Enumerations of the Hamiltonian walks on a cubic sublattice

1994 J. Phys. A: Math. Gen. 27 6231

(http://iopscience.iop.org/0305-4470/27/18/030)

J. Phys. A: Math. Gen. 27 (1994) 6231–6236. Printed in the UK

Enumerations of the Hamiltonian walks on a cubic sublattice

Vijay S Pande†, Chris Joerg‡, Alexander Yu Grosberg†§ and Toyoichi Tanaka†

† Department of Physics and Center for Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

‡ Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 7 April 1994, in final form 6 May 1994

Abstract. A massively parallel supercomputer was used to exhaustively enumerate all of the Hamiltonian walks for simple cubic sublattices of four different sizes (up to $3 \times 4 \times 4$). The behaviour of the logarithm of the number of walks was found to be linear in the number of vertices in the lattice. The linear fit is shown to agree also with the asymptotic limit of the Flory mean field theoretical estimate. Thus, we suggest that the fit obtained yields the number of walks for any size fragment of the cubic lattice to logarithmic accuracy. The significance of this result to the validity of polymer models is also discussed.

A Hamiltonian walk is defined to be a walk over some graph such that each vertex is visited once and only once. In general, Hamiltonian walks are known to be one of the most challenging and important issues in graph theory. As for graphs of cubic sublattices, exhaustive enumeration of Hamiltonian walks is especially important in the physics of heteropolymers. Indeed, Hamiltonian walks on the sublattices are naturally identified with maximally compact conformations of polymer chains. In heteropolymers, such as proteins, there may be one single conformation, which is practically fully compact and which strongly dominates the partition function of the system. Thus, Monte Carlo sampling is not sufficient in this case, and exhaustive enumeration of conformations is required.

This was first performed by Shakhnovich and Gutin [1] when enumerating the 103 346 Hamiltonian walks on the $3 \times 3 \times 3$ cubic sublattice in order to verify the phase transition of heteropolymers predicted analytically. The fact that delicate effects of the analytic theory were reproduced shows that even a small sublattice can be an effective model. However, there are some properties not present in the $3 \times 3 \times 3$ case, such as pseudo-knots. Thus, enumeration of even the $3 \times 3 \times 4$ case (which includes pseudo-knots) can shed light on new physical properties.

The enumeration algorithm is formulated as follows. We can consider any lattice in terms of the graph connecting the lattice sites. Consider all of the (not necessarily self-avoiding) walks of length N on an infinite lattice of coordination number z. At each lattice point, we have z possible different directions to travel in order to reach a new site. These walks can be described as a tree of N levels with z branches at each node, each corresponding to a possible choice of direction to the next site. The enumeration of the possible walks is merely the counting of the number of branches of length N of this 'ideal' tree. We now impose the condition that the lattice is finite, say $l \times m \times n$. We must now

§ On leave from the Institute of Chemical Physics, Russian Academy of Sciences, Moscow 117977, Russia.

0305-4470/94/186231+06$19.50

remove the branches of the ideal tree which corresponds to walks that are not contained in the new boundaries (for example, the walk consisting of N steps in a single direction is no longer in the set of possible walks when l, m, n are less than N). The addition of the constraint of self-avoidance further removes branches from the tree. We study the case of Hamiltonian walks, i.e. in the above notation $N = l \times m \times n$. Thus, the enumeration of all of the Hamiltonian walks is the counting of the number of branches of length N of this new 'restricted' tree.

In order to ascertain which sub-branches of the original ideal tree are removed, we must follow down the sub-branches of the ideal tree until we reach the end of the branch. A branch ends either when the walk is of length N or when there are no other possible sub-branches (for example, when a self-avoiding walk blocks itself off). Now, we back up one level of the tree and continue the procedure on a new sub-branch. In this way, all of the branches of the tree are exhaustively traversed in a very systematic manner.

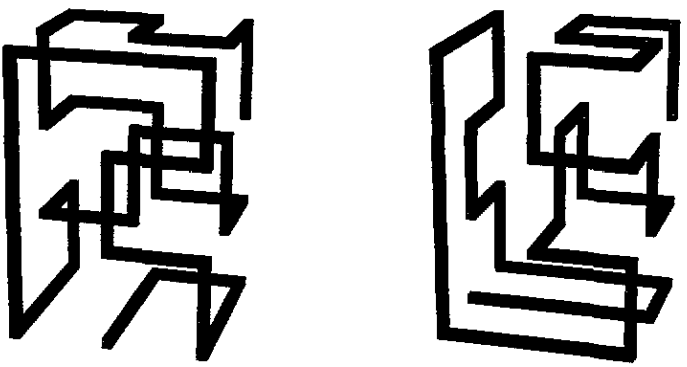

Figure 1. These two 36-site walks are related by mirror symmetry. Thus, only one is included in the enumeration procedure.

Using the prescription above, there will be some walks related by symmetry (e.g. rotations and reflections). For example, consider the two walks shown in figure 1. They are related by symmetry, in this case a reflection. We do not wish to include both of these walks, so we used 'starting paths' which break all possible symmetries. We start enumeration, i.e. the traversal of the ideal tree, only for those sub-branches of the last node in each starting path. In this way, we remove branches related by symmetry. There are in fact many starting paths necessary for several reasons: (i) there are several different points (unrelated by symmetry) where one can start the walk; (ii) there are many symmetries to break. Therefore, we have devised an algorithm to generate these starting paths. This algorithm will be discussed in the appendix.

Note that we have neglected one transformation: the reversal of the start and end of the walk. For heteropolymers, we want to include walks related by this symmetry, as the polymer sequences are not invariant with respect to sequence reversal. However, this may not be appropriate for other applications of Hamiltonian walks and should therefore be addressed accordingly. We also note that the arguments presented here and in the appendix can be easily modified to handle unusual lattices, such as unvisitable sites (used to model a 'target site' in polymer models), lattice dislocations, and other lattice aberrations, since unusual lattice topologies can be easily described in terms of the graph connecting the sites and the symmetries relating orientations of this graph.

The number of Hamiltonian walks increases exponentially with the number of vertices, so in order to gain the necessary computational speed to calculate the number of walks on larger sublattices, we employed two techniques. The most significant technique utilized was the use of a massively parallel computer (128-node Thinking Machines CM-5) and a parallel version of the tree enumeration algorithm. This parallel version used the method of 'continuation-passing threads' [3], i.e. which includes a random work-stealing scheduler able to assign subtrees to different processors and dynamically pass work (i.e. sub-branches to enumerate) to inactive processors as necessary. The throughput of the parallel algorithm was found to scale linearly with the number of processors.

The second technique used was the addition of simple checks to see if we can end the search down a branch early. Each time a node is added to the walk, we check each neighbour of that node to see if it is surrounded by nodes which have already been visited. If so, then the node can never be visited, and if that node has not yet been visited, then the partial path produced so far can never lead to a valid walk; thus, we do not need to search down this path any further. Also, we keep track of how many unvisited nodes have only one unvisited neighbour. Clearly, in a successful walk, such a node must be the last node of the walk. So if we ever find two such nodes, we can safely stop the search down this partial path. These 'blocked neighbour' checks provided one to two orders of magnitude speed improvement over prior algorithms.

These two improvements yielded sufficient computational power to enumerate the Hamiltonian walks on the $3 \times 3 \times 4$ and $3 \times 4 \times 4$ sublattices. The results are summarized in tables 1–3.

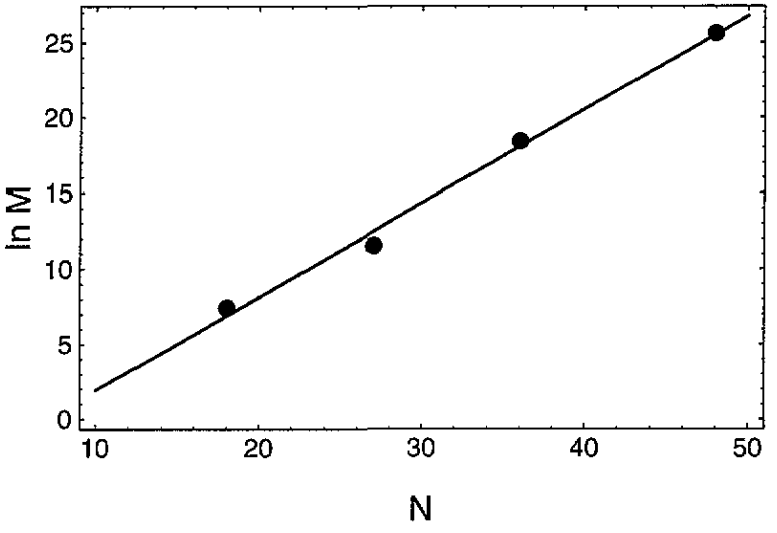

Figure 2. Logarithm of the number of walks (M) versus the number of sites (N), for $N = 18$, 27, 36, 48. We see that the curve is essentially linear.

With four lattice sizes ($N = 18, 27, 36, 48$), it may be possible to see some trend in the number of walks (M) as a function of the number of lattice sites (N). In figure 2, the natural logarithm of the number of walks is plotted versus N. We fitted a linear relation of the form

$$\ln M = \alpha + \beta N \tag{1}$$

Table 1. Summary of enumeration data, where N is the number of sites and M is the number of walks unrelated by symmetry.

N	M	CPU time†	Starting paths
18	1 711	$\ll 1$ s	27
27	103 346	0.2 s	35
36	84 731 192	5 min	816
48	134 131 827 475	64 h	3579

† CPU time given for 128-node CM-5.

Table 2. Number of Hamiltonian walks for a $3 \times 3 \times 4$ cubic sublattice for each different starting point unrelated by symmetry. We use the following convention for numbering sites on an $l \times m \times n$ sublattice: $p(x, y, z) = x + ly + lmz$.

Site type	Starting site	Number of walks
Corner	0	28 186 048
Short edge	1	13 648 609
Long edge	9	16 166 505
Small face	4	5 298 397
Large face	10	18 287 284
Inside	13	3 144 349
Total		84 731 192

Table 3. Number of Hamiltonian walks for a $3 \times 4 \times 4$ cubic sublattice for each different starting point unrelated by symmetry.

Site type	Starting site	Number of walks
Corner	0	31 323 890 329
Short edge	1	9 646 363 521
Long edge	3	42 177 079 725
Small face	4	21 460 509 753
Large face	17	20 883 800 432
Inside	16	8 640 183 715
Total		134 131 827 475

with $\alpha = -4.3 \pm 1.2$ and $\beta = 0.62 \pm 0.04$ (R^2 of the fit: 0.99). Note that while it is trivial to calculate the number of walks for $N < 18$ (i.e. $N = 8$ and 12), the inclusion of these points does not alter (within error) the linear fit or the arguments to follow; however, as discreteness effects should become great in these cases, we exclude them. Thus, we find that this fit works well for the region of small $N \leqslant 48$.

On the other hand, the Flory [2] mean field calculation of the entropy of polymer melt is known to be applicable to the estimation of the number of compact globular conformations in the $N \to \infty$ limit. Indeed, the conceptual foundation of the Flory treatment is the restriction imposed on the addition of new monomers within the constraints of the avoidance of occupied sites and chain connectivity. This kind of argument is equally applicable to both a macroscopic melt of different long chains, and a large globule of one single chain, as the two systems differ only in the contributions of the independent chains mixing entropy, which is negligible in the long-chain melt, and of surface effects, which are negligible in the thermodynamic limit. Therefore, in the $N \to \infty$ limit we have the estimate

$$M \approx \left(\frac{z-1}{e}\right)^N \tag{2}$$

where z is the coordination number of the lattice. The question is, however, how large N should be to validate this approximation. This problem is similar in spirit to the nature of the convergence of other mean field theories, or even the central limit theorem.

It turns out that in fact equations (1) and (2) agree very well, thus validating the extrapolation of equation (1) for the entire region of $N \to \infty$. We can formally transform equation (2) into equation (1) by saying that

$$z = 1 + \exp[1 + \beta + \alpha/N]. \tag{3}$$

In the $N \to \infty$ limit, we have $z = 1 + \exp[1 + \beta]$. Using our fit for β, we calculate $z = 1 + \exp[\alpha] = 6.1 \pm 0.2$, which compares well with the exact value of 6 for the simple cubic lattice. As equation (1) agrees with the results of exact enumeration in the regime $N \approx \mathcal{O}(10^2)$ as well as the Flory theory in the $N \to \infty$ limit, we suggest that equation (1) may be used to derive the number of walks for arbitrary N to logarithmic accuracy.

It is worthwhile to note that the point for $N = 27$ in figure 2 is definitely below the interpolation straight line. This might be related to the fact that this is the case for a maximally symmetric cubic shape. We are indebted to Dr A Gutin for the comment on a similar effect on the 2D lattice [6].

Thus, in terms of models of polymers, the polymeric entropy of small cubic lattice polymer models seems to be valid at least to the mean field approximation, and therefore the results which rely heavily on the nature of the conformations, such as heteropolymer theory, obtained from even small lattice models have some physical meaning. As one examines longer chains, the system starts to exhibit other physical properties, such as the presence of pseudo-trefoils in 36-mers [7] and more complicated topologies in larger sublattices. However, in these cases the effect of the lattice model in modelling of polymer topology, for example, is unclear.

In conclusion, as equation (1) yields $M \approx 2 \times 10^{15}$ for $N = 64$, it seems that the enumeration of the Hamiltonian walks on the $4 \times 4 \times 4$ sublattice is several orders of magnitude out of reach using our current algorithm and supercomputer power. However, perhaps this estimate is slightly pessimistic, as sublattices with a cubic shape are expected to have less conformations than predicted by our fit. Also, the case $N = 48$, while possible to enumerate, is still extremely CPU time-consuming and therefore cannot be used routinely in any current polymer modelling scheme. However, enumeration of $N = 36$ is not very CPU time-consuming. Furthermore, there are fundamental differences between the previously enumerated case of $N = 27$ and $N = 36$, such as the presence of pseudo-knots. Thus, the use of the case $N = 36$ will allow much richer modelling of the thermodynamics of lattice polymers [8]. Finally, while the cases $N = 64$ and greater cannot even be enumerated at present, hopefully the estimate on the number of conformations given will be useful, for example in the analysis of Monte Carlo kinetics on cubic lattices [8, 9].

Acknowledgments

This work was supported by NSF (DMR 90-22933) and NEDO of Japan. VSP acknowledges the support of an NSF Fellowship. AYG acknowledges the support of a Kao Fellowship. Some computations were performed on the MIT LCS Connection Machine CM-5 (Project SCOUT: ARPA contract MDA972-92-J-1032).

Appendix. Enumeration of starting paths

We wish to enumerate the different paths which completely break all of the symmetries. First, we must enumerate all of the symmetries. Consider all of the vertices of the graph to be numbered consecutively. Any transformation (e.g. rotation, mirror inversion, etc.) can be expressed as a permutation of these indices. The number of transformations, and therefore permutations, is calculated as follows in terms of the number of ways a d-dimensional hypercube can be reoriented: (i) we first have the symmetry by the number of corners of the cube (2^d); (ii) next, once we choose a corner to fix, we have $d!$ ways to choose how we arrange the edges (for example, for $d = 3$, we have three ways to place the first edge, leaving two ways to place the second). Thus, there are $d!2^d$ transformations for a simple hypercubic lattice in d dimensions.

To generate the starting paths, we traverse the tree and compare sub-branches for symmetries. At each node, we transform the trajectories formed by each sub-branch using all of the enumerated transformations (i.e. applying all of the permutations). If any transformation can map one sub-branch into another, than the sub-branches are related by symmetry, and we can discard one of them. The remaining sub-branches themselves will now be enumerated using the same procedure. If none of the sub-branches are related by symmetry, and if all of the $d!2^d$ symmetries have been broken by the current path, then the current path is a starting path, and we can backtrack and continue the enumeration with the unexplored branches.

For example, consider a walker starting from the corner of a cube. It is at the top of the tree of Hamiltonian walks. It now has three possible paths, but each path can be transformed into the other by a mirror symmetry. Thus, we can discard two of the sub-branches, choose the third, and continue the process. When none of the sub-branches of a given node are related by symmetry, then each sub-branch is a starting path. Then the walker backs up one level of the tree in order to traverse through the sub-branches left behind.

The enumeration of all of the walks and the enumeration of the starting paths are deeply related. Each traverse the ideal tree, only differing when the walk has completed and when sub-branches are to be discarded.

References

[1] Shakhnovich E and Gutin A 1990 *J. Chem. Phys.* **93** 5967
[2] Flory P J 1953 *Principles of Polymer Chemistry* (New York: Cornell University Press)
[3] Halbherr M, Zhou Y and Joerg C 1994 *Proc. 2nd Int. Workshop on Massive Parallelism: Hardware, Software, and Applications (Capri, Italy)*
[4] Pande V, Grosberg A Yu and Tanaka T 1994 *Proc. Natl Acad. Sci. USA* (in press)
[5] Shakhnovich E and Gutin A 1993 *Proc. Natl Acad. Sci. USA* **90** 7195
[6] Lau K F and Dill K A 1989 *Macromolecules* **22** 3986
[7] Diao Y 1994 *J. Stat. Phys.* **74** 1247
[8] Pande V, Grosberg A Yu and Tanaka T 1994 *J. Chem. Phys.* (in press)
[9] Sali A, Shakhnovich E and Karplus M 1994 *J. Mol. Biol.* **235** 1614