Solution for a bipartite Euclidean traveling-salesman problem in one dimension

Sergio Caracciolo,[*] Andrea Di Gioacchino,[†] Marco Gherardi,[‡] and Enrico M. Malatesta[§]
Dipartimento di Fisica, University of Milan and INFN, Via Celoria 16, I-20133 Milan, Italy

(Received 5 February 2018; published 9 May 2018)

The traveling-salesman problem is one of the most studied combinatorial optimization problems, because of the simplicity in its statement and the difficulty in its solution. We characterize the optimal cycle for every convex and increasing cost function when the points are thrown independently and with an identical probability distribution in a compact interval. We compute the average optimal cost for every number of points when the distance function is the square of the Euclidean distance. We also show that the average optimal cost is not a self-averaging quantity by explicitly computing the variance of its distribution in the thermodynamic limit. Moreover, we prove that the cost of the optimal cycle is not smaller than twice the cost of the optimal assignment of the same set of points. Interestingly, this bound is saturated in the thermodynamic limit.

DOI: 10.1103/PhysRevE.97.052109

I. INTRODUCTION

Given N cities and $N(N-1)/2$ values that represent the cost paid for traveling between all pairs of them, the traveling-salesman problem (TSP) consists of finding the tour that visits all the cities and finally comes back to the starting point with the least total cost to be paid for the journey. The TSP is the archetypal problem in combinatorial optimization [1]. Its first formalization probably can be traced back to the Austrian mathematician Karl Menger, in the 1930s [2], and it still raises interest today. As it belongs to the class of NP-complete problems (see Karp and Steele in [1]), the study of the TSP could shed light on the famous P vs NP problem [3]. Many problems in various fields of science (computer science, operational research, genetics, engineering, electronics, and so on) and in everyday life (lacing shoes, Google maps queries, food deliveries, and so on) can be mapped on a TSP or a variation of it; see, for example, chapter 3 in Ref. [4] for a nonexhaustive list. Interestingly, the complexity of the TSP seems to remain high even if we try to modify the problem. For example, the Euclidean TSP, where the costs to travel from cities are the Euclidean distances between them, remains NP-complete [5]. The bipartite TSP, where the cities are divided in two subsets and the tour has to alternate between them, is NP-complete, too, as its Euclidean counterpart. It is well known that the statistical properties of the optimal solution of problems in combinatorial optimization can be related to the zero temperature behavior of corresponding disordered statistical mechanics models [6–9] when a class of problems and a probability distribution for the different instances are defined.

Previous investigations of some of us suggested that the Euclidean matching problem is simpler to deal with in its bipartite version. This idea encouraged us to consider the bipartite TSP, starting from the one-dimensional case that is fully analyzed here.

The manuscript is organized as follows: In Sec. II we define the TSP and its variants we are interested in. We shall introduce a representation of the model, which is novel as far as we know, in terms of a couple of permutations. In this way we also establish a very general connection between the bipartite TSP and a much simpler model, which is in the P complexity class, the assignment problem. Always using our representation, in Sec. III we can provide the explicit solution of the problem for every instance of the disorder (that is, for every position of the points) in the one-dimensional case when the cost is a convex and increasing function of the Euclidean distance between the cities. In Sec. IV we exploit our explicit solution to compute the average optimal cost for an arbitrary number of points, when they are chosen with uniform distribution in the unit interval, and we present a comparison with the results of numerical simulations. In Sec. V we discuss the behavior of the cost in the thermodynamic limit of an infinite number of points. Here the results can be extended to more general distribution laws for the points. In Sec. VI we give our conclusions.

II. THE MODEL

Given a generic (undirected) graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$, a *cycle* of length k is a sequence of edges $e_1, e_2, \ldots, e_k \in \mathcal{E}$ in which two subsequent edges e_i and e_{i+1} share a vertex for $i = 1, \ldots, k$ where, for $i = k$ the edge e_{k+1} must be identified with the edge e_1. On a bipartite graph each cycle must have an even length. The cycle is *Hamiltonian* when the visited vertices are all different and the cardinality of the set of vertices $|\mathcal{V}|$ is exactly k for $k > 2$. In other terms, a Hamiltonian cycle is a closed path visiting all the vertices in $\mathcal{V}$ only once. The determination of the existence of a Hamiltonian cycle is an NP-complete problem (see Johnson and Papadimitriou in [1]). A graph that contains a Hamiltonian cycle is called a Hamiltonian graph. The complete graph with N vertices $\mathcal{K}_N$ is Hamiltonian for

[*]sergio.caracciolo@mi.infn.it
[†]andrea.digioacchino@unimi.it
[‡]marco.gherardi@mi.infn.it
[§]enrico.m.malatesta@gmail.com

$N > 2$. The bipartite complete graph with $N + M$ vertices $\mathcal{K}_{N,M}$ is Hamiltonian for $M = N > 1$.

Let us denote by $\mathcal{H}$ the set of Hamiltonian cycles of the graph $\mathcal{G}$. Let us suppose now that a weight $w_e > 0$ is assigned to each edge $e \in \mathcal{E}$ of the graph $\mathcal{G}$. We can associate with each Hamiltonian cycle $h \in \mathcal{H}$ a total cost,

$$E(h) := \sum_{e \in h} w_e. \tag{1}$$

In the (weighted) Hamiltonian cycle problem we search for the Hamiltonian cycle $h \in \mathcal{H}$ such that the total cost in (1) is minimized, i.e., the optimal Hamiltonian cycle $h^* \in \mathcal{H}$ is such that

$$E(h^*) = \min_{h \in \mathcal{H}} E(h). \tag{2}$$

When the N vertices of $\mathcal{K}_N$ are seen as cities and the weight for each edge is the cost paid to cover the route distance between the cities, the search for h^* is called the *traveling-salesman problem* (TSP). For example, consider when the graph $\mathcal{K}_N$ is embedded in $\mathbb{R}^d$, that is, for each $i \in [N] = \{1,2,\dots,N\}$ we associate a point $x_i \in \mathbb{R}^d$, and for $e = (i,j)$ with $i,j \in [N]$ we introduce a cost which is a function of their Euclidean distance $w_e = |x_i - x_j|^p$ with $p \in \mathbb{R}$. When $p = 1$, we obtain the usual *Euclidean* TSP. Analogously for the bipartite graph $\mathcal{K}_{N,N}$ we will have two sets of points in $\mathbb{R}^d$, that is the red $\{r_i\}_{i \in [N]}$ and the blue $\{b_i\}_{i \in [N]}$ points and the edges connect red with blue points with a cost,

$$w_e = |r_i - b_j|^p. \tag{3}$$

When $p = 1$, we obtain the usual *bipartite* Euclidean TSP. The simplest way to introduce randomness in the problem is to consider the weights w_e independent and identically distributed random variables. In this case the problem is called *random* TSP and has been extensively studied by disordered system techniques such as replica and cavity methods [7,10–15] and by a rigorous approach [16]. In the random Euclidean TSP [17–21], instead, the positions of the points are generated at random and as a consequence the weights will be correlated. The typical properties of the optimal solution are of interest, and in particular the average optimal cost,

$$\overline{E} := \overline{E(h^*)}, \tag{4}$$

where we have denoted by a bar the average over all possible realization of the disorder.

A. Representation in terms of permutations

We shall now restrict to the complete bipartite graph $\mathcal{K}_{N,N}$. Let $\mathcal{S}_N$ be the group of permutation of N elements. For each $\sigma,\pi \in \mathcal{S}_N$, the sequence for $i \in [N]$,

$$\begin{aligned} e_{2i-1} &= (r_{\sigma(i)}, b_{\pi(i)}), \\ e_{2i} &= (b_{\pi(i)}, r_{\sigma(i+1)}), \end{aligned} \tag{5}$$

where $\sigma(N+1)$ must be identified with $\sigma(1)$, defines a Hamiltonian cycle. More properly, it defines a Hamiltonian cycle with starting vertex $r_1 = r_{\sigma(1)}$ with a particular orientation, that is,

$$h[(\sigma,\pi)] := (r_1 b_{\pi(1)} r_{\sigma(2)} b_{\pi(2)} \cdots r_{\sigma(N)} b_{\pi(N)}) = (r_1 C), \tag{6}$$

where C is an open walk that visits once all the blue points and all the red points with the exception of r_1. Let C^{-1} be the open walk in the opposite direction. This defines a new, dual, couple of permutations that generate the same Hamiltonian cycle,

$$h[(\sigma,\pi)^\star] := (C^{-1} r_1) = (r_1 C^{-1}) = h[(\sigma,\pi)], \tag{7}$$

since the cycle $(r_1 C^{-1})$ is the same as $(r_1 C)$ (traveled in the opposite direction). By definition,

$$h[(\sigma,\pi)^\star] = (r_1 b_{\pi(N)} r_{\sigma(N)} b_{\pi(N-1)} r_{\sigma(N-1)} \cdots b_{\pi(2)} r_{\sigma(2)} b_{\pi(1)}). \tag{8}$$

Let us introduce the cyclic permutation $\tau \in \mathcal{S}_N$, which performs a left rotation, and the inversion $I \in \mathcal{S}_N$, that is, $\tau(i) = i + 1$ for $i \in [N-1]$ with $\tau(N) = 1$ and $I(i) = N + 1 - i$. In the following we shall denote a permutation by using the second raw in the usual two-raw notation, that is, for example, $\tau = (2,3,\dots,N,1)$ and $I = (N, N-1, \dots, 1)$. Then,

$$h[(\sigma,\pi)^\star] = h[(\sigma \circ \tau \circ I, \pi \circ I)]. \tag{9}$$

There are $N!\,(N-1)!/2$ Hamiltonian cycles for $\mathcal{K}_{N,N}$. Indeed the couples of permutations are $(N!)^2$ but we have to divide them by $2N$ because of the N different starting points and the two directions in which the cycle can be traveled.

B. Comparison with the assignment problem

From (5) and weights of the form (3), we get an expression for the total cost,

$$E[h[(\sigma,\pi)]] = \sum_{i \in [N]} [|r_{\sigma(i)} - b_{\pi(i)}|^p + |r_{\sigma\circ\tau(i)} - b_{\pi(i)}|^p]. \tag{10}$$

Now we can reshuffle the sums and we get

$$\begin{aligned} &E[h[(\sigma,\pi)]] \\ &\quad = \sum_{i \in [N]} |r_i - b_{\pi\circ\sigma^{-1}(i)}|^p + \sum_{i \in [N]} |r_i - b_{\pi\circ\tau^{-1}\circ\sigma^{-1}(i)}|^p \\ &\quad = E[m(\pi \circ \sigma^{-1})] + E[m(\pi \circ \tau^{-1} \circ \sigma^{-1})], \end{aligned} \tag{11}$$

where $E[m(\lambda)]$ is the total cost of the assignment m in $\mathcal{K}_{N,N}$ associated with the permutation $\lambda \in \mathcal{S}_N$,

$$E[m(\lambda)] = \sum_{i \in [N]} |r_i - b_{\lambda(i)}|^p. \tag{12}$$

The duality transformation (9), that is,

$$\sigma \to \sigma \circ \tau \circ I, \tag{13}$$

$$\pi \to \pi \circ I, \tag{14}$$

interchanges the two matchings because

$$\begin{aligned} \mu_1 &:= \pi \circ \sigma^{-1} \to \pi \circ I \circ I \circ \tau^{-1} \circ \sigma^{-1} \\ &= \pi \circ \tau^{-1} \circ \sigma^{-1}, \end{aligned} \tag{15a}$$

$$\begin{aligned} \mu_2 &:= \pi \circ \tau^{-1} \circ \sigma^{-1} \to \pi \circ I \circ \tau^{-1} \circ I \circ \tau^{-1} \circ \sigma^{-1} \\ &= \pi \circ \sigma^{-1}, \end{aligned} \tag{15b}$$

where we used

$$I \circ \tau^{-1} \circ I = \tau. \tag{16}$$

The two matchings corresponding to the two permutations μ_1 and μ_2 have no edges in common and therefore each vertex will appear twice in the union of their edges. Remark also that

$$\mu_2 = \mu_1 \circ \sigma \circ \tau^{-1} \circ \sigma^{-1}, \tag{17}$$

which means that μ_1 and μ_2 are related by a permutation which has to be, as it is τ^{-1}, a unique cycle of length N. It follows that, if h^* is the optimal Hamiltonian cycle and m^* is the optimal assignment,

$$E[h^*] \geqslant 2\, E[m^*]. \tag{18}$$

In the case of the Euclidean assignment the scaling of the average optimal cost is known in every dimension and for every $p > 1$ [22]:

$$\overline{E[\mu^*]} \sim \begin{cases} N^{1-\frac{p}{2}} & d = 1; \\ N^{1-\frac{p}{2}} (\log N)^{\frac{p}{2}} & d = 2; \\ N^{1-\frac{p}{d}} & d > 2. \end{cases} \tag{19}$$

The scaling is anomalous in 1 and 2 dimensions, differently from the matching problem on the complete graph $\mathcal{K}_N$ where in any dimension the scaling with the number of points is always $N^{1-\frac{p}{d}}$. Indeed, also for the monopartite Euclidean TSP (that is on $\mathcal{K}_N$) in [17] it has been shown that for $p = 1$, in a finite region, with probability 1, the total cost scales according to $N^{1-\frac{p}{d}}$ in any dimension.

III. SOLUTION IN $d = 1$ FOR ALL INSTANCES

Here we shall concentrate on the one-dimensional case, where both red and blue points are chosen uniformly in the unit interval [0,1]. In our analysis we shall make use of the results for the Euclidean assignment problem in one dimension of [23] which have been obtained when in (3) is set $p > 1$. In this work it is shown that sorting both red and blue points in increasing order, the optimal assignment is defined by the identity permutation $\mathbb{1} = (1,2,\ldots,N)$. From now on, we will assume $p > 1$ and that both red and blue points are ordered, i.e., $r_1 \leqslant \cdots \leqslant r_N$ and $b_1 \leqslant \cdots \leqslant b_N$. Let

$$\tilde{\sigma}(i) = \begin{cases} 2i-1 & i \leqslant (N+1)/2 \\ 2N-2i+2 & i > (N+1)/2, \end{cases} \tag{20}$$

and

$$\begin{aligned} \tilde{\pi}(i) &= \tilde{\sigma} \circ I(i) = \tilde{\sigma}(N+1-i) \\ &= \begin{cases} 2i & i < (N+1)/2 \\ 2N-2i+1 & i \geqslant (N+1)/2, \end{cases} \end{aligned} \tag{21}$$

where the couple $(\tilde{\sigma},\tilde{\pi})$ will define a Hamiltonian cycle $\tilde{h} \in \mathcal{H}$. More precisely, according to the correspondence given in (5), it contains the edges for even N,

$$\tilde{e}_{2i-1} = \begin{cases} (r_{2i-1}, b_{2i}) & i \leqslant N/2 \\ (r_{2N-2i+2}, b_{2N-2i+1}) & i > N/2, \end{cases} \tag{22a}$$

$$\tilde{e}_{2i} = \begin{cases} (b_{2i}, r_{2i+1}) & i < N/2 \\ (b_N, r_N) & i = N/2 \\ (b_{2N-2i+1}, r_{2N-2i}) & N/2 < i < N \\ (b_1, r_1) & i = N, \end{cases} \tag{22b}$$

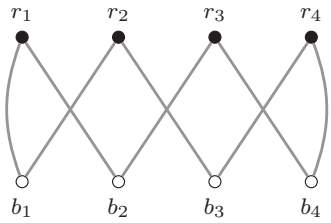

FIG. 1. The optimal Hamiltonian cycle $\tilde{h}$ for $N = 4$ blue and red points chosen in the unit interval and sorted in increasing order.

while for N odd,

$$\tilde{e}_{2i-1} = \begin{cases} (r_{2i-1}, b_{2i}) & i < (N-1)/2 \\ (r_N, b_N) & i = (N-1)/2 \\ (r_{2N-2i+2}, b_{2N-2i+1}) & i > (N-1)/2, \end{cases} \tag{23a}$$

$$\tilde{e}_{2i} = \begin{cases} (b_{2i}, r_{2i+1}) & i < (N-1)/2 \\ (b_{2N-2i+1}, r_{2N-2i}) & (N-1)/2 < i < N \\ (b_1, r_1) & i = N. \end{cases} \tag{23b}$$

The main ingredient of our analysis is the following.

Proposition III.1. For a convex and increasing cost function the optimal Hamiltonian cycle is provided by $\tilde{h}$.

This cycle is the analogous of the *criss-cross* solution introduced by Halton [24] (see Fig. 1). In his work, Halton studied the optimal way to lace a shoe. This problem can be seen as a peculiar instance of a two-dimensional bipartite Euclidean TSP with the parameter which tunes the cost $p = 1$. One year later, Misiurewicz [25] generalized Halton's result giving the least restrictive requests on the two-dimensional TSP instance to have the criss-cross cycle as the solution. Other generalizations of these works have been investigated in more recent papers [26,27]. We will show that the same criss-cross cycle has the lowest cost for the Euclidean bipartite TSP in one dimension, provided that $p > 1$. To do this, we will prove in a novel way the optimality of the criss-cross solution, suggesting two moves that lower the energy of a tour and showing that the only Hamiltonian cycle that cannot be modified by these moves is $\tilde{h}$.

We shall make use of the following moves in the ensemble of Hamiltonian cycles. Given $i,j \in [N]$ with $j > i$ we can partition each cycle as

$$h[(\sigma,\pi)] = (C_1 r_{\sigma(i)} b_{\pi(i)} C_2 b_{\pi(j)} r_{\sigma(j+1)} C_3), \tag{24}$$

where the C_i are open paths in the cycle, and we can define the operator R_{ij} that exchanges two blue points $b_{\pi(i)}$ and $b_{\pi(j)}$ and reverses the path between them as

$$\begin{aligned} h[R_{ij}(\sigma,\pi)] &:= (C_1 r_{\sigma(i)} [b_{\pi(i)} C_2 b_{\pi(j)}]^{-1} r_{\sigma(j+1)} C_3) \\ &= \left(C_1 r_{\sigma(i)} b_{\pi(j)} C_2^{-1} b_{\pi(i)} r_{\sigma(j+1)} C_3\right). \end{aligned} \tag{25}$$

Analogously by writing

$$h[(\sigma,\pi)] = (C_1 b_{\pi(i-1)} r_{\sigma(i)} C_2 r_{\sigma(j)} b_{\pi(j)} C_3), \tag{26}$$

we can define the corresponding operator S_{ij} that exchanges two red points $r_{\sigma(i)}$ and $r_{\sigma(j)}$ and reverses the path between them:

$$\begin{aligned} h[S_{ij}(\sigma,\pi)] &:= (C_1 b_{\pi(i-1)} [r_{\sigma(i)} C_2 r_{\sigma(j)}]^{-1} b_{\pi(j)} C_3) \\ &= \left(C_1 b_{\pi(i-1)} r_{\sigma(j)} C_2^{-1} r_{\sigma(i)} b_{\pi(j)} C_3\right). \end{aligned} \tag{27}$$

Two couples of points $(r_{\sigma(k)},r_{\sigma(l)})$ and $(b_{\pi(j)},b_{\pi(i)})$ have the same orientation if $(r_{\sigma(k)}-r_{\sigma(l)})(b_{\pi(j)}-b_{\pi(i)})>0$. Note that as we have ordered both sets of points this means also that $(\sigma(k),\sigma(l))$ and $(\pi(j),\pi(i))$ have the same orientation.

Then

Lemma 1. Let $E[(\sigma,\pi)]$ be the cost defined in (10). Then $E[R_{ij}(\sigma,\pi)]-E[(\sigma,\pi)]>0$ if the couples $(r_{\sigma(j+1)},r_{\sigma(i)})$ and $(b_{\pi(j)},b_{\pi(i)})$ have the same orientation and $E[S_{ij}(\sigma,\pi)]-E[(\sigma,\pi)]>0$ if the couples $(r_{\sigma(j)},r_{\sigma(i)})$ and $(b_{\pi(j)},b_{\pi(i-1)})$ have the same orientation.

Proof.

$$E[R_{ij}(\sigma,\pi)]-E[(\sigma,\pi)] = w_{(r_{\sigma(i)},b_{\pi(j)})}+w_{(b_{\pi(i)},r_{\sigma(j+1)})}-w_{(r_{\sigma(i)},b_{\pi(i)})}-w_{(b_{\pi(j)},r_{\sigma(j+1)})}, \tag{28}$$

and this is the difference between two matchings which is positive if the couples $(r_{\sigma(j+1)},r_{\sigma(i)})$ and $(b_{\pi(j)},b_{\pi(i)})$ have the same orientation (as shown in [23,28] for a weight which is an increasing convex function of the Euclidean distance). The remaining part of the proof is analogous. ■

Lemma 2. The only couples of permutations (σ,π) with $\sigma(1)=1$ such that both $(\sigma(j+1),\sigma(i))$ have the same orientation as $(\pi(j),\pi(i))$ and $(\pi(j),\pi(i-1))$ and $(\sigma(j),\sigma(i))$, for each $i,j\in[N]$ are $(\tilde\sigma,\tilde\pi)$ and its dual $(\tilde\sigma,\tilde\pi)^\star$.

Proof. We have to start our Hamiltonian cycle from $r_{\sigma(1)}=r_1$. Next we look at $\pi(N)$, if we assume now that $\pi(N)>1$, there will be a j such that our cycle would have the form $(r_1C_1r_{\sigma(j)}b_1C_2b_{\pi(N)})$, if we assume $j>1$ then $(1,\sigma(j))$ and $(\pi(N),1)$ have the opposite orientation, so that necessarily $\pi(N)=1$. In the case $j=1$ our Hamiltonian cycle is of the form (r_1b_1C), that is, (b_1Cr_1), and this is exactly of the other form if we exchange red and blue points. We assume that it is of the form (r_1Cb_1); the other form would give, at the end of the proof, $(\tilde\sigma,\tilde\pi)^\star$. Now we shall proceed by induction. Assume that our Hamiltonian cycle is of the form $(r_1b_2r_3\cdots x_kCy_k\cdots b_3r_2b_1)$ with $k<N$, where x_k and y_k are, respectively, a red point and a blue point when k is odd and vice versa when k is even. Then y_{k+1} and x_{k+1} must be in the walk C. If y_{k+1} it is not the point on the right of x_k the cycle has the form $(r_1b_2r_3\cdots x_ky_sC_1y_{k+1}x_l\cdots y_k\cdots b_3r_2b_1)$ but then (x_l,x_k) and (y_{k+1},y_s) have opposite orientation, which is impossible, so that $s=k+1$, that is, the point on the right of x_k. Where is x_{k+1}? If it is not the point on the left of y_k the cycle has the form $(r_1b_2r_3\cdots x_ky_{k+1}\cdots y_lx_{k+1}C_1x_s\cdots y_k\cdots b_3r_2b_1)$, but then (x_s,x_{k+1}) and (y_k,y_l) have the opposite orientation, which is impossible, so that $s=k+1$, that is, the point on the left of y_k. We have now shown that the cycle has the form $(r_1b_2r_3\cdots y_{k+1}Cx_{k+1}\cdots b_3r_2b_1)$ and can proceed until C is empty. ■

The case with $N=3$ points is explicitly investigated in Appendix.

Now that we have understood what is the optimal Hamiltonian cycle, we can look in more detail at what are the two matchings which enter in the decomposition we used in (11). As $\tilde\pi=\tilde\sigma\circ I$ we have that

$$I=\tilde\sigma^{-1}\circ\tilde\pi=\tilde\pi^{-1}\circ\tilde\sigma. \tag{29}$$

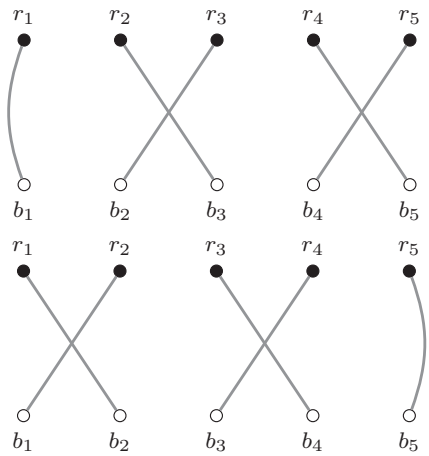

FIG. 2. Decomposition of the optimal Hamiltonian cycle $\tilde h$ for $N=5$ in two disjoint matchings $\tilde\mu_2$ and $\tilde\mu_1$.

As a consequence both permutations associated with the matchings appearing in (11) for the optimal Hamiltonian cycle are involutions:

$$\tilde\mu_1\equiv\tilde\pi\circ\tilde\sigma^{-1}=\tilde\sigma\circ I\circ\tilde\sigma^{-1}=\tilde\sigma\circ\tilde\pi^{-1} =[\tilde\pi\circ\tilde\sigma^{-1}]^{-1}, \tag{30a}$$

$$\tilde\mu_2\equiv\tilde\pi\circ\tau^{-1}\circ\tilde\sigma^{-1}=\tilde\sigma\circ I\circ\tau^{-1}\circ I\circ\tilde\pi^{-1} =[\tilde\pi\circ\tau^{-1}\circ\tilde\sigma^{-1}]^{-1}, \tag{30b}$$

where we used (16). This implies that those two permutations have at most cycles of period two, a fact which reflects a symmetry by exchange of red and blue points.

When N is odd it happens that

$$I\circ\tilde\sigma\circ I=\tilde\sigma\circ\tau^{-\frac{N-1}{2}}, \tag{31}$$

so that

$$I\circ\tilde\pi\circ I=I\circ\tilde\sigma\circ I\circ I=\tilde\sigma\circ\tau^{-\frac{N-1}{2}}\circ I =\tilde\pi\circ I\circ\tau^{-\frac{N-1}{2}}\circ I=\tilde\pi\circ\tau^{\frac{N-1}{2}}. \tag{32}$$

It follows that the two permutations in (30a) and (30b) are conjugate by I,

$$I\circ\tilde\pi\circ\tau^{-1}\circ\tilde\sigma^{-1}\circ I=\tilde\pi\circ\tau^{\frac{N-1}{2}}\circ\tau\circ\tau^{\frac{N-1}{2}}\circ\tilde\sigma^{-1} =\tilde\pi\circ\tilde\sigma^{-1}, \tag{33}$$

so that, in this case, they have exactly the same numbers of cycles of order 2. Indeed we have

$$\tilde\mu_1=(2,1,4,3,6,\ldots,N-1,N-2,N), \tag{34a}$$

$$\tilde\mu_2=(1,3,2,5,4,\ldots N,N-1), \tag{34b}$$

and they have $\frac{N-1}{2}$ cycles of order 2 and one fixed point. See Fig. 2 for the case $N=5$.

In the case of even N the two permutations have not the same number of cycles of order 2; indeed one has no fixed point and the other has two of them. More explicitly,

$$\tilde\mu_1=(2,1,4,3,6,\ldots,N,N-1), \tag{35a}$$

$$\tilde\mu_2=(1,3,2,5,4,\ldots N-1,N-2,N). \tag{35b}$$

See Fig. 3 for the case $N=4$.

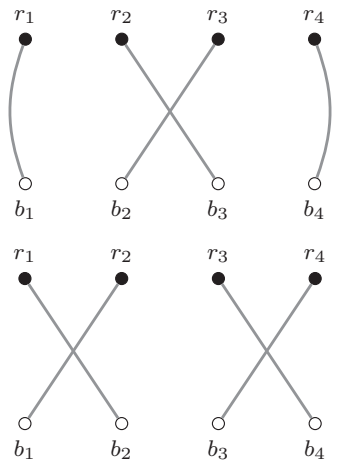

FIG. 3. Decomposition of the optimal Hamiltonian cycle $\tilde{h}$ for $N = 4$ in the two disjoint matchings $\tilde{\mu}_2$ and $\tilde{\mu}_1$.

IV. EVALUATION OF THE COST

Here we will evaluate the cost of the optimal Hamiltonian cycle $\tilde{h}$ for $\mathcal{K}_{N,N}$,

$$E_N(\tilde{h}) = |r_1 - b_1|^p + |r_N - b_N|^p + \sum_{i=1}^{N-1}[|b_{i+1} - r_i|^p + |r_{i+1} - b_i|^p]. \quad (36)$$

Assume that both red and blue points are chosen according to the law ρ and let

$$\Phi_\rho(x) := \int_0^x ds\, \rho(s) \quad (37)$$

be its *cumulative*. The probability that among N randomly located points the kth point is in the interval $(x,x + dx)$, is given by

$$\Pr{}_\rho[x_k \in dx] = k \binom{N}{k} \Phi_\rho^{k-1}(x)[1 - \Phi_\rho(x)]^{N-k} \rho(x)\, dx. \quad (38)$$

In particular for $k = 1$,

$$\Pr{}_\rho[x_1 \in dx] = N\,[1 - \Phi_\rho(x)]^{N-1} \rho(x)\, dx, \quad (39)$$

and $k = N$,

$$\Pr{}_\rho[x_N \in dx] = N\, \Phi_\rho^{N-1}(x) \rho(x)\, dx. \quad (40)$$

Given two sequences of N points, the probability for the difference ϕ_k in the position between the $(k + 1)$th and the kth points is

$$\Pr{}_\rho[\phi_k \in d\phi] = k(k+1) \binom{N}{k} \binom{N}{k+1} d\phi_k \times \int dx\, dy\, \rho(x)\, \rho(y) \delta(\phi_k - y + x)\, \Phi_\rho(y)\, [1 - \Phi_\rho(x)], \times [\Phi_\rho(x)\Phi_\rho(y)]^{k-1}[(1 - \Phi_\rho(x))(1 - \Phi_\rho(y))]^{N-k-1}. \quad (41)$$

Let us now focus on the simple case in which the law ρ is flat, then $\Phi_\rho(x) = x$.

$$\overline{|r_1 - b_1|^p} = N^2 \int_0^1 dx\, dy\, [(1 - x)(1 - y)]^{N-1} |x - y|^p = N^2 \int_0^1 dx\, dy\, (xy)^{N-1} |x - y|^p = \overline{|r_N - b_N|^p}. \quad (42)$$

For $p = 2$,

$$\overline{|r_1 - b_1|^2} = \frac{2N}{(N + 1)^2(N + 2)}, \quad (43)$$

and

$$\overline{|b_{k+1} - r_k|^2} = \overline{|r_{k+1} - b_k|^2} = k(k + 1) \binom{N}{k} \binom{N}{k + 1} \times \int_0^1 dx\, dy\, (x-y)^2 y(1-x)(xy)^{k-1}[(1-x)(1-y)]^{N-k-1} = \frac{2(k + 1)(N - k + 1)}{(N + 1)^2(N + 2)}, \quad (44)$$

and

$$\sum_{k=1}^{N-1} \frac{2(k + 1)(N - k + 1)}{(N + 1)^2(N + 2)} = \frac{1}{3} \frac{(N + 6)(N - 1)}{(N + 1)(N + 2)}. \quad (45)$$

In conclusion, the average cost for the flat distribution and $p = 2$ is exactly

$$\overline{E_N^{(2)}} = \frac{2}{3} \frac{N^2 + 4N - 3}{(N + 1)^2}. \quad (46)$$

If we recall that for the assignment the average optimal total cost is exactly $\frac{1}{3} \frac{N}{N+1}$, the difference between the average optimal total cost of the bipartite TSP and twice the assignment is

$$\frac{2}{3}\left[\frac{N^2 + 4N - 3}{(N + 1)^2} - \frac{N}{N + 1}\right] = \frac{1}{3} \frac{N - 1}{(N + 1)^2} \geqslant 0, \quad (47)$$

and vanishes for infinitely large N. Note that the limiting value is reached from above for the TSP and from below for the assignment. We plot in Fig. 4 the numerical results of the average optimal cost for different number of points.

It is also interesting to look at the contribution from the two different matchings in which we have subdivided the optimal Hamiltonian cycle. In the case of N odd we have for one of them the average cost,

$$\frac{2N}{(N + 1)^2(N + 2)} + 2 \sum_{k=1}^{\frac{N-1}{2}} \frac{4k(N - 2k + 2)}{(N + 1)^2(N + 2)} = \frac{1}{3} \frac{N^2 + 4N - 3}{(N + 1)^2}, \quad (48)$$

and also for the other,

$$\frac{2N}{(N + 1)^2(N + 2)} + 2 \sum_{k=1}^{\frac{N-1}{2}} \frac{2(2k + 1)(N - 2k + 1)}{(N + 1)^2(N + 2)} = \frac{1}{3} \frac{N^2 + 4N - 3}{(N + 1)^2}. \quad (49)$$

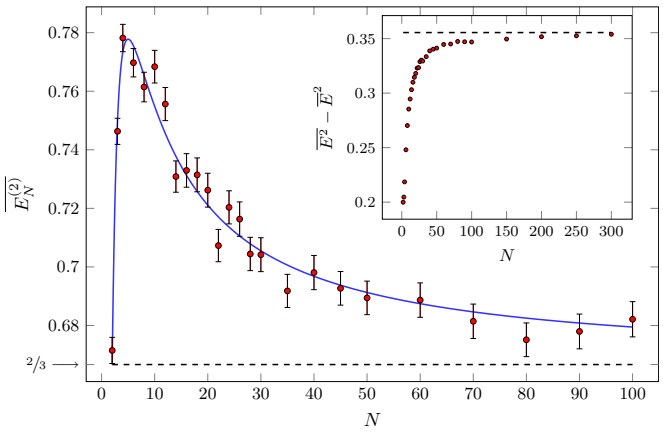

FIG. 4. Numerical results for $\overline{E_N^{(2)}}$ for several values of N. The continuous line represents the exact prediction given in (46) and the dashed line gives the value for infinitely large N. For every N we have used 10^4 instances. In the inset we show the numerical results for the variance of the cost $E_N^{(2)}$ obtained using the exact solution provided by (20) and (21). The dashed line represents the theoretical large N asymptotic value. Error bars are also plotted but they are smaller than the mark size.

In the case of N even we have for the matching with two fixed points the average cost,

$$\frac{4N}{(N+1)^2(N+2)} + 2\sum_{k=1}^{\frac{N-2}{2}} \frac{2(2k+1)(N-2k+1)}{(N+1)^2(N+2)} = \frac{1}{3}\frac{N^2+4N-6}{(N+1)^2}, \tag{50}$$

while for the other with no fixed points,

$$2\sum_{k=1}^{\frac{N-2}{2}} \frac{4k(N-2k+2)}{(N+1)^2(N+2)} = \frac{1}{3}\frac{N^2+4N}{(N+1)^2}, \tag{51}$$

which then has a cost higher at the order N^{-2}.

V. ASYMPTOTIC ANALYSIS FOR THE OPTIMAL AVERAGE COST

Motivated by the preceding discussion, one can try to perform a more refined analysis in the thermodynamic limit. In the asymptotic regime of large N, in fact, only the term with a sum on i in (36) will contribute, and each of the two terms will provide an equal optimal matching contribution. Proceeding as in the case of the assignment [23,29], one can show that the random variables ϕ_k defined above Eq. (41) converge (in a weak sense specified by Donsker's theorem) to $\phi(s)$, which is a difference of two Brownian bridge processes [30].

One can write the rescaled average optimal cost as

$$\overline{E_p} \equiv \lim_{N\to\infty} N^{\frac{p}{2}-1}\,\overline{E_N^{(p)}}, \tag{52}$$

where we have denoted with a bar $\overline{\cdot}$ the average over all the instances. By starting at finite N with the representation (41), the large N limit can be obtained setting $k = Ns + \frac{1}{2}$ and introducing the variables ξ, η, and φ such that

$$x = s + \frac{\xi}{\sqrt{N}}, \quad y = s + \frac{\eta}{\sqrt{N}}, \quad \phi_k = \frac{\varphi(s)}{\sqrt{N}}, \tag{53}$$

in such a way that s is kept fixed when $N \to +\infty$. Using the fact that

$$\Phi_\rho^{-1}(x) \approx \Phi_\rho^{-1}\left(s + \frac{\xi}{\sqrt{N}}\right) = \Phi_\rho^{-1}(s) + \frac{\xi}{\sqrt{N}\left(\rho\circ\Phi_\rho^{-1}\right)(s)}, \tag{54}$$

we obtain, at the leading order,

$$\begin{aligned}
&\Pr[\varphi(s)\in d\varphi] \\
&= d\varphi \iint \delta\left(\varphi - \frac{\eta-\xi}{\rho\left(\Phi_\rho^{-1}(s)\right)}\right)\frac{\exp\left(-\frac{\xi^2+\eta^2}{2s(1-s)}\right)}{2\pi s(1-s)}d\xi\, d\eta \\
&= \frac{\left(\rho\circ\Phi_\rho^{-1}\right)(s)}{\sqrt{4\pi s(1-s)}}\exp\left\{-\frac{\left[\left(\rho\circ\Phi_\rho^{-1}\right)(s)\right]^2}{4s(1-s)}\varphi^2\right\}d\varphi,
\end{aligned} \tag{55}$$

that implies that

$$\begin{aligned}
\overline{E_p} &= 2\int_0^1 \overline{|\varphi(s)|^p}\,ds \\
&= 2\int_0^1 ds\,\frac{s^{\frac{p}{2}}(1-s)^{\frac{p}{2}}}{\left[\left(\rho\circ\Phi_\rho^{-1}\right)(s)\right]^p}\int_{-\infty}^{+\infty} d\varphi\,|\varphi|^p\frac{\exp\left[-\frac{\varphi^2}{4}\right]}{\sqrt{4\pi}} \\
&= \frac{2^{1+p}}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)\int_0^1 ds\,\frac{s^{\frac{p}{2}}(1-s)^{\frac{p}{2}}}{\left[\left(\rho\circ\Phi_\rho^{-1}\right)(s)\right]^p} \\
&= \frac{2^{1+p}}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)\int_0^1 dx\,\frac{\Phi_\rho^{\frac{p}{2}}(x)(1-\Phi_\rho(x))^{\frac{p}{2}}}{\rho^{p-1}(x)}.
\end{aligned} \tag{56}$$

In the particular case of a flat distribution the average cost converges to

$$\overline{E_p} = \frac{2^{1+p}}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)\int_0^1 ds\,[s(1-s)]^{\frac{p}{2}} = 2\,\frac{\Gamma\left(\frac{p}{2}+1\right)}{p+1}, \tag{57}$$

which is two times the value of the optimal matching. For $p = 2$ this gives $\overline{E_2} = 2/3$, according to the exact result (46). Formula (55) becomes

$$p_s(x) = \overline{\delta(\varphi(s)-x)} = \frac{e^{-\frac{x^2}{4s(1-s)}}}{\sqrt{4\pi s(1-s)}}, \tag{58}$$

and similarly (see, for example, Appendix A in [29]) it can be derived that the joint probability distribution $p_{t,s}(x,y)$ for $\varphi(s)$ is (for $t < s$) a bivariate Gaussian distribution,

$$\begin{aligned}
p_{t,s}(x,y) &= \overline{\delta(\varphi(t)-x)\,\delta(\varphi(s)-y)} \\
&= \frac{e^{-\frac{x^2}{4t}-\frac{(x-y)^2}{4(s-t)}-\frac{y^2}{4(1-s)}}}{4\pi\sqrt{t(s-t)(1-s)}}.
\end{aligned} \tag{59}$$

This allows one to compute, for a generic $p > 1$, the average of the square of the rescaled optimal cost,

$$\overline{E_p^2} = 4\int_0^1 dt\int_0^1 ds\,\overline{|\varphi(s)|^p|\varphi(t)|^p}, \tag{60}$$

which is 4 times the corresponding one of a bipartite matching problem.

In the case $p = 2$, the average in Eq. (60) can be evaluated by using the Wick theorem for expectation values in a Gaussian distribution,

$$\overline{E_2^2} = 4\int_0^1 ds \int_0^s dt \int_{-\infty}^{\infty} dx\, dy\, p_{t,s}(x,y)\, x^2 y^2 = \frac{4}{5}, \quad (61)$$

and, therefore,

$$\overline{E_2^2} - \overline{E_2}^2 = \frac{16}{45} = 0.3\bar{5}. \quad (62)$$

This result is in agreement with the numerical simulations (see inset of Fig. 4) and proves that the rescaled optimal cost is not a self-averaging quantity.

VI. CONCLUSION AND PERSPECTIVES

In this work we studied the random Euclidean bipartite TSP in one dimension using a weight function which is a power p of the Euclidean distance between red and blue points. The complete bipartite graph is a special case of a more general problem. The motivation of this choice is twofold: On one hand in the one-dimensional case we have been able to address clearly the connection between this problem and the assignment and on the other hand we expect the bipartite TSP to be more easily tractable than its monopartite counterpart in more than one dimension. Traveling-salesman problems on bipartite graphs may also turn out in practical situations (for instance, a vehicle needing to visit a set of destinations and a set of charging stations). We provide an explicit solution in the convex case $p > 1$, giving the best cycle for each disorder instance of the problem. This allowed us to compute explicitly the average optimal cost when $p = 2$ and for every number of points N. Interestingly, the value of the average optimal cost turned out to be twice the average optimal cost of the assignment problem. In the continuum limit we were also able to find the average optimal cost for generic exponent p, using the relation of the one-dimensional assignment with the Brownian bridge process [23]. In the same thermodynamic limit we computed the variance of the distribution of the optimal costs; since we get a nonvanishing result, we deduce that the average optimal cost is not a self-averaging quantity. This feature is present also in the case of the assignment problem, where the average optimal cost has been shown to be self-averaging only in $d > 2$ [31].

In the field of combinatorial optimization problems, especially in mean field cases (i.e., where the random variables are not correlated), the theory of spin glasses and disordered systems can be used to calculate statistical properties of the optimal solution analytically [8]. In such cases this approach also sheds light on the design of new algorithms to find solutions [9]. However, it is not clear in general how to apply these techniques (beyond expanding around the mean field case [32,33]), when correlations play an important role, as happens when the graph is embedded in Euclidean spaces. For other problems besides the TSP, analysis of the one-dimensional case has enabled progress in the study of higher-dimensional cases [34]. As a consequence, a relevant question is whether the relations we obtained in one dimension continue to exist also in $d > 1$, where the bipartite TSP is an NP-complete problem. Recently, we computed exactly the cost and a two-point correlation function in $d = 2$ for the assignment problem [22,34,35]. The investigation of the connections between these two combinatorial optimization problems is material for future work.

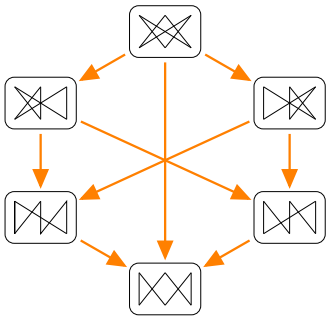

FIG. 5. The whole diagram describing the $N = 3$ case. In the squared boxes the various cycle configurations are represented. Lower boxes correspond to lower costs. All the possible moves suggested in Lemma 1 are represented by orange arrows.

ACKNOWLEDGMENT

The authors are grateful to R. Capelli for useful advice regarding the simulations performed.

APPENDIX: THE CASE $N = 3$

In the case $N = 2$ there is only one Hamiltonian cycle, that is, $\tilde{h}$. The first nontrivial case is $N = 3$. There are six Hamiltonian cycles. If we fix the starting point to be r_1 there are only two possibilities for the permutation σ of the red points, that is, (1,2,3) and (1,3,2). One is the dual of the other. We can restrict to the (1,3,2) by removing the degeneracy in the orientation of the cycles. Indeed $\tilde{\sigma}$ is exactly (1,3,2) according to (20). With this choice the six cycles are in correspondence with the permutations $\pi \in \mathcal{S}_3$ of the blue points. We sort in increasing order both the blue and red points. We have

$$E(\pi) = |r_1 - b_{\pi(1)}|^p + |r_1 - b_{\pi(3)}|^p + |r_3 - b_{\pi(2)}|^p + |r_3 - b_{\pi(1)}|^p + |r_2 - b_{\pi(3)}|^p + |r_2 - b_{\pi(2)}|^p. \quad (A1)$$

The optimal solution is $\tilde{\pi} = (2,3,1)$. The permutations (1,3,2) and (3,2,1) have always a greater cost than $\tilde{\pi}$, indeed the corresponding cycles are $(r_1 b_1 r_3 b_3 r_2 b_2)$ and $(r_1 b_3 r_3 b_2 r_2 b_1)$, where we have colored in orange the path that, according to Lemma 1, can be reversed to lower the total cost. Doing this we obtain the optimal cycle in both cases. Notice that, since we can label each cycle using only the π permutation, we can restrict ourselves to moves that only involve blue points. Since there are three blue points, these moves will always reverse paths of the form $b_i r_j b_k$, so they correspond simply to a swap in the permutation π. Therefore our moves cannot be used to reach the optimal cycle from every starting cycle. A diagram showing all the possible moves is shown in Fig. 5. In conclusion, the cost function makes $\mathcal{S}_3$ a *poset* with an absolute minimum and an absolute maximum. The permutation (2,3,1)

is preceded by both (1,3,2) and (3,2,1), which cannot be compared between them, but both precede (1,2,3) and (3,1,2), which cannot be compared between them. (2,1,3) is the greatest element.

We compute the average costs for all the permutations. Using the same techniques used in Sec. IV, we get, for the $p = 2$ case,

$$\begin{aligned}\overline{E[(2,3,1)]} = \tfrac{3}{4} < \overline{E[(1,3,2)]} = \overline{E[(3,2,1)]} = \tfrac{7}{8} \\ < \overline{E[(1,2,3)]} = \overline{E[(3,1,2)]} = \tfrac{9}{8} \\ < \overline{E[(2,1,3)]} = \tfrac{5}{4}. \end{aligned} \tag{A2}$$

[1] E. Lawler, D. Shmoys, A. Kan, and J. Lenstra, *The Traveling Salesman Problem* (John Wiley & Sons, New York, 1985).
[2] K. Menger, Ergebnisse eines mathematischen kolloquiums **2**, 11 (1932).
[3] http://www.claymath.org/millennium-problems/p-vs-np-problem.
[4] G. Reinelt, *The Traveling Salesman: Computational Solutions for TSP Applications* (Springer-Verlag, Berlin, 1994).
[5] C. H. Papadimitriou, Theor. Comput. Sci. **4**, 237 (1977).
[6] S. Kirkpatrick, C. Gelatt, and M. Vecchi, Science **220**, 671 (1983).
[7] N. Sourlas, Europhys. Lett. **2**, 919 (1986).
[8] M. Mézard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications* (World Scientific Publishing, Singapore, 1987), Vol. 9.
[9] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).
[10] J. Vannimenus and M. Mézard, J. Phys. Lett. **45**, L1145 (1984).
[11] H. Orland, J. Physique Lett. **46**, 763 (1985).
[12] M. Mézard and G. Parisi, J. Phys. **47**, 1285 (1986).
[13] M. Mézard and G. Parisi, Europhys. Lett. **2**, 913 (1986).
[14] W. Krauth and M. Mézard, Europhys. Lett. **8**, 213 (1989).
[15] S. Ravanbakhsh, R. Rabbany, and R. Greiner, Advances in Neural Information Processing Systems **1**, 289 (2014).
[16] J. Wastlund, Acta Mathematica **204**, 91 (2010).
[17] J. Beardwood, J. H. Halton, and J. M. Hammersley, Proc. Cambridge Philos. Soc. **55**, 299 (1959).
[18] M. Steele, Ann. Probability **9**, 365 (1981).
[19] R. M. Karp and M. Steele, *The Traveling Salesman Problem* (John Wiley and Sons, New York, 1985).
[20] A. G. Percus and O. C. Martin, Phys. Rev. Lett. **76**, 1188 (1996).
[21] N. J. Cerf, J. H. B. de Monvel, O. Bohigas, O. C. Martin, and A. G. Percus, J. Phys. I **7**, 117 (1997).
[22] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro, Phys. Rev. E **90**, 012118 (2014).
[23] E. Boniolo, S. Caracciolo, and A. Sportiello, J. Stat. Mech. (2014) P11023.
[24] J. H. Halton, Math. Intell. **17**, 36 (1995).
[25] M. Misiurewicz, Math. Intell. **18**, 32 (1996).
[26] B. Polster, Nature (London) **420**, 476 (2002).
[27] A. García and J. Tejel, Eur. J. Oper. Res. **257**, 429 (2017).
[28] R. McCann, Proc. R. Soc. A **455**, 1341 (1999).
[29] S. Caracciolo and G. Sicuro, Phys. Rev. E **90**, 042112 (2014).
[30] S. Caracciolo, M. P. D'Achille, and G. Sicuro, Phys. Rev. E **96**, 042102 (2017).
[31] J. Houdayer, J. B. de Monvel, and O. Martin, Eur. Phys. J. B **6**, 383 (1998).
[32] M. Mézard and G. Parisi, J. Phys. **49**, 2019 (1988).
[33] C. Lucibello, G. Parisi, and G. Sicuro, Phys. Rev. E **95**, 012302 (2017).
[34] S. Caracciolo and G. Sicuro, Phys. Rev. Lett. **115**, 230601 (2015).
[35] S. Caracciolo and G. Sicuro, Phys. Rev. E **91**, 062125 (2015).