

ARTICLE OPEN



Experimental quantum kernel trick with nuclear spins in a solid

Takeru Kusumoto^{1,7}, Kosuke Mitarai^{1,2,3,4,7} , Keisuke Fujii^{1,2,4,5}, Masahiro Kitagawa^{1,2} and Makoto Negoro^{2,4,6}

The kernel trick allows us to employ high-dimensional feature space for a machine learning task without explicitly storing features. Recently, the idea of utilizing quantum systems for computing kernel functions using interference has been demonstrated experimentally. However, the dimension of feature spaces in those experiments have been smaller than the number of data, which makes them lose their computational advantage over explicit method. Here we show the first experimental demonstration of a quantum kernel machine that achieves a scheme where the dimension of feature space greatly exceeds the number of data using ¹H nuclear spins in solid. The use of NMR allows us to obtain the kernel values with single-shot experiment. We employ engineered dynamics correlating 25 spins which is equivalent to using a feature space with a dimension over 10¹⁵. This work presents a quantum machine learning using one of the largest quantum systems to date.

npj Quantum Information (2021)7:94; <https://doi.org/10.1038/s41534-021-00423-0>

INTRODUCTION

Quantum machine learning is an emerging field that has attracted much attention recently. The major algorithmic breakthrough was an algorithm invented by Harrow–Hassidim–Lloyd¹. This algorithm has been further developed to more sophisticated machine learning algorithms^{2,3}. However, a quantum computer that is capable of executing those algorithms is yet to be realized. At present, noisy intermediate-scale quantum (NISQ) devices⁴, which consist of several tens or hundreds of noisy qubits, are the most advanced technology. Although their performance is limited compared to the fault-tolerant quantum computer, simulation of the NISQ devices with 100 qubits and sufficiently high gate fidelity are beyond the reach for the existing supercomputer and classical simulation algorithms^{5–7}. This fact motivates us to explore its power for solving practical problems.

Many NISQ algorithms for machine learning have been proposed in recent works^{8–17}. Almost all of the algorithms require us to evaluate an expectation value of an observable, which is sometimes troublesome to measure by sampling, for example with superconducting or trapped-ion qubits. On the other hand, NMR can evaluate the expectation value with a one-shot experiment owing to its use of a vast number of duplicate quantum systems. It is, therefore, a great testbed for those algorithms. A major weakness of NMR is that its initialization fidelity is quite low; at the thermal equilibrium of room temperature, the proton spins can effectively be described with a density matrix $\rho_{\text{eq}} = \frac{1}{2}(I + \epsilon I_z)$ with $\epsilon \approx 10^{-5}$. Nevertheless, ensemble spin systems can exhibit complex quantum dynamics that are classically intractable. For example, the dynamical phase transition between localization and delocalization has been observed in polycrystalline adamantane along with tens of correlated proton spins¹⁸. Discrete time-crystalline order has been observed in disordered¹⁹ and ordered^{20,21} spin systems. Also, it has been shown that general dynamics of NMR are classically

intractable under certain complexity conjecture^{22,23}. These facts strengthen our motivation to use NMR for NISQ algorithms.

In this work, we employ NMR for machine learning. Specifically, we implement the kernel-based algorithm which utilizes the quantum state as a feature vector and is a variant of theoretical proposals^{8,9,24}. Ref. ²⁵ has recently proved that this approach has a rigorous advantage over classical computation in a specific task. The experimental verification has been provided in refs. ^{15,26} using either superconducting qubits or the photonic system. Our strategy to use the NMR is advantageous in that we can estimate the value of the kernel, which is the inner product of two quantum states, by single-shot experiments. Moreover, the dimension of the Hilbert space employed in this work greatly exceeds the number of training data. We perform simple regression and binary classification tasks using the dynamics of nuclear spins in a polycrystalline adamantane sample, and observe that the performance of the trained model becomes better as more spins are involved in the dynamics. Also, to carry out the performance analysis of our approach without the inevitable effect of noise in experiments, we present numerical simulations of 20 spin dynamics, which well agree with the experimental results. We employ one of the largest quantum systems to date for a quantum machine learning experiment in this work with the single-shot setting that enables the use of a large quantum system.

RESULTS

Kernel methods in machine learning

In machine learning, one is asked to extract some patterns, or features, in a given dataset^{3,27}. It is sometimes useful to pre-process them beforehand to achieve the objective. For example, a speech recognition task might become easier when we work in the frequency domain; in this case, the useful pre-processing would be the Fourier transform. The space in which such pre-processed data live is called feature space. For a given set of data

¹Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka, Japan. ²Center for Quantum Information and Quantum Biology, Osaka University, Toyonaka, Osaka, Japan. ³QunaSys Inc., Bunkyo, Tokyo, Japan. ⁴JST, PRESTO, Kawaguchi, Saitama, Japan. ⁵RIKEN Center for Quantum Computing (RQC), Wako, Saitama, Japan. ⁶Institute for Quantum Life Science, National Institutes for Quantum and Radiological Science and Technology, Inage-Ku, Chiba, Japan. ⁷These authors contributed equally: Takeru Kusumoto, Kosuke Mitarai. ✉email: mitarai@qc.ee.es.osaka-u.ac.jp

$\{\mathbf{x}_i\}_{i=1}^{N_d} \subset \mathbb{R}^D$, a feature space mapping $\phi(\mathbf{x})$ constructs the data in the feature space $\{\phi(\mathbf{x}_i)\}_{i=1}^{N_d} \subset \mathbb{R}^{D_f}$. The feature map has to be carefully taken to maximize the performance in e.g. classification tasks.

Kernel methods are a powerful tool in machine learning. It uses a distance measure of two inputs defined as a kernel, $k: \mathbb{R}^{D_f} \times \mathbb{R}^{D_f} \rightarrow \mathbb{R}$. For example, a kernel can be defined as an inner product of two feature vectors:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j). \quad (1)$$

Many machine learning models, such as support vector machine or linear regression, can be constructed using the kernel only, that is, we do not have to explicitly hold $\phi(\mathbf{x})$. This dramatically reduces the computational cost when the feature $\phi(\mathbf{x})$ lives in a high-dimensional space. The direct approach that computes and stores $\phi(\mathbf{x})$ for all training data \mathbf{x} would demand the computational resource that is at least proportional to N_d and $D_f \gg N_d$ just for storing them, whereas the kernel method only requires us to store a $N_d \times N_d$ kernel matrix with its (i, j) elements being $k(\mathbf{x}_i, \mathbf{x}_j)$.

Implementing kernel by NMR

In NMR, we can prepare a data-dependent operator $A(\mathbf{x}_i)$ by applying a data-dependent unitary transformation $U(\mathbf{x}_i)$ on the initial z-magnetization $I_z = \sum_{\mu=1}^n I_{z,\mu}$, that is, $A(\mathbf{x}_i) = U(\mathbf{x}_i) I_z U^\dagger(\mathbf{x}_i)$. Here, $I_{\alpha,\mu}$ ($\alpha = x, y, z$) is the α -component of the spin operator of the μ -th spin and n represents the number of spins. $A(\mathbf{x}_i)$ with a sufficiently large n is generally intractable by classical computers^{22,23}. We employ this operator $A(\mathbf{x}_i)$ as a feature map $\phi_{\text{NMR}}(\mathbf{x}_i)$. $A(\mathbf{x}_i)$ can be regarded as a vector, for example, by expanding $A(\mathbf{x}_i)$ as a sum of Pauli operators. For an n -spin-1/2 system, $A(\mathbf{x}_i)$ is a vector in \mathbb{R}^{4^n} . The dynamics of NMR can involve tens of spins maintaining its coherence^{18,28–30}, which means we can employ an approximately $4^{O(10)}$ dimensional feature vector for machine learning. Although the high-dimensional feature space does not always mean superiority in machine learning tasks, the fact that we can work with the feature space which has been intractable with a classical computer motivates us to explore its power.

The kernel method opens up a way to exploit $A(\mathbf{x}_i)$ directly for machine learning purposes. While we cannot evaluate each element of $A(\mathbf{x}_i)$ because it takes an exponential amount of time, we can evaluate the inner product of two feature vector $A(\mathbf{x}_i)$ and $A(\mathbf{x}_j)$ efficiently. The Frobenius inner product between two operators $A(\mathbf{x}_i)$ and $A(\mathbf{x}_j)$ is $\text{Tr}(A(\mathbf{x}_i)A(\mathbf{x}_j))$, which we employ as our NMR kernel function $k_{\text{NMR}}(\mathbf{x}_i, \mathbf{x}_j)$ in this work. Noting that,

$$\text{Tr}(A(\mathbf{x}_i)A(\mathbf{x}_j)) = \text{Tr}(U^\dagger(\mathbf{x}_j)U(\mathbf{x}_i)I_zU^\dagger(\mathbf{x}_i)U(\mathbf{x}_j)I_z), \quad (2)$$

and also at the thermal equilibrium, the density matrix of spin systems is $\rho_{\text{eq}} \approx \frac{1}{2^n}(I + \epsilon I_z)$ assuming $\epsilon \ll 1$, we obtain,

$$\text{Tr}(A(\mathbf{x}_i)A(\mathbf{x}_j)) \propto \text{Tr}(U^\dagger(\mathbf{x}_j)U(\mathbf{x}_i)\rho_{\text{eq}}U^\dagger(\mathbf{x}_i)U(\mathbf{x}_j)I_z). \quad (3)$$

The right hand side can be measured experimentally by first evolving the system at thermal equilibrium with $U(\mathbf{x}_i)$ and then with $U^\dagger(\mathbf{x}_j)$, and finally measuring I_z . A similar protocol is also used for measuring out-of-time-ordered correlator (OTOC)^{31,32}, which is considered as a certain complexity measure of quantum many-body systems.

Experimental

We propose to use $U(\mathbf{x})$ for an input $\mathbf{x} = \{x_j\}_{j=1}^D$ that takes the form of,

$$U(\mathbf{x}) = e^{-iH(x_D)\tau} \dots e^{-iH(x_2)\tau} e^{-iH(x_1)\tau}, \quad (4)$$

where τ is a hyperparameter that determines the performance of this feature map, and $H(x_j)$ is an input-dependent Hamiltonian

(Fig. 1b). In this work, we choose $H(x_j)$ to be

$$H(x_j) = e^{-ix_j I_z} \sum_{\mu < \nu} d_{\mu\nu} (I_{y,\mu} I_{y,\nu} - I_{x,\mu} I_{x,\nu}) e^{ix_j I_z}, \quad (5)$$

where $I_\alpha = \sum_{\mu} I_{\alpha,\mu}$. The Hamiltonian $H(0)$ can approximately be constructed from the dipolar interaction among nuclear spins in solids with a certain pulse sequence^{18,33,34} shown in Fig. 1b. See “Methods” for details. Shifting the phase of the pulse by x provides us $H(x)$ for general x . This Hamiltonian with $x_j=0$ created in adamantane has been shown to have a delocalizing feature in refs. ^{18,28–30}, which makes it appealing as we wish to involve as many spins as possible in the dynamics.

To illustrate character of the kernel function, we show the shape of the kernel for one-dimensional input x obtained with this sequence setting $\tau = N\tau_1$ where $\tau_1 = 60 \mu\text{s}$ and $N = 1, 2, \dots, 6$ as Fig. 1c. In this experiment, the overall unitary dynamics applied to the system is $U^\dagger(x_j)U(x_i) = e^{iH(x_j)\tau} e^{-iH(x_i)\tau} = e^{-ix_j I_z} e^{iH(0)\tau} e^{-i(x_i-x_j)I_z} e^{-iH(0)\tau} e^{ix_i I_z}$. Since the initial state has only I_z element and the final observable is also I_z , we can omit $e^{ix_i I_z}$ and $e^{-ix_j I_z}$ operations in the experiment. We, therefore, show the value of the kernel as a function of $x_i - x_j$ in Fig. 1c. The decay of the intensity of the signal with increasing N is due to decoherence. The number of spins involved in this dynamics can be inferred by the Fourier transform of the signal in Fig. 1c with respect to $x_i - x_j$. This is because the z-rotation $e^{-i(x_i-x_j)I_z}$ after the forward dynamics $e^{-iH(0)\tau}$ induces a phase $e^{-im'(x_i-x_j)}$ to terms in the form of I_{\pm}^m which appears in the density matrix of the system, where m and m' are integers, $|m'| \leq m$, and $I_{\pm} = I_x \pm iI_y$. This Fourier spectra is called multiple-quantum spectra^{18,30,31}, from which we can extract a cluster size K , which can be considered as the effective number of spins involved in the dynamics, by fitting it with Gaussian distribution $\exp(-m^2/K)$ ^{18,30,33}. We show the Fourier transform of the measured NMR kernel (Fig. 1c) along with corresponding K in Fig. 1d. We can conclude that, for $N > 5, 6$, the dynamics involve 25 spins, which corresponds to a feature space dimension of $4^{25} \approx 10^{15}$.

One-dimensional regression task

As the first demonstration, we perform the one-dimensional kernel regression task using the kernel shown in Fig. 1c. To evaluate the nonlinear regression ability of the kernel, we use $y = \sin(2\pi x/50)$ and $y = \frac{\sin(2\pi x/50)}{2\pi x/50}$, which will be referred to as sine and sinc function, respectively. We randomly drew 40 samples of x from $[-45, 45]$ (in degrees) to construct the training data set which consists of the input $\{x_j\}_{j=1}^{40}$ and the teacher $\{y_j\}_{j=1}^{40}$ calculated at each x_j . The NMR kernel $k_{\text{NMR}}(x_i, x_j)$ is measured for each pair of data to construct the model by kernel ridge regression²⁷. We let the model predict y for 64 x 's including the training data. The regularization strength was chosen to minimize the mean squared error of the result at the 64 evaluation data.

The result for the sine function is shown in Figure 2a, b. That for the sinc function is shown in “Methods”. Figure 3a shows the accuracy of learning evaluated by the mean squared error between the output from the trained model and true function. We see that the regression accuracy tends to increase with a larger N . For all N , the dimension of the feature space exceeds the number of data as the dynamics involve over 5 spins which corresponds to the Hilbert space dimension of 4^5 (see Fig. 1d). However, because of the deteriorating signal-to-noise ratio, the result also gets noisy with increasing N .

To certify the trend without the effect of noise, we conducted numerical simulations of 20-qubit dynamics with all-to-all random interaction in the form of Eq. (5), whose coupling strength, $d_{j\ell}$, are drawn uniformly from $[-1, 1]$. We set the evolution time $\tau = 0.01, 0.02, \dots, 0.06$. Details of the numerical simulations are described in Methods. The result for the sine function is shown as Fig. 2g–i. For

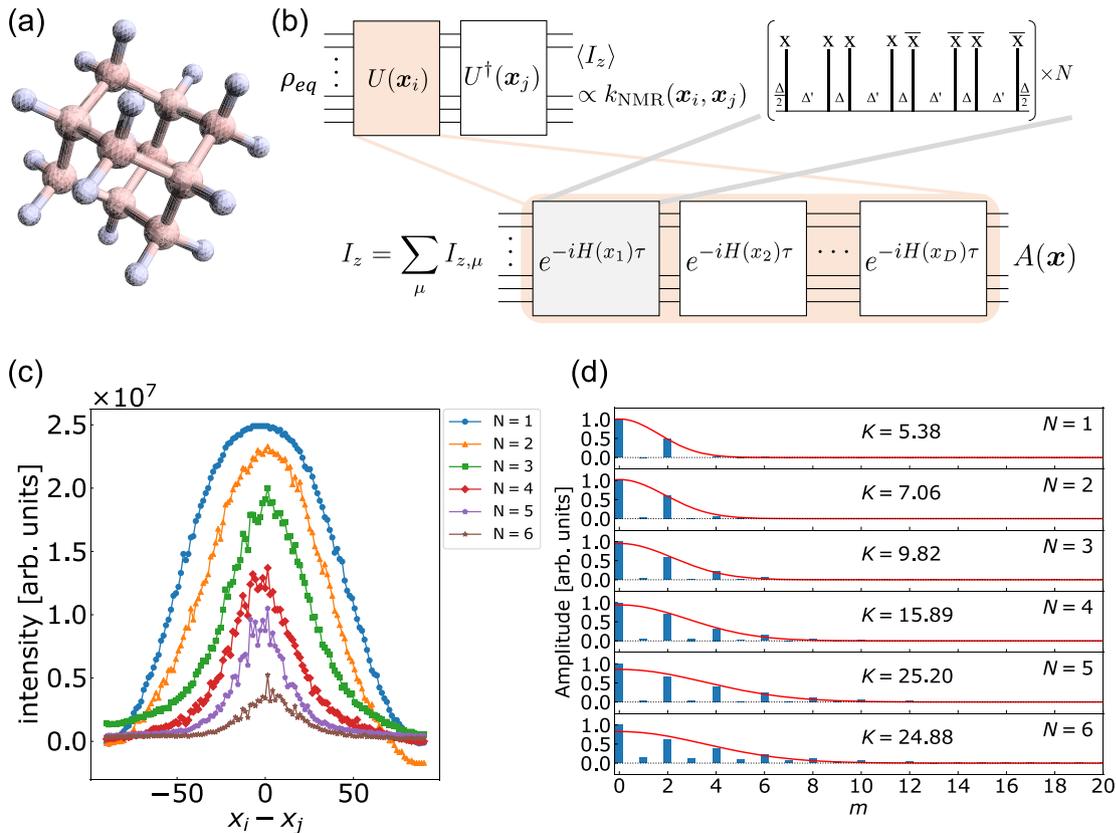


Fig. 1 Experimental setups. **a** Adamantane molecule. The white and pink balls represent hydrogen and carbon atoms, respectively. **b** Quantum circuit and pulse sequence employed in this work to realize evaluation of $k_{\text{NMR}}(\mathbf{x}_i, \mathbf{x}_j)$. The top right shows a pulse sequence for $x_1 = 0$, where X and \bar{X} represent $+\pi/2$ and $-\pi/2$ pulses along x -axis, respectively. In the experiment, Δ and Δ' are set $3.5 \mu\text{s}$ and $8.5 \mu\text{s}$, respectively. **c** NMR kernel employed in this work. **d** Fourier transform of (c) which corresponds to the obtained ^1H multiple-quantum spectra for $N = 1$ to $N = 6$. Red line is Gaussian $a \exp(-m^2/K)$ fitted to the spectra using a and K as fitting parameters. The Gaussian is fitted only to even m data points. K is the cluster size of the system^{18,30,33}.

the sinc function, we place the result in “Methods”. The mean squared error of the prediction evaluated in the same manner is shown in Fig. 3b. We can see the performance gets better with increasing τ , which corresponds to increasing N in the experiment. This certifies the trend observed in the NMR experiment.

Two-dimensional classification task

As the second demonstration, we implement two-dimensional classification tasks. We employ the hard-margin kernel support vector machine²⁷ and its implementation in scikit-learn³⁵ for this task. We use two data sets which we call “circle” and “moon” data set generated by functions available in scikit-learn package³⁵. They are depicted as dots in Fig. 4. A data set consists of pairs of input data point $(x_1, x_2) \in [-45, 45]^2$ and its label $y \in \{-1, +1\}$. For experimental simplicity, we have modified the generated (x_1, x_2) to be on a lattice. The NMR kernels with $N = 1, 2, 3$ are utilized in this task. We again conducted numerical simulations with the same setting as the previous section along with the experiment with $\tau = 0.03, 0.06, 0.09$. The values of this numerical kernel are shown in the “Methods”.

The results are shown in Fig. 4. Also for this classification task, the dimension of feature space exceeds the number of data already at $N = 1$. Note that the evolution time, in this case, is doubled compared to the previous demonstration where the input is one-dimensional (see Eq. (4)), that is, the evolution time of $N = 1$ for two-dimensional input is equivalent to that of $N = 2$ for one-dimensional input. We also note that, for the moon dataset with $N = 1$ experimental NMR kernel, the kernel matrix was

singular, and we did not obtain a reliable result. We reason this to the broadness of the kernel at $N = 1$.

For the circle data set with the experimental kernel, the best performance is achieved when $N = 1$. We believe this is because the task of classifying the circle dataset was too easy and did not need large feature space. This dataset can be completely classified, for example, by merely mapping two-dimensional data (x_1, x_2) to three-dimensional feature space $(x_1, x_2, x_1^2 + x_2^2)$. As for the moon dataset, the performance increased with increasing N . This can be attributed to the increasing dimension of the feature space.

DISCUSSION

In the one-dimensional regression task, we observed the trend of better performance with longer evolution time. This can be explained by the shape of the kernel generated by the NMR dynamics, which is shown in Fig. 1c. As mentioned earlier, this experiment is essentially the Loschmidt echo, and the shape of the signal sharpens as the evolution time increases. The sharpness of the kernel can directly be translated to the representability of the model as it can be in the popular Gaussian kernel because this property allows the machine to distinguish different data more clearly. However, it also causes overfitting problems if the data points are sparse. The most extreme case is when we use a delta function as a kernel, where every training point is learned with the perfect accuracy while the trained model fails to predict for unknown inputs. In our experimental case, we did not observe any overfitting problem, which means that our training samples were

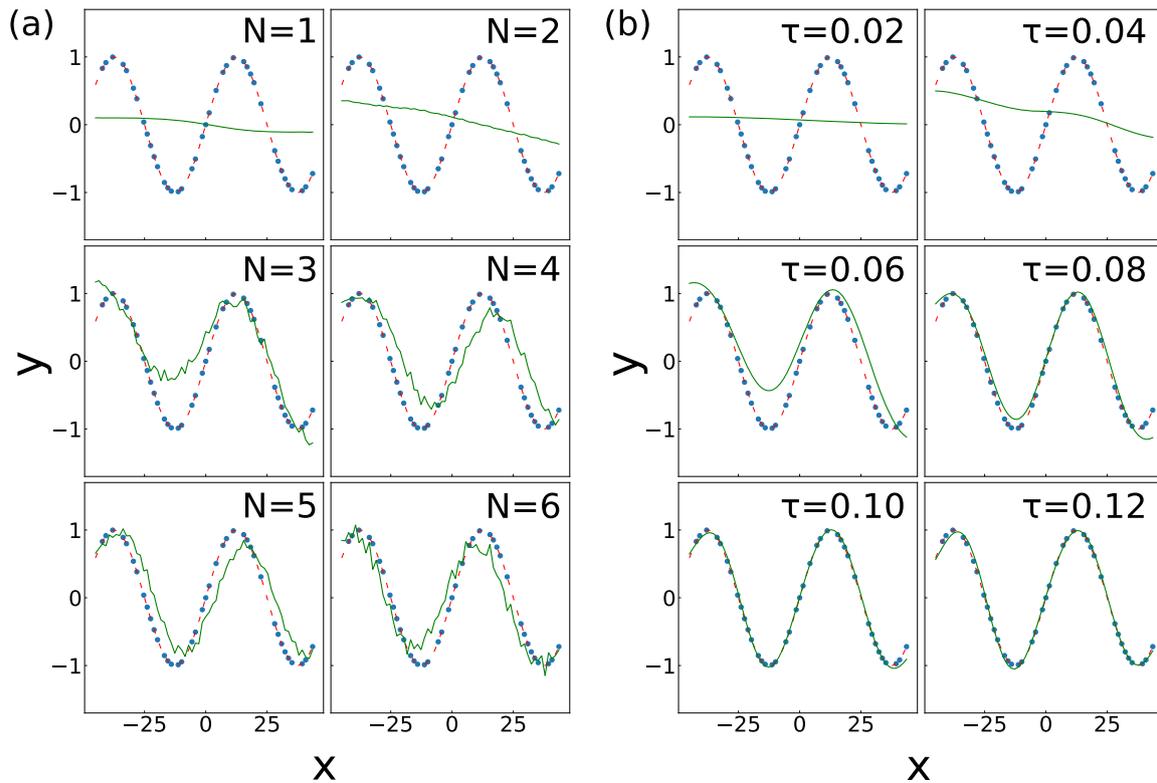


Fig. 2 Demonstration of one-dimensional regression task of $y = \sin(2\pi x/50)$. **a**, **b** shows results for NMR kernel and numerically simulated kernel, respectively. The blue dots, green lines, red dashed lines represent the training data, the prediction of the trained model, and the function from which the function the training data are sampled.

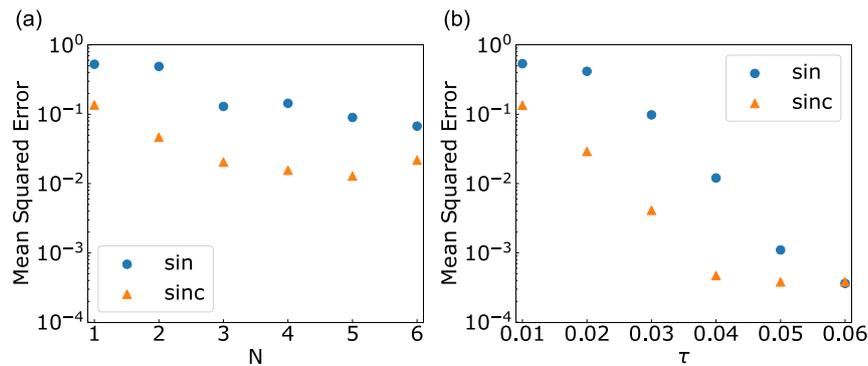


Fig. 3 Mean squared error of the trained models for the regression task of sine and sinc functions. **a**, **b** shows results for experimental NMR kernel and numerically simulated kernel, respectively.

dense enough for the sharpness of the kernel utilized in the model, and thus we observed an increasing performance from the improved representability of the kernel with longer evolution time. For the classification task, we observed the dataset-dependent trend with respect to the evolution time in the performance. We suspect that there is an optimal evolution time for this kind of task, which should be explored in future works. We believe the performance can also be improved by employing ensemble learning approach³⁶.

We note that the shape of the kernel resembles the Gaussian kernel which is widely employed in many machine learning tasks. In fact, we can obtain better results using the Gaussian kernel as shown in “Methods”. More concretely, it can achieve mean squared error less than 10^{-5} for the regression tasks and hinge loss of order 10^{-5} for the classification tasks without fine tuning of a hyperparameter. While this fact casts a shadow on the

usefulness of the NMR kernel used in this work, it is also true that the NMR dynamics evolved by general Hamiltonian cannot be simulated classically under certain complexity conjecture^{22,23}. This leaves a possibility that the NMR kernel performs better than classical kernel in some specific cases. We also note that there is a quantum kernel that is rigorously advantageous over classical ones in a certain machine learning task²⁵. More experiments using different Hamiltonians are required to test whether the “quantum” kernel has any advantage in machine learning tasks over widely used conventional kernels.

The previous research^{18,30,33} has shown that the cluster size can be brought up to over 1000 with more precise experiments. This indicates that the solid-state NMR is a platform well-suited to experimentally analyze the performance of kernel-based quantum machine learning. To extend the experiment to larger cluster sizes, we have to resolve the low initialization fidelity of nuclear spin

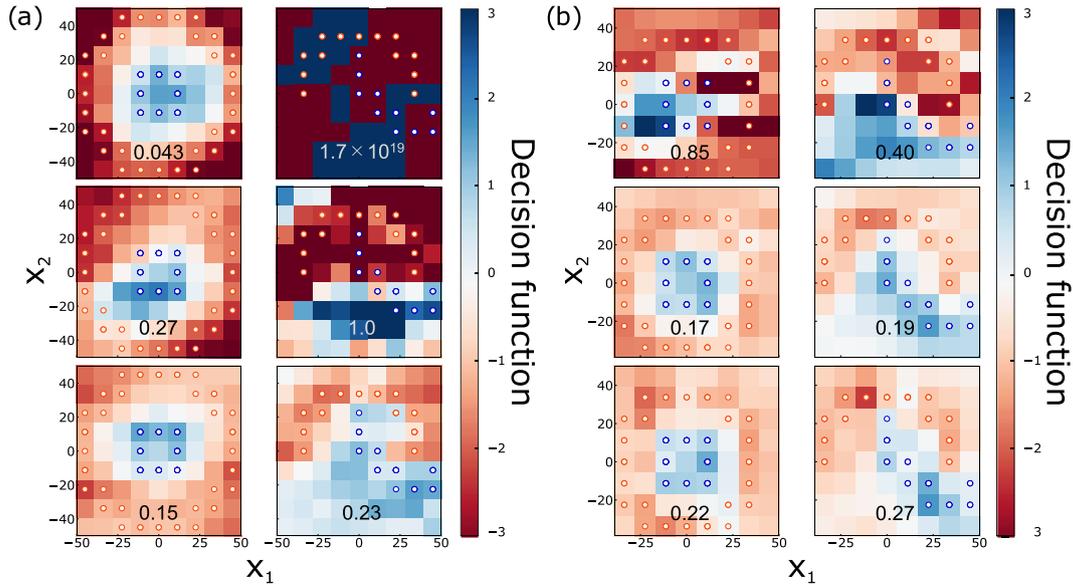


Fig. 4 Demonstration of simple classification tasks. **a** Classification with the NMR kernel. Left and right panels respectively show the results for “circle” and “moon” data set. The red and blue dots represent inputs (x_1, x_2) with label $y = +1$ and $y = -1$, respectively. The background color indicates the decision function of the trained model. The number in each figure is the hinge loss after training defined as $\frac{1}{N_s} \sum_{i=1}^{N_s} \max\{1 - \lambda_i y_i, 0\}$ where N_s is the number of training data, and λ_i and y_i are the output from the trained model and the teacher data corresponding to a training input \mathbf{x}_i . Top, middle, and bottom panels are the results with $N = 1, 2, 3$ NMR kernels, respectively. **b** Results from the numerically simulated quantum kernel. Top, middle, and bottom panels are the results with simulated kernel with $\tau = 0.03, 0.06, 0.09$. All the other notations follow that of (a).

systems, which is a major challenge faced by the NMR quantum information processing in general. We believe this can be overcome by the use of e.g., a technique called dynamic nuclear polarization^{37,38} where we initialize the nuclear spin qubits by transferring the polarization from initialized electrons.

To conclude, we proposed and experimentally tested a quantum kernel constructed by data-dependent complex dynamics of a quantum system. Experimentally, we used complex dynamics of 25 ^1H spins in adamantane to compute the kernel. This allowed us to achieve a scheme where the dimension of feature space greatly exceeds the number of data. We also stress that the spin dynamics of NMR is generally intractable by a classical computer^{22,23}. Machine learning models for one-dimensional regression tasks and two-dimensional classification tasks were constructed with the proposed kernel. The experimental and numerical results showed similar results. Experiments along with numerical simulation also showed that the performance of the model tended to increase with longer evolution time, or equivalently, with a larger number of spins involved in the dynamics for certain tasks. It would be interesting to export this method to more quantum-oriented machine learning tasks. For example, one may be able to distinguish two dynamical phases of spin systems, such as localized and delocalized phases demonstrated in ref. 18, with the kernel support vector machine employed in this work. More experiments are needed to verify the power of this “quantum kernel” approach, but our results can be thought of as one of the baselines of this emerging field.

Note added—After the initial submission of this work, we became aware of related works on quantum kernel methods^{39,40}.

METHODS

Experimental details

A pulse sequence to realize $H(0)$ is given in Supplementary Fig. 1. In the experiment, we set the length of $\pi/2$ pulse, τ_p , to 1.5 μs . For the waiting period, we used $\Delta' = 2\Delta + \tau_p$ with $\Delta = 3.5 \mu\text{s}$, which makes the evolution time for a cycle, τ_1 , 60 μs . By repeating the sequence for N times, we can

effectively evolve the spins with $e^{-iH(x_0)\tau}$ for $\tau = N\tau_1$. NMR spectroscopy with polycrystalline adamantane sample was performed at room temperature with OPENCORE NMR⁴¹, operating at a resonant frequency of 400.281 MHz for ^1H nucleus observation.

Detail of numerical simulation

We drew the interaction strength, $d_{\mu\nu}$, from uniform distribution on $[-1, 1]$ for all μ, ν . The evolution according to the Hamiltonian $H(x)$ is approximated by the first-order Trotter formula, that is,

$$e^{-iH(x)\tau} \approx e^{-i\Delta z} \left[\prod_{\mu < \nu} e^{-i\tau d_{\mu\nu} (I_{y,\mu} I_{y,\nu} - I_{x,\mu} I_{x,\nu}) / M} \right]^M e^{i\Delta z}. \quad (6)$$

We set $\tau = 0.01, 0.02, \dots, 0.06$ and $\tau/M = 0.001$ in the simulation. In order to reduce the computational cost, we set $A(\mathbf{x}) = U(\mathbf{x}) \prod_{\mu} (\frac{1}{2} + I_{z,\mu}) U^\dagger(\mathbf{x})$, which allows us to evaluate $\text{Tr}(A(\mathbf{x}_i)A(\mathbf{x}_j))$ by computing $|\langle 0 | U^\dagger(\mathbf{x}_j) U(\mathbf{x}_i) | 0 \rangle|^2$, where $|0\rangle$ is the ground state of I_z . Since we can compute this quantity by simulating dynamics of a 2^{20} -dimensional state vector, it is significantly easier than computing $A(\mathbf{x}) = U(\mathbf{x}) I_z U^\dagger(\mathbf{x})$ where we would need to simulate dynamics of $2^{20} \times 2^{20}$ matrices. All simulations are performed with a quantum circuit simulator Qulacs⁴².

One-dimensional regression task

We show the results for the regression task of sinc function performed with the experimental NMR kernel and that of numerical simulations in Supplementary Fig. 2.

Comparison with Gaussian kernel

Here, we perform the same machine learning tasks with Gaussian kernel $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|)$, where $\gamma > 0$ is a hyperparameter.

First, we perform kernel regressions of sine and sinc functions using a Gaussian kernel with $\gamma = 0.02$. Results are shown as Supplementary Fig. 3. Mean squared errors of the results are 1.3×10^{-6} and 6.1×10^{-8} for sine and sinc functions, respectively.

Secondly, we perform kernel support vector machine on circle and moon datasets using a Gaussian kernel with $\gamma = 0.5$. Results are shown as Supplementary Fig. 4. Hinge losses of the results are 7.9×10^{-5} and 5.2×10^{-5} for circle and moon datasets, respectively.

Kernel from the numerical simulations

The kernel computed from numerical simulations for one-dimensional input is shown as Supplementary Fig. 5. It is shown as a function of $x - x' \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. We can observe the similar features as the experimental one, such as the sharpening of the kernel with increasing evolution time.

Kernel for two-dimensional data

Let $\mathbf{x} = (x_1, x_2)$ and $\mathbf{x}' = (x'_1, x'_2)$ be two data points with which we wish to evaluate the kernel $k(\mathbf{x}, \mathbf{x}') = k(\{x_1, x_2\}, \{x'_1, x'_2\})$. We note that our experimental kernel satisfies the equality: $k(\{x_1, x_2\}, \{x'_1, x'_2\}) = k(\{x_1 - x'_1, x_2 - x'_2\}, \{x'_1 - x'_2, 0\})$. With this in mind, we define $P_1 = x_1 - x'_1$, $P_2 = x_2 - x'_2$, $P_3 = x'_1 - x'_2$. The value of the experimental and simulated kernel are sliced by the value of P_3 and shown in Supplementary Figs. 6–8 and Supplementary Figs. 9–11, respectively.

DATA AVAILABILITY

Data are available upon reasonable requests.

CODE AVAILABILITY

Program codes are available upon reasonable requests.

Received: 8 June 2020; Accepted: 1 May 2021;

Published online: 08 June 2021

REFERENCES

- Harrow, A. W., Hassidim, A. & Lloyd, S. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.* **103**, 150502 (2009).
- Biamonte, J. et al. Quantum machine learning. *Nature* **549**, 195 (2017).
- Schuld, M. & Petruccione, F. *Supervised Learning with Quantum Computers* (Springer, 2018).
- Preskill, J. Quantum Computing in the NISQ era and beyond. *Quantum* **2**, 79 (2018).
- Boixo, S. et al. Characterizing quantum supremacy in near-term devices. *Nat. Phys.* **14**, 595–600 (2018).
- Villalonga, B. et al. Establishing the quantum supremacy frontier with a 281 Pflop/s simulation. *Quantum Sci. Technol.* **5**, 034003 (2020).
- Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
- Mitarai, K., Negoro, M., Kitagawa, M. & Fujii, K. Quantum circuit learning. *Phys. Rev. A* **98**, 032309 (2018).
- Schuld, M. & Killoran, N. Quantum machine learning in feature hilbert spaces. *Phys. Rev. Lett.* **122**, 040504 (2019).
- Benedetti, M. et al. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Inf.* **5**, 45 (2019).
- Dallaire-Demers, P.-L. & Killoran, N. Quantum generative adversarial networks. *Phys. Rev. A* **98**, 012324 (2018).
- Cong, I., Choi, S. & Lukin, M. D. Quantum convolutional neural networks. *Nat. Phys.* **15**, 1273–1278 (2019).
- Liu, J.-G. & Wang, L. Differentiable learning of quantum circuit Born machines. *Phys. Rev. A* **98**, 062324 (2018).
- Huggins, W., Patil, P., Mitchell, B., Whaley, K. B. & Stoudenmire, E. M. Towards quantum machine learning with tensor networks. *Quantum Sci. Technol.* **4**, 024001 (2019).
- Havlicek, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**, 209–212 (2019).
- Fujii, K. & Nakajima, K. Harnessing disordered-ensemble quantum dynamics for machine learning. *Phys. Rev. Appl.* **8**, 024030 (2017).
- Ghosh, S., Opala, A., Matuszewski, M., Paterek, T. & Liew, T. C. H. Quantum reservoir processing. *npj Quantum Inf.* **5**, 35 (2019).
- Álvarez, G. A., Suter, D. & Kaiser, R. Localization-delocalization transition in the dynamics of dipolar-coupled nuclear spins. *Science* **349**, 846–848 (2015).
- Choi, S. et al. Observation of discrete time-crystalline order in a disordered dipolar many-body system. *Nature* **543**, 221–225 (2017).
- Rovny, J., Blum, R. L. & Barrett, S. E. Observation of discrete-time-crystal signatures in an ordered dipolar many-body system. *Phys. Rev. Lett.* **120**, 180603 (2018).
- Rovny, J., Blum, R. L. & Barrett, S. E. ^3P NMR study of discrete time-crystalline signatures in an ordered crystal of ammonium dihydrogen phosphate. *Phys. Rev. B* **97**, 184301 (2018).

- Morimae, T., Fujii, K. & Fitzsimons, J. F. Hardness of classically simulating the one-clean-qubit Model. *Phys. Rev. Lett.* **112**, 130502 (2014).
- Fujii, K. et al. Impossibility of classically simulating one-clean-qubit model with multiplicative error. *Phys. Rev. Lett.* **120**, 200502 (2018).
- Farhi, E. & Neven, H. Classification with quantum neural networks on near term processors. *arXiv:1802.06002*. Preprint at <http://arxiv.org/abs/1802.06002> (2018).
- Liu, Y., Arunachalam, S. & Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *arXiv:2010.02174*. Preprint at <http://arxiv.org/abs/2010.02174> (2020).
- Bartkiewicz, K. et al. Experimental kernel-based quantum machine learning in finite feature space. *Sci. Rep.* **10**, 12356 (2020).
- Bishop, C. M. *Pattern Recognition and Machine Learning (Information Science and Statistics)* <https://www.ware.org/ref/a/0387310738> (Springer, 2011).
- Krojanski, H. G. & Suter, D. Scaling of decoherence in wide NMR quantum registers. *Phys. Rev. Lett.* **93**, 090501 (2004).
- Krojanski, H. G. & Suter, D. Decoherence in large NMR quantum registers. *Phys. Rev. A* **74**, 062319 (2006).
- Álvarez, G. A. & Suter, D. NMR quantum simulation of localization effects induced by decoherence. *Phys. Rev. Lett.* **104**, 230403 (2010).
- Gärttner, M. et al. Measuring out-of-time-order correlations and multiple quantum spectra in a trapped-ion quantum magnet. *Nat. Phys.* **13**, 781–786 (2017).
- Wei, K. X., Ramanathan, C. & Cappellaro, P. Exploring localization in nuclear spin chains. *Phys. Rev. Lett.* **120**, 070501 (2018).
- Baum, J., Munowitz, M., Garroway, A. N. & Pines, A. Multiple-quantum dynamics in solid state NMR. *J. Chem. Phys.* **83**, 2015–2025 (1985).
- Warren, W. S., Weitekamp, D. P. & Pines, A. Theory of selective excitation of multiple-quantum transitions. *J. Chem. Phys.* **73**, 2084–2099 (1980).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
- Nakajima, K., Fujii, K., Negoro, M., Mitarai, K. & Kitagawa, M. Boosting computational power through spatial multiplexing in quantum reservoir computing. *Phys. Rev. Appl.* **11**, 034021 (2019).
- Tateishi, K. et al. Room temperature hyperpolarization of nuclear spins in bulk. *Proc. Natl Acad. Sci. USA* **111**, 7527–7530 (2014).
- Henstra, A., Lin, T.-S., Schmidt, J. & Wenckebach, W. High dynamic nuclear polarization at room temperature. *Chem. Phys. Lett.* **165**, 6–10 (1990).
- Otten, M. et al. Quantum machine learning using gaussian processes with performant quantum kernels. *arXiv:2004.11280*. Preprint at <http://arxiv.org/abs/2004.11280> (2020).
- Peters, E. et al. Machine learning of high dimensional data on a noisy quantum processor. *arXiv:2101.09581*. Preprint at <http://arxiv.org/abs/2101.09581> (2021).
- Takeda, K. OPENCORE NMR: Open-source core modules for implementing an integrated FPGA-based NMR spectrometer. *J. Magn. Reson.* **192**, 218–229 (2008).
- Suzuki, Y. et al. Qulacs: a fast and versatile quantum circuit simulator for research purpose. *arXiv:2011.13524*. Preprint at <https://arxiv.org/abs/2011.13524> (2020).

ACKNOWLEDGEMENTS

K.M. is supported by JSPS KAKENHI No. 19J10978, 20K22330, and JST PRESTO Grant No. JPMJPR2019. K.F. is supported by KAKENHI No.16H02211, JST PRESTO JPMJPR1668, JST ERATO JPMJER1601, and JST CREST JPMJCR1673. M.N. is supported by JST PRESTO JPMJPR1666. This work is supported by MEXT Quantum Leap Flagship Program (MEXT Q-LEAP) Grant Number JPMXS0118067394 and JPMXS0120319794. We also acknowledge support from JST COI-NEXT program.

AUTHOR CONTRIBUTIONS

T.K. and K.M. contributed equally to this work. T.K. and M.N. conducted the experiments. K.M. conceived the idea and designed the experiments. K.F. provided the program code for simulations. M.K. and M.N. supervised the project. All authors contributed to the manuscript.

COMPETING INTERESTS

K.M., K.F., M.K., and M.N. own stock/options of QunaSys Inc.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41534-021-00423-0>.

Correspondence and requests for materials should be addressed to K.M.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give

appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021