Digitally Stimulated Raman Passage by Deep Reinforcement Learning

Iris Paparelle[a,b,d], Lorenzo Moro[a,c,d], Enrico Prati[a]

[a] *Istituto di Fotonica e Nanotecnologie, Consiglio Nazionale delle Ricerche, Piazza Leonardo da Vinci 32, 20133 Milano, Italy; enrico.prati@cnr.it*

[b] *ENS Paris Saclay, 61, avenue du Président Wilson 94235 Cachan Cedex, France*

[c] *Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Colombo 81, I-20133 Milano, Italy*

[d] *The authors equally contributed to this work*

Abstract

Preparing an arbitrary preselected coherent superposition of quantum states finds widespread application in physics, including initialization of trapped ion and superconductor qubits in quantum computers. Both fractional and integer stimulated Raman adiabatic passage involve smooth Gaussian pulses, designed to grant adiabaticity, so to keep the system in an eigenstate constituted only of the initial and final states. We explore an alternative method for discovering appropriate pulse sequences based on deep reinforcement learning algorithms and by imposing that the control laser can be only either on or off instead of being continuously amplitude-modulated. Despite the adiabatic condition is violated, we obtain fast and flexible solutions for both integer and fractional population transfer. Such method, consisting of a Digital Stimulated Raman Passage (D-STIRaP), proves to be particularly effective when the system is affected by dephasing therefore providing an alternative path towards control of noisy quantum states, like trapped ions and superconductor qubits.

Keywords: D-STIRaP, fractional D-STIRaP, deep reinforcement learning

1. Introduction

Stimulated Raman adiabatic passage (STIRAP [1, 2]) and its fractional version (f-STIRAP [3]) are well known methods for optical manipulation of atomic quantum states by laser pulses. Both are used in quantum information processing (QIP) for coherent control of trapped ions. STIRAP and f-STIRAP proved to be efficient tools for optical [4] and hyperfine [5] qubit manipulation [6] respectively.

STIRAP can flip a qubit between its base states, by allowing the population transfer between two discrete quantum states. In quantum information such method is applied for instance to $^{40}Ca^{+}$ ions trapped in a segmented linear Paul trap, with the Rabi frequencies $\Omega_{3D_{3/2}\longrightarrow 4P_{3/2_{max}}}$=$2\pi \cdot 100 MHz$ and $\Omega_{3D_{5/2}\longrightarrow 4P_{3/2_{max}}}$=$2\pi \cdot 250 MHz$ [4], used as optical qubits. Instead, fractional STIRAP can be advantageously used to create an excited state of the qubit insensitive to magnetic field fluctuations and whose coherence exceeds the coherence time of the bare state qubit by three orders of magnitude [6], by creating a coherent superposition of the two states. This technique was used in the experiment by Timoney et al. (2011) [5] to construct the excited state of a $^{171}Yb^{+}$ trapped ion hyperfine qubit. Furthermore, STIRAP has been applied to superconductive three levels systems [7], including strongly coupled systems [8].

Quantum systems can be controlled by machine learning, such as supervised learning in the determination of high-fidelity gates and the optimization of quantum memories by dynamical decoupling [9]. More recently Deep Reinforcement Learning (DRL) has been proposed to maintain a physical system in its equilibrium condition [10] and achieve a different equilibrium state, such as Coherent Transport by Adiabatic Passage (CTAP) [11, 12]. DRL algorithms can identify strategies for achieving a goal without prior knowledge of a system [13] and have been therefore chosen to short-circuit the analytical approach in this work.

The adiabatic condition is fundamental in the analytical Gaussian solution, but we exploit a pulse sequence where it may be violated. The DRL algorithm is asked to find digital on-off pulse sequences to ensure the population transfer in shorter times and satisfactory fidelity. Therefore, we will refer to integer and fractional Digital Stimulated Raman Passage (D-STIRaP) to indicate the complete transfer of the system from the initial state $|1\rangle$ to the final one $|3\rangle$

and the creation of a coherent superposition of $|1\rangle$ and $|3\rangle$ respectively. In the past, there are examples such as from Ref. [14] of pulses discretized to a number of possible values, for which the definition of Digital-STIRAP has already been used. Differently from such examples, here D- corresponds to the Digital values of only on (maximum value) and off.

STIRAP is mathematically equivalent to CTAP [15, 16, 17, 18, 19, 20] which has already been treated by using deep reinforcement learning [11, 12] and has inspired the current work. There, a three quantum dots solid state system is considered for the shuttling of a single spin across a quantum chip. Consistently, in the case of D-STIRaP the goal is to transfer amplitude from an initial state $|1\rangle$ to a final state $|3\rangle$, without involving a state $|2\rangle$ to which both are coupled, by using a sequence of two laser pulses: the Stokes pulse Ω_S, coupling $|2\rangle$ with $|3\rangle$ and the Pump pulse Ω_P, coupling $|1\rangle$ with $|2\rangle$ [3]. Similarly, in the case of fractional D-STIRaP, the goal is to create a superposition of $|1\rangle$ and $|3\rangle$ [3]. The fidelity of the protocol is defined as ρ_{33}, i.e. the probability density of being in $|3\rangle$.

The fundamental idea underlying (fractional-)STIRAP is to keep the system in the eigenstate associated to the null eigenvalue called "dark state", where $|2\rangle$ is not involved. An adiabatic evolution is therefore used to keep the system in such state, while the superposition of $|1\rangle$ with $|3\rangle$ can be controlled by the amplitude of the two pulses [3]. In this work, we demonstrate that DRL can obtain the same results despite the violation of the adiabatic condition. An example of so-called shortcut to adiabaticity for STIRAP processes is known [21], but it requires not trivial complex tunneling frequencies. Thefore, techniques from optimal control and machine learning have been recently considered important to explore [2].

Since speed and resilience to disturbances are fundamental properties in quantum computing, DRL is a natural option, as already proved in the case of CTAP [11]. In the following, in partial analogy with the case of continuously driven CTAP, we show that the digital pulses discovered by the DRL algorithm are able to achieve a satisfactory population transfer. D-STIRaP and fractional

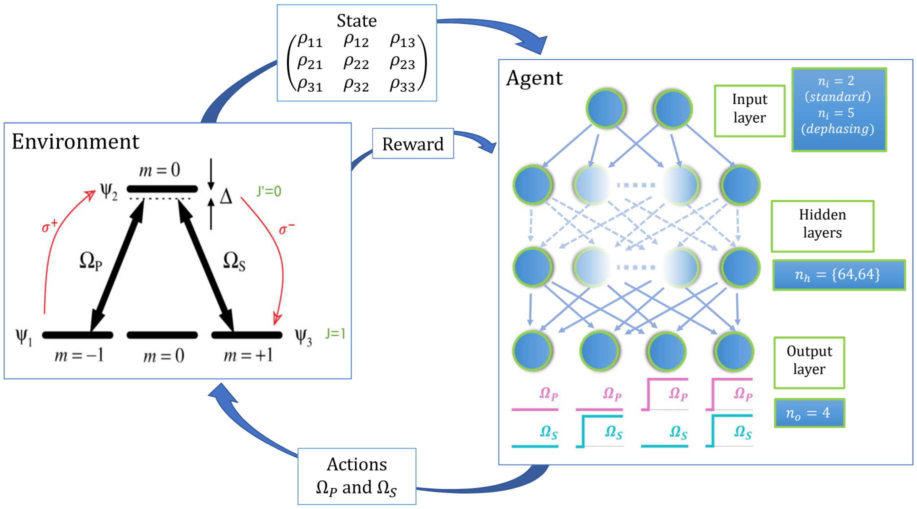


Figure 1: **Deep Reinforcement Learning architecture**: the *environment* corresponds to the physical system, while the *agent* consists of a deep neural network with two hidden layers of 64 neurons each. At each interaction (one time step) the agent observes the current state of the environment and it determines an action to take. After the environment has evolved into a new state, the reward function is returned to the agent. The deep neural network receives as input n_i modulus of elements of the density matrix. The highest activated neuron in the output layer n_o tells which lasers are on during the next time step.

D-STIRaP pulse sequences drive the system faster than the analytical solutions, but more importantly they can be adapted with no additional effort to detuning (no two-photon resonance [6]) and dephasing (phase relaxation [6]). In both cases above, there is no analytical approach to replace Gaussian pulses.

The implementation of DRL is based on an agent-environment system as summarized in Figure 1. More specifically the Proximal Policy Optimization (PPO) algorithm has been chosen.

In the Methods Section, the physical system and the DRL algorithm are explained. The Results and Discussion Section concerns the simulation of integer and fractional D-STIRaP, including the case of evolution affected by disturbances as controlled by the DRL. The last Section summarize the Conclusions.

2. Methods

2.1. The Hamiltonian formulation of STIRAP

STIRAP has been extensively discussed elsewhere [3, 11]. It is a three states system, where only one of them is coupled with both the others and their coupling can be externally controlled. The time evolution of the density matrix $\rho(t)$ is governed by a master equation:

$$i\hbar\frac{d}{dt}\rho(t) = [H(t), \rho(t)] \quad ; \quad H(t) = \hbar\begin{pmatrix} 0 & \frac{1}{2}\Omega_P(t) & 0 \\ \frac{1}{2}\Omega_P(t) & \Delta & \frac{1}{2}\Omega_S(t) \\ 0 & \frac{1}{2}\Omega_S(t) & \delta \end{pmatrix} \qquad (1)$$

where $\hbar\Delta = E_2 - E_1 - \hbar\omega_P$ and $\hbar\delta = E_3 - E_1 - \hbar\omega_P + \hbar\omega_S$, while ω_P and ω_S are the carrier frequencies. Δ and δ are called one-photon detuning and two-photon detuning respectively.

A non vanishing value of Δ does not affect the population transfer. Instead, in the ideal case δ must be null in order to ensure the availability of the "dark state" [6] which is expressed in terms of $|1\rangle$ and $|3\rangle$, but not of $|2\rangle$. In this case, it is possible to apply successful analytical solutions to achieve (f-)STIRAP.

However, detuning can occur due to mismatches between system energy differences (Bohr transition frequencies) and photon energies (field carrier frequencies) [6]. In Subsection 3.3 the case of non-null detuning is treated.

In turn, the coupling with the environment is a source of decoherence. In the case of D-STIRaP and fractional D-STIRaP, both dephasing and phase relaxation can significantly affect the system. The treatment of this disturbance requires solution of the Liouville equation for the density matrix [6]:

$$i\hbar\frac{d}{dt}\rho(t) = [H(t), \rho(t)] - iD(t) \quad ; \quad D(t) = \hbar\begin{pmatrix} 0 & \gamma_{12}\rho_{12}(t) & \gamma_{13}\rho_{13}(t) \\ \gamma_{12}\rho_{21}(t) & 0 & \gamma_{23}\rho_{23}(t) \\ \gamma_{13}\rho_{31}(t) & \gamma_{23}\rho_{32}(t) & 0 \end{pmatrix} \qquad (2)$$

In the following, we postulate the approximation that $\gamma_{nm} = \Gamma \; \forall (n, m)$.

Dephasing affects STIRAP by destroying the coherence between states $|1\rangle$ and $|3\rangle$ and thus it leads to the depopulation of the "dark state" [6] (see Subsection 3.3). In order to exploit the Hamiltonian framework as an environment for a deep learning agent, the time evolution of the system has been implemented via the QuTiP [22, 23] library.

2.2. The deep reinforcement learning architecture for solving D-STIRaP

Deep Reinforcement Learning (DRL) is a set of techniques that exploit artificial neural networks to learn behavior in sequential decision-making problems [24, 25, 26]. Such techniques are highly effective when no prior knowledge about the dynamics of the system is available or when the control problem is too complex to be addressed by classical optimal-control algorithms.

The basic principles and notions in the theory of reinforcement learning are based upon the idea of interactions between a decision maker called *agent* and a controlled system named *environment*. The latter is everything that cannot be controlled directly by the former. In the case of D-STIRaP, the environment corresponds to the physical system, involving the Hamiltonian and the density matrix.

The agent and the environment interact at discrete time steps as shown in Figure 1. At every time step, the agent observes the current state of the environment and it performs actions according to a policy function that fully determines its behavior. Therefore, the environment evolves changing its state and a reward signal is returned to the agent to update its policy. The fundamental units of the learning process are the episodes, during which the agent tries to maximize the sum of the rewards obtained at each time-step, i.e. the *cumulative reward*, changing its policy. Episodes are interrupted when an ending condition is reached. It happens when either the number of steps exceeds a quantity defined by the user, or the agent learns how to solve the problem. Figure 2 shows the cumulative rewards earned during the learning process in the case of D-STIRaP without disturbances.

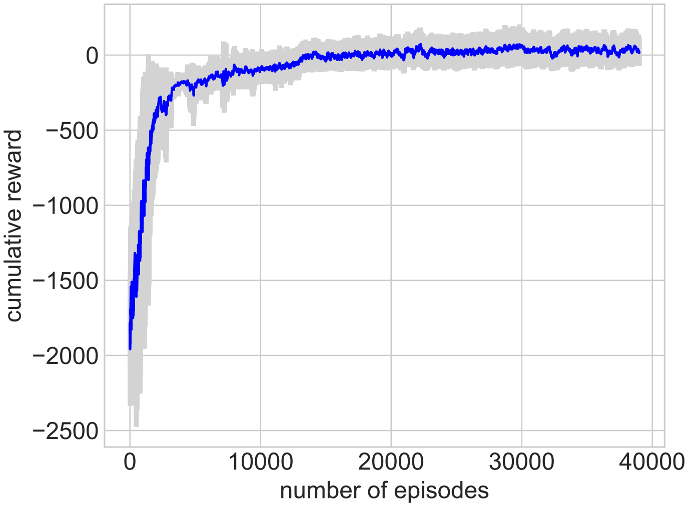


Figure 2: **Cumulative reward evolution during a learning process** in the case of D-STIRaP without disturbances. A moving average on a 30 episodes window (horizontal average) was previously done on 8 independent runs to reduce the noisy effect. The blue line represents the average (vertical) of these smoothed cumulative rewards and the grey line represent its standard deviation. The learning process effectively succeeds after about 25 000 episodes, even if the major learning breakthroughs are done in the first 3000 episodes.

STIRAP is considered as a textbook technique in quantum control [2]. If we consider deep neural networks as discrete-time nonlinear dynamical systems employed as an optimization algorithms, the training processes required when a controller is included can be formulated as an optimal control problem [27].

This kind of approach based on a DRL control system has successfully been exploited in a recent work [11] for the CTAP environment. Such work have inspired some key elements of the learning architecture, due to the similarity between CTAP and STIRAP, such as the shape of the reward function and the state of the environment as described in the following.

The time given to solve the problem t_{max} and the maximum number of steps $nmax_{step}$ have an impact on the learning process. The former should be neither too short, as DRL will not be able to solve the problems, nor too long, as it would slow down the learning procedure. The most effective hyperparameters used for the learning process have been summarized in Table 1.

At each time step, the agent chooses and applies one between four distinct actions. Such actions correspond to four possible combinations generated by having on and off the two lasers, i.e. the Pump pulse $\Omega_P(t)$ and the Stokes pulse $\Omega_S(t)$ respectively. At each time step, of duration $t_{max}/nmax_{step}$, both amplitudes can take only the values 0 or $\Omega_0 \propto 1/t_{max}$. Ω_0 affects the speed of the population transfer from state $|1\rangle$ to state $|3\rangle$, as shown in Figure 3.

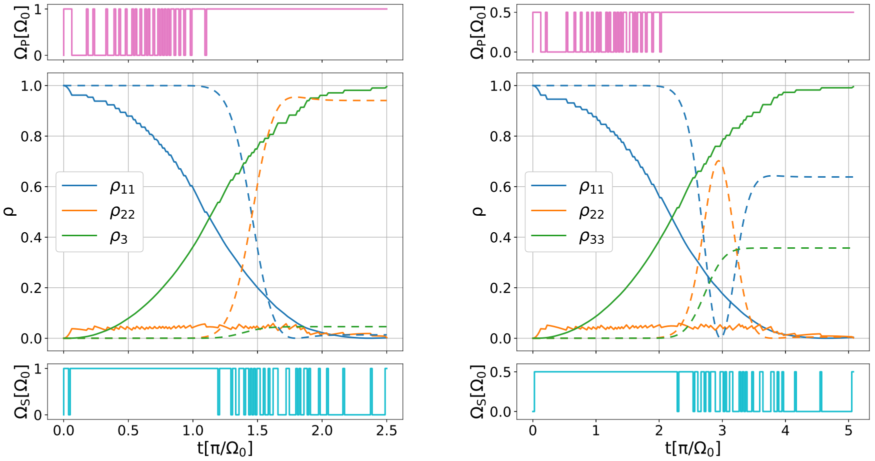
The agent receives the modulus of the elements of the density matrix and the values of the pulse at every time step. As already proven [11], taking separately the real and the imaginary part of the off-diagonal elements is not needed. Furthermore, in the ideal case the information carried by the values of ρ_{33} and ρ_{22} are sufficient. In the dephasing case the upper non diagonal terms are added as input, i.e. $|\rho_{12}|$, $|\rho_{13}|$ and $|\rho_{23}|$.

The reward function $r(t)$ used in this work has been inspired by [11] and has the following expression:

$$r(t) = f\Big(\rho_{11}(t), \rho_{22}(t), \rho_{33}(t)\Big) + A(t) + B(t) \tag{3}$$

where f is a real function which has a global maximum if $\rho_{22} = 0$ and ρ_{33}, ρ_{11} are equal to the aimed final population. $A(t)$ and $B(t)$ are additional punishing or rewarding terms which may intervene when the state of the system becomes too critical. We want the agent to find the sequence of pulses that keeps ρ_{22} close to 0, while transforming (ρ_{33}, ρ_{11}) from (0,1) to the aimed values.

In the framework of deep reinforcement learning, the policy followed by the agent is approximated by an artificial neural network. In this work, we utilized a feed-forward neural network with two hidden layers of 64 neurons each for both fractional and integer D-STIRaP. No other network architectures have been investigated since this neural network is relatively small in size and has achieved excellent results in all the actual simulations. Therefore, we exploit the Proximal Policy Optimization (PPO) algorithm [28] to train agents capable of finding digital pulses sequences to achieve both integer and fractional population transfer between states. PPO has been chosen in favour of Trusted



(a) Digital pulses $\in (0, \Omega_0)$

(b) Digital pulses $\in (0, \Omega_0/2)$

Figure 3: **Comparison between D-STIRaP and Gaussian pulses for fast transfer times.** While Gaussian pulses (dotted lines) heavily fail to transfer the population if the time interval is significantly below 50-60π/Ω_{max}, digital pulses successfully achieve the transfer at the cost of a small non-zero probability of occupying the state $|2\rangle$ of the order of a percent. The probability density of being in $|1\rangle$, $|2\rangle$ and $|3\rangle$ are plotted in blue, orange and green respectively, while the Stokes and Pump pulses are represented in pink and cyan.

Region Policy Optimization (TRPO) exploited in [11], due to several reasons: PPO performs comparably or better than state-of-the-art approaches [28] while being much simpler to implement and tune. Moreover, it allows using multiple agents at the same time, significantly decreasing the training time compared to TRPO.

The D-STIRaP environment has been implemented using Python 3.7.0 language, while we exploited the Stable-Baseline 2.6.0 Python module [29] to implement the deep reinforcement agent. All the hyperparameters used by PPO algorithm are the default ones (see Stable-Baseline documentation for further details).

3. Results and Discussion

3.1. *Control of integer D-STIRaP*

In the case of integer D-STIRaP we exploited the reward function proposed for the CTAP environment [11]:

$$r(t) = -1 - \rho_{22}(t) + \rho_{33}(t) + A(t) + B(t) \quad \text{where} \quad \begin{cases} A(t) = 10 \cdot H(\rho_{33}(t) - 0.97) \\ B(t) = -e^{6\rho_{22}} \cdot H(\rho_{22}(t) - 0.05) \end{cases} \tag{4}$$

where the ending conditions are:

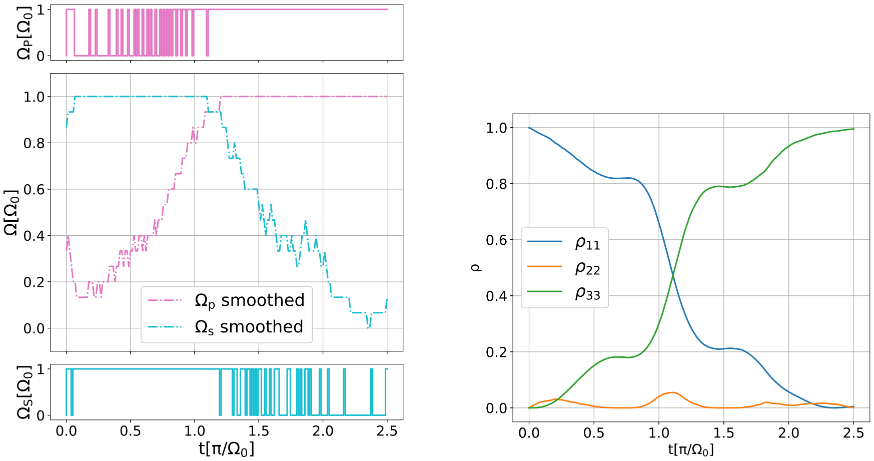
- the number of steps exceeds an arbitrary limit of time steps $nmax_{step}$;
- ρ_{33} exceeds 0.995. In such case an additional positive reward of 100 is given.

The results obtained by the DRL agent with no disturbances ($\Delta = \delta = \Gamma = 0$) are shown in Figure 3. The pulse sequences found by DRL achieve the population transfer in roughly $2.5 \cdot \pi/\Omega_0$, reaching a fidelity of 99.5% and keeping ρ_{22} close to 0.05 before dropping to zero. Such solution is approximately 20 times faster than the analytical one consisting of two Gaussian pulses in a counterintuitive sequence [6, 3], where the transfer is achieved after a time of $160/\Omega_0$ only.

The digital pulses found by the agent are still conceptually close to both the Gaussian [3] and to the DRL-based [11] continuous pulses, as shown by their moving average in Figure 4. In order to intuitively relate the successful drive achieved by the digital pulses, we show that they act as if their average is responsible of the effect on a longer timescale. In fact, they not only resemble the counterintuitive sequence, i.e. the Stokes pulse starts before the Pump pulse, but they are capable to induce the STIRAP.

3.2. *Control of fractional D-STIRaP*

In the case of fractional D-STIRaP a generalized reward function is required to take into account the final ratio between the populations according to a



(a) Digital pulses $\in \{0, \Omega_0\}$ and their moving average (15 values window).

(b) Probability density of being in state $|1\rangle$, $|2\rangle$ and $|3\rangle$.

Figure 4: **Smoothing of D-STIRaP pulses to recover STIRAP**. Solid lines in (a) represent the Stokes and Pump digital pulses and dashed lines its moving average. The counterintuitive sequence can be recognized. Such pulses achieve STIRAP with a 99.5% fidelity as shown in (b).

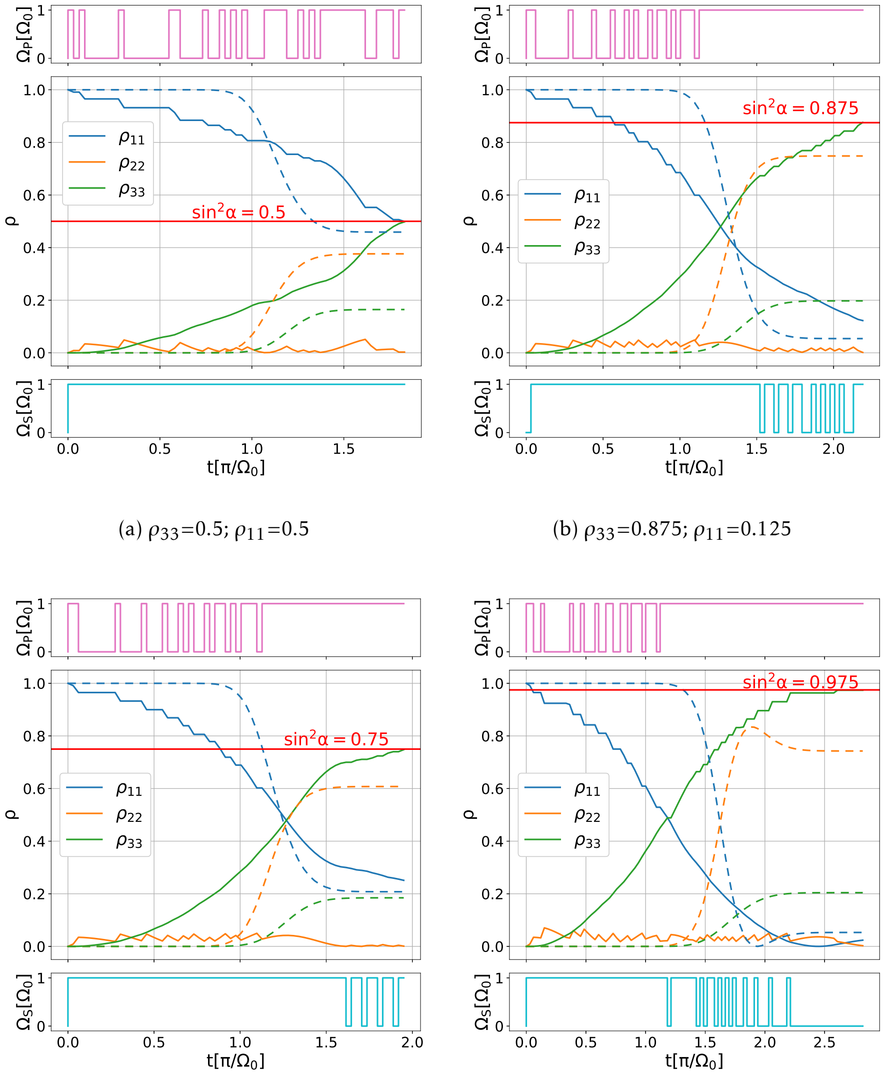
parameter. The reward function is defined as:

$$r(t) = -1 - (\rho_{11}(t) - cos^2\alpha)^2 - (\rho_{33}(t) - sin^2\alpha)^2 + A(t) + B(t) \tag{5}$$

where α is the parameter controlling the ratio of ρ_{11} and ρ_{33} [3], $A(t) = 0$ and $B(t) = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$. The ending conditions are:

- the number of steps exceeds an arbitrary limit of time steps $nmax_{step}$;
- $\left|\rho_{33} - sin^2\alpha\right| < \varepsilon$ and $\left|\rho_{11} - cos^2\alpha\right| < \varepsilon$ with $\varepsilon = 0.25\%$.

Figure 5 reports the results obtained with no disturbances ($\Delta = \delta = 0$) for four different angles α. The agent was able to achieve a precision of 0.25% on (ρ_{11}, ρ_{33}) and to keep ρ_{22} close to 0.05. Similarly to the integer D-STIRaP, the DRL pulse sequences reach the desired coherent superposition of states $|1\rangle$ and $|3\rangle$ in approximately $2\pi/\Omega_0$, outperforming standard pulses.



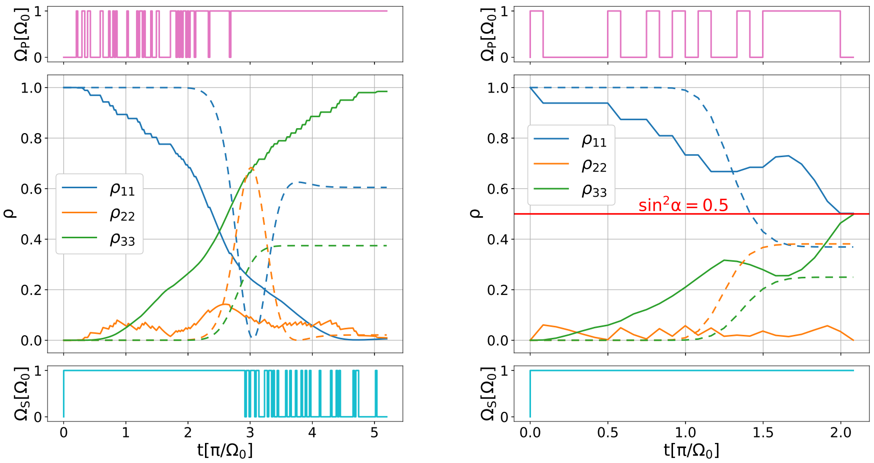
(a) ρ_{33}=0.5; ρ_{11}=0.5

(b) ρ_{33}=0.875; ρ_{11}=0.125

(c) ρ_{33}=0.75; ρ_{11}=0.25

(d) ρ_{33}=0.975; ρ_{11}=0.025

Figure 5: **Comparison of fractional D-STIRaP without disturbances** achieved by Gaussian pulses (dashed lines) versus pulses found by DRL algorithm (solid lines) for different values of α. The goal is to reach the desired population of $\rho_{11} = cos^2\alpha$ and $\rho_{33} = sin^2\alpha$ represented by the solid red line. Despite allowing a maximum time t_{max} of $6\pi/\Omega_0$, the DRL managed to reach the goal in shorter times. In such time the analytical pulses do not achieve D-STIRaP, while the DRL pulses are able to maintain ρ_{22} close to 0 and to transfer the state from $|1\rangle$ to the desired coherent superposition of $|1\rangle$ and $|3\rangle$.



(a) D-STIRaP with detuning $\Delta = 0.15\Omega_0$ and $\delta = 0.15\Omega_0$

(b) Fractional D-STIRaP with $\alpha = \pi/4$ and $\Delta = 0.15\Omega_0$ and $\delta = 0.15\Omega_0$

Figure 6: **Comparison of D-STIRaP and fractional D-STIRaP with detuning** ($\alpha = \pi/4$) achieved by Gaussian pulses (dashed lines) vs. pulses found by DRL algorithm (solid lines). The maximum time t_{max} allowed was $10\pi/\Omega_0$ for D-STIRaP and $8\pi/\Omega_0$ for fractional D-STIRaP, but DRL managed to reach the goal in shorter time. In such time the analytical pulses do not achieve D-STIRaP, while the DRL pulses succeed in transferring amplitude from state $|1\rangle$ to the desired coherent superposition of $|1\rangle$ and $|3\rangle$.

3.3. Resilience to disturbances

When sources of disturbance act on the quantum system, the known pulse control sequence is inadequate to preserve the transfer. Moreover, no analytical methods to derive suitable sequences are known. However, deep reinforcement algorithms such PPO can be exploited to address such cases. Table 1 summarizes the results obtained by the agent, including the cases of detuning and dephasing, fractional D-STIRaP and their reward functions.

3.3.1. Detuning

The population transfer as a function of time in case of detuning ($\Delta = \delta = 0.15 \cdot \Omega_0$) is shown in Figure 6. DRL achieves (fractional)D-STIRaP without being perturbed by two-photon detuning in a time of $4\pi/\Omega_0$, which is comparable to the unperturbed case.

3.3.2. Dephasing

Figure 7 shows the population transfer in the case of dephasing ($\Delta = \delta = 0$). Two regimes, consisting of a weak ($\Gamma = 0.01\Omega_0$) and a strong ($\Gamma = 0.1\Omega_0$) dephasing, are displayed.

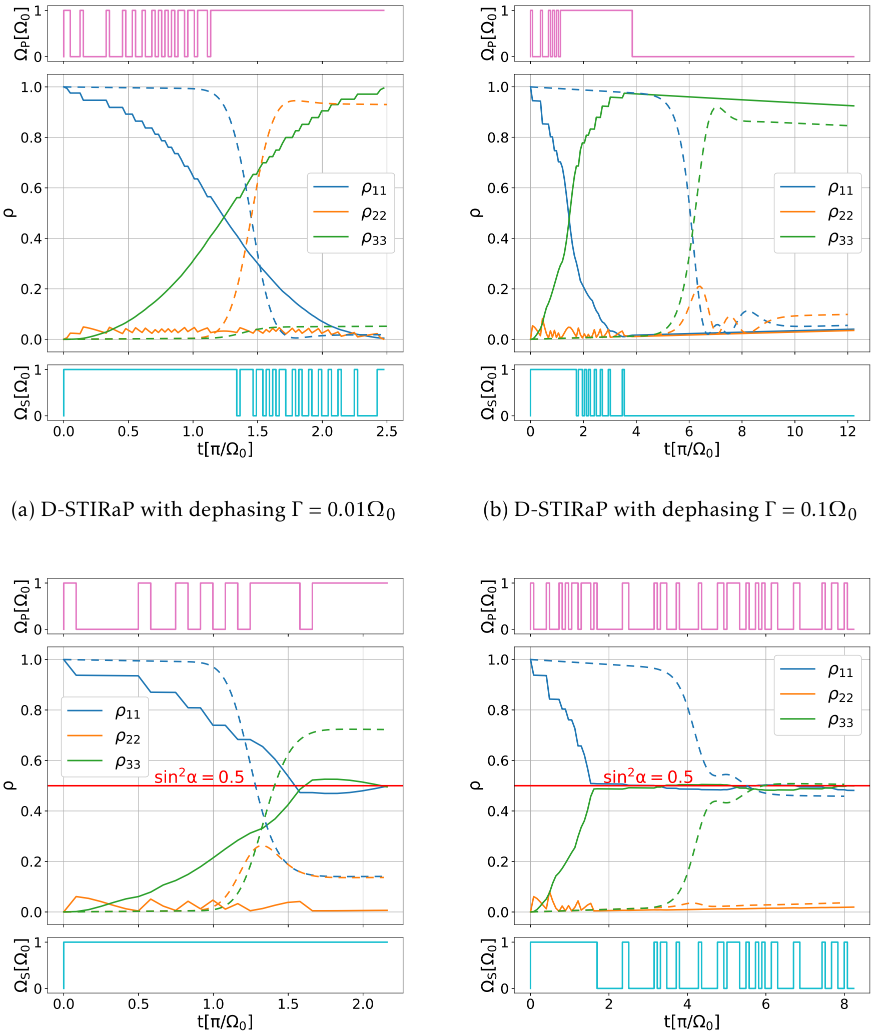
In such case, the non-diagonal terms of the density matrix are non-null. Therefore, $\left|\rho_{13}\right|$, $\left|\rho_{12}\right|$ and $\left|\rho_{23}\right|$ were added to ρ_{22} and ρ_{33} as inputs of the neural network. No additional terms were considered, since the density matrix is Hermitian and the non-diagonal terms belong to $\mathbb{R}$ or to $\mathbb{C} \setminus \mathbb{R}$.

In the case of weak dephasing ($\Gamma = 0.01\Omega_0$), DRL achieves both D-STIRaP and fractional D-STIRaP, bringing ρ_{33} to the desired proportion with an accuracy of 0.25% while keeping ρ_{22} close to 0. The transfer is reached faster than the analytical approach without disturbances.

When a strong dephasing ($\Gamma = 0.1\Omega_0$) occurs, it is possible to appreciate the long-term effect of the disturbance, i.e. the populations start to converge towards $\rho_{11} = \rho_{22} = \rho_{33} = 1/3$. The DRL algorithm finds pulse sequences that reach a fidelity of 97%, higher than Gaussian pulses, which achieve a fidelity of 91.5% only. Moreover, the probability of being in state $|2\rangle$ during the transfer is lower than using Gaussian pulses. In fact, the former is $0.28 \cdot \pi/\Omega_0$, while the latter is approximately $0.62 \cdot \pi/\Omega_0$. The fact that the population is inverted before stabilizing to the final value is an accidental feature which may or may not happen depending on the time the training is stopped.

3.4. General discussion

The deep reinforcement learning agent learns suitable pulse sequences that are able to achieve (fractional)D-STIRaP. Ideally we want the transfer to be as fast as possible and to keep the state $|2\rangle$ not populated. The latter condition is especially relevant in the case of Zeeman qubit, where the state $|2\rangle$ can naturally decay failing the transfer. However, it is preferable a sequence of pulses



(a) D-STIRaP with dephasing $\Gamma = 0.01\Omega_0$

(b) D-STIRaP with dephasing $\Gamma = 0.1\Omega_0$

(c) Fractional D-STIRaP with $\alpha = \pi/4$ and dephasing $\Gamma = 0.01\Omega_0$

(d) Fractional D-STIRaP with $\alpha = \pi/4$ and dephasing $\Gamma = 0.1\Omega_0$

Figure 7: **Comparison of D-STIRaP and fractional D-STIRaP** ($\alpha = \pi/4$) with strong and weak dephasing achieved by Gaussian pulses (dotted lines) vs. pulses found by DRL algorithm (solid lines). In case of strong dephasing, DRL did not manage to satisfy the strict conditions imposed ($\rho_{33} \leq 0.995$ and a margin of ϵ=0.5% for fractional D-STIRaP) but it was anyway able to find better solutions than the analytical approach.

Environment	Input	Reward	Additional terms	Ending condition	$\mathbf{t_{max}}$ $\mathbf{nmax_{step}}$
Standard D-STIRaP (Fig. 3)	ρ_{22}, ρ_{33}	$-1 - \rho_{22}(t) + \rho_{33}(t)$	$A = 10 \cdot H(\rho_{33}(t) - 0.97)$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\rho_{33} > 0.995$	$2.5\pi/\Omega_0$ 200
Standard D-STIRaP Detuning (Fig. 6)	ρ_{22}, ρ_{33}	$-1 - \rho_{22}(t) + \rho_{33}(t)$	$A = 10 \cdot H(\rho_{33}(t) - 0.97)$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\rho_{33} > 0.995$	$9\pi/\Omega_0$ 400
Standard D-STIRaP Weak Dephasing (Fig. 7)	$\rho_{22}, \rho_{33}, \lvert\rho_{12}\rvert, \lvert\rho_{13}\rvert, \lvert\rho_{23}\rvert$	$-1 - \rho_{22}(t) + \rho_{33}(t)$	$A = 0$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\rho_{33} > 0.995$	$5\pi/\Omega_0$ 200
Standard D-STIRaP Strong Dephasing (Fig. 7)	$\rho_{22}, \rho_{33}, \lvert\rho_{12}\rvert, \lvert\rho_{13}\rvert, \lvert\rho_{23}\rvert$	$-1 - \rho_{22}(t) + \rho_{33}(t)$	$A = 0$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\rho_{33} > 0.995$	$12\pi/\Omega_0$ 160
Fractional D-STIRaP (Fig. 3)	ρ_{22}, ρ_{33}	$-1 - (\rho_{11}(t) - cos^2\alpha)^2 - (\rho_{33}(t) - sin^2\alpha)^2$	$A = 0$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\lvert\rho_{33} - sin^2\alpha\rvert < \varepsilon$ $\lvert\rho_{11} - cos^2\alpha\rvert < \varepsilon$ with $\varepsilon = 0.0025$	$3\pi/\Omega_0$ 100
Fractional D-STIRaP Detuning (Fig. 6)	ρ_{22}, ρ_{33}	$-1 - (\rho_{11}(t) - cos^2\alpha)^2 - (\rho_{33}(t) - sin^2\alpha)^2$	$A = 0$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\lvert\rho_{33} - sin^2\alpha\rvert < \varepsilon$ $\lvert\rho_{11} - cos^2\alpha\rvert < \varepsilon$ with $\varepsilon = 0.0025$	$4\pi/\Omega_0$ 100
Fractional D-STIRaP Strong and Weak dephasing (Fig. 7)	$\rho_{22}, \rho_{33}, \lvert\rho_{12}\rvert, \lvert\rho_{13}\rvert, \lvert\rho_{23}\rvert$	$-1 - (\rho_{11}(t) - cos^2\alpha)^2 - (\rho_{33}(t) - sin^2\alpha)^2$	$A = 0$ $B = -e^{6\rho_{22}(t)} \cdot H(\rho_{22}(t) - 0.05)$	$\lvert\rho_{33} - sin^2\alpha\rvert < \varepsilon$ $\lvert\rho_{11} - cos^2\alpha\rvert < \varepsilon$ with $\varepsilon = 0.0025$	$8\pi/\Omega_0$ 100

Table 1: Summary of the main parameters used to train the agent for each D-STIRaP environment. The ending condition *number of step exceed* $nmax_{step}$ is not reported because it is common to all environments. An additional reward of +100 is returned if the episode ends before exceeding $nmax_{step}$ steps.

that achieve a faster population transfer even at the cost of slightly populating the state $|2\rangle$.

It is worth noting that the Gaussian pulses are able to achieve the transfer maintaining ρ_{22} closer to 0 than DRL pulses, if they are applied with a longer t_{max} [3], e.g. of $40\pi/\Omega_0$. Unfortunately, in real systems such ideal condition can not be precisely met due to the disturbances that affect the transfer. For instance, when a dephasing source acts during the process, the result of DRL can be significantly better despite the small non vanishing occupation of $|2\rangle$.

The trade-off between occupation of ρ_{22} and the transition time is irrelevant, when considering hyperfine qubits, because $|2\rangle$ is not a decaying state.

In the context of an actual implementation of a (fractional)D-STIRaP guided by DRL, it is not necessary to know the density matrix at every time-step. The main idea is to extract Γ experimentally from the system beforehand, then run a simulation to get the sequences of pulses and finally applying them to the actual experiment.

While managing digital pulses required for the control of superconductive qubits looks straightforward, one may wonder if the time scale required to switch on/off lasers is compatible with the control trapped ions. If we consider gate operations of the order of tens to hundred of microseconds (see for instance Ref. [30] related to $^{40}\mathrm{Ca}^{+}$), such digital control is made possible by modern electro-acoustic modulators which have rise and fall time of the order of 2 ns. Superconductive qubits such as transmons are generally operated under global adiabatic condition as it shows that the transfer efficiency can be improved by making the pulses longer, which is in turn limited by its decoherence [7]. Violating the adiabatic condition and achieveing therefore significantly faster pulse sequences limits the time during which such decohernce occurs. Dephasing may arise differently depending on the kind of superconductive qubit and is less critical but still present. For instance, the dephasing is minimized for the Quantronium in correspondance of the symmetry point but a selection rule prevents to implement the STIRAP there [31]. The solution is to detune away from such symmetry point, at the cost of introducing some unavoidable degree of dephasing. For what concerns trapped ions qubits, dephasing is one of the three ion trap specific error types, together with overrotation and crosstalk [32]. Here again the use of D-STIRaP shortens the population transfer time therefore reducing the global effect acting because of dephasing. Our method can be extended to other kind of disturbance such as population decay set by a finite T_1 time as already discussed in Ref. [11].

4. Conclusions

We propose a population transfer pulsing method alternative to STIRAP, which relies on a Deep Reinforcement Learning (DRL) agent to find very fast digital pulse sequences. We refer to this method as integer or fractional Digital Stimulated Raman Passage (D-STIRaP).

Although the condition of adiabaticity is violated, the population transfer is successful and faster than using continuous amplitude-modulated pulses. In fact, DRL finds sequences of two-valued laser amplitude intensities corresponding to turning on and off two laser-beams, achieving population transfer within 0.5% accuracy significantly faster than the analytical approach. More importantly, DRL is able to find sequences even in the presence of disturbances such as detuning and dephasing. In the latter case, D-STIRaP is highly more efficient than continuously STIRAP.

More generally, artificial intelligence [33] can be applied successfully in quantum information processing and in particular reinforcement learning in the field of trapped ions and superconducting quantum computing, where both D-STIRaP and fractional D-STIRaP can be used to manipulate quantum states.

[1] A. F. Kockum, A. Miranowicz, V. Macrì, S. Savasta, F. Nori, Deterministic quantum nonlinear optics with single atoms and virtual photons, Physical Review A 95 (6) (2017) 063849.

[2] K. Bergmann, H.-C. Nägerl, C. Panda, G. Gabrielse, E. Miloglyadov, M. Quack, G. Seyfang, G. Wichmann, S. Ospelkaus, A. Kuhn, et al.,

Roadmap on STIRAP applications, Journal of Physics B: Atomic, Molecular and Optical Physics 52 (20) (2019) 202001.

[3] N. Vitanov, K. Suominen, B. Shore, Creation of coherent atomic superpositions by fractional stimulated Raman adiabatic passage, Journal of Physics B: Atomic, Molecular and Optical Physics 32 (18) (1999) 4535.

[4] J. L. Sørensen, D. Møller, T. Iversen, J. B. Thomsen, F. Jensen, P. Staanum, D. Voigt, M. Drewsen, Efficient coherent internal state transfer in trapped ions using stimulated Raman adiabatic passage, New Journal of Physics 8 (11) (2006) 261.

[5] N. Timoney, I. Baumgart, M. Johanning, A. Varón, M. B. Plenio, A. Retzker, C. Wunderlich, Quantum gates and memory using microwave-dressed states, Nature 476 (7359) (2011) 185.

[6] N. V. Vitanov, A. A. Rangelov, B. W. Shore, K. Bergmann, Stimulated Raman adiabatic passage in physics, chemistry, and beyond, Reviews of Modern Physics 89 (1) (2017) 015006.

[7] K. Kumar, A. Vepsäläinen, S. Danilin, G. Paraoanu, Stimulated Raman adiabatic passage in a three-level superconducting circuit, Nature communications 7 (2016) 10628.

[8] M. Stramacchia, A. Ridolfo, G. Benenti, E. Paladino, F. Pellegrino, G. Maccarrone, G. Falci, Speedup of adiabatic multiqubit state-transfer by ultrastrong coupling of matter and radiation, arXiv preprint arXiv:1904.04141.

[9] M. August, X. Ni, Using recurrent neural networks to optimize dynamical decoupling for quantum memory, Physical Review A 95 (1) (2017) 012335.

[10] T. Fösel, P. Tighineanu, T. Weiss, F. Marquardt, Reinforcement learning with neural networks for quantum feedback, Physical Review X 8 (3) (2018) 031084.

[11] R. Porotti, D. Tamascelli, M. Restelli, E. Prati, Coherent transport of quantum states by deep reinforcement learning, Communications Physics 2 (1) (2019) 61.

[12] R. Porotti, D. Tamascelli, M. Restelli, E. Prati, Reinforcement learning based control of coherent transport by adiabatic passage of spin qubits, in: Journal of Physics: Conference Series, Vol. 1275, IOP Publishing, 2019, p. 012019.

[13] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[14] J. A. Vaitkus, A. D. Greentree, Digital three-state adiabatic passage, Physical Review A 87 (6) (2013) 063820.

[15] A. D. Greentree, J. H. Cole, A. Hamilton, L. C. Hollenberg, Coherent electronic transfer in quantum dot systems using adiabatic passage, Physical Review B 70 (23) (2004) 235317.

[16] J. H. Cole, A. D. Greentree, L. Hollenberg, S. D. Sarma, Spatial adiabatic passage in a realistic triple well structure, Physical Review B 77 (23) (2008) 235418.

[17] E. Ferraro, M. De Michielis, M. Fanciulli, E. Prati, Coherent tunneling by adiabatic passage of an exchange-only spin qubit in a double quantum dot chain, Physical Review B 91 (7) (2015) 075435.

[18] E. Prati, T. Shinada, Single-atom nanoelectronics, CRC Press, 2013.

[19] E. Prati, A. Morello, Quantum information in silicon devices based on individual dopants, Single-Atom Nanoelectronics (2013) 5–39.

[20] D. Rotta, F. Sebastiano, E. Charbon, E. Prati, Quantum information density scaling and qubit operation time constraints of CMOS silicon-based quantum computer architectures, npj Quantum Information 3 (1) (2017) 26.

[21] X. Chen, I. Lizuain, A. Ruschhaupt, D. Guéry-Odelin, J. Muga, Shortcut to adiabatic passage in two-and three-level atoms, Physical review letters 105 (12) (2010) 123003.

[22] J. R. Johansson, P. Nation, F. Nori, QuTiP: An open-source Python framework for the dynamics of open quantum systems, Computer Physics Communications 183 (8) (2012) 1760–1772.

[23] N. Shammah, S. Ahmed, N. Lambert, S. De Liberato, F. Nori, Open quantum systems with local and collective incoherent processes: Efficient numerical simulations using permutational invariance, Physical Review A 98 (6) (2018) 063815.

[24] S. Tognetti, S. M. Savaresi, C. Spelta, M. Restelli, Batch reinforcement learning for semi-active suspension control, in: 2009 IEEE Control Applications,(CCA) & Intelligent Control,(ISIC), IEEE, 2009, pp. 582–587.

[25] M. Y. Niu, S. Boixo, V. N. Smelyanskiy, H. Neven, Universal quantum control through deep reinforcement learning, in: AIAA Scitech 2019 Forum, 2019, p. 0954.

[26] A. Castelletti, F. Pianosi, M. Restelli, A multiobjective reinforcement learning approach to water resources systems operation: Pareto frontier approximation in a single run, Water Resources Research 49 (6) (2013) 3476–3486.

[27] G.-H. Liu, E. A. Theodorou, Deep learning theory review: An optimal control and dynamical systems perspective, arXiv preprint arXiv:1908.10920.

[28] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347.

[29] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford,

J. Schulman, S. Sidor, Y. Wu, Stable baselines, `https://github.com/hill-a/stable-baselines` (2018).

[30] F. Schmidt-Kaler, H. Häffner, M. Riebe, S. Gulde, G. P. Lancaster, T. Deuschle, C. Becher, C. F. Roos, J. Eschner, R. Blatt, Realization of the Cirac–Zoller controlled-NOT quantum gate, Nature 422 (6930) (2003) 408.

[31] A. La Cognata, P. Caldara, D. Valenti, B. Spagnolo, A. D'ARRIGO, E. Paladino, G. Falci, Effect of low-frequency noise on adiabatic passage in a superconducting nanocircuit, International Journal of Quantum Information 9 (supp01) (2011) 1–15.

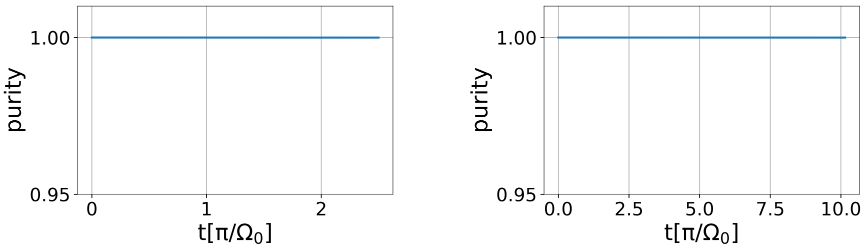
[32] D. M. Debroy, M. Li, S. Huang, K. R. Brown, Logical performance of 9 qubit compass codes in ion traps with crosstalk errors, arXiv preprint arXiv:1910.08495.

[33] E. Prati, Quantum neuromorphic hardware for quantum artificial intelligence, in: Journal of Physics: Conference Series, Vol. 880, IOP Publishing, 2017, p. 012018.

Appendices

Appendix A. Purity of the states

This Appendix is devoted to quantitatively account for the time evolution of the purity i.e. $\mathrm{Tr}\,|\rho|^2$. The purity is traced as function of time for all the simulations discussed throughout the Sections.



(a) Digital pulses $\in (0, \Omega_0)$

(b) Digital pulses $\in (0, \Omega_0/2)$

Figure A.8: Purity of the states as function of time in the case of D-STIRAP as represented in Figure 3.

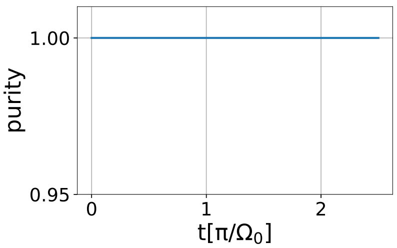
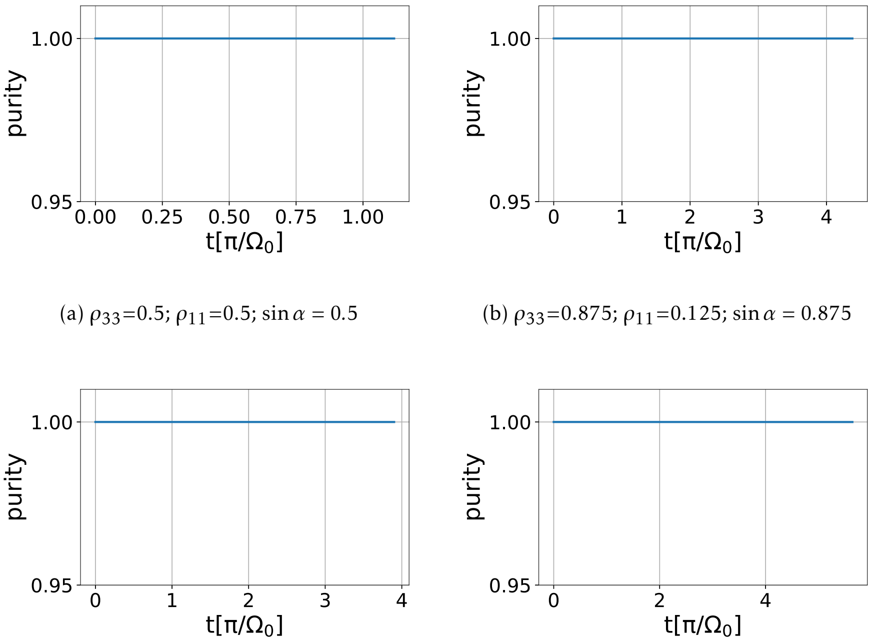


Figure A.9: Purity of the states as function of time of time in the case of D-STIRAP as represented in Figure 4b.



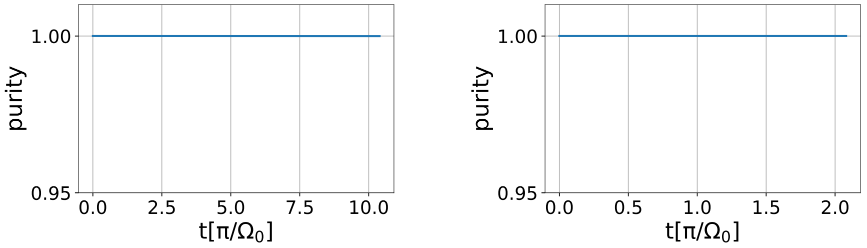
(a) ρ_{33}=0.5; ρ_{11}=0.5; $\sin\alpha = 0.5$

(b) ρ_{33}=0.875; ρ_{11}=0.125; $\sin\alpha = 0.875$

(c) ρ_{33}=0.75; ρ_{11}=0.25; $\sin\alpha = 0.75$

(d) ρ_{33}=0.975; ρ_{11}=0.025; $\sin\alpha = 0.975$

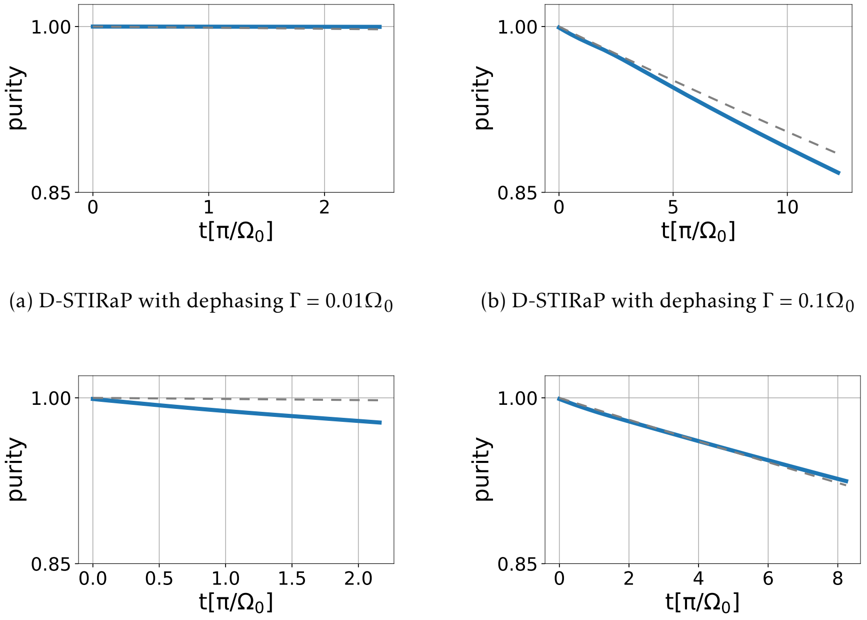
Figure A.10: Purity of the states as function of time in the case of fractional D-STIRAP as represented in Figure 5.



(a) D-STIRaP with detuning $\Delta = 0.15\Omega_0$ and $\delta = 0.15\Omega_0$

(b) Fractional D-STIRaP with $\alpha = \pi/4$ and $\Delta = 0.15\Omega_0$ and $\delta = 0.15\Omega_0$

Figure A.11: Purity of the states as function of time in the case of D-STIRAP and fractional D-STIRAP with detuning as represented in Figure 6.



(a) D-STIRaP with dephasing $\Gamma = 0.01\Omega_0$

(b) D-STIRaP with dephasing $\Gamma = 0.1\Omega_0$

(c) Fractional D-STIRaP with $\alpha = \pi/4$ and dephasing $\Gamma = 0.01\Omega_0$

(d) Fractional D-STIRaP with $\alpha = \pi/4$ and dephasing $\Gamma = 0.1\Omega_0$

Figure A.12: Purity of the states as function of time in the case of D-STIRAP and fractional D-STIRAP (blue lines) in the case of weak (a,b) and strong dephasing (b,d) respectively as represented in Figure 7, together with the corresponding function $e^{-\Gamma\cdot t}$ (dashed gray).