Quantum targeted energy transfer through machine learning tools

I. Andronis,[1, *] G. Arapantonis,[1, 2, †] G. D. Barmparis,[1] and G. P. Tsironis[1, 3]

[1]*Department of Physics, University of Crete, Heraklion 70013, Greece*
[2]*William H. Miller III Department of Physics & Astronomy,*
Johns Hopkins University, Baltimore, Maryland 21218, USA
[3]*John A. Paulson School of Engineering and Applied Sciences,*
Harvard University, Cambridge, Massachusetts 02138, USA
(Dated: March 20, 2023)

In quantum targeted energy transfer, bosons are transferred from a certain crystal site to an alternative one, utilizing a nonlinear resonance configuration similar to the classical targeted energy transfer. We use a novel computational method based on machine learning algorithms in order to investigate selectivity as well as efficiency of the quantum transfer in the context of a dimer and a trimer system. We find that our method identifies resonant quantum transfer paths that allow boson transfer in unison. The method is readily extensible to larger lattice systems involving nonlinear resonances.

I. INTRODUCTION

Nonlinear dynamical systems are notoriously difficult to study analytically and in most cases, one needs to resort to numerical methods for their analysis [1]. In the classical realm, they are typically described mathematically through coupled nonlinear differential equations that very scarcely admit exact solutions. In this case, one resorts to direct numerical integration of the equations of motion. In the quantum domain, on the other hand, even though the equations are linear, one needs to engage a very large part of the Hilbert space in order to find good, yet approximate, solutions [2]. Direct numerical methods, either in the classical or quantum domain, are relatively straightforward, yet might fail in large or strongly coupled systems. The recent widespread of Machine Learning (ML) techniques and their implementation in the domain of dynamical systems aim to both facilitate and also improve the discovery process in these systems [3]. The aim of the present work is to use techniques motivated by ML and obtain results that would be otherwise very complex to derive. The specific model we work with is that of Targeted Energy Transfer (TET) that was inspired by energy transfer processes in chlorophyll [4]. We have two targets here; the first one is to show how ML-motivated techniques may be superior to standard numerical methods when applied in quantum complex systems and the second is to find explicit results that would be otherwise much more difficult to obtain.

In the semiclassical TET model, we focus on resonant exciton transfer between non-identical molecules [4]. In each molecule a single energy state participates in the process, thus we have non-identical energy states coupled together via a non-zero transfer matrix element. In the simplest case of two molecules, we deal with a non-degenerate linear dimer system. Due to the energy mis-

match, the exciton transfer from the first site to the second is non-resonant and thus occurs only partially. In order to make the transfer resonant we need to add local interaction with additional degrees of freedom, such as phonons. In the anti-adiabatic approximation, this procedure introduces effective qubic nonlinearities [5]. In TET, complete resonant transfer is restored for specific nonlinearity parameter configurations linking the local interaction with anti-adiabatic phonons and the actual energy mismatch. The analysis is done within the context of the discrete nonlinear Schrödinger (DNLS) equation, a ubiquitous model for a large class of nonlinear phenomena [6],[7]. A number of analytical results known for DNLS dimers are particularly useful for ML implementations in nonlinear systems [8–10].

While in the case of the semiclassical TET dimer system the resonant transfer regime can be found analytically, similar analysis in larger systems is an arduous task. In order to bypass this difficulty, ML-based approaches have been introduced that enable the discovery of nonlinear resonances in a straightforward way [11, 12]. The method was tested both in the TET dimer and also in other analytically known DNLS equation results. Also, it was applied to the case of a TET trimer model. This ML approach found readily the trimer resonances that were very difficult to obtain differently. The implementation of ML in this semiclassical regime shows that its application can be very beneficial.

The next challenge is that of addressing the fully quantum TET regime; this is the aim of the present article. When the TET dimer is quantized with bosonic degrees of freedom a more general resonant condition, which involves the number of quanta, arises [13]. When at resonance, these bosons may transfer collectively from the first site to the second in a way similar to the semiclassical TET, although with rates depending on the boson number and the energy difference. As in the semiclassical case, in the fully quantum TET, the resonant condition can be found analytically in the dimer model [13], but any extension to larger systems is prohibitive analytically and involves a high computational cost. We show in the

present work that the implementation of ML methods can help in overcoming these difficulties and be able to obtain readily the required resonant transfer properties.

The structure of the present article is thus the following. In the next section II, we introduce a general form of the DNLS equation that upon quantization leads to fully quantum TET models. Once we define clearly the problem in the quantum case, we discuss in the following section III the ML technique we use. More specifically, we discuss the choice of the Loss Function (LF) that enables the analysis of the resonant transfer both in the dimer and also more generally to arbitrary chains. In section IV we detail the optimization method, and how the quantum resonant paths are found for different boson numbers. Section V is the central section of the article; we not only recover the exact dimer results but also apply the method to a trimer configuration. This gives not only the new result of the resonant paths but also shows that the method is fully able to investigate in detail the specifics of the resonant transfer. Finally, in section VI we conclude and summarize the findings of the work and comment on possible extensions.

II. FROM THE SEMICLASSICAL TO THE FULLY QUANTUM TET

It is known that the semiclassical DNLS equation can be derived from a classical Hamiltonian through the use of Hamilton's equations [5]. This classical DNLS Hamiltonian is

$$H = \sum_{k=1}^{f} \omega_k |\psi_k|^2 + \frac{1}{2}\chi_k |\psi_k|^4 - \lambda \sum_{k=1}^{f-1} (\psi_k^* \psi_{k+1} + \psi_{k+1}^* \psi_k), \quad (1)$$

where ω_k and χ_k denote the frequency and nonlinearity parameter of the oscillator at site k respectively. Also, $(\psi_k^*, i\psi_k)$ is a pair of conjugate variables, and the parameter λ is the coupling among neighboring sites. For simplicity, we assume that it is the same between every pair of adjacent oscillators. In order to move to the quantum mechanical case we need to focus on the Bose-Hubbard operator [14]. The DNLS model for the quantum domain may be seen to arise also from the Bose-Hubbard model by using the time-dependent variational principle [15]. A simple way to quantize the Hamiltonian of Eq. (1) is by substituting ψ_k^*, ψ_k with the creation and annihilation operators $a_k^\dagger, a_k$ respectively, as expressed in the second quantization formalism [16]. These operators obey to the commutation relations $[a_k, a_m^\dagger] = \delta_{km}$, and $[a_k, a_m] = 0$, where δ_{km} is the Kronecker delta. Therefore, the Hamiltonian operator in the quantum case becomes

$$\hat{H} = \sum_{k=1}^{f} \omega_k \hat{N}_k + \frac{1}{2}\chi_k \hat{N}_k^2 - \lambda \sum_{k=1}^{f-1} (\hat{a}_k^\dagger \hat{a}_{k+1} + \hat{a}_{k+1}^\dagger \hat{a}_k), \quad (2)$$

where $\hat{N}_k = \hat{a}_k^\dagger \hat{a}_k$ is the boson number operator for the site k. The dimension of the Hilbert space $\mathcal{H}_N$ for this

problem is finite. Each state corresponds to an allowed configuration of N indistinguishable bosons occupying f distinguishable sites-nonlinear oscillators, with repetitions. Thus the dimension of the Hilbert space is

$$\mathcal{D} = \frac{(N+f-1)!}{N!(f-1)!}.$$

A basis associated with that problem is the one composed of the so-called Fock states $|n\rangle \equiv |n_1, n_2, \ldots, n_f\rangle$, where $n_1, n_2, \ldots, n_f$ is the number of bosons at each respective site $1, 2, \ldots, f$ at the state indexed as n. We discuss the procedure of labeling the Fock states in section III. The occupation numbers $\{n_i\}$ are restricted to $\sum_i n_i = N$, while $n_i = 0, 1, \ldots, N$. Additionally, the Fock states are orthonormal, meaning

$$\langle n|m\rangle = \delta_{n_1 m_1} \ldots \delta_{n_f m_f}.$$

Continuing, the actions of the operators $\hat{a}_k, \hat{a}_k^\dagger, \hat{N}_k$ on each component $|n\rangle$ of the basis are described by

$$\hat{a}_k |\ldots, n_k, \ldots\rangle = \sqrt{n_k} |\ldots, n_k - 1, \ldots\rangle, \quad (3a)$$

$$\hat{a}_k^\dagger |\ldots, n_k, \ldots\rangle = \sqrt{n_k + 1} |\ldots, n_k + 1, \ldots\rangle, \quad (3b)$$

$$\hat{N}_k |\ldots, n_k, \ldots\rangle = n_k |\ldots, n_k, \ldots\rangle. \quad (3c)$$

III. DETERMINATION OF AN APPROPRIATE LOSS FUNCTION

We now focus on the choice of an appropriate loss function. We assume that the *donor* site has the lowest energy with nonlinearity parameter χ_D while the highest energy site is the *acceptor* site with nonlinearity parameter χ_A. We further assume that all the bosons are placed initially (at time $t = 0$) to the donor, and investigate the TET configurations that allow the complete transfer of these bosons to the acceptor site. To achieve this, we employ an algorithm relying on the same principle as the ones presented in [11, 17], where an optimization algorithm is used to minimize a quantity that is defined as the LF. The LF is usually associated with some physical parameters and thus, the problem becomes one where the algorithm, has to tune the parameters.

The first step towards defining the LF is to construct a numerical scheme for calculating the matrix elements of the Hamiltonian in Eq. (2). While this is usually a simple process, there are special intricacies, with the main one being the labeling of the Fock states. We manage to overcome this obstacle by implementing a similar technique to [18], where the authors rank the states in *lexicographic* order and assign indices $1, 2, \ldots, \mathcal{D}$ to each configuration of bosons among the sites. For instance, assuming $N = 2$ and $f = 3$, state 1 corresponds to $|1\rangle = |2, 0, 0\rangle$, state 2 corresponds to $|2\rangle = |1, 1, 0\rangle$ and so on, assigning every state-configuration to a distinct index.

Under this indexing policy, it is now straightforward to compute the elements $\hat{H}_{ij}$ similarly to [13], considering that

$$\hat{H}_{ij} = \langle i| \hat{H} |j\rangle \overset{(2)}{=} T_1 + T_2, \tag{4a}$$

$$T_1 \equiv \langle i| \sum_{k=1}^{f} \left[\omega_k \hat{N}_k + \frac{1}{2}\chi_k(\hat{N}_k)^2 \right] |j\rangle, \tag{4b}$$

$$T_2 \equiv -\lambda \langle i| \sum_{k=1}^{f-1} \left[\hat{a}_k^\dagger \hat{a}_{k+1} + \hat{a}_{k+1}^\dagger \hat{a}_k \right] |j\rangle. \tag{4c}$$

Calculating T_1 is simple since it involves the simple action of the boson number operator on state $|j\rangle$, as shown in Eq. (3c). That said, Eq. (4b) is equivalent to

$$T_1 = \sum_{k=1}^{f} \left[\omega_k j_k + \frac{1}{2}\chi_k(j_k)^2 \right] \delta_{ij},$$

where j_k stands for the number of bosons on site k for the state $|j\rangle$. Evaluating the second term is nontrivial because it involves the action of the operators $\hat{a}_k^\dagger \hat{a}_{k+1}$, $\hat{a}_{k+1}^\dagger \hat{a}_k$. The consecutive action of these operators can be explored by referring to Eq. (3a) and Eq. (3b). The operators $\hat{a}_k^\dagger$, $\hat{a}_k$ create and annihilate a boson at a given site k respectively. However, when a pair of operators like $\hat{a}_k^\dagger \hat{a}_{k+1}$ (or $\hat{a}_{k+1}^\dagger \hat{a}_k$) acts on a Fock state $|j\rangle$, it creates a boson at the site k (or $k+1$) but also destroys a boson at the site $k+1$ (or k). Thus, their action on a Fock state conserves the total number of bosons and the resulting state is going to be, up to a constant, another state in $\mathcal{H}_N$. Specifically,

$$\langle i| \hat{a}_{k+1}^\dagger \hat{a}_k |j\rangle = \sqrt{j_k(j_{k+1}+1)}\delta_{ip} \equiv C_k^{(p)}\delta_{ip},$$

$$\langle i| a_k^\dagger a_{k+1} |j\rangle = \sqrt{j_{k+1}(j_k+1)}\delta_{im} \equiv D_k^{(m)}\delta_{im},$$

where the two new states $|p\rangle, |m\rangle$ are

$$|p\rangle = |j_1,\ldots,j_k-1,j_{k+1}+1\ldots,j_f\rangle,$$

$$|m\rangle = |j_1,\ldots,j_k+1,j_{k+1}-1\ldots,j_f\rangle.$$

Combining the above yields

$$T_2 = -\lambda \sum_{k=1}^{f-1} \left[C_k^{(p)}\delta_{ip} + D_k^{(m)}\delta_{im} \right].$$

The matrix representation of the Hamiltonian can be produced and subsequently, the eigenstates and eigenvalues can be calculated. Everything is developed in *Python*, using the *Tensorflow* [19] library.
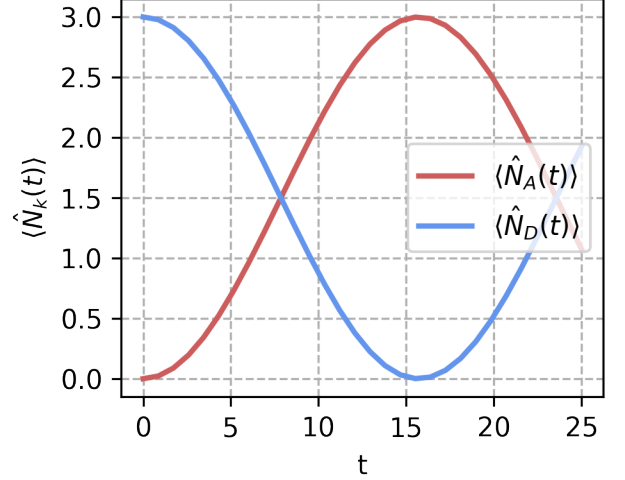

FIG. 1. Time evolution of the expectation value of the number of bosons for the two sites of the dimer under the parameters $(\chi_A, \chi_D, \omega_A, \omega_D, \lambda, N, \texttt{maxt}) = (-2,2, 3, -3, 0.1, 3, 25)$. The blue line denotes the donor's expectation value, while the red one the acceptor's.

The initial distribution of bosons $|\Psi(0)\rangle$ can be expanded to the basis of the eigenstates $|\psi_i\rangle$

$$|\Psi(0)\rangle = \sum_{i=1}^{\mathcal{D}} C_i |\psi_i\rangle, \quad C_i = \langle \psi_i|\Psi(0)\rangle.$$

Also, these eigenstates can be expanded to the basis of the Fock states

$$|\psi_i\rangle = \sum_{j=1}^{\mathcal{D}} b_{j,i} |j\rangle, \quad b_{j,i} = \langle j|\psi_i\rangle.$$

We can now express the time evolution of the initial distribution $|\Psi(0)\rangle$ by applying the time evolution operator $\hat{U}(t) = e^{-\mathbf{i}\hat{H}t}$:

$$|\Psi(t)\rangle = e^{-\mathbf{i}\hat{H}t} |\Psi(0)\rangle = \sum_{i,j}^{\mathcal{D}} C_i b_{j,i} e^{-\mathbf{i}E_i t} |j\rangle, \tag{7}$$

where E_i is the i-th eigenvalue and $\mathbf{i}$ is the imaginary unit. Similarly, the time evolution of the average number of bosons at the site k is given by

$$\langle \hat{N}_k(t)\rangle = \langle \Psi(t)| \hat{N}_k |\Psi(t)\rangle. \tag{8}$$

Combining Eq. (7) and Eq. (8), $\langle \hat{N}_k \rangle$ can be assessed in the following way

$$\langle \hat{N}_k(t)\rangle = \sum_{i,j,n}^{\mathcal{D}} j_k C_n^* C_i b_{j,n}^* b_{j,i} e^{\mathbf{i}(E_n - E_i)t}. \tag{9}$$

We use Eq. (9) in order to compute the LF for the quantum TET problem. Specifically, we time-evolve Eq. (9)

for the acceptor energy level until a pre-defined time `maxt`. During this period, the oscillator of the fth site (acceptor) has completed a few oscillations. The next step in computing the LF is to extract the maximum value from that time evolution. Concluding, the LF is defined as

$$LF = N - max\{\langle \hat{N}_f(t) \rangle\} = N - max\{\langle \hat{N}_A(t) \rangle\}. \quad (10)$$

In Fig. 1 one can observe the characteristic oscillatory behavior, in the dimer system, when the complete transfer occurs.

In the present work, we fix the frequency of the oscillators and optimize for the nonlinearity parameters of the oscillators. While the LF might appear not to have any explicit connection to the nonlinearity parameters we want to optimize for, we can use Tensorflow's $GradientTape$ to compute the derivatives with respect to these parameters. In that way, we keep track of gradient information in terms of the trainable variables throughout the whole process described in this section; from creating the Hamiltonian to calculating $\langle \hat{N}_k(t) \rangle$. Using this information, an optimizer like $ADAM$ [20] can now update the parameters accordingly, so that the LF is minimized, signifying complete transfer. It is important to note that, while many other optimization algorithms (simulated annealing [21], particle swarm [22], differential evolution [23, 24]) were tried on this problem, we were not able to produce adequate results with none of them.

The parameter `maxt` is of major importance, because of the oscillation of the bosons between the sites. If it is not set large enough, TET can be missed since the system would not have time to complete an oscillation, while it also has to be small enough, so that precious computational time is saved. Thus, it has to be large enough to obtain at least one complete oscillation, producing this way essential information. It is also observed that the period of each oscillator is proportional to λ^{-1}, so as one would expect, changes to these parameters should be made concurrently.

IV.   OPTIMIZATION OF THE LOSS FUNCTION

Once we define the LF we may now proceed with its minimization through the procedure outlined in Fig. 2. We emphasize the first step of this process graph, i.e. a proper update method, which is essential for the present work. We noticed by making several computational runs that optimizers did not work very well when initialized with a random set of parameters; this is due to the extreme selectivity of the TET resonant condition. Specifically, the parameter space that the LF maps to, exhibits slowly varying gradients everywhere, except for areas close to the optimal parameters for transfer, where relatively large gradients are present.

Our approach for bypassing this problem relies on exploring a wide range of the parameter space at the same time. To be more specific, the optimization procedure is
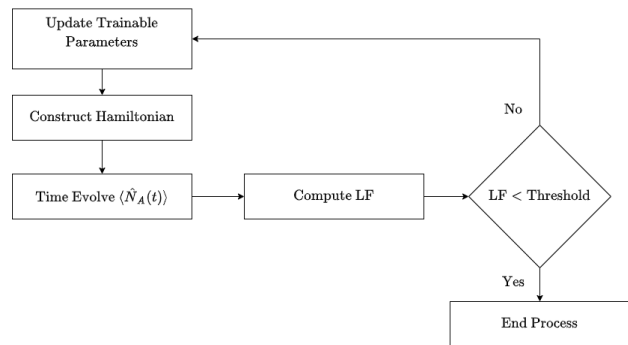

FIG. 2. Schematic representation of the parameter optimization procedure.

comprised of two distinct phases. First, many *test optimizers* are assigned some initial guesses and left to run simultaneously, as different processes, until a stopping condition is reached. Then we gather the LF and the parameters that each test optimizer yielded at the last iteration. For the second part of the optimization procedure, a *main optimizer* is employed. Its initial guesses are the parameters of the test optimizer with the best performance (i.e smallest LF). The purpose of this additional step is to further minimize the LF, if possible.

Two methods were developed for defining the initial guesses of the test optimizers, given that we choose the range of the parameter space they will investigate. In the first approach, which is represented by Fig. 3a, we define a *grid* of points in the parameter space and use the lattice points as initial guesses for the test optimizers. In the second method, we split the parameter space into a number of regions and chose random initial combinations of parameters from each region. On the one hand, the former method has a high computational cost, since it requires a large number of optimizers running at the same time, but has a greater potential to derive the optimal parameters. On the other hand, the latter is fast but less robust. It is useful in cases of stronger coupling, as the gradients of the parameter space smooth out. If the first run of any method does not produce favorable results we have the option of redefining the limits of the parameter space around the best parameters provided by the test optimizers. Nevertheless, choosing one of the above methodologies relies on the problem at hand. In our case, both of them produce accurate results. Moreover, we need to disambiguate that the main optimizer is an optional step, that aims to corroborate TET, by producing a LF lower than the one deduced from the test optimizers.

Regardless of whether we use the grid or the splitting into regions method, the trainable parameters are updated in the same way for both the test optimizers and the main optimizer. Given the initial guesses of $\{\chi_k\}$, $k = 1, 2, \ldots, f$, the nonlinearity parameters after the mth iteration are updated as following:
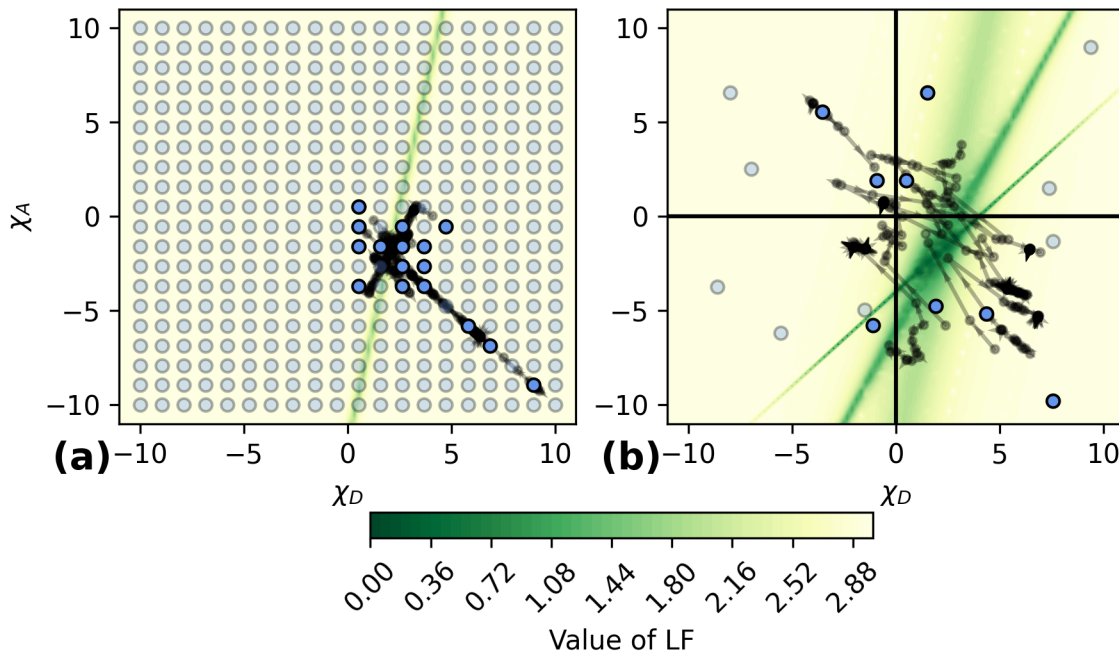
FIG. 3. Defining the initial guesses of the test optimizers. **(a)** An example using the *grid* method, where we define 400 different initial guesses for the same number of test optimizers. **(b)** Representative example of the second method, where splitting of the parameter space into regions is employed. The black lines represent the boundaries among the four different regions from which 16 different guesses are sampled. In both graphs, the test optimizers that were able to derive a LF < 1 are displayed with normal opacity and their trajectories, while the rest are faded. Both figures refer to a dimer system with $(N, \omega_A, \omega_D, \texttt{maxt}) = (3, 3, -3, 25)$. For the purposes of explaining the difference between the two methods, the system presented in the left figure has a coupling parameter of $\lambda = 0.1$, while for the system on the right figure $\lambda = 1$.

$$\chi_j^{(m+1)} = \chi_j^{(m)} - \alpha \vec{\nabla}_j(LF),$$

where the gradient is computed with Tensorflow's *GradientTape*, as discussed in section III. The learning rate α is a positive real number that defines the rate of change at each iteration while moving towards the minimum of the LF. With the new set of nonlinearity parameters, the optimization procedure moves to the next iteration and it will be interrupted either due to slow convergence or because of reaching maximum iterations.

The threshold parameter of Fig. 2 has a dual role. On one hand, it checks whether the nonlinearity parameters of the current iteration yield a LF close to zero, signifying TET. On the other hand, it determines the slow convergence of the algorithm, and its role is to pause the optimization procedure when there is no significant improvement in minimizing the LF. Thus, its value should be small enough (close to zero) to manifest TET, but still nonzero because otherwise, the stochastic optimization procedure we introduce will lead to an infinite loop. The latter problem is also resolved by terminating the optimization procedure after reaching a predefined maximum number of iterations.

The *Python* code implementing this procedure is located in our GitHub repository [25].

V. RESULTS

A. TET Quantum Dimer

In the preceding section, we describe the optimization technique, including the possible alternatives. We may now apply this scheme to the dimer realm and test our method from the semiclassical limit to the fully quantum one, using the grid method described earlier. As we discussed in section I, the optimal parameters for this case are already known for both the semiclassical [4] and the quantum regime [13]:

$$\chi_D = -\chi_A = \frac{\omega_A - \omega_D}{N}. \tag{11}$$

Our method is successful in obtaining the TET configurations. We fix the frequencies of the donor and acceptor to $\omega_D = -3$ and $\omega_A = 3$ respectively, the coupling parameter to $\lambda = 0.1$, while the time evolution of Eq. (9) for
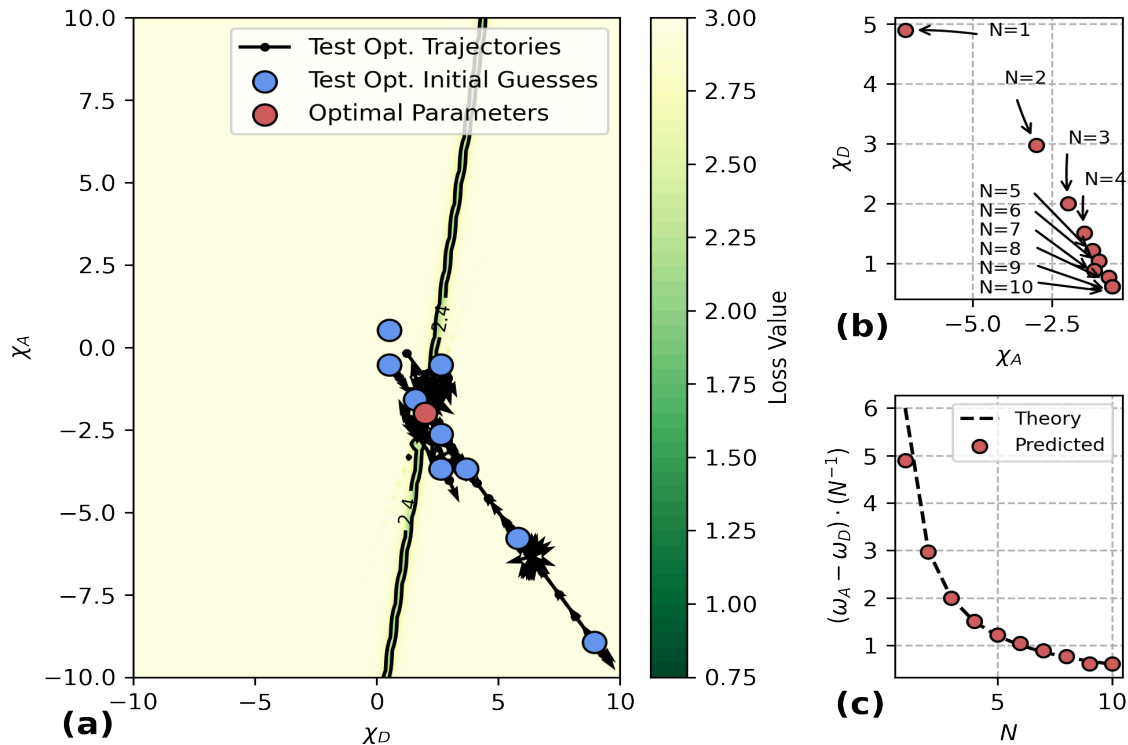
FIG. 4. Optimization Results. (a) The trajectories and initial guesses of our algorithm, using the grid method, for $N = 3$ bosons. The optimal parameters found -by the main optimizer- in this example are $(\chi_A, \chi_D) = (-1.99, 2)$. Only the initial guesses of the test optimizers that resulted in a LF smaller than 0.5 are displayed. (b) Parameters for TET from the quantum limit to the semiclassical limit, as deduced from the grid method and the main optimizer, in the dimer system. (c) Comparison of the predicted nonlinearity parameters $\chi_D = -\chi_A$, illustrated in the figure as red dots, with the reference value derived from Eq. (11) represented in the figure as a dashed line. The system's constant parameters are $(\omega_A, \omega_D, \lambda, \mathtt{maxt}) = (3, -3, 0.1, 25)$.

the acceptor is performed until $\mathtt{maxt} = 25$. The Fig. 4a displays the outcome of the optimization procedure for $N = 3$ bosons, with the optimal nonlinearity parameters being:

$$(\chi_D, \chi_A) = (2, -1.99).$$

The illustrated test optimizers produce a LF lower than the threshold mentioned in Fig. 2, which is set arbitrarily (in this case 0.5). The results validate the sensitivity of each optimizer to the initial guesses.

We keep the same values for $\omega_A, \omega_D, \lambda$, and apply the grid method for a variety total of bosons in the dimer system. The outcome of the optimization scheme is shown in Fig. 4b, while we compare our results with Eq. (11) in Fig. 4c. In every case, the proposed method succeeds in identifying the TET paths. It is worth mentioning that we can deduce the same results with the method of splitting the parameter space.

Moreover, our analysis proves that the fully quantum case of $N = 1$ boson is of special interest since TET occurs for a whole set of nonlinearity parameters χ_A, χ_D instead of a single, very limited configuration. Specifically, we identify this set as

$$\chi_D = \chi_A + 2(\omega_A - \omega_D). \quad (12)$$

This result is in agreement with previous analytical calculations. To be more specific, Maniadis et al. in [13] prove that the condition for having TET is for the detuning function to vanish. The latter is defined as the variation of the energy of the oscillators during a transfer:

$$\epsilon = [H_D(N) + H_A(0)] - [H_D(i) + H_A(N - i)]. \quad (13)$$

In Eq. (13) $i = 0, 1, \ldots, N$ while H_D, H_A are the donor and acceptor parts of the Hamiltonian in Eq. (2). In this case, any nonzero λ can raise the degeneracy of the system and complete TET occurs in the limiting of zero coupling. We can easily prove that the detuning function of Eq. (13) for one boson vanishes under the parameters of Eq. (12).

For all the cases where the detuning function vanishes, the Hamiltonian becomes quadratic of the bosons operators:

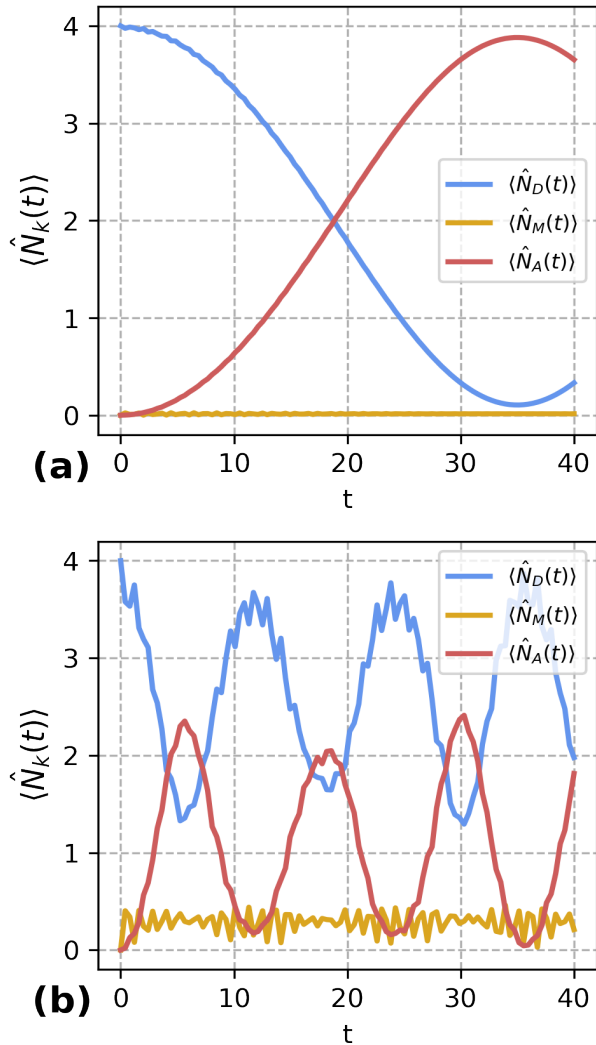$$\hat{H} = \hat{H}_D(N) - \lambda(a_D^\dagger a_A + a_A^\dagger a_D). \quad (14)$$

(a)

(b)

FIG. 5. Time evolution of the average number of bosons for the three sites of the trimer system. **(a)** The time evolution of the expectation values of the boson number operators for non-linearity parameters that produce near complete TET where $(\chi_A, \chi_M, \chi_D, \omega_D, \omega_M, \omega_A, \lambda, N, \texttt{maxt}) = (-1.5, -38.39, 1.5, -3, -3, 3, 1, 4, 40)$. **(b)** The time evolution of the expectation values of the boson number operators for a system with parameters that don't produce complete TET, where $(\chi_A, \chi_M, \chi_D, \omega_D, \omega_M, \omega_A, \lambda, N, \texttt{maxt}) = (-1.5, 1.5, 1.5, -3, -3, 3, 1, 4, 40)$. For both figures, the blue, yellow, and red lines refer to the expectation value of the boson number operator for the donor, middle site, and acceptor respectively.

B.   TET Quantum Trimer

The trimer system has been investigated in [11, 26] in the context of a single electron or boson. We aim to expand this investigation to the arbitrary case of N bosons. While the method is able to optimize the nonlinearity parameters of all three sites, it proved computationally consuming. To circumvent this, we set the acceptor and
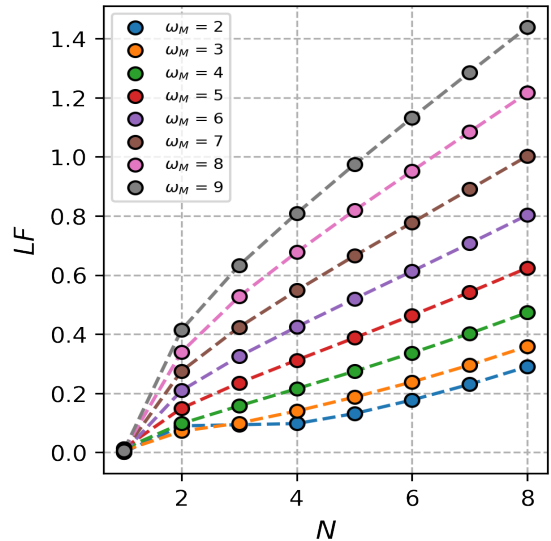

FIG. 6. Lowest value of the LF in a trimer system with the parameters displayed on the figure. The other system parameters are $(\omega_A, \omega_D, \lambda, \texttt{maxt}) = (\omega_M\text{-}1, \text{-}\omega_M\text{+}1, 1, 40)$.

donor sites to their dimer values, and optimize the parameter of the middle layer (labeled as "M"). Similarly to the dimer case, we can set an arbitrary threshold for the LF (LF< 0.2) and begin the optimization process.

We observe that for TET to occur in the quantum realm, the middle layer had to exhibit extremely high nonlinearity. One of the trimer systems examined in this section has the following properties: It is strongly coupled, $\lambda = 1$, with frequencies $\omega_D = 3$, $\omega_M = -3$, $\omega_A = -3$ and a maximum number of bosons $N = 4$ initially at the donor site.

The nonlinearity parameter required to have TET, in this case, is $\chi_M \approx 38.39$, much higher than that of Eq. (11), compared to the parameters of the donor and acceptor sites $\chi_D = -\chi_A = 1.5$. It is important to mention that the opposite value $\chi_M \approx -38.39$ will produce a similar LF value, but still small enough to exceed the threshold and stop the iterative process. We carry out the same procedure for a variety of system parameters, as seen in Fig. 7.

We display the absolute value of χ_M as both positive and negative values produce the desired result. We observe that the relation between χ_M and the frequencies or the number of bosons of the system seems to differ from the dimer case, as there appears to be a greater correlation with the former rather than the latter. In the graph some points seem to differ a lot from others, a fact that is attributed to an optimization process that was not able to minimize the loss function enough before being terminated by the rules we defined earlier in section IV. We also have plotted the minimum value of the LF that we observed in Fig. 6. The time evolution of the boson number operators for this system is shown in Fig. 5. As we can see in subplot (a) the expectation value of the
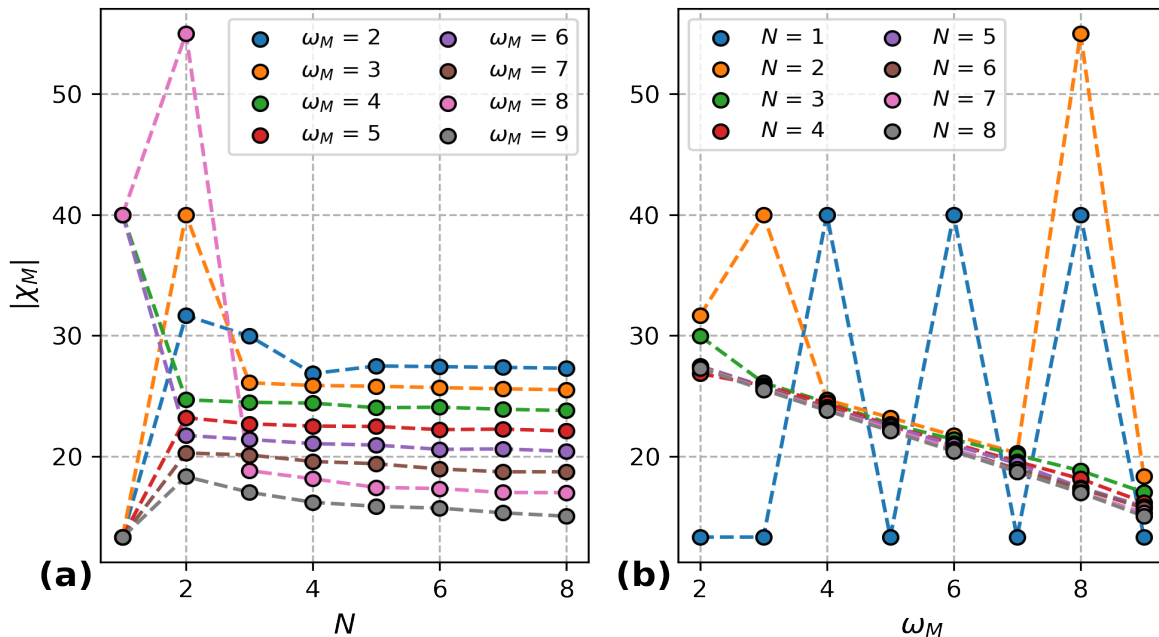
FIG. 7. Trimer Results. Absolute values of the nonlinearity parameter of the middle layer with respect to **(a)** different values of the maximum number of bosons in the system, **(b)** the frequency of the middle layer ω_M. In both cases we use the grid method while the system's constant parameters for both figures are $(\omega_A, \omega_D, \lambda, \texttt{maxt}) = (\omega_M-1, -\omega_M+1, 1, 40)$.

.

boson number operator for the middle layer, in the optimal case, appears to be zero for all time steps which indicates that bosons do not stay on this site for any significant amount of time, or at all. However, this is not observed in cases where the nonlinearity parameters are set to nonoptimal values, as seen in subplot (b), where the expectation value is nonzero. It is apparent that the oscillation frequency of the number operator is larger in the non-resonant system and that complete TET cannot be achieved.

VI. DISCUSSION

In this work, we introduced a ML method in the context of a quantum many-body system and showed that its efficient implementation can produce results that are very hard to obtain with more conventional methods. We focused on the quantized version of the DNLS equation with arbitrary local energies and nonlinearities and addressed the question of optimal transfer in between different sites for the dimer and trimer cases. Since the fully quantum transfer dimer case is known analytically we compared our method with these results and showed perfect agreement. This successful comparison between analytics and ML methods shows that the latter can be used confidently in more complex cases where results are not known. Subsequently, we applied the method to the trimer case that cannot be solved analytically. Our

method enabled a detailed search showing the specifics of the resonant transfer, the different parameter regimes as well as the transfer efficiencies. In terms of physics, we found in the trimer system that in the nonresonant transfer regime from donor to acceptor sites the intermediate state retains some of the probability. In the resonant case, on the other hand, the intermediate site is essentially not populated. This shows that this site acts as some form of a barrier between the donor and acceptor sites that can be completely bypassed in the fully resonant regime. It is noteworthy to point out that the bosons move in unison over to the acceptor site showing a very interesting collective behavior in the transfer.

The collective boson transfer can be investigated also in more general chains with a larger number of sites. The computational challenge is now larger since the dimensionality of the system becomes large and the calculation of the Hamiltonian and the evolution of $\langle \hat{N}_A(t) \rangle$ slows down. In this regime, one needs to explore other loss functions that could improve scalability and/or implementation of meta-learning methods described in [27]. In this work, we tested also alternative optimizers such as ones with momentum. We found that they were more efficient in finding the resonant transfer parameter regime but were highly dependent on hyper-parameters that needed to be also optimized. Finally, it is possible that variants of the gradient descent algorithm might help in reducing the computational cost[28].

The phenomenon of the collective transfer of bosons

in the trimer case opens up very interesting new questions on the interplay of nonlinearity and disorder in the fully quantum regime for more extended systems. The ML method provided in the present work can be readily generalized to this case and be utilized to investigate this very exciting problem with applications in condensed matter physics as well as quantum optics.

AUTHOR CONTRIBUTIONS

I. Andronis and G. Arapantonis contributed equally to this work.

[1] E. Ott, *Chaos in Dynamical Systems*, 2nd ed. (Cambridge University Press, 2002).

[2] V. M. Akulin, *Dynamics of Complex Quantum Systems*, 2nd ed. (Springer Dordrecht, 2014).

[3] G. Neofotistos, M. Mattheakis, G. D. Barmparis, J. Hizanidis, G. P. Tsironis, and E. Kaxiras, Frontiers in Physics **7**, 10.3389/fphy.2019.00024 (2019).

[4] G. Kopidakis, S. Aubry, and G. P. Tsironis, Phys. Rev. Lett. **87**, 165501 (2001).

[5] D. Hennig and G. Tsironis, Physics Reports **307**, 333 (1999).

[6] J. Eilbeck, P. Lomdahl, and A. Scott, Physica D: Nonlinear Phenomena **16**, 318 (1985).

[7] V. M. N. Kenkre, *Interplay of Quantum Mechanics and Nonlinearity*, 1st ed. (Springer Cham, 2022).

[8] V. M. Kenkre and D. K. Campbell, Phys. Rev. B **34**, 4959 (1986).

[9] G. Tsironis and V. Kenkre, Physics Letters A **127**, 209 (1988).

[10] G. Tsironis, Physics Letters A **173**, 381 (1993).

[11] G. D. Barmparis and G. P. Tsironis, JOSA B, Vol. 38, Issue 9, pp. C120-C126 **38**, C120 (2021).

[12] G. P. Tsironis, G. D. Barmparis, and D. K. Campbell, International Journal of Modern Physics B **36**, 2240001 (2022), https://doi.org/10.1142/S021797922240001X.

[13] P. Maniadis, G. Kopidakis, and S. Aubry, Physica D: Nonlinear Phenomena **188**, 153 (2004).

[14] H. A. Gersch and G. C. Knollman, Physical Review **129**, 959 (1963).

[15] L. Amico and V. Penna, Phys. Rev. Lett. **80**, 2189 (1998).

[16] T. P. Pearsall, *Quantum Photonics*, 2nd ed. (Springer International Publishing, 2020).

[17] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, Nature Reviews Physics **3**, 422 (2021).

[18] J. M. Zhang and R. X. Dong, European Journal of Physics **31**, 591 (2010).

[19] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from tensorflow.org.

[20] D. P. Kingma and J. L. Ba, 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings 10.48550/arxiv.1412.6980 (2014).

[21] D. Bertsimas and J. Tsitsiklis, Statistical Science **8**, 10 (1993).

[22] J. Kennedy and R. Eberhart, in *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4 (1995) pp. 1942–1948 vol.4.

[23] J. Brest, S. Greiner, B. Boskovic, M. Mernik, and V. Zumer, IEEE transactions on evolutionary computation **10**, 646 (2006).

[24] S. M. Elsayed, R. A. Sarker, and D. L. Essam, in *2011 IEEE Congress of Evolutionary Computation (CEC)* (IEEE, 2011) pp. 1041–1048.

[25] J. Andronis and G. Arapantonis, Quantum-targeted-energy-transfer, https://github.com/JAndronis/Quantum-Targeted-Energy-Transfer (2022).

[26] S.Aubry and G.Kopidakis, A non adiabatic theory for ultrafast and catalytic transfer of electrons at low temperature, in *Localization and Energy Transfer in Nonlinear Systems* (WORLD SCIENTIFIC, 2003) pp. 1–27.

[27] K. Li and J. Malik, Learning to optimize (2016), arXiv:1606.01885v1.

[28] S. Ruder, An overview of gradient descent optimization algorithms (2016), arXiv:1609.04747.