ADVANCED QUANTUM TECHNOLOGIES
www.advquantumtech.com

Experimental Quantum Embedding for Machine Learning

*Ilaria Gianani, Ivana Mastroserio, Lorenzo Buffoni, Natalia Bruno, Ludovica Donati, Valeria Cimini, Marco Barbieri, Francesco S. Cataliotti, and Filippo Caruso**

The classification of big data usually requires a mapping onto new data clusters which can then be processed by machine learning algorithms by means of more efficient and feasible linear separators. Recently, Lloyd et al. have advanced the proposal to embed classical data into quantum ones: these live in the more complex Hilbert space where they can get split into linearly separable clusters. Here, these ideas are implemented by engineering two different experimental platforms, based on quantum optics and ultra-cold atoms, respectively, where we adapt and numerically optimize the quantum embedding protocol by deep learning methods, and test it for some trial classical data. A similar analysis is also performed on the Rigetti superconducting quantum computer. Therefore, it is found that the quantum embedding approach successfully works also at the experimental level and, in particular, we show how different platforms could work in a complementary fashion to achieve this task. These studies might pave the way for future investigations on quantum machine learning techniques especially based on hybrid quantum technologies.

1. Introduction

Since ancient times transferring acquired knowledge has induced mankind to develop solutions by applying established results in different contexts. This tendency remains valid even for today's most sophisticated technologies. Indeed, the current need of processing large amount of data and the availability of supercomputers has thus fostered an innovative take on programming: rather than trying and structuring the database so that a computer can walk through it, we now mimic processes of natural intelligence.[1–3] Algorithms are then able to act as quasi-conscious agents, thanks to the introduction of machine learning. These days, its applications are ubiquitous in everyday life, ranging from domotic systems, to autonomous cars, to face and voice recognition, and to medical diagnostics.[4–6] Artificial neural networks are one of the most common solutions[7,8] whose exploitation targets, among the others, deep learning—the classification of data according to processes akin to those occurring in the brain and reinforcement learning—finding the optimal strategy for a given task in complex environments.[9]

In classification problems, it is highly desirable that data can be sorted into distinct categories with tight dividing borders.[3,8] For this purpose, however, intensive preprocessing is often necessary on the original data for machine learning algorithms to perform efficiently. In the case of two-class problems, one would aim at achieving a geometrical representation of the data in order to establish a dividing hyperplane. All points sitting on either side of such a plane would be assigned to one class, for example, cats or dogs, implementing what is known as a linear classifier. However, it is not granted that the natural structure of the data suits these needs. Boundary may be obvious in the geometric representation, for example, the human-perceived distinction between cats and dogs, but, when these are applied to the original data, a complicated, nonlinear classifier results. For instance, scalar data are represented as points on a line, but the classification system may require to group disjoint intervals.

Linear classifiers are a preferable option as they are easier to find, however, they require to address the nontrivial task of embedding the original data in the appropriate space. Such a preprocessing, in turn, should not be too complex and resource intensive, with the risk of thwarting the benefits of an efficient classification. Quantum mechanics can provide an intriguing solution: even in the simplest instance, the natural representation of a quantum bit is the Bloch sphere, rather than the single-dimensional geometry of classical data. Data are thus naturally embedded into a much larger Hilbert space, and tight, separate clusters are formed, which can then be easily (quantum) recognized by a linear classifier.

I. Gianani, V. Cimini, M. Barbieri
Dipartimento di Scienze
Università degli Studi Roma Tre
Rome 00146, Italy

I. Mastroserio, L. Buffoni, N. Bruno, L. Donati, F. S. Cataliotti, F. Caruso
LENS & Dipartimento di Fisica e Astronomia
Università di Firenze
Sesto Fiorentino I-50019, Italy
E-mail: filippo.caruso@lens.unifi.it

I. Mastroserio, N. Bruno, M. Barbieri, F. S. Cataliotti
Istituto Nazionale di Ottica (CNR-INO)
Largo Enrico Fermi 6, Florence 50125, Italy

I. Mastroserio
Dipartimento di Fisica Ettore Pancini
Università degli Studi di Napoli Federico II
Napoli, Italy

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/qute.202100140

DOI: 10.1002/qute.202100140

A proposal on utilizing quantum computers for embedding has recently been advanced in ref. [10]. This comes in the framework of an exchange of concepts and methods between machine learning and quantum information.[11] Indeed, there have been demonstrations of the benefits of machine learning approaches to analyze data generated by quantum experiments,[12–19] to improve the performance of quantum sensors,[20,21] to Bayesian parameter estimations,[22] to the classification of non-Markovian noise,[23] and to the design of optical experiments.[24,25] Small gate-model devices and quantum annealers have been used to perform quantum heuristic optimization[26–30] and to solve classification problems.[31–34] Such devices have even been used with good promises in the context of unsupervized and reinforcement learning.[35–37] These recent applications on noisy intermediate size quantum (NISQ)[38] devices have depicted machine learning as a good candidate to harness the power of existing quantum technologies, albeit noisy and imperfect.

The exact detail on the performance of quantum embedding will depend on the degree of control on the actual system, thus on the level of experimental imperfections specific to the solutions adopted. This does affect, foremost, the size of the classical data set which can realistically be manipulated. In this article, we present an extensive experimental study of quantum embedding carried out on multiple platforms. In particular, we investigate how the protocol can be tailored to ultra-cold atoms, photonics, and via cloud available NISQ computers. Starting from a single prescription, we implement three different experiments, highlighting requirements and tolerances of each one for this task. The specific features come into play in a complementary fashion, hence supporting the promising idea of hybrid quantum technologies for future quantum machine learning applications.

2. Results

In our investigation we explore the application of quantum embedding in the simple, but illustrative instance of a single-qubit embedder. This is carried out in two steps: first, we identify the optimal quantum circuit, based on the classical data to be classified, using an iterative routine. Here the optimization is carried out off-line on a classical computer. Second, the circuit is implemented in three different architectures to explore how different sources of noise and imperfections impact the realization.

Quantum embedding is the representation of classical points x from a data domain X as a quantum feature state $|x\rangle$. Either the full embedding, or part of it, can be facilitated by a *quantum feature map*, that is, a quantum circuit $\Phi(x)$ that depends on the input. If the circuit has additional parameters θ that are adaptable, $\Phi(x) = \Phi(x, \theta)$, the quantum feature map can be trained via optimization. If we have classical data points a_i from class A, and b_j from class B, we want their quantum embeddings $|a_i\rangle$ and $|b_j\rangle$ to be as separated as possible in the Hilbert space. The process is pictorially represented in **Figure 1**. The approach is similar in spirit to the classical support vector machines (SVMs) commonly used in machine learning to perform classification.[39] SVMs map complex data (i.e., nonlinearly separable) via a nonlinear kernel into a high dimensional space where the data can be easily (linearly) classified by an hyperplane. The aim of this learning procedure is thus to be able, given new data point never seen during training, to correctly place it in the correct class by means

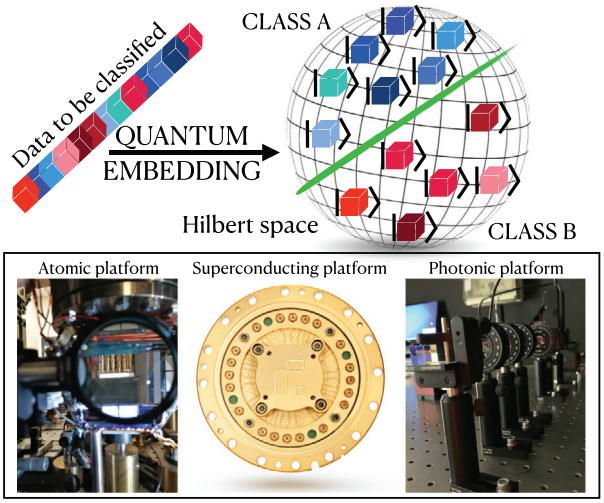

Figure 1. Pictorial view of the quantum embedding process where classical data (originally living in a highly complex set) can be embedded into the larger (Hilbert) space of quantum states that possibly belong to tight, more distant, and linearly separable clusters. Here we perform several experimental tests (based on atomic, superconducting, and photonic platforms, respectively) of this quantum embedding, theoretically proposed in ref. [10], in order to successfully demonstrate it at the experimental level and also to exploit the complementarity of the exploited platforms that is really crucial toward practical and more feasible hybrid quantum technologies.

of a suitable high dimensional embedding. This need for generalization to new data rules out trivial solution such as mapping all the points of a class to an extreme of the Bloch sphere, because these trivial mappings suppose the knowledge of the class, while the method we want to learn will correctly embed the point in its class without supplying any further knowledge. From the mathematical side, the determination of the optimal parameters θ of the embedding is based on computing the overlaps $|\langle a_i|b_j\rangle|^2$, $|\langle a_i|a_{i'}\rangle|^2$, $|\langle b_j|b_{j'}\rangle|^2$ for all members of the two classes, and minimizing the so-called cost function C.

$$C = 1 - \frac{1}{2}\left(\sum_{i,i'}|\langle a_i|a_{i'}\rangle|^2 + \sum_{j,j'}\left|\langle b_j|b_{j'}\rangle\right|^2\right) + \sum_{i,j}\left|\langle a_i|b_j\rangle\right|^2 \quad (1)$$

This amounts to both maximizing the Hilbert–Schmidt norm between the two classes and minimizing it within each one. This optimization procedure starts from random guesses and then leads to the optimal embeddings $|a_i\rangle$, $|b_j\rangle$, hence the associated optimal quantum circuit is found. In our example, the classical data set is a collection of ten scalars ϕ chosen in the interval $[-\pi, \pi]$, arranged in two classes with five elements each (see Supporting Information for details). The optimal training is achieved by using the open-source software Pennylane[40] and the quantum circuit introduced in ref. [10], based on a sequence of rotations on non-commuting axes. Note that the particular choice of the circuit is completely arbitrary as long as the operators are non-commuting, indeed it might be the case that some quantum circuits generate better embeddings for particular problems by taking advantages of symmetries in the data set. The qubit is first initialized to an input-independent state $SH|0\rangle = (|0\rangle + i|1\rangle)/\sqrt{2}$, where H is the Hadamard gate, and S is the phase gate. The following sequence

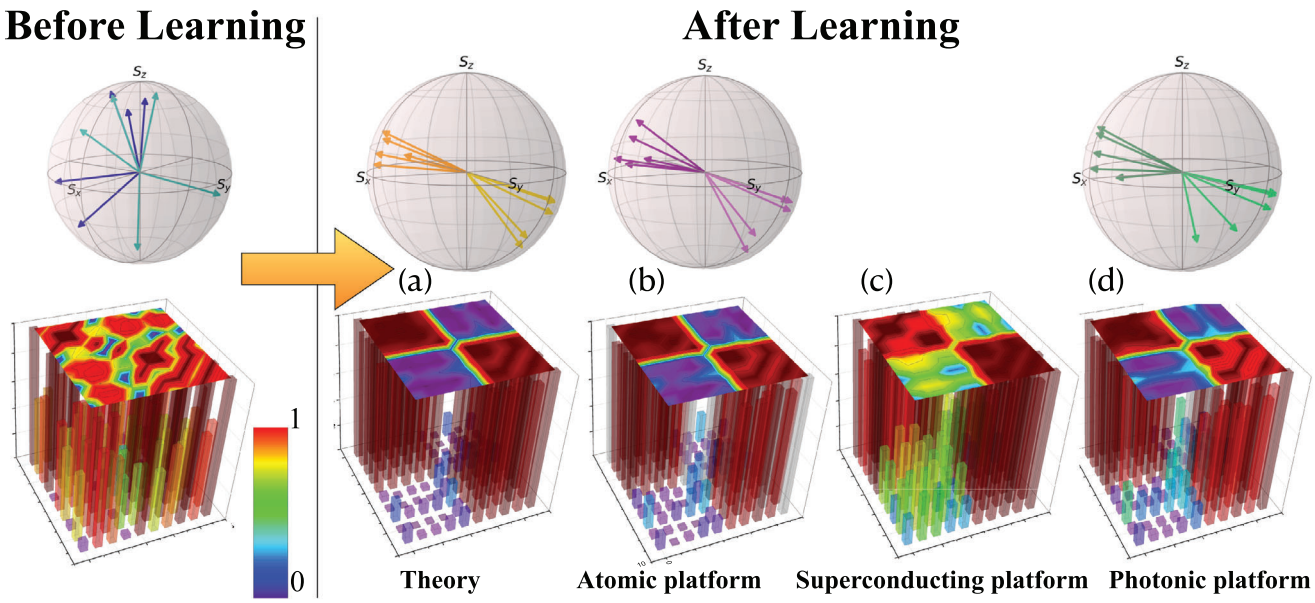

Figure 2. Top panel: Bloch sphere representation of the embedded quantum states that are randomly distributed (before learning) and clustered in two families (after learning), as provided by (a) theory (i.e., the simulation of the embeddings on a classical computer) and by the experimental results obtained on the (b) atomic and (d) photonic platforms. The corresponding Gram matrices for the set of ten embedded quantum states (before and after learning) are shown in the bottom panel. Here, the results achieved with a superconducting platform (c) are also shown. For the superconducting platform we were not able to extract the Bloch vectors due to the limited controls given by accessing it remotely and the consequent difficulty to perform a quantum state tomography that was, on the other hand, possible on the other two platforms.

of rotations around the X and Z axes of the Bloch sphere is then applied

$$\{R_X(\phi), R_Z(\theta_1), R_X(\phi), R_Z(\theta_2), R_X(\phi), R_Z(\theta_3), R_X(\phi)\} \tag{2}$$

in order to construct the quantum state $|\phi\rangle$. The circuit parameters to be optimized are thus the rotation angles $\theta = \{\theta_1, \theta_2, \theta_3\}$. This sequence of operations provides the optimal strategies for encoding we used in our experiments, with specific adaptations to the requisites of each architecture.

The performance of the embedding is conveniently captured by the Gram matrix, containing all the scalar products between the embedded states. **Figure 2** shows how optimization of the parameters θ does lead to the presence of two clusters of quantum states, as clearly illustrated by the Bloch sphere representation. As the system is trained, we can take advantage of the Gram matrix to control how well it has learned to separate the points in the Hilbert space. The training is done taking two data points, embedding them into two separate qubits and using a third qubit to perform a SWAP test between the two embeddings, hence giving us the overlaps we need to compute the cost function in Equation (1). The parameters of the embedding circuit are then updated by gradient descent using the automatic differentiation capabilities of Pennylane. After a few hundreds of gradient descent steps (i.e., training steps), the cost function reaches a minimum and the optimal encoding circuit is found[10]—as illustrated in Figure 2a for the optimally embedded (theoretical) states. This procedure is completely general, and allows extension to a more complex data set (see Supporting Information) and to rotations around different axes. Therefore, the scheme is flexible and can be manipulated in order to account for the specific needs of different experimental platforms.

The manipulation of atomic internal states is naturally suited to the above scheme, since qubit operations in (2) are indeed rotations realized by sequences of control pulses. The first experiment is therefore performed on a Bose–Einstein condensate (BEC) of $\approx 10^5$ ^{87}Rb atoms, evolving according to an external microwave driving field. The qubit is identified as the two-level system $\{|0\rangle \equiv |F = 2, m_F = 0\rangle; |1\rangle \equiv |F = 1, m_F = 0\rangle\}$ formed by rubidium hyperfine ground states (see Figure B1, Supporting Information). The classical information to be encoded is mapped in the rotation angles of the Bloch vector of our two-level system, around the non-commuting axes S_x and S_z of the Bloch sphere.

The Bloch vector, representing the atomic qubit, evolves under the rotating wave approximation, according to the unitary transformation.

$$U_\Omega(\Omega, \delta, t) = exp\left(-\frac{i|\vec{\Omega}|t}{2}\vec{n}_\Omega \cdot \vec{\sigma}\right) \tag{3}$$

with $|\vec{\Omega}| = \sqrt{\Omega^2 + \delta^2}$ being the generalized Rabi frequency and the Rabi vector direction $\vec{n}_\Omega \equiv (\frac{\Omega}{|\vec{\Omega}|}, 0, -\frac{\delta}{|\vec{\Omega}|})$ identifying the rotation axis. Here, $\vec{\sigma}$ is the vector of the Pauli matrices, δ is the detuning of the microwave frequency from the atomic resonance, while Ω is the Rabi frequency of the pulse. The embedding parameters, θ and ϕ, are related to the experimental frequencies Ω and δ and the interaction time t as

$$\begin{aligned} \theta &= \delta t \\ \phi &= \Omega t \end{aligned} \tag{4}$$

To realize the embedding of sequence (2) we adapt the embedding procedure to the experimental implementation, for each of the ten classical inputs ϕ. Indeed, due to experimental constraints that fundamentally limit the interaction time between the atomic qubit and the microwave field, we are only allowed to perform a few rotations instead of the seven required by Equation (2). Taking into account this limitation we numerically compute, for each embedded point, three new angles $\{\varphi_1, \vartheta, \varphi_2\}$ that allow us to effectively perform the same global rotation but using less operations and thus respecting our constraints. Hence, the sequence of rotations (2) reduces to the following three unitary transformations.

$$\{U_\Omega(\Omega, \delta, \tau_1), U_\Omega(0, \delta, T), U_\Omega(\Omega, \delta, \tau_2)\} \tag{5}$$

The total evolution length is determined by the sum of each interaction time duration $\tau_1 + T + \tau_2$. Their values can be found in Appendix B, Table B1. While Ω and δ are fixed by the experimental setting, the only free parameter that we optimize for the embedding is the interaction time t. The total evolution of Equation (5) is implemented by a quasi-resonant microwave rectangular waveform that couples the two-level system. The rectangular waveform is composed of a first pulse that realizes the evolution $U_\Omega(\Omega, \delta, \tau_1)$. It is implemented with the microwave switched on ($\Omega \neq 0$) thus letting the qubit to precess with an angular frequency equal to the generalized Rabi frequency $|\vec{\Omega}|$ around the axis $\vec{n}_\Omega$ about an angle $\varphi_1 = \Omega\tau_1$. A subsequent free evolution, with the microwave switched off ($\Omega = 0$), lets the system rotate, according to $U_\Omega(0, \delta, T)$, around the S_z axis of an angle $\vartheta = \delta T$. Similarly, a final pulse, with the microwave switched on, drives the evolution $U_\Omega(\Omega, \delta, \tau_2)$ thus allowing the qubit to precess around $\vec{n}_\Omega$ of an angle $\varphi_2 = \Omega\tau_2$. Note that, since the detuning is fixed during the evolution, when the microwave pulse is on the system rotates around the axis $\vec{n}_\Omega$ that forms an angle of $\arctan(\delta/\Omega)$ with the S_x axis. The non-orthogonality of the two rotation axes $\vec{n}_\Omega$ and S_z is taken into account when computing the interaction times to realize the sequence of rotations in (5). To reconstruct the final embedded state after the total evolution, we perform a state tomography as explained in detail in the Supporting Information. The sources of error and imperfections in the implementation of the protocol on the atomic experimental platform come from the limited accuracy and stability of the measured Rabi frequency and detuning used to set the desired rotations. As explained in the Supporting Information, this induces an error of about 0.1 in the measurement of the components S_x, S_y, and S_z of the final collective-spin vector $\vec{S}$. The experimental findings in terms of the Bloch vectors and Gram matrix are reported in Figure 2b where a clear clusterization of the reconstructed states can be seen. The fidelity between the theoretical prediction and the experimental result is on average better than 0.99 (see Figure B2, Supporting Information). We can safely conclude that the embedding procedure is faithfully working with the atomic platform.

Different platforms, however, do not necessarily have the same features and constraints. We evidence this by taking as an example the case of the superconducting chip of Rigetti[41] named Aspen-8. It is composed of a lattice of 30 superconducting qubits in a ladder-like configuration. The qubits are controllable by the action of single- and two-qubit gates giving this platform all the characteristics of an universal quantum computer. We thus use this platform to deploy again the same embedding in Equation (2) but on a completely different device with different noise profile and constraints. Each circuit is sampled 2000 times for each of the 100 data points necessary to build the Gram matrix, using Rigetti's cloud service. As shown in Figure 2c, the Gram matrix (even if it is sensibly noisier than our other tests) clearly exhibits the separation boundary between the two classes. The advantage of this experiment is that it could be performed without the need of an ad hoc lab; it has been carried out remotely by just reserving some time on the Rigetti system and programming it.[42] The entire set of experiments performed on this platform to get these results has taken a total time of around 5 min to run.

The examples above detail results obtained on platforms with static material qubits. It is also of interest to perform the quantum embedding with moving qubits, that is, photons. Thus, we have also investigated quantum embedding in an optical experiment by generating photon pairs at the degenerate wavelength 810 nm via parametric down conversion realizing a heralded single-photon source; the input state of the photons is set to $|H\rangle$. A direct application of the sequence (2) would result impractical and resource consuming. Instead, the embedding is performed by applying a single unitary

$$\tilde{U}(\phi; \vec{n}) = \cos(\phi/2)\sigma_0 - i\sin(\phi/2)(\vec{n} \cdot \vec{\sigma}) \tag{6}$$

where $\vec{n} = (n_x, n_y, n_z)$ denotes the associated rotation axis (the actual used values are in the Supporting Information). This represents the compiled effect of the original sequence, written as a single operation.

We have then adopted an alternative approach, based on decomposing the unitary (6) and the SH initialization by means of a series of three wave plates—a quarter wave plate (Q1), a half wave plate (H2), and a second quarter wave plate (Q3), whose angles are associated to the parameters ϕ and $\vec{n}$ as

$$\begin{aligned}
\theta_{Q1} &= \frac{1}{2}\left(-\arctan\left(\frac{n_z}{n_x}\right) - \arctan\left(n_y \tan\left(\frac{\phi}{2}\right)\right)\right) \\
\theta_{H2} &= \frac{1}{2}\left(-\arcsin\left(n_x\sqrt{\frac{n_z^2}{n_x^2}+1}\sin\left(\frac{\phi}{2}\right)\right) - \arctan\left(\frac{n_z}{n_x}\right)\right) \\
\theta_{Q3} &= \frac{1}{2}\left(-\arctan\left(\frac{n_z}{n_x}\right) - \arctan\left(n_y \tan\left(\frac{\phi}{2}\right)\right)\right)
\end{aligned} \tag{7}$$

This is more suited to photon polarization. On the one hand, this comes at the cost of losing a direct mapping of each term in the sequence (2) into a physical object. It also introduces the requirement of controlling a larger class of operations than rotations on two specified axes. On the other hand it reduces the number of required elements, leveraging on the availability of ready-made hardware components. Verification of the outputs is carried out by means of quantum state tomography, operated by a further half wave plate/quarter wave plate sequence and a polarizer.

The reconstructed Gram matrix is reported in Figure 2d. The two main sources of imperfection are the limited accuracy in the setting of the wave plate axis ($\pm 1^\circ$ for the encoding plates, $\pm 0.25^\circ$ for the tomography plates), and deviations of the imparted

phase shifts from the target values $\pi/2$ or π. Despite these deviations, clustering of the states in two classes is clearly observed. Once again the fidelity between the experimental results and the theoretical predictions is on average above 0.96 (see Figure C2, Supporting Information) accounting for a more than satisfactory agreement.

3. Conclusions

The Gram matrices reveal that all the architectures we have considered indeed give satisfactory results in realizing the quantum embedding. To make this observation more quantitative, here we provide a theoretical bound, based on the observed fidelities, for the number of different classes and/or different points that can be embedded on a single qubit. Our bound is derived from geometrical constraints of the embedding on the sphere. The minimal assumption we can make for the embedding to be successful is to avoid overlapping between different classes. These are associated with distinct spherical sectors defined by its central angle Θ; in our example, the two sectors correspond to two separate halves of the sphere ($\Theta = \pi$). By simple geometrical arguments, the surface occupied by a single spherical sector is given by $2\pi[1 - \cos(\Theta/2)]$. Since we want the sectors to be non-overlapping, and the total available surface of our Bloch sphere is 4π, we can formulate a geometrical constraint considering the number of classes N as follows.

$$2\pi N\left[1 - \cos\left(\frac{\Theta}{2}\right)\right] \leq 4\pi \tag{8}$$

This equation thus computes the tightness of the embedding given a number N of classes to embed.

A similar geometric reasoning can also be applied to give an upper bound, albeit not a tight one, on the dimension of data set which can effectively be embedded. To do this we will use the fidelity F between the experimental states and the corresponding targets. While the ideal state is associated to a unique direction in the Bloch sphere, accounting for the imperfections of the actual prepared state leads us to consider a spherical surface $2\pi(1 - F)$ as the actual geometric measure of the volume occupied by a state on the Bloch sphere prepared with fidelity F with respect to the ideal target state. Thus, the maximum number of points allowed is

$$N_{\max} \leq \frac{4\pi}{2\pi(1 - F)} \tag{9}$$

Our experiments demonstrate that a fidelity exceeding 0.9 can be routinely achieved on all platforms we have explored. This determines a maximal embedding capacity $N_{\max} \simeq 20$ as a conservative estimate. This reasoning can be obviously generalized to the 2^n-dimensional hypersphere in the case of multi-qubit embedding. We also remark that the bounds given in this section are not tight ones, nevertheless they are useful gauges to make theoretical estimates on the tightness of embedding and the maximum number of points that can be embedded using a certain platform.

Our experimental investigation demonstrates how quantum embedding techniques may suit radically different approaches to qubit encoding and manipulation—by pulses as for cold atom, by quantum logic circuits as for the Rigetti machine, or by compiled operations as for photons. Such a versatility shows promises for future interconnected systems on hybrid architectures, with specialized hardware for storage, processing, and distribution of quantum data. Finally, the potential advantages of representing classical data on quantum systems include not only the possibility to simplify a classification problem as experimentally demonstrated in this work, but also the ability to speed up any processing of the classical data such as, among others, the quantum parallelism to search through a database, feature extraction, image segmentation, and edge detection. Indeed, combining quantum machine learning and quantum image processing is expected to allow to potentially solve real-world problems that are very challenging via classical supercomputers, especially in the case of large volumes of data in various domains ranging from sociology to economy, from geography to biomedicine.

4. Experimental Section

Atomic Platform: Starting from a room temperature gas in ultra-high vacuum conditions a BEC of ^{87}Rb atoms were produced in 8 s by laser cooling followed by evaporative cooling in a magnetic micro-trap realized with an atom chip.[43] Quantum degeneracy was achieved by ramping down the frequency of a radio-frequency field integrated on the chip. This procedure yielded a BEC of $\approx 10^5$ atoms in $|F = 2, m_F = 2\rangle$ (see Supporting Information), at a critical temperature of 0.5 μ K and a distance of 300 μ m between the atomic cloud and the chip surface. After switching off the magnetic trap, atoms were let to expand for 1 ms to strongly reduce the effects of atomic collisions. Subsequently, a constant magnetic bias field of 6.179 G was applied to lift the magnetic degeneracy of the hyperfine states and to define the quantization axis of the system. The opposite sign of the Landé factors in the two hyperfine ground levels isolated the $|F = 2\rangle \rightarrow |F = 1\rangle$ two-level system microwave clock transition chosen as the qubit for the embedding. To transfer all the atoms in the initial state of the experiment $|0\rangle = |F = 2, m_F = 0\rangle$, a frequency modulated radio-frequency pulse designed with an optimal control strategy was applied.[44] An external microwave antenna at 6.834 GHz was then exploited to drive the clock transition inducing an average Rabi oscillation frequency of $2\pi \times 38(2)$ kHz. After all the manipulations to realize the embedding of the states, the number of atoms in each of the m_F Zeeman states were finally recorded, of the $|F = 1\rangle$ and $|F = 2\rangle$ hyperfine levels, applying a Stern–Gerlach method.[45] The different m_F states were let to spatially separate by means of an inhomogeneous magnetic field applied along the quantization axis and then, after 23 ms of free expansion, a standard absorption imaging sequence was executed. The sequence was repeated twice, once with light resonant with the $|F = 2\rangle$ state and immediately after with light resonant to the $|F = 1\rangle$ state.

Photonic Platform: Degenerate 810 nm photon pairs were generated from a 50 mW CW laser at 405 nm through Type I spontaneous parametric downconversion using a 3 mm BBO crystal. Both photons were filtered with bandpass filters (FWHM = 7.3 nm) and single mode fibers. One photon acted as a trigger and was directly coupled to an avalanche photodiode detector, while the other underwent the transformation $\tilde{U}(\phi; \vec{n})$ imparted by the sequence of quarter–half–quarter wave plates described in the main text. Quantum state tomography was then performed by collecting coincidence counts in correspondence to the projections along the polarization directions horizontal, vertical, diagonal, anti-diagonal, right-circular, and L-circular. For each state, ≈20k coincidence events were collected, distributed among the six projectors.

Appendix A: Theoretical Details

The classical data set is reported in **Figure A1**. This data set cannot be linearly separated in 1D representation, thus making the quantum embedding a suitable resource to classify the two classes. In our case the

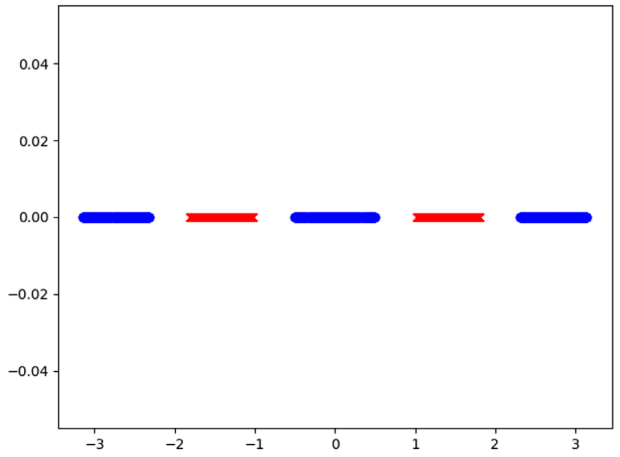

Figure A1. Data set to be classified using the embedding. The data set is 1D and it is not linearly separable. The classes (blue) and (red) have been normalized to live in the interval $[-\pi, \pi]$.

training is performed on 1000 points and the cost function is optimized as explained in the main text by using the simulation software Pennylane. The peculiarity of Pennylane is that it integrates a suite of quantum simulations (that we use to implement the circuit) with common machine learning tools such as TensorFlow and Pytorch. Leveraging the remarkable automatic differentiation and optimization capabilities of these libraries our procedure is able to converge to a minimum of our cost function in 200 iterations taking only a few minutes of computational time. We use the TensorFlow backend with the Adam optimizer.

Once the optimization is done, we sample ten more data points (not exploited for the training) in order to test the generalization capabilities of the embedding. These ten test points are all correctly classified and they subsequently serve as a benchmark for all the experimental platforms. After this "learning" phase has been carried out on an ideal circuit, we deploy the same embedding on different experimental platforms to test its robustness to real-world deployment scenarios. Since different experiments have different capabilities, each deployment requires to "recompile" the embedding unitary transformation to fit with the specifications of each device while keeping the learned parameters fixed. The different constraints are extensively explained in the following sections.

Appendix B: Atomic Platform

Experimental Procedure: The experimental setup used for the atomic experiment is shown in **Figure B1**. The embedding is realized by letting the Bloch vector evolve from the initial state $|0\rangle = |F = 2, m_F = 0\rangle$ to the final embedded state $|x\rangle$ through the set of rotations in (5) dictated by the embedding. The latter are implemented designing a microwave rectangular wave, whose duration is fixed by the angles $\varphi_i = \Omega\tau_i$, with $i = 1, 2$, and $\vartheta = \delta T$, where $\Omega = 2\pi\times 38(2)$ kHz is the Rabi frequency and $\delta = 2\pi\times 6.57(4)$ kHz the detuning of the microwave frequency from the atomic resonance. The interaction time with microwave switched on (τ_1, τ_2) and with microwave switched off (T) are reported in **Table B1**.

At the end of the evolution for each final state we measure the number of atoms $N_{|1\rangle}$ in state $|1\rangle$ and $N_{|0\rangle}$ in state $|0\rangle$. From the relative population $P_{|1\rangle} = N_{|1\rangle}/(N_{|0\rangle} + N_{|1\rangle})$, we get the z-component of the state vector defined as $\langle S_z\rangle = (N_{|0\rangle} - N_{|1\rangle})/(N_{|0\rangle} + N_{|1\rangle})$, obtaining

$$\langle S_z\rangle = 1 - 2P_{|1\rangle} \tag{B1}$$

To measure the other components $\langle S_x\rangle$ and $\langle S_y\rangle$ of the final state, we project them onto the z-axis of the Bloch sphere by adding further rotations to the one computed to measure $\langle S_z\rangle$. Experimentally, we realize them by adding another sequence of rectangular wave pulses of suitable duration.

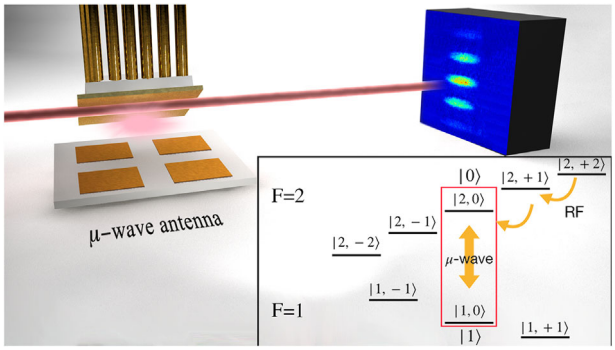

Figure B1. Left: Pictorial representation of the atom chip placed inside the vacuum science cell to achieve the BEC and manipulate the atoms' internal dynamics. An external microwave antenna, shown in the image below the chip, is used to drive the evolution of the qubit and to realize the embedding. Note that the drawing is not to scale. Right: Typical absorption image of the atomic clouds after Stern–Gerlach separation. All Zeeman substates can be simultaneously detected. In the inset we show the involved level scheme with yellow thin lines representing the radio-frequency transitions while thick lines represent the microwave transition.

Table B1. Interaction times with the microwave switched on (τ_1, τ_2) and off (T) for the atomic experiment, in order to generate the ten optimally embedded quantum states.

State	$\tau_1[\mu s]$	$T[\mu s]$	$\tau_2[\mu s]$
1	19	36	8
2	45	51	7
3	20	37	1
4	20	38	3
5	19	28	3
6	30	47	38
7	32	20	8
8	4	7	35
9	4	25	12
10	7	10	6

In this way we retrieve $\langle S_x\rangle$ and $\langle S_y\rangle$ by the measured relative population with Equation (B1). By doing so, we implement the quantum state tomography of the embedded states.

Accuracy and Stability: The experimental realization of the embedding protocol with atoms is fundamentally limited by the accuracy and stability with which we measure the Rabi frequency and the radio-frequency detuning necessary to set the desired rotations. Therefore, in order to estimate this accuracy, we have performed Rabi oscillations and a Ramsey interferometer up to a duration of 300 μs, much longer that any of the embedding sequences. We have hence measured a fluctuation of the order of 1% on Rabi and detuning frequencies. Furthermore, the frequency stability, presumably due to microwave power fluctuations for the Rabi frequency and due to magnetic field fluctuations for the detuning, have been evaluated measuring the Allan variance for a total time interval of 3 h. The results have allowed us to estimate a fluctuation of $2\pi\times$ 1.5 kHz and of $2\pi\times$ 71 Hz for Rabi frequency and detuning, respectively, evaluated on an average time of 115 s corresponding to the time needed for five repeated measurement of each embedded state. The linear uncertainty propagation realized with a simulation of the experiment results in an evaluated error of 0.1 on the measurement of the components $\langle S_x\rangle$, $\langle S_y\rangle$, and $\langle S_z\rangle$ of the final collective-spin vector. The fidelities between the predicted states and the reconstructed experimental ones, together with error bars computed via uncertainty propagation, are shown in **Figure B2**.

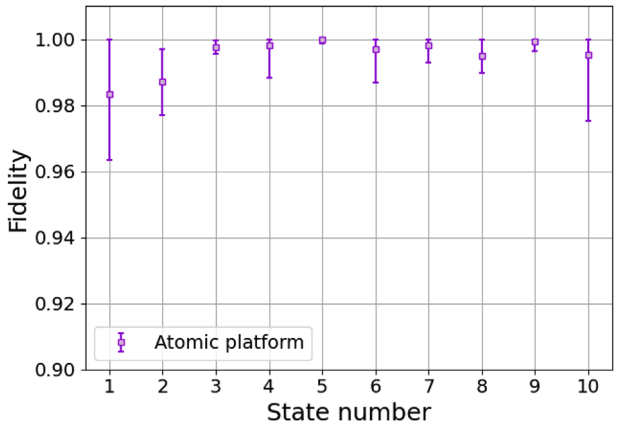

Figure B2. Fidelity between the predicted state and the one experimentally reconstructed in the atomic experiment by measuring the three $\vec{S}$ components after applying the quantum embedding circuit, calculated for the ten validation states. Each measurement of the components is repeated five times and the uncertainty on the fidelity is obtained via error propagation. The error bar for the state 5 is smaller than the marker.

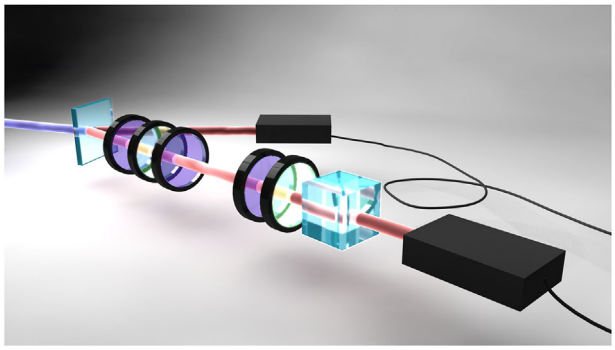

Figure C1. Optical experimental setup: SPDC through a 3 mm Type I BBO crystal generates degenerate photon pairs at 800 nm. One photon is used for heralding, while the other is rotated with a quarter wave plate–half wave plate–quarter wave plate and then quantum state tomography is performed with a quarter wave plate, half wave plate, and a polarizing beam splitter. Coincidence counts are then collected as reported in the main text.

Appendix C: Photonic Platform

The experimental setup used for the photonic experiment is shown in **Figure C1**. A heralded single photon source is set up, based on parametric down conversion, and a three-plate arrangement is used to impart a generic transform, while tomography adopts the standard technique of projecting on three mutually unbiased bases.

The embedding parameters dictating the rotation for each of the ten states are reported in **Table C1**. The fidelities between the predicted states and the reconstructed ones are shown in **Figure C2**. Uncertainties have mainly statistical origin, due to fluctuations in the number of collected counts. In order to account for those, we have performed a Monte Carlo routine, simulating 300 experiments: in each of them, counts are generated by bootstrapping on the observed values, assuming Poissonian statistics, as customary. The uncertainty on the fidelity is thus evaluated as the standard deviation of the fidelities with the target state observed in the Monte Carlo replicas.

Table C1. Embedding parameters for the photonic experiment.

State	ϕ	n_x	n_y	n_z
1	0.668	0.667	0.143	0.731
2	1.986	−0.423	0.460	−0.781
3	2.111	−0.510	0.379	−0.772
4	2.408	0.619	0.240	0.748
5	1.301	−0.405	0.914	0.034
6	4.258	0.418	0.908	−0.006
7	4.367	0.247	0.969	0.026
8	3.549	−0.475	0.847	0.239
9	4.379	0.197	0.980	0.036
10	3.762	−0.433	0.877	0.208

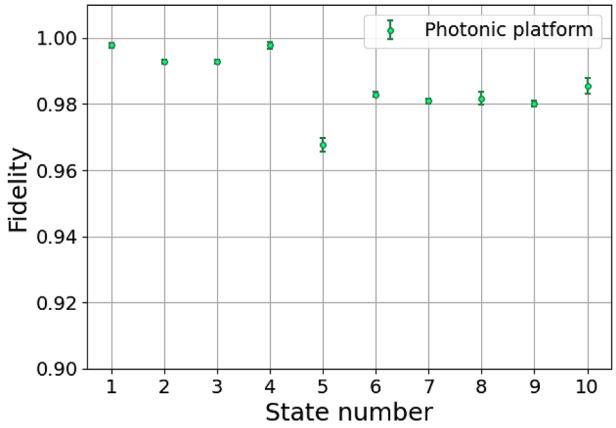

Figure C2. Fidelity between the predicted state and the experimentally reconstructed one in the photonic experiment, obtained by quantum state tomography after applying the quantum embedding circuit, calculated for the ten validation states.

Acknowledgements

F.C. was financially supported by the European Union's Horizon 2020 research and innovation programme under FET-OPEN Grant Agreement No. 828946 (PATHOS). N.B. was financially supported by the Qombs Project, FET Flagship on Quantum Technologies grant no. 820419.

Open Access Funding provided by Universita degli Studi di Firenze within the CRUI-CARE Agreement.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

I.G., V.C., and M.B. carried out the photonic experiment; I.M., N.B., L.D. and F.S.C. carried out the atomic experiment; F.C. and L.B. led and carried out the theoretical work. L.B. performed the numerical optimizations and the analysis of the results from the Rigetti machine. F.C. conceived the whole project. F.C., M.B., and F.S.C. supervized the project. All authors contributed to the discussion, analysis of the results, and the writing of the manuscript.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

experimental quantum technologies, noisy intermediate size quantum devices, quantum machine learning, quantum optics, ultra-cold atoms

Received: November 4, 2021
Revised: March 7, 2022
Published online: June 9, 2022

[1] F. Rosenblatt, *Psychol. Rev*. **1958**, *65*.
[2] J. Schmidhuber, *Neural Networks* **2015**, *61*.
[3] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data mining, Inference, and Prediction*, Springer Science & Business Media, Berlin **2009**.
[4] S. Grigorescu, B. Trasnea, T. Cocias, G. Macesanu, *J. Field Robot.* **2020**, *37*, 362.
[5] N. Sebe, I. Cohen, A. Garg, T. S. Huang, *Machine Learning in Computer Vision*, Springer Science & Business Media, Berlin **2005**.
[6] A. Graves, A. Mohamed, G. Hinton, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, **2013**, pp. 6645–6649, http://doi.org/10.1109/ICASSP.2013.6638947.
[7] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, *Deep Learning*, MIT Press, Cambridge, MA **2016**.
[8] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, Berlin **2006**.
[9] R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA **2018**.
[10] S. Lloyd, M. Schud, A. Ijaz, J. Izaac, N. Killoran, arXiv:2001.03622 **2020**.
[11] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining*, Academic Press, Cambridge, MA **2014**.
[12] J Gao, L.-F. Qiao, Z.-Q. Jiao, Y.-C. Ma, C.-Q. Hu, R.-J. Ren, A.-L. Yang, H. Tang, M.-H. Yung, X.-M. Jin, *Phys. Rev. Lett.* **2018**, *120*, 240501.
[13] A. Rocchetto, S. Aaronson, S. Severini, G. Carvacho, D. Poderini, I. Agresti, M. Bentivegna, F. Sciarrino, *Sci. Adv.* **2019**, *5*, eaau1946.
[14] G. Torlai, B. Timar, E. P. L. van Nieuwenburg, H. Levine, A. Omran, A. Keesling, H. Bernien, M. Greiner, V. Vuletic, M. D. Lukin, R. G. Melko, M. Endres, *Phys. Rev. Lett.* **2019**, *123*, 230504.
[15] A. M. Palmieri, E. Kovlakov, F. Bianchi, D. Yudin, S. Straupe, J. D. Biamonte, S. Kulik, *npj Quantum Inf.* **2020**, *6*, 20.
[16] T. Giordani, A. Suprano, E. Polino, F. Acanfora, L. Innocenti, A. Ferraro, M. Paternostro, N. Spagnolo, F. Sciarrino, *Phys. Rev. Lett.* **2020**, *124*, 160401.
[17] E. S. Tiunov, V. V. Tiunova (Vyborova), A. E. Ulanov, A. I. Lvovsky, A. K. Fedorov, *Optica* **2020**, *7*, 448.
[18] V. Cimini, M. Barbieri, N. Treps, M. Walschaers, V. Parigi, *Phys. Rev. Lett.* **2020**, *125*, 160504.
[19] V. Gebhart, M. Bohmann, K. Weiher, N. Biagi, A. Zavatta, M. Bellini, E. Agudelo, arXiv:2101.07112 **2021**.
[20] S. Paesani, A. A. Gentile, R. Santagati, J. Wang, N. Wiebe, D. P. Tew, J. L. O'Brien, M. G. Thompson, *Phys. Rev. Lett.* **2017**, *118*, 100503.
[21] V. Cimini, I. Gianani, N. Spagnolo, F. Leccese, F. Sciarrino, M. Barbieri, *Phys. Rev. Lett.* **2019**, *123*, 230502.
[22] S. P. Nolan, A. Smerzi, L. Pezzé, arXiv:2006.02369 **2020**.
[23] S. Martina, S. Gherardini, F. Caruso, arXiv:2101.03221 **2021**.
[24] M. Krenn, M. Malik, M. R. Fickler, R. Lapkiewicz, R. A. Zeilinger, *Phys. Rev. Lett.* **2016**, *116*, 090405.
[25] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, H. J. Briegel, *Proc. National Acad. Sci.* **2018**, *115*, 1221.
[26] T. Kadowaki, H. Nishimori, *Phys. Rev. E* **1998**, *58*, 5355.
[27] J. Brooke, T. F. Rosenbaum, G. Aeppli, *Nature* **2001**, *413*, 610.
[28] E. Farhi, J. Goldstone, S. Gutmann, arXiv:1411.4028 **2014**.
[29] A. Peruzzo, J. McClean, P. Shadbolt, M. Yung, X. Zhou, P. J. Love, A. Aspuru-Guzik, J. L. O'brien, *Nat. Commun.* **2014**, *5*, 4213.
[30] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, J. M. Gambetta, *Nature* **2017**, *549*, 7671.
[31] H. Neven, V. S. Denchev, G. Rose, W. G. Macready, arXiv:0811.0416 **2008**.
[32] V. S. Denchev, N. Ding, S. V. N. Vishwanathan, H. Neven, arXiv:1205.1148 **2012**.
[33] K. L. Pudenz, D. A. Lidar, *Quantum Inf. Process.* **2013**, *12*, 5.
[34] A. Mott, J. Job, J. Vlimant, D. Lidar, M. Spiropulu, *Nature* **2017**, *550*, 7676.
[35] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. S. Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, C. A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, R. S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, B. R. Johnson, M. Reagor, M. P. da Silva, C. Rigetti, arXiv:1712.05771 **2017**.
[36] W. Vinci, L. Buffoni, H. Sadeghi, A. Khoshaman, E. Andriyash, M. H. Amin, *Mach. Learn.: Sci. Tech.* **2020**, *1*, 4.
[37] L. Buffoni, F. Caruso, *EPL* **2020**, *132*, 60004.
[38] J. Preskill, *Quantum* **2018**, *2*, 79.
[39] J. A. K. Suykens, J. Vandewalle, *Neural Process. Lett.* **1999**, *9*, 293.
[40] V. Bergholm, J. Izaac, M. Schuld, C. Gogolin, C. Blank, K. McKiernan, N. Killoran, arXiv:1811.04968.
[41] M. Reagor, C. B. Osborn, N. Tezak, A. Staley, G. Prawiroatmodjo, M. Scheer, N. Alidoust, E. A. Sete, N. Didier, M. P. da Silva, E. Acala, J. Angeles, A. Bestwick, M. Block, B. Bloom, A. Bradley, C. Bui, S. Caldwell, L. Capelluto, R. Chilcott, J. Cordova, G. Crossman, M. Curtis, S. Deshpande, T. El Bouayadi, D. Girshovich, S. Hong, A. Hudson, P. Karalekas, K. Kuang, et al., *Sci. Adv.* **2018**, *4.2*, eaao3603.
[42] R. S. Smith, M. J. Curtis, W. J. Zeng, arXiv:1608.03355 **2016**.
[43] J. Petrovic, I. Herrera, P. Lombardi, F. Schäfer, F. S. Cataliotti, *New J. Phys.* **2013**, *15*, 043002.
[44] C. Lovecchio, F. Schäfer, S. Cherukattil, M. Alì Khan, I. Herrera, F. S. Cataliotti, T. Calarco, S. Montangero, F. Caruso, *Phys. Rev. A* **2016**, *93*, 010304(R).
[45] W. Gerlach, O. Stern, *Z. Phys.* **1922**, *9*, 353.