMS
	0.990	0.962	0.995	0.973	0.990	Corr
	0.978	0.922	0.990	0.945	0.979	$R^2$
IST	<b>0.011</b>	<b>0.011</b>	<b>0.003</b>	0.013	0.009	RMS
	<b>0.995</b>	<b>0.981</b>	<b>0.998</b>	0.981	0.992	Corr
	<b>0.989</b>	<b>0.962</b>	<b>0.997</b>	0.962	0.984	$R^2$
DPS	<b>0.011</b>	<b>0.011</b>	0.006	<b>0.010</b>	<b>0.007</b>	RMS
	0.994	<b>0.981</b>	0.996	<b>0.990</b>	<b>0.996</b>	Corr
	0.988	<b>0.962</b>	0.991	<b>0.979</b>	<b>0.991</b>	$R^2$

described in Algorithm 3, using the coefficient of determination ( $R^2$ ) as the reward function and Adam for gradients [68].

Figure 4.4 reports the mean reward and the acceptance rate of MCMC proposals during training with DPS as the reconstruction algorithm. The spectra for training are generated on the fly according to the parameters shown in Table A.1 with the number of peaks fixed to 100. We found it beneficial to restrict spin interactions to  $k$ -nearest neighbors. With all-to-all connectivity, the Markov chain starts to mix poorly midway training, producing highly correlated samples and collapsing to a single mask.

**Table 4.3:** Training specifications for Ising model (DPS/IST)

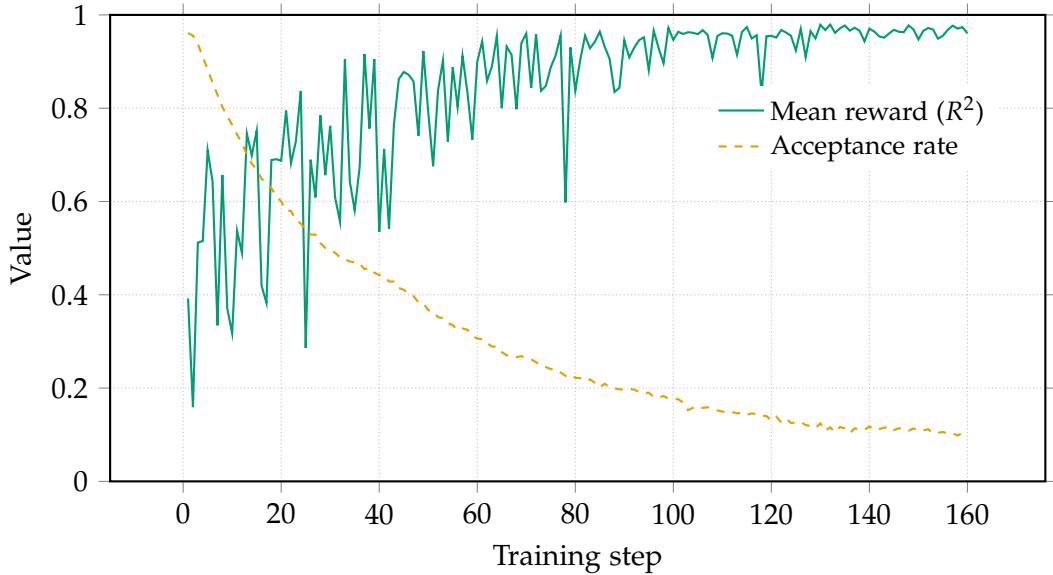
Component	Specification
Nearest neighbors	32
MC steps	$350 \times 127$
Temperature	2.5
Training steps	160 / 500
Learning rate	0.005
L2-regularization	0.01
Batch size	8 / 32

#### 4.2.1 DPS as the reconstruction algorithm

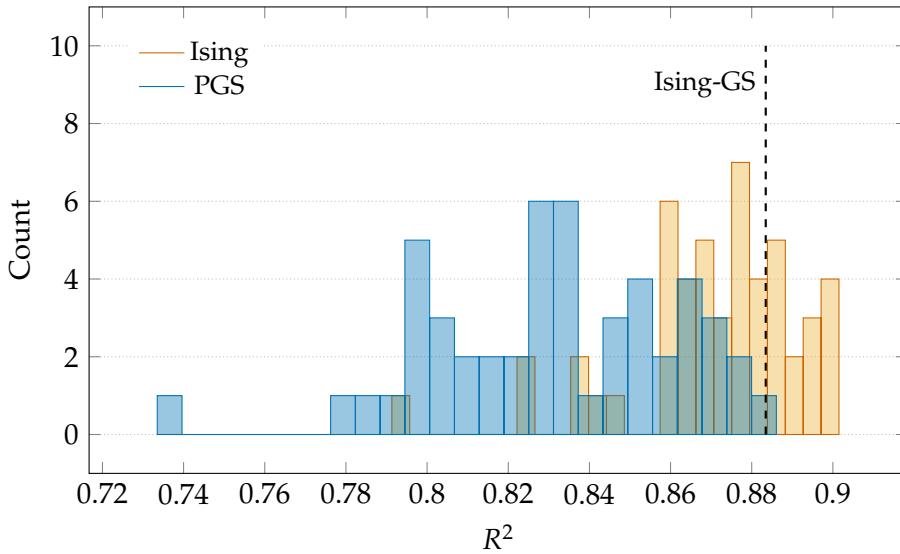
**Data and schedules.** To evaluate the effectiveness of the learned Ising masks for DPS, we benchmark their reconstruction performance against PGS masks at the same sampling density generated with the tool of HMS. We compare the reconstructions on the experimental MALT1 spectrum across 50 seeds. Here, we use the average over four runs to recover the original spectrum.

**Results.** Figure 4.5 shows the distribution of  $R^2$  scores using DPS as the reconstruction algorithm. Reconstructions obtained with Ising-trained masks consistently achieve higher  $R^2$  values than those from PGS masks. To establish a performance bound, we solve for the global optimum of the Ising model (see Eq. (2.34)), using the integer optimization solver Gurobi [69], and denote the corresponding mask as the Ising ground state (Ising-GS).

Among 50 PGS masks, only one attains comparable performance to the Ising-GS ( $R^2 > 0.88$ ). The Ising-GS, thus, provides a deterministic alternative to stochastic mask sampling.



**Figure 4.4:** Mean reward ( $R^2$ ) and acceptance rate of the MCMC proposals as a function of training step with DPS as the reconstruction algorithm.



**Figure 4.5:** Histogram of reconstruction quality ( $R^2$ ) of MALT1 using DPS averaging over four runs across 50 seeds for Ising and Poisson-Gap (PGS) masks. The dashed line corresponds to the  $R^2$  value of the Ising ground-state mask.

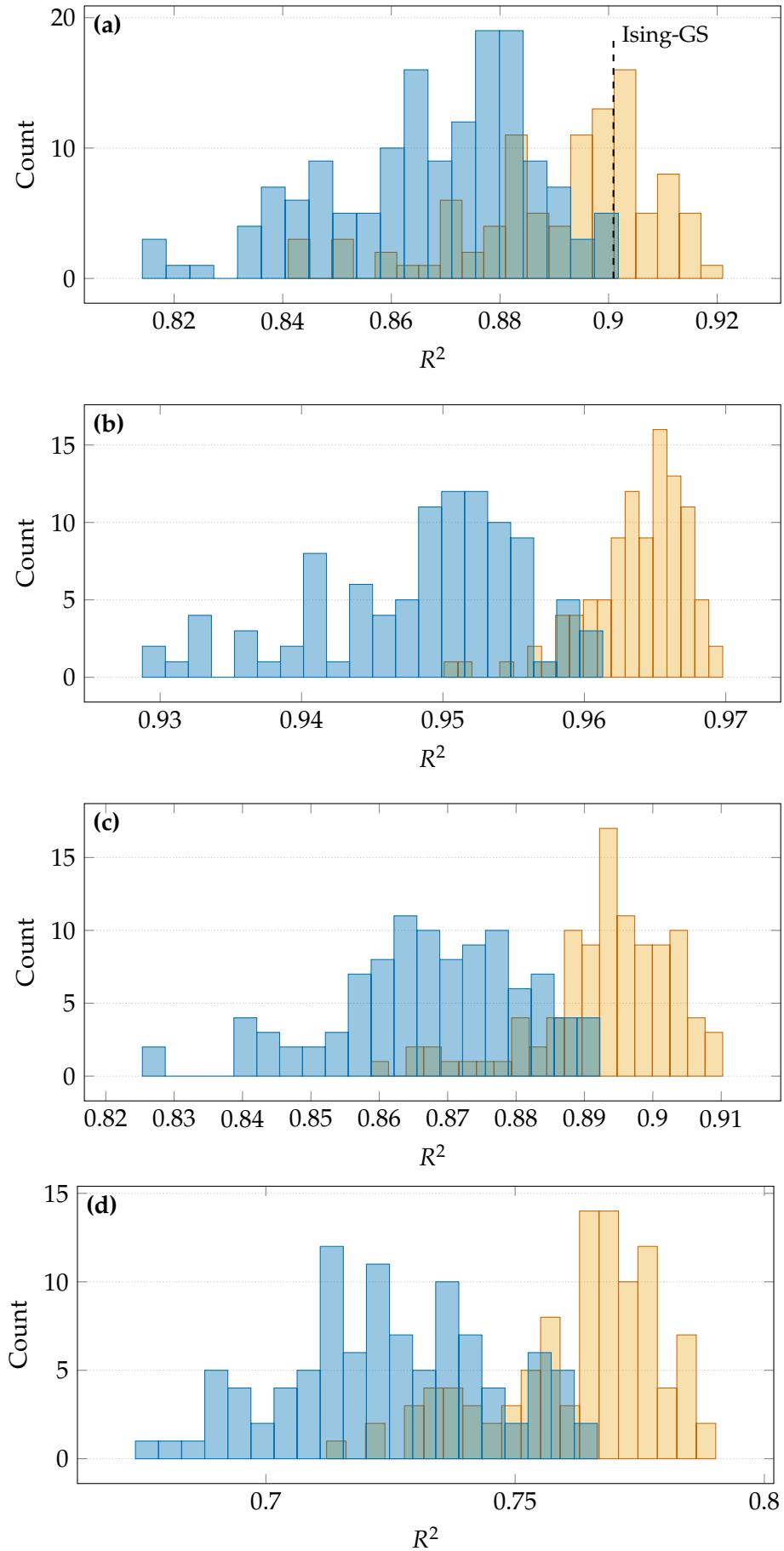
#### 4.2.2 IST as the reconstruction algorithm

**Data and schedules.** We conduct an analogous experiment with IST as the reconstruction algorithm given by Algorithm 4. We compare Ising masks against PGS masks across 100 random seeds. In addition to MALT1, we also evaluate performance on synthetic spectra containing 100, 250, and 500 peaks. Except for the peak count, the synthetic spectra are

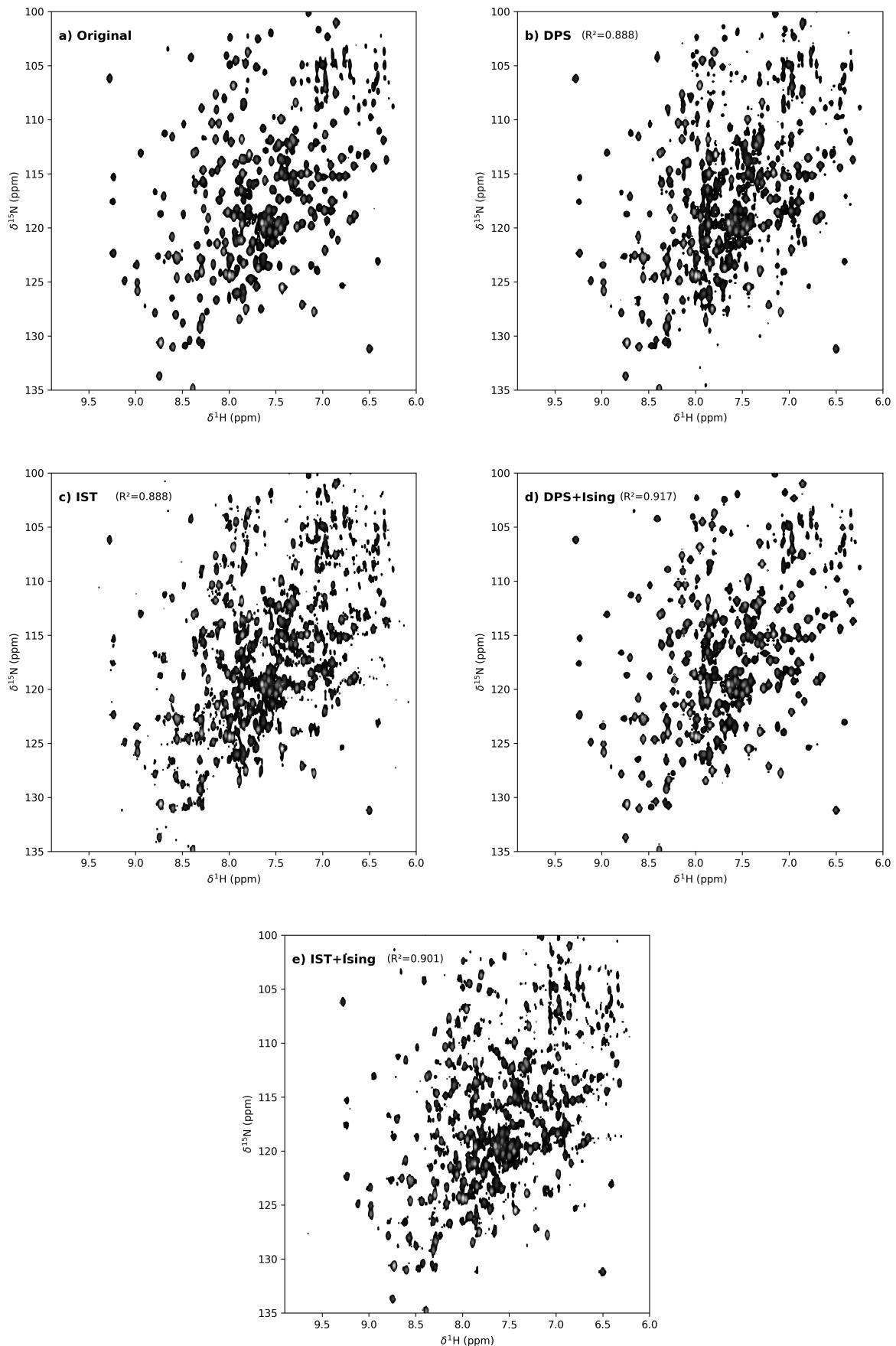
generated under the same conditions as the training data, with ten spectra reconstructed per mask amounting to 1000 different spectra in total for each experiment.

**Results.** For MALT1 (Fig. 4.6a), only three PGS masks exceed the  $R^2 > 0.90$  achieved by the Ising-GS. On synthetic spectra, Ising masks consistently outperform PGS regardless of peak count. The case with 100 peaks (Fig. 4.6b) represents a fully in-distribution setting, where the spread of the Ising distribution is especially narrow. While Ising masks outperform PGS across all four settings, the overlap between the two histograms is largest for MALT1. A likely explanation is distributional shift: synthetic spectra are simulated with ideal Lorentzians and uniformly random peak placement, whereas in MALT1 peaks are clustered.

Lastly, we present the reconstructions of MALT1 in Fig. 4.7 at undersampling level of 12 out of 128 points. Panels b) and c) show DPS and IST reconstructions obtained with a PGS mask, and panels d) and e) show reconstructions obtained with their respective Ising-GS masks. This sampling density is very low for a protein of the size of MALT1 and leads to degraded reconstructions for both methods under PGS masks. In contrast, the DPS + Ising-GS combination provides the best reconstruction, achieving the highest  $R^2$  score and producing the visually cleanest spectrum.



**Figure 4.6:** Histograms of reconstruction quality ( $R^2$ ) using IST across 100 seeds for Ising and PGS masks using  
a) MALT1 spectrum b) synthetic spectra with 100 peaks c) 250 peaks d) 500 peaks. In b), c) and d) we generate  
ten spectra per mask. Note that the histograms have different scales.



**Figure 4.7:** Reconstructions of the MALT1 spectrum: (a) fully sampled reference, (b) DPS reconstruction with PGS mask at 12/128, (c) IST reconstruction with the same PGS mask, (d) DPS reconstruction with Ising-GS mask at 12/128 and (e) IST reconstruction with Ising-GS mask.

## Chapter 5

---

# Discussion

---

This study investigated the use of diffusion models for reconstructing non-uniformly sampled (NUS) NMR spectra and the optimization of undersampling schedules.

### 5.1 Reconstruction with Diffusion Posterior Sampling

For the reconstruction task, we applied a modified version of diffusion posterior sampling (DPS) within the virtual-echo representation of NMR spectra. We compared DPS against both an unconditional deep neural network WNN [43] and a compressed sensing algorithm iterative soft thresholding (IST). Results were evaluated on  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra of three proteins of increasing molecular weight and reconstruction difficulty: Ubiquitin, Azurin, and MALT1.

In our experiments, DPS and IST achieve nearly identical performance for the lighter proteins Ubiquitin and Azurin for low and moderate undersampling levels. As spectra of Ubiquitin and Azurin are sparse with clearly isolated peaks, compressed sensing algorithms such as IST remain highly effective.

For MALT1, a large protein with a highly dense spectrum, DPS gave the best results across all reference metrics (RMSE, correlation,  $R^2$ ), while also producing visibly cleaner spectra at very low undersampling rates. These results show that generative priors can outperform traditional reconstruction algorithms in the most challenging cases with densely clustered spectra containing a large number of peaks.

Both DPS and IST clearly outperformed WNN, with the exception of Azurin at 16.4 % undersampling, even though the comparison was conducted on the fixed Poisson-gap (PGS) schedules used to train WNN. Unlike supervised methods, which must be trained for a specific undersampling density – and in the case of WNN, even a fixed sampling schedule – diffusion models offer far greater flexibility. Therefore, our results indicate that the unconditional generative approach to NUS reconstruction is considerably more appealing than supervised learning.

This flexibility comes at a considerable computational cost. The generative process in DPS is inherently stochastic and typically requires averaging over multiple runs to obtain stable results. Despite the use of a comparatively lightweight network architecture, inference remains substantially slower than IST. As a result, the current approach is not yet practical for 3D or 4D NMR applications without access to high-performance GPU resources. A potential next step is to use alternative diffusion pipelines with fewer sampling steps [6],

adopt deterministic sampling [70] or use a specific acceleration strategy of diffusion models for inverse problems [71].

We trained the diffusion model exclusively on synthetic spectra generated with randomly sampled parameters. Incorporating experimental data or synthetic spectra generated with chemically meaningful parameter distributions could further improve performance. The synthetic data used here was adapted from [43, 67], with parameters tailored specifically to  $^1\text{H}$ - $^{15}\text{N}$  correlation spectra; extending the approach to other types of experimental spectra, such as NOESY, would therefore require a different strategy.

One important limitation of this work is the limited validation in terms of data and metrics. We restricted our experiments to spectra from three proteins, selected to enable a direct comparison with WNN since these same spectra were used in their study; however, a larger dataset would have supported stronger conclusions. In addition, we relied on generic statistical measures ( $R^2$ , RMSE, correlation) that do not necessarily capture the aspects most relevant to practitioners. The utilization of domain-specific evaluation metrics, such as those developed in NUScon [72], would enable more meaningful assessments in future studies.

## 5.2 Optimizing undersampling patterns with the Ising model

We presented a simple yet novel approach for optimizing undersampling patterns for arbitrary reconstruction algorithms by formulating the problem with an Ising model. Training was performed using the REINFORCE policy gradient method [55] with a Leave-One-Out baseline [57], providing a learning signal derived from the entire reconstruction pipeline. Ising masks were optimized for both DPS and IST reconstructions using synthetic data at a low undersampling density of 12 out of 128 points.

Despite the simplicity of this training setup, the learned Ising masks consistently outperformed industry-standard Poisson Gap Sampling (PGS) masks. In experiments on the MALT1 spectrum, we compared 50 (DPS) and 100 (IST) sampled Ising and PGS masks, and in both cases, the Ising masks produced superior reconstructions. Given that MALT1 is a large protein, sampling only 12 out of 128 points would typically result in severely degraded spectra. On synthetic data, the performance difference between Ising and PGS masks was even larger. Future work should integrate experimental training data or synthetic spectra generated with peak parameters that closely match those observed experimentally.

An additional feature of the Ising formulation is that its ground state (the lowest-energy configuration) can be deterministically obtained via integer optimization. The reconstruction quality achieved by this Ising-GS mask was matched by only 1 out of 50 PGS masks with DPS and 3 out of 100 with IST in terms of  $R^2$  score. Among all tested settings, DPS combined with the Ising-GS mask produced the visually cleanest spectrum, outperforming both DPS with PGS masks and IST with the Ising-GS mask. These results indicate that pairing DPS with optimized Ising masks enables reliable reconstructions even in extreme low-sampling regimes.

The success of the simple Ising-based optimization further shows that reinforcement learning techniques can be effectively applied to sampling design. A natural extension of our method is adaptive sampling, where the weights of the Ising model are updated based on the current state of the reconstruction. This setting aligns more closely with a traditional reinforcement learning framework and can be formalized as a Partially Observable Markov Decision Process (POMDP). An additional advantage of the Ising formulation is its ability to encode pairwise correlations between variables, making it well-suited for capturing structured mask

patterns. Another interesting direction for future work is to increase the model order and explore  $k$ -body Ising models, which can capture higher-order interactions.

While our current formulation uses a simple energy-based model, it could in principle be replaced by far more expressive policy networks. The advantage of the Ising model lies in its computational efficiency. For comparison, training a transformer-based policy with REINFORCE for combinatorial optimization tasks such as the traveling salesman problem typically requires on the order of 2,500 batches of 512 instances per epoch for 100 epochs [73]. Although the tasks are not directly comparable, they share a similar spirit—selecting an optimal subset of nodes. In contrast, training our Ising model with DPS required only 8 spectra instances per step for 160 steps. This efficiency makes the approach particularly attractive for 3D NMR, where the benefits of NUS grow by orders of magnitude. Moreover, the ability to train Ising masks on a standard laptop makes them directly accessible to NMR practitioners for their preferred reconstruction algorithms.

A limitation in our method is that the REINFORCE estimator, despite the use of a baseline, remains to have a high variance, and the interpretability of the learned sampling patterns is less principled compared to alternative strategies such as greedy optimization.

---

## Bibliography

---

- [1] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, “Deep unsupervised learning using nonequilibrium thermodynamics,” in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, Eds., ser. Proceedings of Machine Learning Research, vol. 37, PMLR, 2015, pp. 2256–2265.
- [2] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 6840–6851.
- [3] P. Dhariwal and A. Nichol, “Diffusion models beat gans on image synthesis,” in *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., 2021, pp. 8780–8794.
- [4] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” in *International Conference on Learning Representations*, 2021.
- [5] Y. Song and S. Ermon, “Generative modeling by estimating gradients of the data distribution,” in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. Fox, and R. Garnett, Eds., vol. 32, Curran Associates, Inc., 2019.
- [6] T. Karras, M. Aittala, T. Aila, and S. Laine, “Elucidating the design space of diffusion-based generative models,” in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35, Curran Associates, Inc., 2022, pp. 26565–26577.
- [7] B. Kawar, M. Elad, S. Ermon, and J. Song, “Denoising diffusion restoration models,” in *Advances in Neural Information Processing Systems*, vol. 35, Curran Associates, Inc., 2022, pp. 23593–23606.
- [8] H. Chung, B. Sim, D. Ryu, and J. C. Ye, “Improving diffusion models for inverse problems using manifold constraints,” in *Advances in Neural Information Processing Systems*, vol. 35, Curran Associates, Inc., 2022.
- [9] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, “Diffusion posterior sampling for general noisy inverse problems,” in *The Eleventh International Conference on Learning Representations*, 2023.
- [10] J. Song, A. Vahdat, M. Mardani, and J. Kautz, “Pseudoinverse-guided diffusion models for inverse problems,” in *International Conference on Learning Representations*, 2023.

- [11] L. Rout, N. Raoof, G. Daras, C. Caramanis, A. Dimakis, and S. Shakkottai, “Solving linear inverse problems provably via posterior sampling with latent diffusion models,” in *Advances in Neural Information Processing Systems*, vol. 36, Curran Associates, Inc., 2023, pp. 49 960–49 990.
- [12] B. Song, S. M. Kwon, Z. Zhang, X. Hu, Q. Qu, and L. Shen, “Solving inverse problems with latent diffusion models via hard data consistency,” in *The Twelfth International Conference on Learning Representations*, 2024.
- [13] Y. Wang, J. Yu, and J. Zhang, “Zero-shot image restoration using denoising diffusion null-space model,” *The Eleventh International Conference on Learning Representations*, 2023.
- [14] B. Boys, M. Girolami, J. Pidstrigach, S. Reich, A. Mosca, and O. D. Akyildiz, *Tweedie moment projected diffusions for inverse problems*, 2024.
- [15] B. Zhang, W. Chu, J. Berner, C. Meng, A. Anandkumar, and Y. Song, “Improving diffusion inverse problem solving with decoupled noise annealing,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2025.
- [16] M. Mardani, J. Song, J. Kautz, and A. Vahdat, “A variational perspective on solving inverse problems with diffusion models,” in *The Twelfth International Conference on Learning Representations*, 2024.
- [17] Z. Dou and Y. Song, “Diffusion posterior sampling for linear inverse problem solving: A filtering perspective,” in *The Twelfth International Conference on Learning Representations*, 2024.
- [18] G. Daras *et al.*, *A survey on diffusion models for inverse problems*, 2024. arXiv: [2410.00083](#).
- [19] H. Chung, J. Kim, and J. C. Ye, *Diffusion models for inverse problems*, 2025. arXiv: [2508.01975](#).
- [20] Y. Zhu *et al.*, “Denoising Diffusion Models for Plug-and-Play Image Restoration,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2023, pp. 1219–1229.
- [21] H. Liu, J. Xing, M. Xie, C. Li, and T.-T. Wong, *Improved diffusion-based image colorization via piggybacked models*, 2023. arXiv: [2304.11105](#).
- [22] Y. Song, L. Shen, L. Xing, and S. Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *International Conference on Learning Representations*, 2022.
- [23] A. Jalal, M. Arvinte, G. Daras, E. Price, A. G. Dimakis, and J. Tamir, “Robust compressed sensing mri with deep generative priors,” in *Advances in Neural Information Processing Systems*, vol. 34, Curran Associates, Inc., 2021, pp. 14 938–14 954.
- [24] H. Chung and J. C. Ye, “Score-based diffusion models for accelerated mri,” *Medical Image Analysis*, vol. 80, p. 102 479, 2022.
- [25] S. Ravula, B. Levac, A. Jalal, J. I. Tamir, and A. G. Dimakis, *Optimizing sampling patterns for compressed sensing mri with diffusion generative models*, 2023. arXiv: [2306.03284](#).
- [26] H. Zheng *et al.*, *Inversebench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences*, 2025. arXiv: [2503.11043 \[cs.LG\]](#). [Online]. Available: <https://arxiv.org/abs/2503.11043>.
- [27] J. Keeler, “Understanding nmr spectroscopy,” 2002.

- [28] J. Barna, E. Laue, M. Mayger, J. Skilling, and S. Worrall, "Exponential sampling, an alternative method for sampling in two-dimensional nmr experiments," *Journal of Magnetic Resonance* (1969), vol. 73, no. 1, pp. 69–77, 1987.
- [29] D. L. Donoho, I. M. Johnstone, A. S. Stern, and J. C. Hoch, "Does the maximum entropy method improve sensitivity?" *Proceedings of the National Academy of Sciences*, vol. 87, no. 13, pp. 5066–5068, 1990.
- [30] S. Robson, H. Arthanari, S. G. Hyberts, and G. Wagner, "Nonuniform sampling for nmr spectroscopy," in *Biological NMR Part A*, ser. Methods in Enzymology, A. J. Wand, Ed., vol. 614, Academic Press, 2019, pp. 263–291.
- [31] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [32] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [33] D. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [34] I. Drori, "Fast minimization by iterative thresholding for multidimensional nmr spectroscopy," *Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 295–300, 2007.
- [35] K. Kazimierczuk and V. Orekhov, "Accelerated nmr spectroscopy by using compressed sensing," *Angewandte Chemie (International ed. in English)*, vol. 50, pp. 5556–9, Jun. 2011. DOI: [10.1002/anie.201100370](https://doi.org/10.1002/anie.201100370).
- [36] S. G. Hyberts, A. G. Milbradt, A. B. Wagner, H. Arthanari, and G. Wagner, "Application of iterative soft thresholding for fast reconstruction of nmr data non-uniformly sampled with multidimensional poisson gap scheduling," *Journal of biomolecular NMR*, 20012.
- [37] A. S. Stern, D. L. Donoho, and J. C. Hoch, "Nmr data processing using iterative thresholding and minimum 11-norm reconstruction," *Journal of Magnetic Resonance*, vol. 188, no. 2, pp. 295–300, 2007.
- [38] J. Ying, F. Delaglio, D. Torchia, and A. Bax, "Sparse multidimensional iterative lineshape-enhanced (smile) reconstruction of both non-uniformly sampled and conventional nmr data," *Journal of Biomolecular NMR*, 2017.
- [39] S. Gull and G. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686–690, 1978.
- [40] D. Rovnyak, J. Hoch, A. Stern, and G. Wagner, "Resolution and sensitivity of high field nuclear magnetic resonance spectroscopy," *Journal of Biomolecular NMR*, 2004.
- [41] X. Qu *et al.*, "Accelerated nuclear magnetic resonance spectroscopy with deep learning," *Angewandte Chemie (International ed. in English)*, 2020.
- [42] G. Karunanithy and D. F. Hansen, "A versatile deep neural network architecture for nmr spectral reconstruction and virtual decoupling," *Journal of Biomolecular NMR*, 2021.
- [43] A. Jahangiri *et al.*, "Nmr spectrum reconstruction as a pattern recognition problem," *Journal of Magnetic Resonance*, vol. 346, p. 107342, 2023.
- [44] Y. Luo *et al.*, "Deep learning network for nmr spectra reconstruction in time-frequency domain and quality assessment," *Nature Communications* 16, 2025.

- [45] S. Yan, F. Gabellieri, E. Goffinet, F. Castiglione, and T. Launey, "Diffnmr: Advancing inpainting of randomly sampled nuclear magnetic resonance signals," in *Computational Science and Computational Intelligence*, Springer Nature Switzerland, 2025, pp. 301–314.
- [46] S. G. Hyberts, K. Takeuchi, and G. Wagner, "Poisson-gap sampling and forward maximum entropy reconstruction for enhancing the resolution and sensitivity of protein nmr data," *Journal of the American Chemical Society*, vol. 132, no. 7, pp. 2145–2147, 2010.
- [47] S. G. Hyberts and G. Wagner, "High fidelity sampling schedules for nmr spectra of high dynamic range," *Journal of Magnetic Resonance*, vol. 339, p. 107228, 2022.
- [48] C. D. Bahadir, A. Q. Wang, A. V. Dalca, and M. R. Sabuncu, "Deep-learning-based optimization of the under-sampling pattern in mri," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1139–1152, 2020.
- [49] J. Zhang *et al.*, "Extending loupe for k-space under-sampling pattern optimization in multi-coil mri," in *International Workshop on Machine Learning for Medical Image Reconstruction*, Springer, 2020, pp. 91–101.
- [50] J. Ren, K. Shang, Q. Jin, C. Wu, and D. Liang, "Escape the corrupted lines by a learning sampling pattern for mri reconstruction," in *Proceedings of the International Conference on Computing, Machine Learning and Data Science*, ser. CMLDS '24, New York, NY, USA: Association for Computing Machinery, 2024.
- [51] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," in *International Conference on Learning Representations*, 2017.
- [52] B. McKibben, *Mr\_utils: Iterative soft thresholding (ist)*, [https://mr-utils.readthedocs.io/en/latest/modules/cs/thresholding/iterative\\_soft\\_thresholding.html#mr\\_utils.cs.thresholding.iterative\\_soft\\_thresholding.IST](https://mr-utils.readthedocs.io/en/latest/modules/cs/thresholding/iterative_soft_thresholding.html#mr_utils.cs.thresholding.iterative_soft_thresholding.IST), Accessed: 2025-09-26, 2025.
- [53] M. Rani, S. B. Dhok, and R. B. Deshmukh, "A systematic review of compressive sensing: Concepts, implementations and applications," *IEEE Access*, vol. 6, pp. 4875–4894, 2018.
- [54] E. Ising, "Contribution to the Theory of Ferromagnetism," *Zeitschrift für Physik*, vol. 31, pp. 253–258, 1925.
- [55] R. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning* 8, pp. 229–256, 1992.
- [56] M. Bånekstad, J. R. Andersson, S. Mair, and J. Sjölund, "Ising on the graph: Task-specific graph subsampling via the Ising model," in *The Third Learning on Graphs Conference*, 2024. [Online]. Available: <https://openreview.net/forum?id=HFRCBCHHz4>.
- [57] W. Kool, H. van Hoof, and M. Welling, *Buy 4 REINFORCE samples, get a baseline for free!* 2019. [Online]. Available: <https://openreview.net/forum?id=r1lgTGL5DE>.
- [58] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, no. 24, pp. 695–709, 2005.
- [59] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, 2011.
- [60] B. D. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.
- [61] B. Efron, "Tweedie's formula and selection bias," *Journal of the American Statistical Association*, vol. 106, no. 496, pp. 1602–1614, 2011.

- 
- [62] R. de L. Kronig, "On the theory of dispersion of x-rays," *Journal of the Optical Society of America*, vol. 12, no. 6, pp. 547–557, Jun. 1926.
  - [63] H. A. Kramers, "La diffusion de la lumière par les atomes," *Transactions of Volta Centenary Congress*, pp. 545–557, 1927.
  - [64] M. Mayzel, K. Kazimierczuk, and V. Y. Orekhov, "The causality principle in the reconstruction of sparse nmr spectra," *Chemical Commun.*, vol. 50, pp. 8947–8950, 64 2014.
  - [65] F. Barahona, "On the computational complexity of ising spin glass models," *Journal of Physics A: Mathematical and General*, vol. 15, no. 10, p. 3241, 1982.
  - [66] M. Newman and G. Barkema, *Monte Carlo Methods in Statistical Physics*. Clarendon Press., 1999.
  - [67] A. Jahangiri and V. Orekhov, "Beyond traditional magnetic resonance processing with artificial intelligence," *Nature, Communications Chemistry*, vol. 7, 2024.
  - [68] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, 2014.
  - [69] Gurobi Optimization, LLC, *Gurobi Optimizer Reference Manual*, 2024. [Online]. Available: <https://www.gurobi.com>.
  - [70] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," in *International Conference on Learning Representations*, 2021.
  - [71] H. Chung, B. Sim, and J. C. Ye, "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 413–12 422.
  - [72] Y. Pustovalova *et al.*, "Nuscon: A community-driven platform for quantitative evaluation of nonuniform sampling in nmr," *Magnetic Resonance*, vol. 2, no. 2, pp. 843–861, 2021.
  - [73] W. Kool, H. van Hoof, and M. Welling, "Attention, learn to solve routing problems!" In *International Conference on Learning Representations*, 2019. [Online]. Available: <https://openreview.net/forum?id=ByxBFsRqYm>.

## Appendix A

---

# Appendix

---

### A.1 Generating synthetic training data

For training the model we construct a synthetic data set like is typically done in training deep neural networks in NMR. We adapt the data creation process from the MRAI repository. We start by generating  $128 \times 256$  FID signal following Eq. (2.26) with the parameters given by Table A.1, apply a weighting function, zero-fill the indirect dimension to 256 points and take the Fourier transform to obtain  $256 \times 256$  spectrum. We use only the real part of the spectrum and normalize it by dividing by maximum. We note that since we are training a generative prior we do not add phase errors or noise to the spectrum.

**Table A.1:** Parameter ranges for 2D FID (adapted from Jahangiri & Orekhov, 2024).

Parameter	Direct dimension	Indirect dimension
$T \in \mathbb{N}$	256	128
$\omega_n \in \mathbb{R}$	$[-0.5, 0.5]$	$[-0.5, 0.5]$
$\tau_n \in \mathbb{R}$	$[12.8, 128]$	$[256, 1280]$
$\varphi_n \in \mathbb{R}$	0	0
$a_n \in \mathbb{R}$	$[10^{-3}, 1]$	—
$N \in \mathbb{N}$	$1 - 300$	—

### A.2 Iterative Soft Thresholding (IST)

The Iterative Soft Thresholding (IST) algorithm [daubechies2004] alternates between a gradient descent-like step for data consistency and a soft-thresholding operator  $S_\tau$  with threshold  $\tau > 0$  that enforces sparsity. The operator acts componentwise on a vector  $\mathbf{z} \in \mathbb{C}^n$  like

$$S_\tau(z_j) = \text{sgn}(z_j) \max(|z_j| - \tau, 0). \quad (\text{A.1})$$

We adapt the implementation from [52, 53] where thresholds are contracted by a scalar  $\mu \in (0, 1]$  such that  $\tau_t = \mu \tau_{t-1}$ . This gives the IST update rule

$$\mathbf{x}^{(t+1)} = S_{\tau_t} \left( \mathbf{x}^{(t)} + A^\top (y - A\mathbf{x}^{(t)}) \right). \quad (\text{A.2})$$

The full algorithm is given by Algorithm 4

---

**Algorithm 4** Iterative Soft Thresholding (IST)

---

```

1: Input: Forward operator  $A$ , measurement  $y$ , initial threshold  $\tau_0$ , sparsity guess  $k$ ,
contraction  $\mu$ , number of iterations  $M$ , tolerance  $\varepsilon$ 
2: Initialize  $\mathbf{x} = \mathbf{0}$ ,  $\tau_0 = k$  : th largest entry of  $|A^\dagger y|$ 
3: for  $t = 1$  to  $M$  do
4:    $\mathbf{x} = S_{\tau_t}(\mathbf{x} + A^\top(y - A\mathbf{x}))$ 
5:   if  $\|y - A\mathbf{x}\|/\|y\| < \varepsilon$  then break
6:   end if
7:    $\tau_{t+1} = \mu\tau_t$ 
8: end for
9: Return  $\mathbf{x}$ 
```

---

We use following parameters  $k = 120$ ,  $M = 2000$ ,  $\varepsilon = 10^{-8}$ ,  $\mu = 0.995$ .

### A.3 Reconstruction figures

## Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

---

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

**Title of work** (in block letters):

**Authored by** (in block letters):

*For papers written by groups the names of all authors are required.*

**Name(s):**

---

---

---

---

---

**First name(s):**

---

---

---

---

---

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

**Place, date**

**Signature(s)**

---

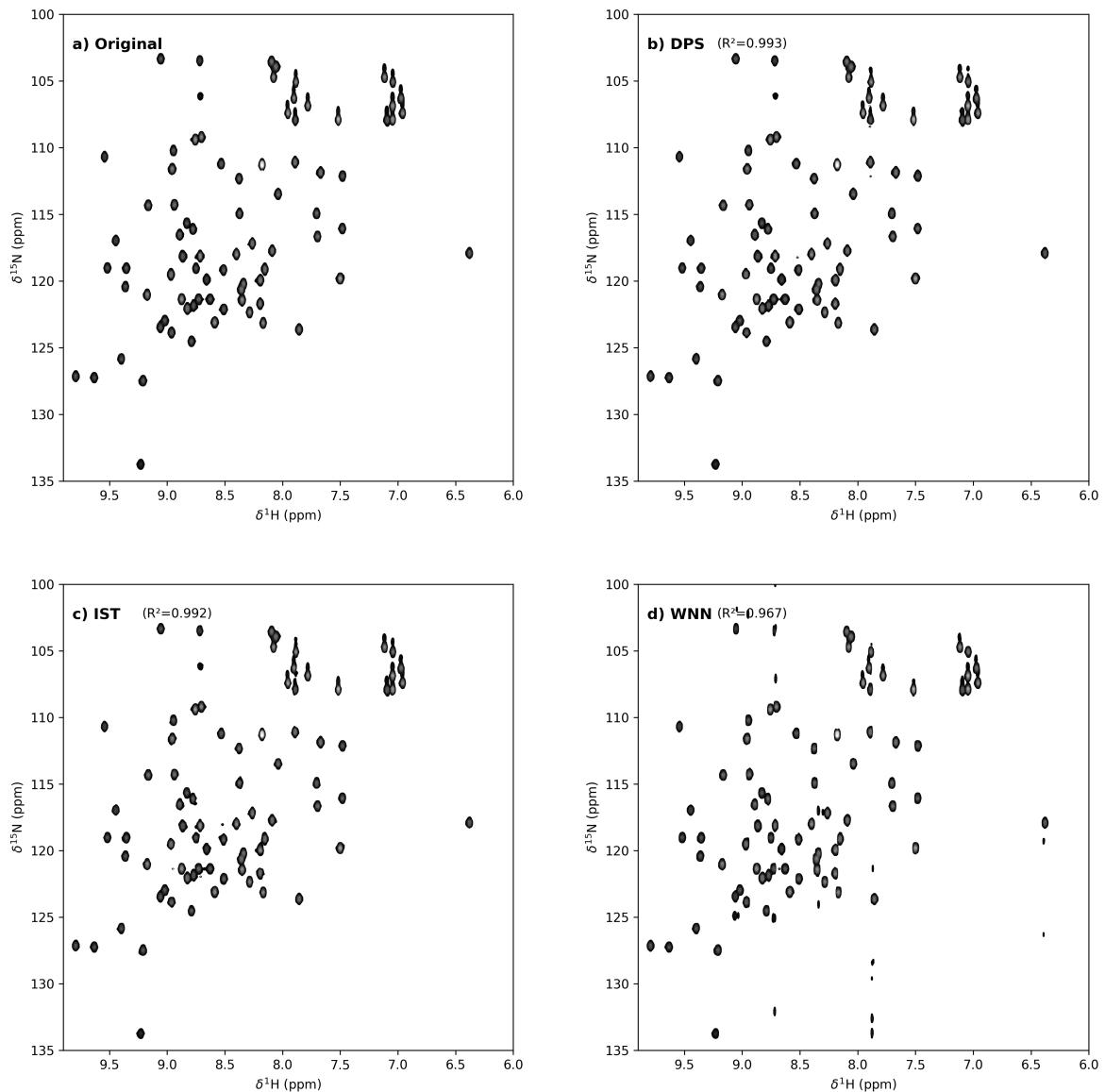
---

---

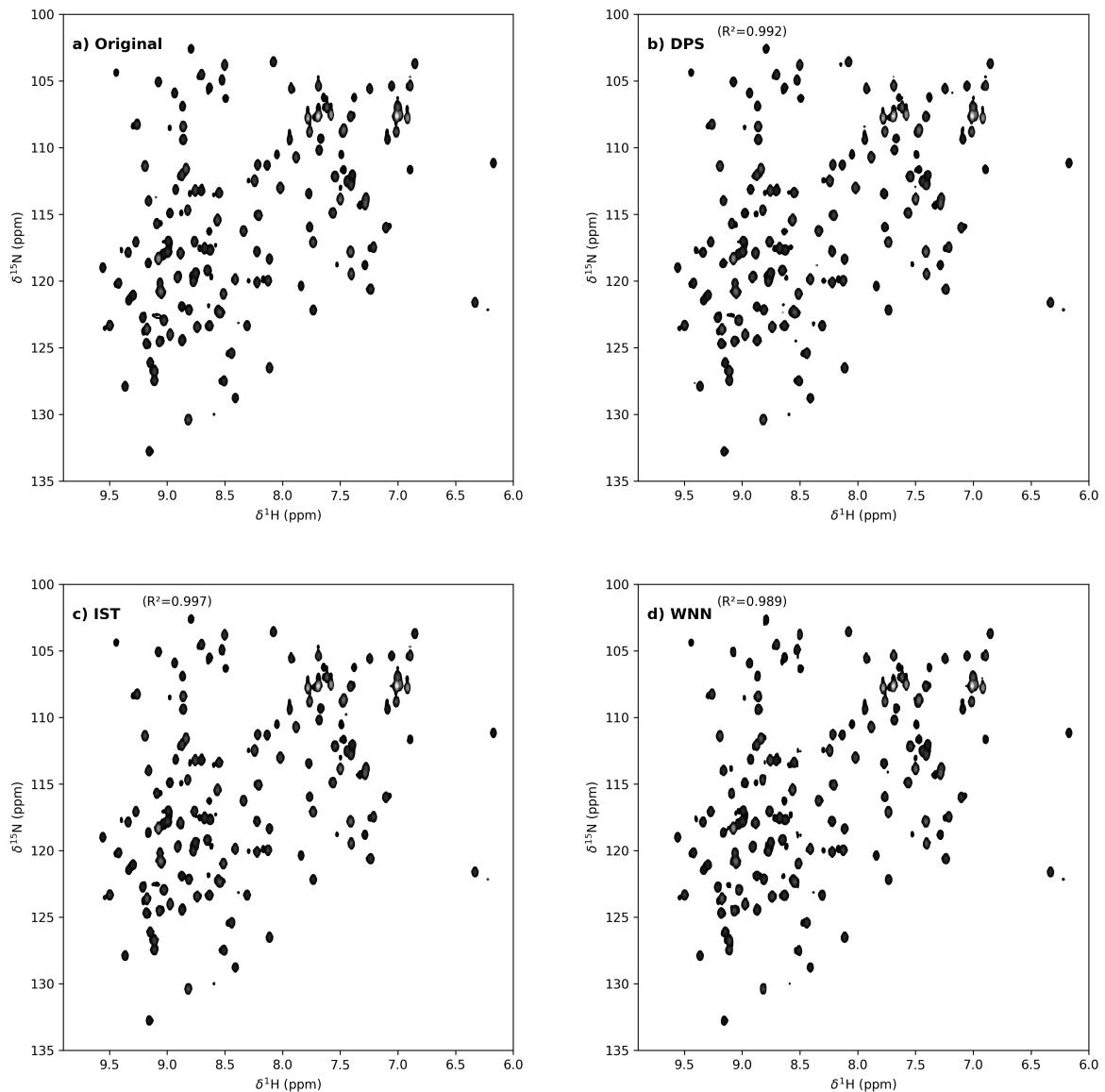
---

---

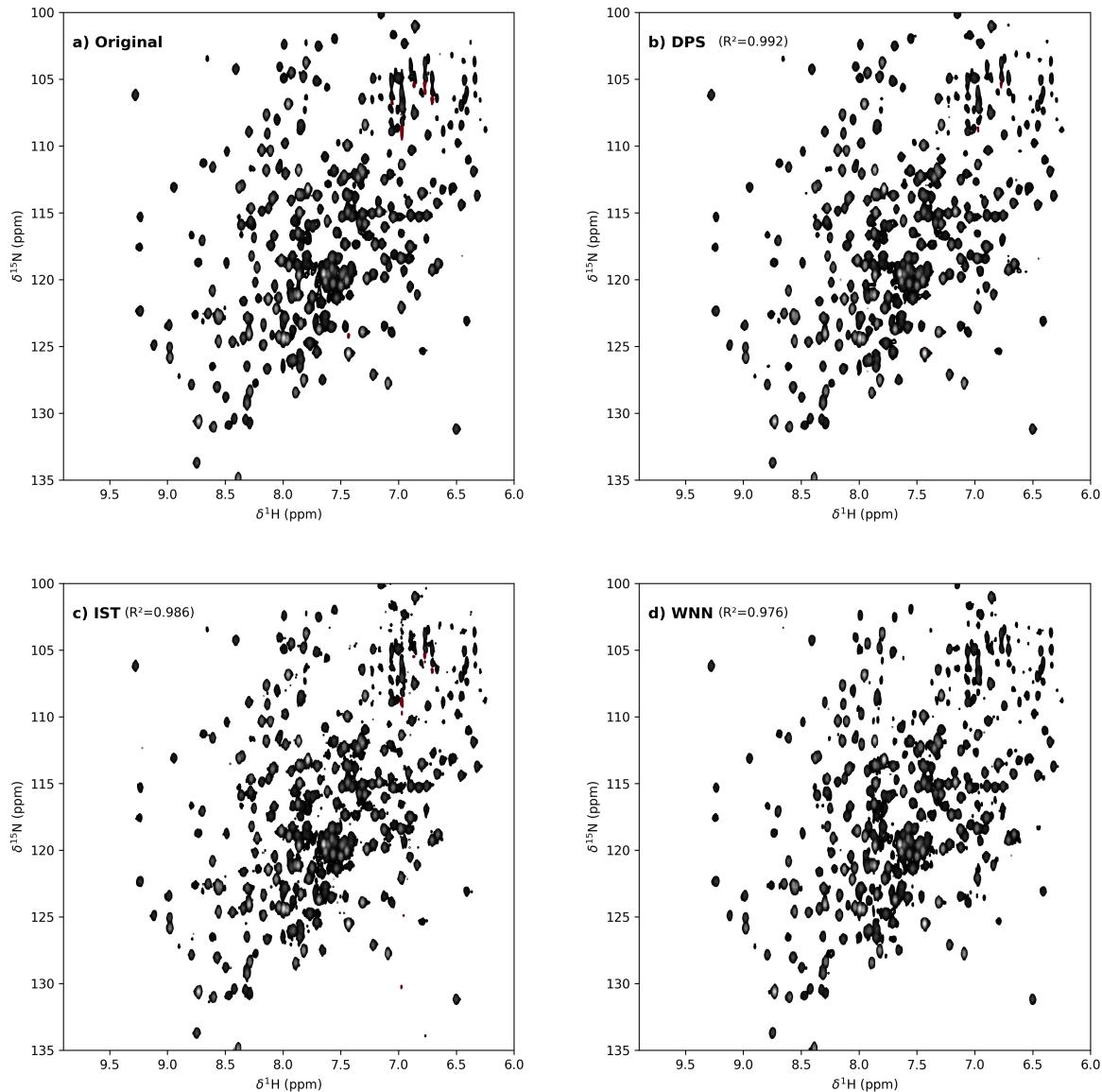
*For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.*



**Figure A.1:** Reconstructed spectra for Ubiquitin undersampled at 11/128 (8.6 %, mask 6): (a) fully sampled reference, (b) DPS reconstruction, (c) IST reconstruction, (d) WNN reconstruction.



**Figure A.2:** Reconstructed spectra for Azurin undersampled at 21/128 (16.4 %, mask 6): (a) fully sampled reference, (b) DPS reconstruction, (c) IST reconstruction, (d) WNN reconstruction.



**Figure A.3:** Reconstructed spectra for MALT1 undersampled at 31/128 (24.2 %, mask 5): (a) fully sampled reference, (b) DPS reconstruction, (c) IST reconstruction, (d) WNN reconstruction.