Quantum approximate Bayesian computation for NMR model inference

Dries Sels,1,2 Hesam Dashti,3 Samia Mora,3,4 Olga Demler,3 and Eugene Demler1

1Department of Physics, Harvard University, 17 Oxford st., Cambridge, MA 02138, USA
2Theory of quantum and complex systems, Universiteit Antwerpen, B-2610 Antwerpen, Belgium
3Division of Preventive Medicine, Brigham and Womens Hospital, Harvard Medical School, 900 Commonwealth Ave., Boston, MA 02215, USA
4Division of Cardiovascular Medicine, Brigham and Women's Hospital, Harvard Medical School, 900 Commonwealth Ave., Boston, MA 02215, USA

(Dated: November 1, 2019)

Recent technological advances may lead to the development of small scale quantum computers capable of solving problems that cannot be tackled with classical computers. A limited number of algorithms has been proposed and their relevance to real world problems is a subject of active investigation. Analysis of many-body quantum system is particularly challenging for classical computers due to the exponential scaling of Hilbert space dimension with the number of particles. Hence, solving problems relevant to chemistry and condensed matter physics are expected to be the first successful applications of quantum computers. In this paper, we propose another class of problems from the quantum realm that can be solved efficiently on quantum computers: model inference for nuclear magnetic resonance (NMR) spectroscopy, which is important for biological and medical research. Our results are based on the cumulation of three interconnected studies. Firstly, we use methods from classical machine learning to analyze a dataset of NMR spectra of small molecules. We perform a stochastic neighborhood embedding and identify clusters of spectra, and demonstrate that these clusters are correlated with the covalent structure of the molecules. Secondly, we propose a simple and efficient method, aided by a quantum simulator, to extract the NMR spectrum of any hypothetical molecule described by a parametric Heisenberg model. Thirdly, we propose an efficient variational Bayesian inference procedure for extracting Hamiltonian parameters of experimentally relevant NMR spectra.

INTRODUCTION

One of the central challenges for quantum technologies during the last few years has been a search for useful applications of near-term quantum machines [11]. While considerable progress has been achieved in increasing the number of qubits and improving their quality [2, 3], in the near future, we expect the number of reliable gates to be limited by noise and decoherence; the so called Noisy Intermediate-Scale Quantum (NISQ) era. As such, hybrid quantum-classical methods have been proposed to make the best out of the available quantum hardware and supplement it with classical computation. Most notably, there has been the development of the Quantum Approximate Optimization Algorithm (QAOA) [4] and the Variational Quantum Eigensolver (VQE) [5]. Both algorithms use the quantum computer to prepare variational states, some of which might be inaccessible through classical computation, but use a classical computer to update the variational parameters. A number of experiments have already been performed, demonstrating the feasibility of these algorithms [6-8], yet their bearing on real world problems remains unclear. In model-based statistical inference one is often faced with similar problems. For simple models one can find the likelihood and maximize it but for complex models the likelihood is typically intractable [9, 10]. Nuclear magnetic resonance (NMR) spectroscopy is a perfect example: there is a good understanding of the type of model that should be used (see equation (1)) and one only needs to determine the appropriate parameters. However, computing the NMR spectrum for a specific model requires performing computations in the exponentially large Hilbert space, which makes it extremely challenging for classical computers. This feature has been one of the original motivations for proposing NMR as a platform for quantum computing [11]. While it has been shown that no entanglement is present during NMR experiments [12, 13], strong correlations make it classically intractable [14]. Its computational power is between classical computation and deterministic quantum computation with pure states [15], which makes it an ideal candidate for hybrid quantum-classical methods. By simulating the model on a quantum computer, it runs efficiently while the remaining inference part is simply solved on a classical computer. One can think of this as an example of quantum Approximate Bayesian Computation (qABC), putting it in the broader scope of quantum machine learning methods [16]. In contrast to most of the proposed quantum machine learning applications, the present algorithm does not require challenging routines such as amplitude amplification [17, 18] or Harrow-Hassidim-Lloyd (HHL) algorithm [19].

NMR-SPECTROSCOPY

NMR spectroscopy is a spectroscopic technique which is sensitive to local magnetic fields around atomic nuclei.
Typically, samples are placed in a high magnetic field while driving RF-transitions between the nuclear magnetic states of the system. Since these transitions are affected by the intramolecular magnetic fields around the atom and the interaction between the different nuclear spins, one can infer details about the electronic and thus chemical structure of a molecule in this way. One of the main advantages of NMR is that it is non-destructive, in contrast to, for example, X-ray crystallography or mass spectrometry. This makes NMR one of the most powerful analytical techniques available to biology, as it is suited for in vivo and in vitro studies. NMR can, for example, be used for identifying and quantifying small molecules in biological samples (serum, cerebrospinal fluid, etc.). On the other hand, NMR experiments have limited spectral resolution and as such face the challenge of interpreting the data, since extracted information is quite convoluted. We only directly observe the magnetic spectrum of a biological sample, whereas our goal is to learn the underlying microscopic Hamiltonian and ultimately identify and quantify the chemical compounds. While this inference is tractable for small molecules, it quickly becomes problematic, making inference a slow and error-prone procedure. The analysis can be simplified by incorporating a priori spectral information in the parametric model. For that purpose, considerable attention has been devoted to determining NMR model parameters for relevant metabolites such as those found in plasma, cerebrospinal fluid and mammalian brains.

In what follows we will be concerned with 1D proton NMR but generalization to other situations is straightforward. For liquid $^1$H-NMR, a Heisenberg Hamiltonian

$$H(\theta) = \sum_{i,j} J_{ij} S_i \cdot S_j + \sum_i h_i S_i^z,$$

yields a reasonable effective description for the nuclear spins, where $\theta$ explicitly denotes the dependence of the Hamiltonian on its parameters $\theta = \{J_{ij}, h_i\}$. Here $J_{ij}$ encodes the interaction between the nuclear spins $S$ and $h_i$ is the effective local magnetic field. Note that this Hamiltonian contains two essential approximations (i) the interactions are chosen to $SU(2)$ invariant and (ii) the local magnetic fields — called chemical shifts in the NMR literature — are unidirectional. The rationale for the latter is that most of these local magnetic fields are caused by diamagnetic screening due to electronic currents induced by the large external magnetic field. This field will tend to oppose the external field and hence be largely uniaxial. For liquid state NMR, the rapid tumbling of the molecules averages out the dipolar coupling between the nuclei, approximately resulting in isotropic exchange interactions between nuclear spins. The fact that the interactions are rotationally invariant, allows us to remove the average (external) field from the Hamiltonian, i.e. $S_{tot}^z = \sum_i S_i^z$ commutes with Hamiltonian and will therefore only shift the NMR spectrum.

Within linear response the evolution of the system subject to a radio frequency $z$-magnetic field is determined by the response function:

$$S(t|\theta) = \text{Tr} \left[ e^{i H(\theta) t} S_{tot}^z e^{-i H(\theta) t} S_{tot}^z \rho_0 \right],$$

where $\rho_0$ denotes the initial density matrix of the system and $S_{tot}^z = \sum S_i^z$. The measured spectrum is simply given by:

$$A(\omega|\theta) = \text{Re} \int_0^\infty dt e^{\gamma t} S(t|\theta),$$

where $\gamma$ is the effective decoherence rate. For room temperature $^1$H-NMR, the initial density matrix can be taken to be an infinite temperature state, i.e.

$$\rho_0 \approx \frac{1}{\text{Tr} \{1\}}.$$

Indeed, even a 20 T magnetic field will only lead to a bare proton resonance frequency of about 900 MHz. In contrast, room temperature is about $40 \text{ THz}$, so for all practical purposes we can consider it equally likely for the spin to be in the excited state or in the ground state. Chemical shifts $h_i$ are of the order of a few parts per million, resulting in local energy shifts of a few kHz, while the coupling or interaction strength $J$ is of the order of a few Hz. Despite these low frequencies and the high temperature of the system, one can typically still infer the parameters due to the small decoherence rate of the proton nuclear spin. Due to the absence of a magnetic quadrupole moment, the protons do not decohere from the electric dipole fluctuations caused by the surrounding water molecules. This gives the proton nuclear spin a coherence time of the order of seconds to tens of seconds, sufficiently long to create some correlations between the various spins. The remaining part of this work is concerned with the question of how to infer the model parameters $J_{ij}$ and $h_i$ of our effective Hamiltonian from a measured spectrum.

### CLASSICAL LEARNING FROM DATA

Given real NMR data, summarized by the experimentally acquired spectrum $A(\omega)$, our goal, in general, is to learn a parametrized generative model which explains how this NMR data is generated. Fortunately, we have a good idea about the physics which allows us to write down a model, i.e. expressions, that is close to reality thereby ensuring a small misspecification error. The drawback however is that the model is analytically intractable and becomes increasingly complex to simulate with increasing number of spins. In the next section we will discuss how to alleviate this problem by using a programmable quantum simulator to simulate the problem.
Cluster 2 goes up if we go from cluster one to cluster four. All spectrum for each of the clusters, as expected the IPR are all very similar. Fig. 2-B shows a representative the spectra of each of the clusters indeed reveals they the structures in the system. In fact a closer look at Fig. 2- A, it clearly shows we managed to find most of all, if they do not, how could we infer any structural chemical shifts and interactions between them. As a testament, the identified molecular sub-structures. Moreover, even clearly find out whether the molecule belongs to any of the KL-loss was 0.145.

Figure 1. Clustering In order to identify whether naturally occurring molecules have some atypical NMR spectrum, we perform a clustering analysis. In panel–A we show the distance between the various NMR spectra, where the Bhat-tacharyya coefficient is used to measure similarity. To obtain a meaningful comparison, spectra are shifted and scaled such that they are all centered around the same frequency and have the same bandwidth. To extract clusters we perform a t-SNE shown in panel–B with perplexity of 10; which is chosen because it has minimal Kullback-Leibler (KL) loss, i.e. the KL-loss was 0.145.

spectra in cluster 1 have the property of containing two large peaks and two small peaks, where the larger peak is about three times higher than the small peak. This is indicative of molecules with a methyl group (CH3) with its protons coupled with a methine proton (CH). One example of such structures can be seen in acetaldehyde oxime (BMRB ID 37; bmse000467) (as shown to the left in Fig. 2-B). The fact that the 3 protons are equivalent results in the 3:1 ratio of the peaks. Molecules from cluster 2 have a sub-structure similar to the 1,4-Benzquinone molecule shown Fig. 2-B. They are highly symmetric and have two pairs of two methine protons (CH) where the protons are on neighboring carbon atoms. The symmetry in the molecule makes the spectrum highly degenerate. In contrast, cluster 3 has molecules where there are two neighboring methylene groups (CH2). The interacting splitting causes a spectrum as shown in Fig. 2-B. Finally, cluster 4 has four inequivalent protons with different chemical shifts and interactions between them. As a results, there is a plethora of possible transitions and the spectrum has an erratic form such as shown in Fig. 2-B. In that sense, cluster 4 is most like a disordered quantum spin chain.

Given a new spectrum of an unknown molecule we can clearly find out whether the molecule belongs to any of the identified molecular sub-structures. Moreover, even if it is not in the dataset, the high degree of clustering will allow us to easily place the spectrum within one of the clusters. Since we know the spin matrix θi for each of the molecules in the dataset, we have a rough estimate of what the Hamiltonian parameters are and where the protons are located with respect to each other. However, there is still a lot of fine structure within clusters, in

a measure for the total number of transitions that contribute to the spectrum. At least 4 well defined clusters are identified. Using the clusters as indicated in Fig. 4-B, we can sort the molecules per cluster and have a look at the spectra. The sorted distance matrix is shown in Fig. 2-A, it clearly shows we managed to find most of the structures in the system. In fact a closer look at the spectra of each of the clusters indeed reveals they are all very similar. Fig. 2-B shows a representative spectrum for each of the clusters, as expected the IPR goes up if we go from cluster one to cluster four. All
QUANTUM MODEL SIMULATION

While our model is microscopically motivated, thereby capturing the spectra very well and allowing for a physical interpretation of the model parameters, it has the drawback that, unlike simple models such as Lorentzian mixture models [38, 39], there is no analytic form for the spectrum in terms of the model parameters. Moreover, even simulating the model becomes increasingly complex when the number of spins increase. Before we solve the inference problem, let us present an efficient algorithm to further improve the Hamiltonian parameter estimation.

Figure 2. NMR Spectra By clustering the molecules according to the Hellinger distance t-SNE clusters we can reorganize the distance matrix as shown in panel A. For each of the clusters, we look at the different spectra, which indeed show great similarity. A representative spectrum for each of the clusters is shown in panel B, where the spectra are labeled according to the t-SNE clusters shown in Fig. 1-B. In addition, we show an example small molecule out of this cluster next to the associated spectrum. The atoms and interactions responsible for the shown portions of the spectra are indicated in blue and red arrows respectively.

particular in clusters 3 and 4, as can be seen in Fig. 2 A. In what remains, we are concerned with finding an algorithm to further improve the Hamiltonian parameter estimation.

Figure 2. NMR Spectra By clustering the molecules according to the Hellinger distance t-SNE clusters we can reorganize the distance matrix as shown in panel A. For each of the clusters, we look at the different spectra, which indeed show great similarity. A representative spectrum for each of the clusters is shown in panel B, where the spectra are labeled according to the t-SNE clusters shown in Fig. 1-B. In addition, we show an example small molecule out of this cluster next to the associated spectrum. The atoms and interactions responsible for the shown portions of the spectra are indicated in blue and red arrows respectively.

particular in clusters 3 and 4, as can be seen in Fig. 2 A. In what remains, we are concerned with finding an algorithm to further improve the Hamiltonian parameter estimation.

QUANTUM MODEL SIMULATION

While our model is microscopically motivated, thereby capturing the spectra very well and allowing for a physical interpretation of the model parameters, it has the drawback that, unlike simple models such as Lorentzian mixture models [38, 39], there is no analytic form for the spectrum in terms of the model parameters. Moreover, even simulating the model becomes increasingly complex when the number of spins increase. Before we solve the inference problem, let us present an efficient method to extract the simulated NMR spectrum on a quantum simulator-computer. The basic task is to extract the spectrum [3] by measuring [2]. Recall that we work at infinite temperature, hence by inserting an eigenbasis of the total z-magnetization $S_z^{tot} = \sum_j m_j | z_j \rangle \langle z_j |$, we find

$$S(t|\theta) = \sum_{j=1}^{2N} m_j \frac{2N}{2N} | z_j \rangle e^{i H(\theta)t} S_z^{tot} e^{-i H(\theta)t} | z_j \rangle,$$

where $m_j$ is the total z-magnetization in the eigenstate $| z_j \rangle$. Consequently, we can extract the spectrum by initializing our system in a product state of z-polarized states after which we quench the system to evolve under the Hamiltonian $H(\theta)$, and then finally measure the expectation value of $S_z^{tot}$ at time $t$. By repeating the procedure for various initial eigenstates and weighting the results by the initial magnetization $m_j$, we obtain an estimate of $S(t|\theta)$. While intuitive and simple, this naive procedure has an exponential sampling cost and is therefore extremely inefficient. Fortunately, due to the massive degeneracy of the $S_z^{tot}$ operator and some remaining $Z_2$ symmetry of the Hamiltonian [1], we can reduce the sampling down to $O(N/2)$ samples rather than $O(2^N)$. The basic idea is to prepare a random state in the subspace of fixed z-magnetization such that the sampling over all the states at fixed magnetization can be replaced by averaging over realization of the random state. Such states can be efficiently prepared using Hamiltonians that scramble information quickly, moreover fluctuations from the mean are exponentially suppressed in $N$ such that it’s sufficient to average over $O(1)$ different Hamiltonians. A detailed analysis is given in the SI. The entire procedure is schematically depicted in Fig. 3. Clearly, it only requires $N$ qubits. Obtaining $S(t|\theta)$ at a fixed time $t$ will require sampling $O(N)$ random initial states with fixed total z-magnetization. These states can be prepared by randomizing initial product states with a fixed $S_z^{tot}$ using a fast scrambling unitary $U_{mix}$, as shown in Fig. 3. Next, we propagate each state with the physical Hamiltonian $H(\theta)$ for fixed time $t$. Since the mixing takes $O(\log^c(N))$ [40, 41], this will take $O(N \times (t + \log^c N))$ with $c$ a constant of $O(1)$. In order to construct the full spectrum, we will have to measure at various times $t$. Given the finite decoherence rate, samples only have to be collected up to a maximum time $O(1/\gamma)$, while the spectral norm of the system typically increases linearly in the number of spins. Taking samples at the Nyquist frequency, we will have to collect $O(N)$ time samples, leading to a final scaling of $O(N^2 \log^c N)$ to obtain a simulated NMR spectrum.
**Method overview** Take a product state $|z_i\rangle$ with a given total magnetization $m_i$, according to the relative fraction of Hilbert space occupied by $m_i$ states $P(z_i)$. This product state is scrambled with a unitary channel that conserves total $Z$ magnetization. After this initial preparation, we evolve the state under the Hamiltonian $H(\theta)$ and measure the $Z$ magnetization at time $t$. By applying a fast Fourier transform we obtain the spectrum which can be used to infer the parameters of the Hamiltonian.

**Variational Bayesian Inference**

Now that we have a procedure of efficiently obtaining spectra of hypothetical molecules, how do we solve the inference problem? The standard approach would be to do maximum likelihood estimation of the parameters given the experimental spectrum or minimize one of the aforementioned cost functions. This cannot be done analytically and the problem can clearly be highly non-convex. Using a quantum simulator, one only obtains a statistical estimate of the cost function and its gradient since we only perform a finite number of measurements. In order to move down the optimization landscape we thus need to resolve the signal from the noise, meaning gradients have to be sufficiently large to be resolved. However, we find extremely small gradients for this problem. Taking for example the Hellinger distance, $D_H$, used to construct Fig. (1), we find the gradient satisfies

$$\left|\partial_\theta D_H^2\right| = \left|\int \frac{d\omega}{2\pi} \sqrt{\frac{A(\omega)}{A(\omega)} \partial_\theta A(\omega)}\right| \leq \sqrt{I_{\theta\theta}},$$

where $I_{\theta\theta}$ is the diagonal component of the Fisher information. The bound simply follows from Cauchy-Schwarz inequality. As shown in the SI, the Fisher information, even for the optimal values, is very small; typically of the order $10^{-4}$ – $10^{-6}$ for our 4 spin molecules. We are thus in a situation of a very shallow rough optimization landscape. The problem is of similar origin as the vanishing gradient problem in quantum neural networks [42]. A gradient free method seems advisable but simple heuristic methods such as simplex or pattern search seem to fail. We therefore adopt a Bayesian approach to update our estimated parameters. Recall Bayes theorem, in the current notation, reads:

$$P(\theta|\omega) = \frac{A(\omega|\theta) P(\theta)}{A(\omega)},$$

where $P(\theta|\omega)$ is the conditional probability to have parameters $\theta$ given that we see spectral weight at frequency $\omega$, $A(\omega|\theta)$ is the NMR spectrum for fixed parameters $\theta$, $P(\theta)$ is the probability to have parameters $\theta$ and $A(\omega)$ is the marginal NMR spectrum averaged over all $\theta$. If we acquire some data, say a new spectrum $A(\omega)$ and we have some prior believe about the distribution $P(\theta)$, we can use it to update our believe about the distribution of the parameters, i.e.

$$P_{t+1}(\theta) = \int \frac{d\omega}{2\pi} A(\omega) \frac{A(\omega|\theta) P(\theta)}{A(\omega)} P_i(\theta),$$

with $A_i(\omega) = \int d\theta A(\omega|\theta)P_i(\theta)$. Note that the above rule indeed conserves positivity and normalization. Moreover, it simply reweights the prior distribution with some weight

$$w_i(\theta) = \int \frac{d\omega}{2\pi} A(\omega) \frac{A(\omega|\theta)}{A_i(\omega)},$$

that is directly related to the log-likelihood, since Jensen inequality gives:

$$\log(w_i(\theta)) \geq \int \frac{d\omega}{2\pi} A(\omega) \log \frac{A(\omega|\theta)}{A_i(\omega)} = \mathcal{L}(\theta) + c,$$

where $\mathcal{L}(\theta)$ is the log-likelihood and $c$ is a constant independent of $\theta$. Consequently, iterating expression $9$ is expected to converge to a distribution of parameters which is highly peaked around the maximum likelihood estimate. While it avoids the use of any gradients, it requires us to sample from the current parameter distribution $P_i(\theta)$. This by itself could become intractable and so we make an additional approximation. In order to be able to sample from the parameter distribution, we approximate it by a normal distribution at every step. That is, given that we have obtained some Monte Carlo samples out of $P_i(\theta)$, we can estimate all the weights $w_i(\theta)$ by simply simulating the model and obtaining $A(\omega|\theta)$ for all the samples. Next, we approximate $P_{t+1}(\theta)$ with a normal distribution that is as close as possible to it, i.e. has minimal KL-distance. The latter is simply the distribution with the same sample mean and covariance as $P_{t+1}(\theta)$. We use an atomic prior, $P_0(\theta) = \sum_i \frac{1}{N} \delta(\theta - \theta_i)$, consisting of all the samples that belong to the same cluster to which the spectrum is identified to belong. The result of this procedure for some randomly chosen test molecules is shown in Fig. (3). We observe steady, albeit noisy, convergence of the molecular spectra. Two sources of noise

![Figure 3](image-url)
by combining or extending the variational method with Hamiltonian Monte Carlo techniques.

It is interesting to extend our technique to other types of experiments. NMR is hardly the only problem where one performs inference on spectroscopic data. For example, one can imagine combining resonant inelastic X-ray scattering (RIXS) data from strongly correlated electron systems, with Fermi-Hubbard simulators based on ultracold atoms. Currently, RIXS data is analyzed by performing numerical studies of small clusters on classical computers (see Ref. for review). A DMFT-based hybrid algorithm was recently proposed in. With cold atoms in optical lattices one may be able to create larger systems and study their non-equilibrium dynamics corresponding to RIXS spectroscopy.

**Acknowledgements.**—DS acknowledges support from the FWO as post-doctoral fellow of the Research Foundation – Flanders. SM is supported by a research grant from the National Heart, Lung, and Blood Institute (K24 HL136852). OD and HD are supported by a research award from the National Heart, Lung, and Blood Institute, (5K01HL135342) and (T32 HL007575) respectively. ED acknowledges support from the Harvard-MIT CUDA, AFOSR-MURI: Photonic Quantum Matter (award FA95501610323), DARPA DRINQS program (award D18AC00014), AFOSR-MURI: Quantum Phases of Matter (grant FA9550-14-1-0035). The authors acknowledge useful discussion with P. Mehta, M. Lukin.

---

**SUMMARY AND OUTLOOK**

Here we have presented a method to improve model inference for NMR with relatively modest amount of quantum resources. Similar to generic generative models such as Boltzmann machines, for which a more efficient quantum version has been constructed, we have constructed an application specific model from which a quantum machine can sample more efficiently than a classical computer. Model parameters are determined through a variational Bayesian approach with an informative prior, constructed by applying t-SNE to a dataset of small molecules. As a consequence of the noisy nature of the generative model, as well as the absence of significant gradients, both the initial bias as well as the derivative free nature of Bayesian inference are crucial to tackling the problem. This situation, however, is generic to any hybrid quantum-classical setting that is sufficiently complicated. A similar approach might thus be useful to improve convergence of QA or VQE, e.g. heuristic optimization strategies for QA have been developed in Ref. Both the classical and quantum part of our approach can be extended further. On the quantum side, one can envision developing more efficient approaches for computing the spectra; trading computational time for extra quantum resources. On the classical side, improvements on the inference algorithm might be possible...
Supplemental Material for:

Quantum approximate Bayesian computation for NMR model inference

**SIMILARITY MEASURES**

To perform clustering or simply find the best fit to a certain spectrum one has to define a measure of distance or equivalently of similarity between different spectra. A priori, there is no unique optimal choice for this and certain measures might be much better suited for the current problem than others. Let’s therefore have a closer look at a few possible distance matrices:

- **Euclidean**:
  \[
  D_{\text{Eucl}}^2(i,j) = \int \frac{d\omega}{2\pi} (A_i - A_j)^2,
  \]

- **Hellinger**:
  \[
  D_{\text{Hell}}^2(i,j) = \frac{1}{2} \int \frac{d\omega}{2\pi} \left( \sqrt{A_i} - \sqrt{A_j} \right)^2,
  \]

- **Jensen-Shannon**:
  \[
  D_{\text{JS}}^2(i,j) = \frac{1}{2} \int \frac{d\omega}{2\pi} \left( A_i \log A_i + A_j \log A_j - (A_i + A_j) \log \frac{A_i + A_j}{2} \right),
  \]

where \(A_i\) is short hand notation denoting \(A_i = A(\omega|\theta_i)\). Note that the spectrum is positive and can be normalized since it satisfies the f-sum rule:

\[
S(0|\theta) = \int \frac{d\omega}{2\pi} A(\omega|\theta) = \frac{\text{Tr} \left[ (S^z_{\text{tot}})^2 \right]}{\text{Tr} [1]} = \frac{N}{4},
\]

hence it makes sense to think of \(A(\omega|\theta)\) (once normalized) as the conditional probability to generate an RF photon given the Hamiltonian \(H(\theta)\). In that respect one might suspect that statistical measures of distance might be better suited then a simple least square error. To check the performance of each of those measures we perform a t-SNE based on each of them and look at the t-SNE loss. Figure S1–A shows the distance matrix between all molecules in the dataset for the 3 different metrics under consideration. First of all, a lot of structure is observed in all three distance metrics. While the Hellinger distance and Jensen-Shannon distance are both qualitatively and quantitatively similar, the Euclidean distance only captures the large distance features well. By squaring the probability distribution, the Euclidean distance effectively only cares about the mode of the distribution, suppressing information about smaller peaks in the absorption spectrum. We observe better clustering for Hellinger and Jensen-Shannon distance, this is also quantified by the increased Kullback-Leibler loss of the Euclidean t-SNE. In fact, at the level of the t-SNE loss, the Hellinger distance performs the best.
Figure S1. **Clustering** In order to identify whether naturally occurring molecules have some atypical NMR spectrum we perform a clustering analysis based on 3 different measures of similarity. In panel–A we show the distance between the various NMR spectra for three different distance metrics. To extract clusters we perform a t-SNE shown in panel–B for each of the metrics respectively. The t-SNE is performed with the same initial seed and perplexity(10) for all plots. The KL-loss for the shown plots was \{0.145, 0.510, 0.299\} for the Hellinger, Euclidean and JS distance respectively.
ARE PHYSICAL MOLECULES TYPICAL?

In order to discuss whether actual spectra are atypical we need to define a notion of likelihood of a given spectra, i.e. we need a measure on the space of molecular parameters \( \theta \). The measure should be unbiased by any knowledge we believe to have about physical molecules. It should only satisfy some basic consistency conditions. One very simply and natural condition is that whatever measure we are sampling from, it ought not to depend on the way we parametrize our model. That is, if one makes a change of variables \( \theta' = f(\theta) \) it shouldn’t change the likelihood of a given molecule since it represents exactly the same data. This parametrization invariance was first argued by Jeffreys [52], and it was shown that the distribution should therefore be proportional to the square root of the determinant of the Fisher information metric (FIM):

\[
P(\theta) \propto \sqrt{\text{det } I_{ij}(\theta)},
\]

where \( I_{ij}(\theta) \) is the FIM:

\[
I_{ij}(\theta) = \int \frac{d\omega}{2\pi} A(\omega|\theta) \frac{\partial \log A(\omega|\theta)}{\partial \theta_i} \frac{\partial \log A(\omega|\theta)}{\partial \theta_j}.
\]

(S3)

In Bayesian inference, \( P(\theta) \) is known as Jeffrey’s prior and is an example of a so called uninformative prior. The question of whether molecular parameters are typical thus becomes a question about the structure of the eigenvalues of the Fisher information metric. Some representative Fisher matrices for physical molecules are shown in Fig. S2. Note that the FIM is generally small and appears to be structured. The structure should become apparent when we look at the eigenvalues of the FIM. These are depicted in Fig. S3. Most molecules indeed seem to have a some eigenvectors – combinations of parameters – that are much more important then others, i.e. have eigenvalues that are exponentially larger then others. Such characteristic has been termed “sloppiness” in the past and it has been shown to arrise naturally in multiparameter mathematical models that probe collective behavior [53, 54]; meaning they can not probe the individual parameters but only have access to some coarse grained observable. NMR spectroscopy can be argued to be in this regime as there is no easy way to directly extract the model parameters from the spectrum. The fact that there are irrelevant combinations of parameters immediatly implies the molecules are unlikely because the determinant of the FIM must be small. In other words, these sloppy parameters represent approximate or possibly even exact symmetries of the molecules. Random models possess no symmetries and so molecules are atypical. Finally note that even the large eigenvalues of the FIM are relatively small, sampling parameters \( \theta_i \) from a normal distribution with zero mean and unit variance results in significantly larger eigenvalues, see red dots in Fig. S3. In fact, the FIM eigenvalues are of \( O(1) \) rather then \( O(10^{-4}) \). Given that we will only have a finite amount of data available to

Figure S2. **FIM Features** Fisher information metric for a typical molecule out of each cluster.
Figure S3. **FIM eigenvalues** Blue dots show the eigenvalues of the FIM for all the molecules contained in the dataset. Red dots show eigenvalues for samples obtained by sampling each of the parameters $\theta_i$ from a normal distribution with unit variance and zero mean.

finally perform the inference, it would be extremely hard to converge to physical model parameters, starting from an uninformative prior. It’s useful to actually take the sloppiness of molecules into account and start from a biased prior that already takes into account the aforementioned clustering.
UNITARY MIXING

Our goal is to extract spectrum \( \{3\} \) by measuring \( \{2\} \) and applying classical Fourier transform. Recall that, at infinite temperature we find

\[
S(t|\theta) = \sum_{j=1}^{2^N} \frac{m_j}{2^N} \langle z_j | e^{iH(\theta)t} | S^z_{\text{tot}} e^{-iH(\theta)t} | z_j \rangle = \sum_{j=1}^{2^N} \frac{m_j}{2^N} \langle z_j(t) | S^z_{\text{tot}} | z_j(t) \rangle , \tag{S4}
\]

where \( m_j \) is the total \( z \)-magnetization in the eigenstate \( |z_j\rangle \). Consequently, we could extract the spectrum by initializing our system in a \( z \)-polarized product state, evolve under the Hamiltonian \( H(\theta) \) for time \( t \) and measure the expectation value of \( S^z_{\text{tot}} \).

Note however that \( S^z_{\text{tot}} \) is hugely degenerate. There are \( 2^N \) states but only \( N + 1 \) different magnetization sectors. We are thus sampling the same magnetization \( m_j \) many times. In this case, a much smarter way to take sample averages is to create random superpositions of states within each fixed magnetization sector. Instead of starting with a product state \( |z\rangle \), let us make a superposition:

\[
|r_M\rangle = \sum_{j: m_j = M} c_j |z_j\rangle , \tag{S5}
\]

where \( c_j \) are complex uniform random number of the sphere defined by \( \sum_j |c_j|^2 = 1 \). Then, if we perform our quantum quench and measure the \( z \)-magnetization we find

\[
\langle r_M | S^z_{\text{tot}}(t) | r_M \rangle = \sum_{i: m_i = M} \sum_{j: m_j = M} c_i^* c_j \langle z_i | S^z_{\text{tot}}(t) | z_j \rangle , \tag{S6}
\]

taking expectation values of the random number \( c_j \) we find

\[
\mathbb{E}_c [\langle r_M | S^z_{\text{tot}}(t) | r_M \rangle] = \frac{1}{\dim H_M} \sum_{i: m_i = M} \langle z_i | S^z_{\text{tot}}(t) | z_i \rangle , \tag{S7}
\]

where \( H_M \) denotes the Hilbert space of fixed \( z \)-magnetization states. Consequently we can rewrite our response function as:

\[
S(t|\theta) = \sum_{j=1}^{N+1} p_j m_j \mathbb{E}_c [\langle r_j(t) | S^z_{\text{tot}} | r_j(t) \rangle] , \tag{S8}
\]

where \( p_j = \dim H_j / 2^N \) is the fraction of computational basis states occupied by magnetization \( m_j \) states. Whether or not we have made any gain depends on how efficient we can get a sample average to approximate the true expectation \( \mathbb{E}_c [\cdot] \). There are two parts to this, one particular— that is how long does it take to prepare such a state— and the other statistical. Let’s start with the latter, in order to see how fast our sample average converges let’s look at the variance of the magnetization

\[
\text{Var} [\langle r_j(t) | S^z_{\text{tot}} | r_j(t) \rangle] = \frac{\text{Tr}_{H_j} [S^z_{\text{tot}}(t)^2]}{(\dim H_j)^2} \leq \frac{N^2}{4 \dim H_j} . \tag{S9}
\]

The inequality simply follows from the fact that the magnetization can never be larger than \( N/2 \). Fluctuations around the mean are thus suppressed by a factor \( 1/\dim H_j \). This clearly solves the sampling problem in our naive expression \( \{S4\} \). The exponential sum is due to magnetization sectors which are exponentially large but all of those will have exponentially small fluctuations around the mean if we simply sample a single random vector in that subspace. Finally note that another factor of 2 can be gained because the entire response is invariant under taking \( S^z_{\text{tot}} \rightarrow -S^z_{\text{tot}} \), consequently sampling either the positive or the negative magnetization subsector is sufficient.

This brings us to the second question, how do we generate those states? First of all note that we do not really need to generate Haar random states, we only need pseudorandom states that appear sufficiently random to guarantee that the same average value of the magnetization and similar variance such that convergence is fast. We expect this to be the case if we simply initialize the system in a product state with fixed \( m_j \) and evolve the system under some unitary
which conserves $S_{i}^{z}$ but other than that has no conserved quantities. A simple Hamiltonian that would serve the purpose is anything of the form

$$H_{\text{mix}} = \sum_{ij} (A_{ij} S_{i}^{z} S_{j}^{z} + B_{ij} (S_{i}^{+} S_{j}^{-} + S_{i}^{-} S_{j}^{+})) + \sum_{i} C_{i} S_{i}^{z}. \quad (S10)$$

To avoid the energy to be conserved one has to change $H_{\text{mix}}$ in time between at least two non-commuting versions. In other words, we can simply make a random circuit out of Ising XY, ZZ-gates and phase shift gates. A good pseudorandom state is obtained once we manage to entangle all the qubits; since we only wish to measure the magnetization which is completely local. Even if we were to do it with a local Hamiltonian, constraint by Lieb-Robinson bounds, we expect maximal entanglement to be generated at a time $t \sim N^{1/d}$ in $d$-dimensions. If we consider coupling between any pair of qubits this is even reduced to $t \sim \log N$. For detailed mathematical analysis regarding the properties and generation of pseudorandom quantum states we refer to [40, 41, 55–58] and references therein.

This is illustrated in Fig. S4. Panel A shows the distance between true uniform sampling and the ensemble averaged state obtained by unitary scrambling of the states in a random $z$-conserving circuit. Clearly, there is very little dependence of the mixing time on the system size. In fact for the small systems under consideration we see almost perfect mixing for circuits of depth 6. In panel B, the variance of the $z$ distribution is shown for different circuits and different system sizes. We clearly see exponential decay of the variance in each sample with system size, moreover the circuit-to-circuit fluctuations also decrease. For sufficiently wide circuits it thus suffices to take a single sample out of a single circuit to estimate the ensemble average.

Figure S4. **Unitary scrambling** For systems of size $N = \{5, 7, 9, 11\}$ we investigate scrambling in the $M_{z} = 1/2$ subsector under random circuits build out of Ising XY and ZZ gates and Phase shift gates. We start from a product state and at each step we make random pairs to apply the Ising gates in parallel after which we do a random local phase-shift gate, i.e. apply $S_{i}^{z}$ over a random angle. Panel A shows the trace distance of the ensemble averaged state (estimated by taking 10 000 random circuit samples) and the infinite temperature uniform distribution. The gray dashed line indicates the noise limit at which we cannot accurately compute the distance because we only have a limited amount of samples. Panel B shows the variance of the $n(z)$ density at a circuit depth of 10 for various system sizes and circuit realizations.