# Further reflections on the one-dimensional packing problem

Richard W. Freedman
*Department of Information Systems, University of Maryland, Baltimore, MD 21228, USA*
and
Fred Gornick
*Department of Chemistry, University of Maryland, Baltimore, MD 21228, USA*

Received 5 March 1992

The one-dimensional packing problem may be stated as follows: When objects of length $L$ are randomly placed on a line of length $N$ until no more placement is possible, how much space remains unoccupied? In a previous paper, the authors showed that, for $L = 2$, the fraction of unoccupied space is dependent on the model governing the placement mechanism. In this paper, these results are extended from the discrete to the continuous case by allowing both $N$ and $L$ to increase, while keeping their ratio constant. The methodology was validated by reproducing the analytical results for limiting cases.

## 1. Introduction

In 1958, the Hungarian mathematician A. Renyi published a seminal paper [1] which addressed the following problem: If non-overlapping objects of length $L$ are placed at random along a line of length $N$, what is the expected value of the wasted space at saturation, i.e. what portion of the line remains uncovered when no spaces equal to or greater than $L$ are available. Solomon and Weiner [2] have reviewed the literature inspired by this problem and have attributed its origin to a model first proposed by the British physical chemist J.D. Bernal for liquids composed of spherically symmetrical molecules. In that model, a specified volume is randomly filled with impenetrable spheres of uniform diameter until no space large enough to accommodate a single sphere remains. The inefficiencies of packing due to the random placement of the spheres results in a certain amount of "wasted" space which, in the absence of attractive or repulsive forces between spheres, corresponds to the increase in volume accompanying the melting of a perfect crystal in which packing efficiency is presumed to be a maximum. Renyi's problem is a one-dimensional version of the one posed by Bernal.

It is of interest to note that, almost twenty years before the publication of Renyi's work, the American chemists Flory [3] and Wall [4,5] considered a model (the FW model) of a chemical reaction in which adjacent substituents along the backbone of a linear polymer molecule are removed in pairs. A given substituent is thus permanently isolated if both of its nearest neighbors have been paired previously. This is clearly a discrete version of the packing problem; the objects are of length two and their end points are confined to integral coordinates. Related problems were treated by Gornick and Jackson [6], and McQuarrie [7]. The former paper considered the crystallization of randomly selected segments of polymer chains. In this case, a sequence is selected only if its length exceeds some critical value. Sequences of sub-critical length bounded by previously selected sequences are thus excluded from the crystal phase and are thus isolated. McQuarrie [7] treated a one-dimensional nucleation and growth problem in which sequential reactions of substituents along a polymer chain are initiated at random, followed by near-neighbor propagation. He then derived a distribution function for lengths of sequences of unreacted substituents as a function of the extent of the reaction. Extensive reviews [8–10] related to sequential reactions along the backbone of polymer chains are available in the chemical literature. For further references to packing problems encountered in the physical sciences and engineering, the reader is referred to an interesting paper by Mackenzie [11].

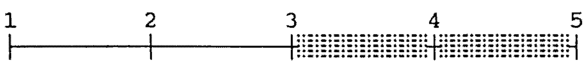
In a previous paper [12], we re-examined the FW model ($L = 2$) and showed that the expected value of the packing efficiency depends on the mechanism governing the pair selection process. By "mechanism" we refer to the specification of individual steps in a process. The FW model corresponds to a one-step mechanism insofar as a pair of unoccupied points on the line is selected at random and the object placed upon them in a single concerted step. In our model, two distinct steps are required: (1) an unoccupied point is chosen, and (2) the object is placed either to the right or to the left of it, if at least one of those placements is possible. The fraction of unoccupied points in the two-step process is shown to be approximately 10% lower than that of the one-step process.[#1] Since the publication of our paper, we have become aware of the work of Evans and Nord [13], who have treated an identical problem by other methods. In the present paper, we generalize our previous results [12] by extrapolating from the discrete to the continuous case. This is accomplished by allowing both the length of the object $L$ and the length of the line $N$ to increase without bound while keeping their ratio fixed.

A forthcoming review paper by Evans [14] will encompass all of the aforementioned topics, as well as others dealing with random and cooperative sequential adsorption.

[#1] It has been brought to our attention that our results, as well as those of FW, were reported by Page [15] and Downton [16] in 1959 and 1961, respectively. These authors were no doubt unaware of the work of Flory and Wall, which appeared in the chemical literature twenty years earlier. Our failure to acknowledge these contributions, although inadvertent, is regrettable.

## 2. Comparison of the one- and two-step models

In the interest of clarity, we reiterate the essential features of both the one- and two-step models and generalize our previous results (i.e. $L = 2$) for any integral values of $L$ and $N$. In both models, objects are placed so that their end points are at integral coordinates. Thus, for a line of length $N$ which is marked off in coordinates from 1 on the left to $N + 1$ on the right, and an object of length $k$, we define the $i$th location as the one in which the line segment lies between coordinates $i$ and $i + k$, denoted as $[i, i + k]$. As an illustration, the third location for an object of length 2 on a line of length 4 would be pictured as the patterned segment in fig. 1(a). A more compact, but equivalent, diagram is shown in fig. 1(b).




(a)

[ – – * * ]

(b)

Fig. 1 (a) ( ░ ) The third location of an object of length $k = 2$ on a line of length $N = 4$. (b) Same as (a), where ( – ) is an unoccupied unit line segment and ( * * ) is the third location of length $k = 2$.

We define $S(L, N)$ as the expected value of the amount of unoccupied space on a line of length $N$ when objects of length $L$ are placed upon it to the point of saturation. The reader can readily verify that for $(N = 1, 2, \ldots, L - 1)$ (i.e. the length of the object exceeds the length of the line), $S(L, N) = N$. Furthermore, for $(N = L, \ldots, 2L - 1)$ (i.e. only one object will fit on the line), $S(L, N) = N - L$.

There are $(N - L + 1)$ possible locations for the placement of the first object. Each placement will correspond to a different initial configuration. Let $S(L, i, N)$ be the expected value of the amount of wasted space, given that the first object placed occupies $[i, i + L]$. Then, if $p(L, i, N)$ is the probability of that initial placement, it follows that:

$$S(L, N) = \sum_{i=1}^{N-L+1} p(L, i, N) \ \ S(L, i, N).$$

Consider the one-step model. Here, $p(L, i, N) = 1/(N - L + 1)$ for $i = 1, \ldots, (N - L + 1)$ because all configurations are equally probable. The four possible configurations for the case of $L = 3$, $N = 6$ are shown in fig. 2, where three contiguous asterisks $(***)$ represent the location of the object and each dash (–) represents an unoccupied unit line segment.

(A) [* * * - - -] (B) [- * * * - -]

(C) [- - * * * -] (D) [- - - * * *]

Fig. 2. Possible configurations after placement of the first object (for $L = 3$ and $N = 6$).

From fig. 2, it is clear that in configurations A and D ($i = 1$ and 4), $S(3, i, 6) = 0$ since, in both cases, there remains a space exactly equal to the length of one object and consequently there is no wasted space at saturation. However, in configurations B and C ($i = 2$ and 3), the two remaining spaces are each too small to accommodate another object. Therefore, further occupancy is precluded and three spaces are wasted. As noted above, all four configurations have equal probability ($p(3, i, 6) = 1/4$ for $i = 1, \ldots, 4$), and thus

$$S(3, 6) = (1/4)\,(0) + (1/4)\,(3) + (1/4)\,(3) + (1/4)\,(0) = 1.5 \quad \text{(one-step model)}.$$

Now consider the case of the one-step model when $L = 3$ and $N = 9$. Here, seven equally probable configurations, diagrammed in table 1, result from the placement of the first object. Following the procedure used for $N = 6$, we compute $S(3, 9)$, i.e. the product of the probability of a configuration and the expected value of the amount of wasted space given that particular configuration, summed over all configurations.

The resulting value of $S(3, 9)$ is $(1/7)\,(1.5 + 3 + 3 + 0 + 3 + 3 + 1.5) = 2.1429$. The fractional wasted space $S(L, N)/N$ is thus 0.2381.

In the first step of the two-step model, an arbitrary location $j$ ($j = 1, \ldots, N$), on the line is chosen with uniform probability. In step 2, the object is placed either to the right or to the left of location $j$, depending on whether there is space available. If the line segment $[j - L + 1, j + L - 1]$ is unoccupied, the conditional probabilities of placing the object either on the line segment $[j - L - 1, j]$ or on $[j, j + L - 1]$ are each equal to 1/2. If, however, any of the line segment $[j - L - 1, j]$ is occupied, then placement in that range is impossible. Clearly, a similar statement applies to the line segment $[j, j + L - 1]$.

To illuminate the contrast between the one- and two-step models, let us calculate $S(3, 6)$ and $S(3, 9)$ for the latter. The possible configurations resulting from the placement of the initial object of length $L = 3$ are as shown in fig. 2 (for $N = 6$) and table 1 (for $N = 9$). The values of $S(3, N)$ for the two-step model must be the same as those for the one-step model for line lengths less than or equal to $(2L - 1)$. Hence, $S(3, N) = 1, 2, 0, 1$, and 2 for $N = 1, 2, 3, 4$, and 5, respectively.

Following the rules for the two-step model for $L = 3$ and $N = 6$, let us now calculate the probability of configuration A (see fig. 2), in which locations 1, 2, and 3 are occupied. There are two ways to achieve this configuration: either location 1 is initially chosen and the object placed to the right (with conditional probability of unity), or location 3 is initially chosen and the object placed to the left (with conditional probability of 1/2). Since the a priori probability of choosing any particular initial location is 1/6, the probability Pr(A) of configuration A is

$$\Pr(A) = (1/6)\,(1) + (1/6)\,(1/2) = 0.25 \quad \text{(two-step model)},$$

which is coincidentally equal to Pr(A) for the one-step model. Similar arguments show that the probabilities of the four possible configurations are equal, so that the value of $S(3, 6)$ for the two-step model is calculated as

$$S(3, 6) = (1/4)\,(0) + (1/4)\,(3) + (1/4)\,(3) + (1/4)\,(0) = 1.5 \quad \text{(two-step model)}.$$

It should be noted that although this is the same result as that obtained for the corresponding one-step model, this is not true in general as illustrated in the case of $L = 3$ and $N = 9$. Following the procedure outlined above, the seven possible configurations that result from the placement of the initial object of length 3 in a sequence of length 9 are shown in table 1. Whereas in the case of the one-step model the probabilities of each configuration were equal, that is not the case in the two-step model, as shown in table 1.

Table 1

Possible configurations and probabilities for the two-step model ($L = 3$, $N = 9$).

| $i$ | Configuration | $p(3, i, 9)$ two-step | $S(3, i, 9)$ two-step |
|---|---|---|---|
| 1 | [***------] | (1/9) (1 + 1/2) | $S(3, 6) = 1.5$ |
| 2 | [-***-----] | (1/9) (1 + 1/2) | $S(3, 1) + S(3, 5) = 3$ |
| 3 | [--***----] | (1/9) (1/2 + 1/2) | $S(3, 2) + S(3, 4) = 3$ |
| 4 | [---***---] | (1/9) (1/2 + 1/2) | $S(3, 3) + S(3, 3) = 0$ |
| 5 | [----***--] | (1/9) (1/2 + 1/2) | $S(3, 4) + S(3, 2) = 3$ |
| 6 | [-----***-] | (1/9) (1/2 + 1) | $S(3, 5) + S(3, 1) = 3$ |
| 7 | [------***] | (1/9) (1/2 + 1) | $S(3, 6) = 1.5$ |

Multiplying the above probabilities with the corresponding values of wasted space and summing the results, it is found that the value of $S(3, 9)$ is (1/9) (39/2) = 2.1667. The fractional wasted space is, by definition, $S(L, N)/N$ which, for this case, is 0.2407.

## 3. Results and discussion

The calculation of $S(L, N)$ for arbitrary integer values of $L$ and $N$ does not lend itself to a closed form solution for either the one-step or the two-step models.

Consequently, we have developed the following algorithms for calculating these quantities (the derivations are presented in the appendix):

$$S(L, N) = [1/(N - L + 1)]\,[(N - L)S(L, N - 1) + (2)S(L, N - L)] \quad \text{(one-step)}, \tag{1}$$

$$S(L, N) = (1/N)[(N - 1)S(L, N - 1) + (3)S(L, N - L) - S(L, N - 2L + 1)] \quad \text{(two-step)}. \tag{2}$$

Due to the recursive nature of these algorithms, evaluation of $S(L, N)$ for large values of $L$ and $N$ is very tedious (e.g. more than 100 h of CPU time on a VAX 8600 were required to calculate $S(1000, 10^{**}(6))$ for the two-step model). Therefore, the results of these calculations are summarized in tables 2 and 3. These tables show the expected value of isolated space normalized by line length $N$ for given values of object length $L$ under the protocols of the one-step method (table 2) and the two-step method (table 3). These data, then, are the expected values of the fraction of isolated space $S(L, N)/N$.

In our previous paper [12], we considered the one-step and two-step methods for the discrete case, that is, where $L = 2$ and objects are placed only at integral coordinates. Flory [3] and Wall [4,5] determined analytically that for the one-step model the value of $S(L, N)/N$ equals $\exp(-2)$ ($=0.13534$) in the limit of large $N$ (for $L = 2$). Page [15] obtained an analytical result for the same quantity for the two-step model. His result is 0.12332. These results are in agreement with the values shown in tables 2 and 3 ($L = 2$, $N/L = 1000$).

Now consider the generalization to the continuous case where the placement of the object is not confined to integral coordinates, and the object and line each may be of any length. The continuous case can be simulated from the discrete algorithm by increasing the length of the object, while keeping the ratio of the line length to object length $N/L$ constant. Normalizing the results by the length of the line, we obtain the value of $S(L, N)/N$ in the limit of large $N$ (see tables 2 and 3).

The results of our simulation of the continuous case are consistent with several results obtained analytically. For example, consider the placement of objects of length $L$ upon a continuous line of length $2L$. It is clear that for both the one- and two-step models, the fraction of the unoccupied space on the line upon saturation will be vanishingly close to one half. Similarly, the fraction of isolated space for the continuous case where the ratio of line length to object length equals 3 must approach one third. These results are consistent with those shown in tables 2 and 3 ($N/L = 2, 3$; $L = 1000$). Moreover, Renyi's [1] result for an infinite line (i.e. ratio of line length to object length approaches infinity) for the one-step model ($=0.252$) is closely approximated by our results of 0.2524, shown in table 2, $N/L = 1000$, $L = 1000$ (i.e. $N = 1000000$). We consider these results an indication of the validity of our method.

Table 2

One-step model results: expected value of the fraction of isolated space $S(L, N)/N$ for an object of length $L$ and a line of length $N$.

| Size ratio $N/L$ | Object length $L$: 2 | 5 | 10 | 50 | 100 | 1000 |
|---|---|---|---|---|---|---|
| 2 | 0.1667 | 0.3333 | 0.4091 | 0.4804 | 0.4901 | 0.4990 |
| 3 | 0.1778 | 0.2828 | 0.3117 | 0.3298 | 0.3316 | 0.3332 |
| 4 | 0.1690 | 0.2595 | 0.2873 | 0.3090 | 0.3117 | 0.3141 |
| 5 | 0.1624 | 0.2492 | 0.2765 | 0.2978 | 0.3004 | 0.3027 |
| 10 | 0.1489 | 0.2285 | 0.2535 | 0.2729 | 0.2753 | 0.2774 |
| 30 | 0.1398 | 0.2146 | 0.2381 | 0.2563 | 0.2586 | 0.2606 |
| 50 | 0.1380 | 0.2119 | 0.2350 | 0.2530 | 0.2552 | 0.2572 |
| 100 | 0.1367 | 0.2098 | 0.2327 | 0.2505 | 0.2527 | 0.2547 |
| 500 | 0.1356 | 0.2081 | 0.2309 | 0.2486 | 0.2507 | 0.2527 |
| 1000 | 0.1355 | 0.2079 | 0.2307 | 0.2483 | 0.2505 | 0.2524 |

Table 3

Two-step model results: expected value of the fraction of isolated space $S(L, N)/N$ for an object of length $L$ and a line of length $N$.

| Size ratio $N/L$ | Object length $L$: 2 | 5 | 10 | 50 | 100 | 1000 |
|---|---|---|---|---|---|---|
| 2 | 0.1250 | 0.3500 | 0.4250 | 0.4850 | 0.4925 | 0.4992 |
| 3 | 0.1528 | 0.2911 | 0.3172 | 0.3309 | 0.3322 | 0.3332 |
| 4 | 0.1523 | 0.2614 | 0.2948 | 0.3236 | 0.3273 | 0.3308 |
| 5 | 0.1478 | 0.2526 | 0.2866 | 0.3126 | 0.3157 | 0.3184 |
| 10 | 0.1357 | 0.2318 | 0.2622 | 0.2858 | 0.2887 | 0.2912 |
| 30 | 0.1274 | 0.2177 | 0.2463 | 0.2685 | 0.2712 | 0.2736 |
| 50 | 0.1258 | 0.2149 | 0.2432 | 0.2650 | 0.2677 | 0.2701 |
| 100 | 0.1246 | 0.2128 | 0.2408 | 0.2624 | 0.2651 | 0.2674 |
| 500 | 0.1236 | 0.2111 | 0.2389 | 0.2603 | 0.2630 | 0.2653 |
| 1000 | 0.1234 | 0.2109 | 0.2386 | 0.2600 | 0.2625 | 0.2648 |

Considering next the fraction of isolated space on a line of infinite length following the two-step model, our result is 0.2648 (table 3, $N/L = 1000$, $L = 1000$). Based on the close agreement with known results for the one-step model, we conclude that the results for the two-step model are accurate. We observe that the two-step model for the continuous case yields expected values of the fraction of unoccupied space approximately five percent higher than the one-step model. As we have noted in our previous paper [12], the difference between the two mechanisms comes into play as the average distance between placed objects decreases. Moreover, this average distance decreases continuously as objects are placed on the line. For example, it is easily shown that even for a pre-saturation occupancy level of 20%, the average distance between placed objects is approximately four times the length of the object (=$4L$); at the 50% level, the average distance between placed objects decreases to a single object length. Thus, the effect of the choice of mechanism rapidly becomes appreciable.

## Appendix

Recall that $S(L, N)$ is the expected value of the amount of isolated space remaining upon saturation for objects of length $L$ and a line of length $N$. Let $p(L, i, N)$ be the probability that the first object to be placed on the line occupies $[i, i+L]$.

### TWO-STEP MODEL

It follows from the assumptions of the two-step model that

$$\begin{aligned} p(L, i, N) &= [3/(2N)] && \text{for } i = 1, \dots, L-1; \\ &= (1/N) && \text{for } i = L, \dots, N-(2L)+2; \\ &= [3/(2N)] && \text{for } i = N-(2L)+3, \dots, N-L+1. \end{aligned} \tag{A.1}$$

We observe that the selection of the location for the initial object $[i, i+L]$ splits the line into two line segments of lengths $(i-1)$ and $(N-i-L+1)$. The total amount of isolated space on the line will be the sum of the isolated space in each of the two line segments. Let $S(L, i, N)$ be the expected value of the amount of isolated space remaining upon saturation for objects of length $L$ and a line of length $N$, given that only one object has been placed, in the location $[i, i+L]$. Then,

$$S(L, i, N) = S(L, i-1) + S(L, N-i-L+1). \tag{A.2}$$

Since any of the $(N-L+1)$ possible locations could have been initially selected,

$$S(L, N) = \sum_{i=1}^{N-L+1} p(L, i, N)\ [S(L, i, N)]. \tag{A.3}$$

Combining eqs. (A.2) and (A.3), and noting that $p(L, i, N) = p(L, N-L-i+2)$, it follows that

$$S(L, N) = (2) \sum_{i=1}^{N-L+1} p(L, i, N)\ [S(L, i-1)]. \tag{A.4}$$

Substitution from eq. (A.1) into (A.4) yields

$$(N/2)[S(L,N)] = (3/2)[S(L,0) + S(L,1) + \dots + S(L,L-2)]$$
$$+ [S(L,L-1) + \dots + S(L,N-2L+1)]$$
$$+ (3/2)[S(L,N-2L+2) + \dots + S(L,N-L)]. \quad \text{(A.5)}$$

Equation (A.5) may be rewritten as

$$(N)[S(L,N)] = \left(\sum_{i=0}^{L-2} S(L,i) + \sum_{i=N-2L+2}^{N-L} S(L,i)\right) + (2)\left(\sum_{i=0}^{N-L} S(L,i)\right). \quad \text{(A.6)}$$

Subtracting $(N-1)S(L, N-1)$ from the above and rearranging, we obtain

$$S(L,N) = (1/N)[(N-1)S(L,N-1)$$
$$+ (3)S(L,N-L) - S(L,N-2L+1)] \quad \text{(two-step)}. \quad \text{(A.7)}$$

Equation (A.7) is a general recursive relationship for $S(L, N)$. Since it is obvious that

$$S(L,i) = i \quad \text{for } i = 0, 1, \dots, L-1;$$
$$= i - L \quad \text{for } i = L, L+1, \dots, 2L-1, \quad \text{(A.8)}$$

we can insert $N = 6$ and $L = 3$ into (A.7) and obtain $S(3, 6) = 1.5$, which we have previously shown to be the case.

ONE-STEP MODEL

We observe that the difference between the one-step model and the two-step model resides entirely in the difference in the two expressions for $p(L, i, N)$. The one-step model equation which corresponds to eq. (A.1) is

$$p(L,i,N) = [1/(N-L+1)] \quad \text{for } i = 1, \dots, N-L+1 \quad \text{(one-step)}. \quad \text{(A.1}')$$

Replacing the expression for $p(L, i, N)$ in the two-step model (eq. (A.1)) with the one-step model expression (eq. (A.1′)) and thereafter following the same line of reasoning, we obtain the following recursion relationship for $S(L, N)$:

$$S(L,N) = [1/(N-L+1)][(N-L)S(L,N-1)$$
$$+ (2)S(L,N-L)] \quad \text{(one-step)}. \quad \text{(A.7}')$$

For obvious reasons, the relationships expressed in eq. (A.8) apply to both the one-step and the two-step models.

Equations (A.7) and (A.7′) are the final recursion relations for the two-step and one-step models, corresponding to eqs. (2) and (1), respectively. A comparison

Table 4

The expected number of unoccupied sites after saturation: comparison between the one-step and two-step methods ($L = 3$).

| $N$ | $S(3, N)$ (one-step) (eq. (A.7′)) | $S(3, N)$ (two-step) (eq. (A.7)) |
|---|---|---|
| 1 | 1.000 | 1.000 |
| 2 | 2.000 | 2.000 |
| 3 | 0.000 | 0.000 |
| 4 | 1.000 | 1.000 |
| 5 | 2.000 | 2.000 |
| 6 | 1.500 | 1.500 |
| 7 | 1.600 | 1.429 |
| 8 | 2.000 | 2.000 |
| 9 | 2.143 | 2.167 |
| 10 | 2.275 | 2.179 |

of the values of $S(L, N)$ for these two models for $L = 3$ and for values of $N$ between 1 and 10 is shown in table 4. Note that the values of $S(3, 6)$ and $S(3, 9)$ for both models agree with the values previously determined.

## References

[1] A. Renyi, Magyar Tudomanyos Akad. Mat. Kutato Int. Kozlemenyei 3(1958)109; see Math. Rev. 21(1960)577.
[2] H. Solomon and H. Weiner, Commun. Statistic. Theor. Meth. 15(1986)2571.
[3] P.J. Flory, J. Am. Chem. Soc. 61(1939)1518.
[4] F.T. Wall, J. Am. Chem. Soc. 62(1940)803.
[5] F.T. Wall, J. Am. Chem. Soc. 63(1941)821.
[6] F. Gornick and J.L. Jackson, J. Chem. Phys. 38(1963)1150.
[7] D.A. McQuarrie, J. Appl. Prob. 4(1967)413.
[8] E.A. Boucher, Progr. Polym. Sci. 6(1978)63.
[9] N.A. Plate and G.V. Noah, Adv. Polymer Sci. 31(1979)135.
[10] P. Rempp, Pure Appl. Chem. 46(1076)9.
[11] J.K. Mackenzie, J. Chem. Phys. 37(1962)723.
[12] F. Gornick and R.W. Freedman, J. Math. Chem. 5(1990)265.
[13] J.W. Evans and R.S. Nord, J. Chem. Phys. 93(1990)8397.
[14] J.W. Evans, private communication.
[15] E.S. Page, J. Roy. Statist. Soc. B21(1959)364.
[16] F. Downton, J. Roy. Statist. Soc. B23(1961)207.