Interdisciplinary Information Sciences Vol. 19, No. 1 (2013) 79–83
©Graduate School of Information Sciences, Tohoku University
ISSN 1340-9050 print/1347-6157 online
DOI 10.4036/iis.2013.79

How to Estimate the Number of Self-Avoiding Walks over 10^{100}? Use Random Walks

Nobu C. SHIRAI[1,2,*] and Macoto KIKUCHI[2,1,3]

[1]*Graduate School of Science, Osaka University, Toyonaka, Osaka 560-0043, Japan*
[2]*Cybermedia Center, Osaka University, Toyonaka, Osaka 560-0043, Japan*
[3]*Graduate School of Frontier Biosciences, Osaka University, Suita, Osaka 565-0871, Japan*

Received April 18, 2013; final version accepted May 13, 2013

Counting the number of N-step self-avoiding walks (SAWs) on a lattice is one of the most difficult problems of enumerative combinatorics. Once we give up calculating the exact number of them, however, we have a chance to apply powerful computational methods of statistical mechanics to this problem. In this paper, we develop a statistical enumeration method for SAWs using the multicanonical Monte Carlo method. A key part of this method is to expand the configuration space of SAWs to random walks, the exact number of which is known. Using this method, we estimate a number of N-step SAWs on a square lattice, c_N, up to $N = 256$. The value of c_{256} is $5.6(1) \times 10^{108}$ (the number in the parentheses is the statistical error of the last digit) and this is larger than one googol (10^{100}).

KEYWORDS: self-avoiding walks, statistical enumeration, multicanonical Monte Carlo method, rare event sampling

1. Introduction

Starting from the origin of a lattice and connecting a line to one neighboring site after another, we get a path. If we choose a random direction for each step, the path becomes a random walk. The question we seek to answer is "How many N-step random walks are there on a square lattice?" Since there are four possible directions for each step, the answer is 4^N. If we restrict the path by limiting visits to the same site, we get a self-avoiding walk (SAW) instead of a random walk. We then ask the similar question. "How many N-step SAWs are there on a square lattice?" The answer, however, is not so simple as with random walks. The number of SAWs, c_N, is not known for large values of N. There are some sophisticated enumeration methods for SAWs and exact values for c_N on a square lattice are known only up to $N = 71$ [1]. The enumerated c_N of $N = 71$ is

$$c_{71} = 4\ 190\ 893\ 020\ 903\ 935\ 054\ 619\ 120\ 005\ 916 \sim 4.19 \times 10^{30},$$

which is much larger than Avogadro's constant. In order to go further, we change the problem from calculating all digits of c_N to estimating the first two or three digits of c_N. How large a c_N can we estimate? Here, we set a tentative goal of a googol (10^{100}).

One popular strategy in enumeration is *divide and conquer*. We propose, however, a different strategy to estimate a large c_N, in which we expand the configuration space of SAWs to that of random walks and use knowledge of the number of random walks. In order to explain how to use this exact number, we describe a statistical method of estimating the area of a unit circle using Monte Carlo integration as an example. In Fig. 1(a), we show a unit circle surrounded by a square, the area of which is exactly 4. To calculate the area of the circle, we randomly generate n points on the square using pairs of real random numbers (r_1, r_2) $(-1 \leq r_1, r_2 \leq 1)$ and count the number of pairs which satisfy the condition $x^2 + y^2 \leq 1$. Let this number be n_c. Using n_c and n, we can estimate the area ratio of the circle to the square. Since we know that the area of the square is 4, we can finally estimate the area of the unit circle as $4 \times n_\mathrm{c}/n$. Let us apply this logic to the estimation of c_N.

In Fig. 1(b), we show a Venn diagram of SAWs and random walks. The number of N-step random walks on a d-dimensional hypercubic lattice is exactly known as $(2d)^N$. Instead of generating random points as in the above example, we use Markov chain Monte Carlo (MCMC) to calculate the ratio of the number of SAWs to that of random walks. Once we obtain the ratio, we can estimate c_N by multiplying the ratio by $(2d)^N$.

2010 Mathematics Subject Classification: Primary 65C05, Secondary 82B41.

This work is supported by the Global COE Program Core Research and Engineering of Advanced Materials-Interdisciplinary Education Center for Materials Science, MEXT, Japan.

[*] Corresponding author. E-mail: shirai@cp.cmc.osaka-u.ac.jp

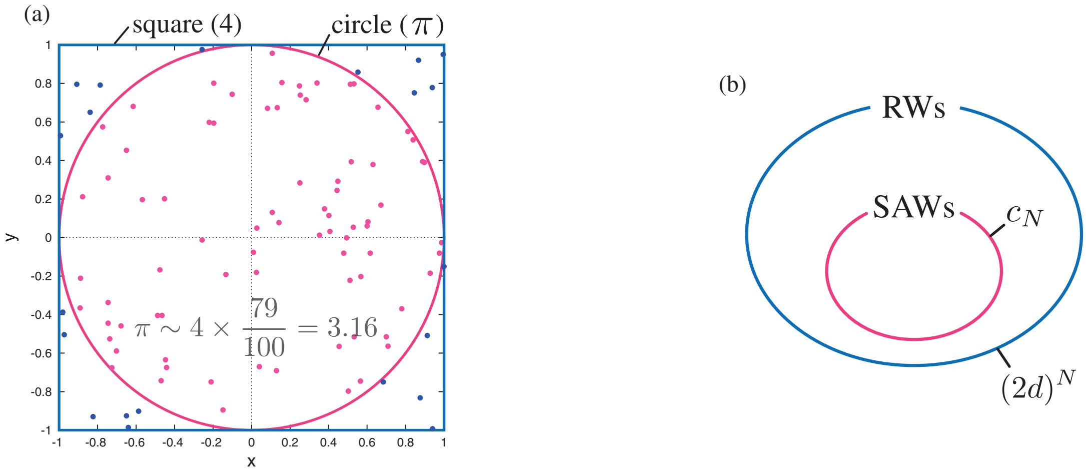

Fig. 1. (a) Estimating the area of the unit circle using Monte Carlo integration. The values in the parentheses are the areas of the shapes. 79 of 100 points were plotted inside the circle (pink dots) and the area of the circle can be estimated as $4 \times 79/100 = 3.16$. (b) A Venn diagram of SAWs and random walks. In order to calculate c_N, we expand the configuration space of SAWs to random walks, the number of which is exactly $(2d)^N$.

2. The Statistical Enumeration Method for Self-Avoiding Walks

2.1 SAWs are rare random walks

We define a SAW on a d-dimensional hypercubic lattice and its mathematical notation. We denote an N-step path as $\omega = (\omega(0), \omega(1), \dots, \omega(N))$, where each of $\omega(i)$ $(i = 0, 1, \dots, N)$ denotes a point which is randomly selected from the lattice. When we impose two conditions, namely (i) $\omega(0)$ is at the origin of the lattice, and (ii) $|\omega(i+1) - \omega(i)| = 1$ $(i = 0, 1, \dots, N-1)$ on ω, ω becomes an N-step random walk. If we add one more condition (iii) $\omega(i) \neq \omega(j)$ $(i \neq j)$ on ω, ω becomes an N-step SAW [2]. It is obvious from the definition of a SAW that the set of N-step SAWs is a subset of that of N-step random walks. SAWs are known as $n \to 0$ limit of n-vector model and have attracted a great interest of physicists in the context of critical phenomena. As we mentioned above, we expand the configuration space of SAWs to that of random walks and try to sample SAWs using MCMC.

Before we move on to the sampling method, we would like to discuss how rare SAWs are. The asymptotic behavior of c_N is conjectured to be

$$c_N = A\mu^N N^{\gamma-1}[1 + o(1)], \tag{2.1}$$

where μ is called the connective constant, which is known to be ~2.64 [1]. γ is a critical exponent, which is thought to be universal. Namely, it depends only on the spatial dimensionality and it is independent of the lattice type. In order to roughly estimate the ratio of the number of SAWs to that of random walks on a square lattice, we use the approximation

$$\frac{c_N}{(2d)^N} \sim \left(\frac{\mu}{4}\right)^N \sim \left(\frac{2.64}{4}\right)^N. \tag{2.2}$$

In the case of $N = 50$, 150, and 250, the ratios are

$$\left(\frac{2.64}{4}\right)^{50} \sim 10^{-9}, \ \left(\frac{2.64}{4}\right)^{150} \sim 10^{-27}, \ \left(\frac{2.64}{4}\right)^{250} \sim 10^{-45}.$$

Even with $N = 50$, SAWs are very rare, and if we randomly generate random walks, we need about 10^9 samples to get one SAW. In the case of $N = 250$, SAWs become extremely rare and we cannot get one SAW by random sampling even if we use the massive computer resources available today. Thus, we need a more efficient sampling method.

2.2 Dividing the configuration space of random walks by the number of intersections

As a first step towards an efficient sampling method, we introduce the number of intersections V as a parameter. In Fig. 2, we show a configuration of a random walk with three intersections, as an example. If k points of a path $(k \geq 3)$ are on the same site, we define that there are $k - 1$ intersections on that site. We divide the configuration space into N parts $(V = 0, 1, 2, \dots, N-1)$ according to the value of V, and configurations with $V = 0$ are all SAWs. If we introduce the energy of random walks by V, we get the modified Domb–Joyce model [3,4]. In the high- and low-temperature limits, this model reduces to random walks and SAWs respectively. From the discussion above, we need the ratio of the

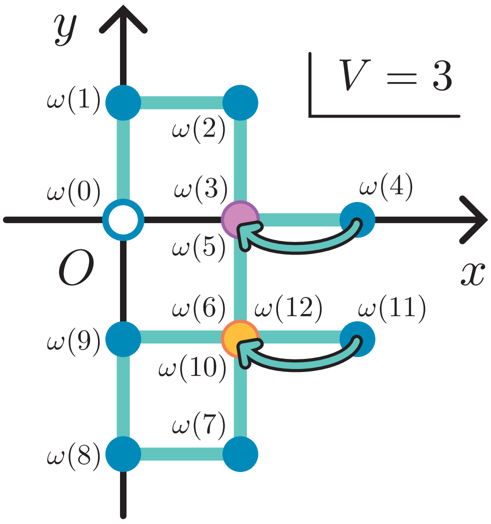

Fig. 2. A 12-step random walk with three intersections. There are one intersection on the purple circle and two intersections on the orange circle. Total intersections V of the path is 3.

number of SAWs to that of random walks to estimate c_N. If we denote the density of states of the modified Domb–Joyce model as $\Omega(V)$, $\Omega(0)/\sum_{V=0}^{N-1}\Omega(V)$ gives the desired ratio, and we can calculate c_N by

$$c_N = (2d)^N \frac{\Omega(0)}{\sum_{V=0}^{N-1}\Omega(V)}. \tag{2.3}$$

In order to calculate c_N accurately, we should ensure the accuracy of not only $\Omega(0)$ but $\Omega(V)$ for the entire range of V. It is difficult to do so, however, because the value of $\Omega(V)$ varies across many orders of magnitude. To overcome these difficulty, we use a powerful computational method of statistical mechanics.

2.3 Multicanonical simulation of the modified Domb–Joyce model

In the next step, we explain how to sample SAWs. We use the multicanonical Monte Carlo method [5, 6] to accurately estimate the density of states of the modified Domb–Joyce model, $\Omega(V)$, over a wide range of V. This calculation is essentially the same as that of the multi-self-overlap ensemble [7, 8] without Hamiltonian. A key to this method is the weight function $W(V)$ that is built up to be proportional roughly to $1/\Omega(V)$ by iterative methods. Frequently-used iterative methods are as follows:

- the multicanonical method [5, 6]
- the entropic sampling method [9]
- the Wang–Landau method [10, 11].

We used the Wang–Landau method. We omit explaining these iterative methods here, and explain the multicanonical method, assuming that $W(V)$ has already been obtained. In the multicanonical method, the transition probability from one path, ω_a $(V = V_a)$, to another path ω_b $(V = V_b)$ is given by

$$p(\omega_a \to \omega_b) = \min\left[\frac{W(V_b)}{W(V_a)}, 1\right], \tag{2.4}$$

where this form of the transition probability is the same as that of the Metropolis method with $W(V)$ instead of the Boltzmann factor. An MCMC simulation using this transition probability generates an ensemble called a multicanonical ensemble. Sampling from this ensemble gives a flat histogram $H(V)$ (in the top left graph of Fig. 3). Dividing $H(V)$ by $W(V)$, we obtain $\Omega(V)$ (in the bottom left graph of Fig. 3). From Eq. (2.3) with $\Omega(V)$ obtained, we can finally calculate c_N (in the bottom right graph of Fig. 3).

3. Results

Using the proposed method above, we calculate c_N with a statistical error up to $N = 256$. Some of the calculated c_N values are shown in Table 1, together with the available exact values ($N \leq 71$) and estimated values given by Prellberg and Krawczyk using flatPERM [12]. The errors were estimated from the standard errors of histograms. The estimated values agree well with exact values and estimated values of flatPERM. We should note that flatPERM achieved longer walks up to $c_{1028} = 1.74 \times 10^{438}$, but due to the lack of error estimation, it is now known how many digits of the values are reliable. The largest value of our estimation is

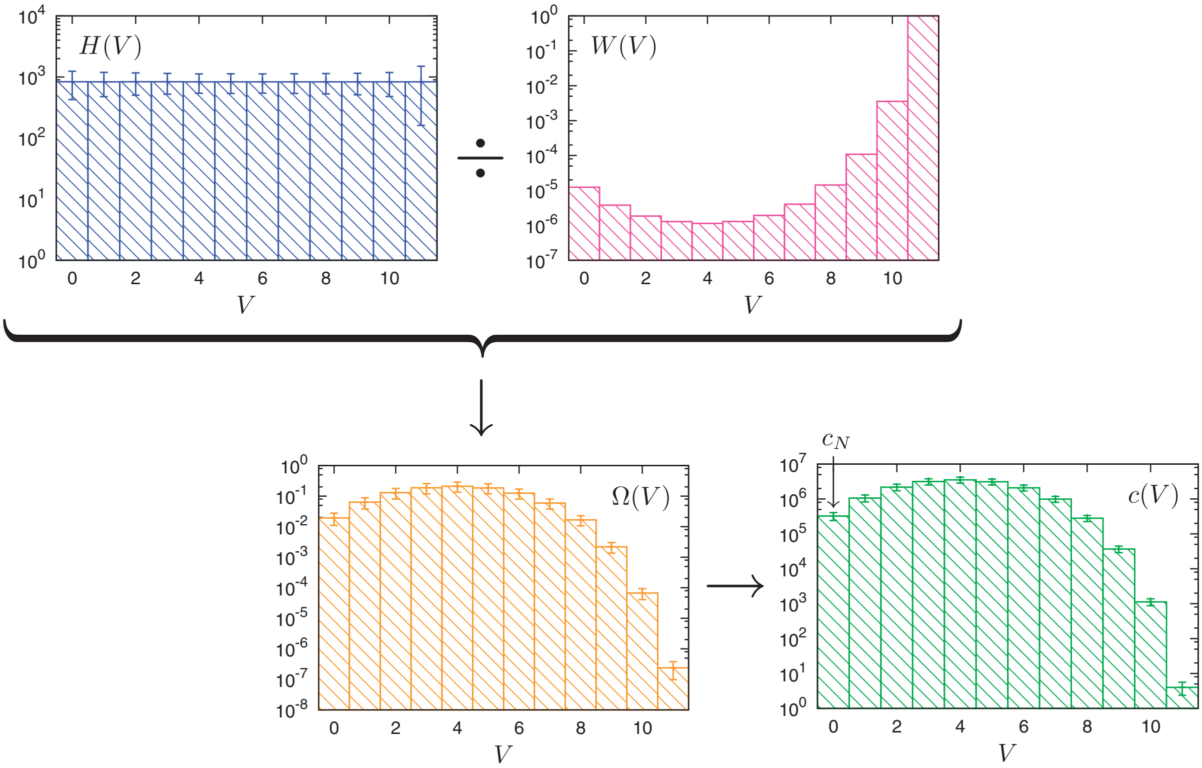

Fig. 3. A procedure of calculating $c(V)$. $c(V)$ is the number of random walks with V intersections, and is calculated by $c(V) = (2d)^N \Omega(V) / \sum_{V=0}^{N-1} \Omega(V)$.

Table 1. List of exact and estimated c_N values on a square lattice.

N	c_N (exact)	c_N (flatPERM [12])	c_N (our method)
4	100		$1.0001(2) \times 10^2$
8	5916		$5.915(2) \times 10^3$
16	1.7245×10^7	1.7265×10^7	$1.724(1) \times 10^7$
32	1.1903×10^{14}	1.1911×10^{14}	$1.186(2) \times 10^{14}$
64	4.5493×10^{27}	4.5529×10^{27}	$4.54(2) \times 10^{27}$
128		5.2970×10^{54}	$5.32(6) \times 10^{54}$
256		5.6700×10^{108}	$5.6(1) \times 10^{108}$

$$c_{256} = 5.6(1) \times 10^{108},$$

and we achieved the goal of a c_N value of over 10^{100}.

Acknowledgments

We would like to thank Dr. T. Prellberg for providing us detailed data of c_N given by flatPERM. This work was supported by the Global COE Program Core Research and Engineering of Advanced Materials-Interdisciplinary Education Center for Materials Science, MEXT, Japan.

REFERENCES

[1] Jensen, I., “Enumeration of self-avoiding walks on the square lattice,” *J. Phys. A: Math. Gen.*, **37**: 5503–5524 (2004).

[2] Madras, N., and Slade, G., The Self-Avoiding Walk, Birkhäuser (1996).

[3] Domb, C., and Joyce, G. S., “Cluster expansion for a polymer chain,” *J. Phys. C: Solid State Phys.*, **5**: 956–976 (1972).

[4] Shirai, N. C., and Kikuchi, M., “Multicanonical simulation of the Domb–Joyce model and the Go model: New enumeration methods for self-avoiding walks,” arXiv:1212.2181 (2012).

[5] Berg, B. A., and Neuhaus, T., “Multicanonical algorithms for first order phase transitions,” *Phys. Lett. B*, **267**: 249–253 (1991).

[6] Berg, B. A., and Neuhaus, T., “Multicanonical ensemble: A new approach to simulate first-order phase transitions,” *Phys. Rev. Lett.*, **68**: 9–12 (1992).

[7] Iba, Y., Chikenji, G., and Kikuchi, M., “Simulation of lattice polymers with multi-self-overlap ensemble,” *J. Phys. Soc. Jpn.*,

67: 3327–3330 (1998).

[8] Chikenji, G., Kikuchi, M., and Iba, Y., “Multi-self-overlap ensemble for protein folding: Ground state search and thermodynamics,” *Phys. Rev. Lett.*, **83**: 1886–1889 (1999).

[9] Lee, J., “New Monte Carlo algorithm: Entropic sampling,” *Phys. Rev. Lett.*, **71**: 211–214 (1993).

[10] Wang, F., and Landau, D. P., “Efficient, multiple-range random walk algorithm to calculate the density of states,” *Phys. Rev. Lett.*, **86**: 2050–2053 (2001).

[11] Wang, F., and Landau, D. P., “Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram,” *Phys. Rev. E*, **64**: 056101 (2001).

[12] Prellberg, T., and Krawczyk, J., “Flat histogram version of the pruned and enriched rosenbluth method,” *Phys. Rev. Lett.*, **92**: 120602 (2004). Only the graph of the logarithm of c_N is given in the paper. The numbers shown in Table 1 were provided by Dr. T. Prellberg as a private communication.