From statistics of regular tree-like graphs to distribution function and gyration radius of branched polymers

Alexander Y Grosberg\textsuperscript{1,2} and Sergei K Nechaev\textsuperscript{3,4}

\textsuperscript{1} Physico-Chimie Curie UMR 168, Institut Curie, PSL Research University, 26 rue d’Ulm, F-75248 Paris Cedex 05, France
\textsuperscript{2} Department of Physics and Center for Soft Matter Research, New York University, 4 Washington Place, New York, NY 10003, USA
\textsuperscript{3} Université Paris—Sud/CNRS, LPTMS, UMR8626, Bât. 100, F-91405 Orsay, France
\textsuperscript{4} P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991, Moscow, Russia

E-mail: sergei.nechaev@gmail.com

Received 2 April 2015, revised 9 July 2015
Accepted for publication 10 July 2015
Published 6 August 2015

Abstract
We consider flexible branched polymer, with quenched branch structure, and show that its conformational entropy as a function of its gyration radius $R$, at large $R$, obeys, in the scaling sense, $\Delta S \sim R^2/(a^2L)$, with $a$ bond length (or Kuhn segment) and $L$ defined as an average spanning distance. We show that this estimate is valid up to at most the logarithmic correction for any tree. We do so by explicitly computing the largest eigenvalues of Kramers matrices for both regular and ‘sparse’ three-branched trees, uncovering on the way their peculiar mathematical properties.

Keywords: branched polymers, Kramers theorem, eigenvalue analysis

(Some figures may appear in colour only in the online journal)

1. Introduction

Essential progress in tailored design of polymer systems with complex architecture puts forward challenging questions of their detailed statistical description. Ensembles of long flexible branched polymers are typical examples of such very demanded systems, with wide application ranging from synthetic chemistry to biophysics of living systems. The RNA molecules provide the key biological example of polymers with branching topology. By means of covalent bonds acting between nucleotides along a chain and non-covalent...
saturating bonds between complementary bases, RNA folds in a peculiar secondary structure which is effectively a branched polymer.

There is enormous literature on RNA; for instance, the view of viral RNA as a branched polymer is considered in the works [1–3]. There is also a large variety of works on branched synthetic polymers, starting from classical study by Zimm and Stockmayer [4]. We will not review all this literature, because we are interested in a specific question which has not been sufficiently considered so far: what is the conformational entropy of a branched polymer in a three-dimensional space, and how does it depend on the overall spatial span (e.g., on the gyration radius), and on the specific arrangement of branches. To be more specific, we imagine a branched polymer, such as a secondary structured RNA, to flex bonds and junctions without re-arranging secondary structure itself, which is referred to as quenched topology, and we are interested in an entropy associated with spatial conformations of this quenched object.

Conformational entropy of this type was first considered by Daoud et al. [5], they expressed it as \( \Delta F / T \sim R^2 / R_0^2 \), where \( R \) and \( R_0 \) have been denoted as ‘actual’ and ‘unperturbed’ gyration radii, respectively. In the work [5], authors examined the influence of excluded volume interactions on swelling of branched polymers, and on properties of branched polymers in solutions; so, for them \( R \) was the average size of a self-avoiding polymer, while \( R_0 \) was considered as ideal. Actually, \( R_0 \) was taken as \( R_0 \sim aN^{1/4} \), i.e., as a mean size of an ideal tree [4], where averaging was supposed to include both spatial fluctuations of a given tree, and all rearrangements of the tree branches, namely, it was considered an annealed tree. Just such double averaging lead to the well known Zimm–Stockmayer law, \( R_0 \sim aN^{1/4} \), for ideal branched polymers [4].

Later on, a similarly looking dimensionless expression

\[
\Delta F / T \sim R^2 / (a^2 L),
\]

was employed in a more subtle context, emphasizing the difference between quenched and annealed branched polymers [6]. In that case, \( a^2 L \) replaced \( R_0^2 \). Obviously, in here, \( L \) can be thought as a length, measured in the number of bonds, of a Gaussian polymer of typical mean-squared size, \( \sim R_0^2 \). More meaningfully, \( L \) was interpreted as an order parameter, coinciding with the average (normalized by \( a^2 \)) spanning distance of a tree. This theory (applied in a variety of contexts [3, 7–9]) implies the necessity to consider conformational entropy of a branched polymer with fixed (quenched) branches, characterized by \( L \). From this point of view, the free energy estimate, \( \Delta F / T \sim R^2 / (a^2 L) \), may seem suspicious. Indeed, if \( L \) is the spanning distance of the tree, we can imagine selecting one particular line of exactly \( L \) bonds in the tree (call it a ‘tree trunk’), and then \( R^2 / (a^2 L) \) is (or seems to be) the free energy of this tree trunk viewed as a linear polymer. In fact, to relate this expression to the free energy of a tree, two aspects have to be taken into account: first, when the tree is stretched (or swollen), then all its branches are stretched, not only the trunk; second, if \( R \) is the gyration radius of the tree, then the gyration radius of the trunk is smaller than \( R \) (because branches provide a lot of mass close to the center).

The said above leads us to the following question. Is the free energy estimate, \( \Delta F / T \sim R^2 / (a^2 L) \), valid for every tree, with \( L \) defined as an average spanning distance, at least in the scaling sense? In this paper, we address just this question, and our answer is positive: we show that such estimate of entropy is valid for any tree, up to at most the logarithmic corrections.

Our arguments are based on the so-called Kramers theorem [10] (see also [11]) and its recent generalization found in [9]. These statements can be formulated as follows. Let us label all bonds of the tree (in arbitrary order) with the index \( x \), \( 1 \leq x \leq N - 1 \), where \( N \) is the number of monomers (tree vertices). For a Gaussian system (without excluded volume, and when ‘bonds’ are long enough compared to the persistence length), each bond vector, \( \mathbf{a} \), is
normally distributed in 3D space, $\sim e^{-3a^2/2a^2}$, with the mean squared bond length, $a^2 = \langle a^2 \rangle$, where the vectors $a_x$ and $a_y$ are independent for any $x \neq y$. Let us further define the $(N - 1) \times (N - 1)$ matrix $G$ with the entries $G_{xy} = M(x)M(y)/N^2$, where $M(x)$ and $M(y)$ are the numbers of tree vertices on the one side of bond $x$, while $M(y)$ is the number of vertices on the other side of bond $y$—see the illustration in figure 1 and more detailed discussion in section 2. In terms of this matrix, the Kramers theorem reads:

**Kramers theorem.** The averaged gyration radius of the tree is given by the trace $\text{tr} G$:

$$\langle R^2/a^2 \rangle = \sum_x G_{x,x} = \frac{1}{N^2} \sum_{x=1}^{N-1} M(x)(N - M(x)).$$  \hspace{1cm} (2)

Here and below, $\langle ... \rangle$ means quenched average, i.e., average over all spatial conformations with fixed tree structure.

**Generalization of Kramers theorem.** For $R \gg \langle R \rangle$, the probability $P(R)$ of a given branched polymer (with quenched tree structure) to have a gyration radius $R$, up to power law corrections, goes as

$$P(R) \big|_{R \to \infty} \sim \exp \left( -\frac{3R^2}{2a^2\lambda_{\text{max}}} \right).$$  \hspace{1cm} (3)

where $\lambda_{\text{max}}$ is the largest eigenvalue of the matrix $G$. In other words, the free energy price for swelling a quenched tree to a large size $R$, up to logarithmic corrections, goes as

$$\Delta F/T \approx 3R^2/\left(2a^2\lambda_{\text{max}}\right).$$  \hspace{1cm} (4)

The former statement (the Kramers theorem itself) can be also found in the textbook [11], while the latter statement (its generalization) was proven recently in [9]. To make the present work self-contained, we reproduce the derivation in appendix A.

The result (4) of the generalized Kramers theorem does not explain the relation between $\lambda_{\text{max}}$ and the internal geometry/topology of the tree itself. In this paper we will establish the following:

- For a regularly $p = 3$-branching tree (a ‘three-dendrimer’) $\lambda_{\text{max}} = c_{\text{tree}}$ with $c_{\text{tree}} \approx 0.957$;
- For a regular sparse tree, where each bond is a linear polymer of length $s$ (measured in the number of bonds), $\lambda_{\text{max}} = sc_{\text{tree}}$.
• For a linear polymer (or a palm tree, with trunk, but without branches), \( \lambda_{\text{max}} = 2 \frac{\pi}{N} \);

• We provide an interpolation of \( \lambda_{\text{max}} \) from ‘dense’ (\( s = 1 \)) to ‘sparse’ \( N > s \gg 1 \) (almost palm) trees in the form \( \lambda_{\text{max}} \approx \tau s \), where \( \tau \) is another constant of order of unity.

• Based on the above examples, we conclude that in general, for any tree, \( \lambda_{\text{max}} = \tilde{c}L \), where \( L \) is the average spanning distance of the tree, and \( \tilde{c} \) does not involve any powers of \( N \), but may include a factor of the order of \( 1/\ln N \).

The last statement makes the bridge between equations (1) and (4), thus proving the correctness of the former and establishing its accuracy and limitations.

2. Largest eigenvalue of Kramers matrix for regular trees

2.1. Trees with branchings at every node

To get an explicit form provided by the estimate (4), we computed the highest eigenvalue of adjacency matrix \( G \) of a regular symmetric tree. For simplicity we consider three-branching trees only, however results can be easily generalized for trees with any branching number, \( p \). To specify the system, suppose that the tree has \( k \) generations, and the total number of vertices is \( N_k \). It is convenient to represent a tree by descending levels (generations) as shown in figure 2.

The distance from the root point, \( O \), is counted in the number of levels. In order to construct the adjacency matrix \( G \) corresponding to a tree, we should enumerate somehow the tree links. Apparently, the most straightforward and naive is the sequential enumeration of links in each level. Namely, we start with a left-most link in the 1st level and enumerate all links in this generation left-to-the-right, then proceed in the same way with the next level, etc. Thus, links are enumerated sequentially first in linear subchains and then—as in (a).

- For a linear polymer (or a palm tree, with trunk, but without branches), \( \lambda_{\text{max}} = N/\pi^2 \);
- We provide an interpolation of \( \lambda_{\text{max}} \) from ‘dense’ (\( s = 1 \)) to ‘sparse’ \( N > s \gg 1 \) (almost palm) trees in the form \( \lambda_{\text{max}} \approx \tau s \), where \( \tau \) is another constant of order of unity.
- Based on the above examples, we conclude that in general, for any tree, \( \lambda_{\text{max}} = \tilde{c}L \), where \( L \) is the average spanning distance of the tree, and \( \tilde{c} \) does not involve any powers of \( N \), but may include a factor of the order of \( 1/\ln N \).

The last statement makes the bridge between equations (1) and (4), thus proving the correctness of the former and establishing its accuracy and limitations.

2. Largest eigenvalue of Kramers matrix for regular trees

2.1. Trees with branchings at every node

To get an explicit form provided by the estimate (4), we computed the highest eigenvalue of adjacency matrix \( G \) of a regular symmetric tree. For simplicity we consider three-branching trees only, however results can be easily generalized for trees with any branching number, \( p \). To specify the system, suppose that the tree has \( k \) generations, and the total number of vertices is \( N_k \). It is convenient to represent a tree by descending levels (generations) as shown in figure 2.

The distance from the root point, \( O \), is counted in the number of levels. In order to construct the adjacency matrix \( G \) corresponding to a tree, we should enumerate somehow the tree links. Apparently, the most straightforward and naive is the sequential enumeration of links in each level. Namely, we start with a left-most link in the 1st level and enumerate all links in this generation left-to-the-right, then proceed in the same way with the next level, etc. Thus, links are enumerated sequentially first in linear subchains and then—as in (a).

- For a linear polymer (or a palm tree, with trunk, but without branches), \( \lambda_{\text{max}} = N/\pi^2 \);
- We provide an interpolation of \( \lambda_{\text{max}} \) from ‘dense’ (\( s = 1 \)) to ‘sparse’ \( N > s \gg 1 \) (almost palm) trees in the form \( \lambda_{\text{max}} \approx \tau s \), where \( \tau \) is another constant of order of unity.
- Based on the above examples, we conclude that in general, for any tree, \( \lambda_{\text{max}} = \tilde{c}L \), where \( L \) is the average spanning distance of the tree, and \( \tilde{c} \) does not involve any powers of \( N \), but may include a factor of the order of \( 1/\ln N \).

The last statement makes the bridge between equations (1) and (4), thus proving the correctness of the former and establishing its accuracy and limitations.

2. Largest eigenvalue of Kramers matrix for regular trees

2.1. Trees with branchings at every node

To get an explicit form provided by the estimate (4), we computed the highest eigenvalue of adjacency matrix \( G \) of a regular symmetric tree. For simplicity we consider three-branching trees only, however results can be easily generalized for trees with any branching number, \( p \). To specify the system, suppose that the tree has \( k \) generations, and the total number of vertices is \( N_k \). It is convenient to represent a tree by descending levels (generations) as shown in figure 2.

The distance from the root point, \( O \), is counted in the number of levels. In order to construct the adjacency matrix \( G \) corresponding to a tree, we should enumerate somehow the tree links. Apparently, the most straightforward and naive is the sequential enumeration of links in each level. Namely, we start with a left-most link in the 1st level and enumerate all links in this generation left-to-the-right, then proceed in the same way with the next level, etc. Thus, links are enumerated sequentially first in linear subchains and then—as in (a).

- For a linear polymer (or a palm tree, with trunk, but without branches), \( \lambda_{\text{max}} = N/\pi^2 \);
- We provide an interpolation of \( \lambda_{\text{max}} \) from ‘dense’ (\( s = 1 \)) to ‘sparse’ \( N > s \gg 1 \) (almost palm) trees in the form \( \lambda_{\text{max}} \approx \tau s \), where \( \tau \) is another constant of order of unity.
- Based on the above examples, we conclude that in general, for any tree, \( \lambda_{\text{max}} = \tilde{c}L \), where \( L \) is the average spanning distance of the tree, and \( \tilde{c} \) does not involve any powers of \( N \), but may include a factor of the order of \( 1/\ln N \).

The last statement makes the bridge between equations (1) and (4), thus proving the correctness of the former and establishing its accuracy and limitations.
For the adjacency matrix $\tilde{G}$ of size $3(21) \times 3(21)$ we have $21 \times 21$ (for $k=3$). In figure 3(a) we provide the explicit form of the matrix $\tilde{G}$ with components $\tilde{G}_{xy} \equiv M(x)M(y)$ for the tree shown in figure 2(a). Note that conventionally the Kramers matrix is normalized by $N_k^2$ (where $N_k$ is the number of vertices), while itself the matrix $G$ is of size $k \times k$ (where $k$ is the number of links).

It is instructive to have a dictionary with necessary definitions:

- the distance from the root point on the tree, counted in the number of levels, is denoted by $j$, the total number of levels in the tree is $k$;
- the total number of vertices, $N_k$, of the $k$-level three-branching tree is $N_k = 3 \times 2^k - 2$;
- the number of vertices, $N_j$, in the subtree starting from some link in the level $j$, is $N_j = 2^{k-j+1} - 1$, where $1 \leq j \leq k$. (In figure 2(a) the subtrees with $N_{j=2, k=3} = 3$ and $N_{j=1, k=3} = 7$ are shown.) The ‘mass,’ $M_j$ of the subtree is counted in the number of vertices of this subtree.

The generic structure of the square $N_k \times N_k$ adjacency matrix $\tilde{G}$ with components $\tilde{G}_{xy} \equiv M(x)M(y)$ is schematically shown in figure 3(b). Recall that $N_k = 3 \times 2^k - 2$ is the number of links, while $N_k^2 = 3(2^k - 1)$ is the number of vertices.

Denote by $U^{(s)} = \{U_1^{(s)}, \ldots, U_k^{(s)}\}$ the eigenvector $s$ of the matrix $G$ corresponding to the eigenvalue $\lambda_s$ ($1 \leq s \leq N_k$), some of eigenvalues and eigenvectors may be degenerated. We noticed that due to the symmetry properties of the adjacency matrix $G$, its leading (maximal) eigenvector $U_{\text{max}}$ has the form

$$U_{\text{max}} = \left\{ U_1^{\text{max}}, U_1^{\text{max}}, \ldots, U_2^{\text{max}}, \ldots, U_k^{\text{max}} \right\}. \quad (5)$$

It seems instructive to show on a particular example of the $N_3 \times N_3 = 21 \times 21$ matrix $\tilde{G}$, presented in figure 3(a), how the set of eigenvectors and eigenvalues looks like.
Solving corresponding secular equation

\[ G \mathbf{U}^{(n)} = \lambda \mathbf{U}^{(n)} \]  

we find that the maximal eigenvalue, \( \lambda_{\text{max}} \), satisfies the cubic equation

\[ 7086 244 \lambda_{\text{max}}^3 - 4934 017 \lambda_{\text{max}}^2 + 291 973 \lambda_{\text{max}} - 3248 = 0 \]

which has three real roots among whose the maximal one is \( \lambda_{\text{max}} \approx 0.632 \). This eigenvalue corresponds to the eigenvector with components

\[ \mathbf{U}_{\text{max}} = \left\{ U_1^{\text{max}}, \ldots; U_2^{\text{max}}, \ldots; U_k^{\text{max}} \right\} \]

where

\[ U_1^{\text{max}} = \frac{17u^2 - 4801u + 39 732}{26 912}; \quad U_2^{\text{max}} = \frac{-u^2 + 337u - 4084}{1856}; \quad U_3^{\text{max}} = 1 \]

and \( u \) is the solution of another cubic equation

\[ u^3 - 337u^2 + 9652u - 51968 = 0. \]

The relevant root of this equation is \( u \approx 306.0139 \), which gives the following components of \( \mathbf{U}_{\text{max}} \):

\[ U_1^{\text{max}} \approx 6.0388; \quad U_2^{\text{max}} \approx 2.9085; \quad U_3^{\text{max}} = 1. \]

Computing eigenvectors \( \mathbf{U}_i \), corresponding to other eigenvalues \( \lambda_i \), we note that each \( \mathbf{U}_i \) has a wavelet structure [12]. In what follows we investigate the structure of the pair \( (\lambda_{\text{max}}, \mathbf{U}_{\text{max}}) \) only.

Plugging the ansatz (5) into the generic equation \( G \mathbf{U}^{(n)} = \lambda \mathbf{U}^{(n)} \), we find the leading eigenvalue \( \lambda_{\text{max}} \) to be also the leading eigenvalue of an exponentially smaller matrix \( B \):

\[ B \mathbf{V}^{\text{max}} = \lambda_{\text{max}} \mathbf{V}^{\text{max}}, \]

where \( \mathbf{V}^{\text{max}} = \{ U_1^{\text{max}}, U_2^{\text{max}}, \ldots, U_k^{\text{max}} \} \), and the matrix \( B \) reads

\[ B = \begin{pmatrix}
    b_{11} & b_{12} & b_{13} & b_{14} & \ldots & b_{1k} \\
    b_{12} & b_{22} & b_{23} & b_{24} & \ldots & b_{2k} \\
    b_{13} & b_{23} & b_{33} & b_{34} & \ldots & b_{3k} \\
    b_{14} & b_{24} & b_{34} & b_{44} & \ldots & b_{4k} \\
    \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
    b_{1k} & b_{2k} & b_{3k} & b_{4k} & \ldots & b_{kk}
  \end{pmatrix};
\]

\[ b_{ij} = 2^{-i-2j} \left( 2^{i+k} - 3 \times 2^{2+i+k} + 3 \times 2^i \right) \times N_i^k \quad (j \geq i). \]
definiteness we take each first row in 1st, 2nd and 3rd levels of \( G \):

\[
\begin{align*}
\begin{array}{l}
\begin{array}{l}
b_{11} = \frac{105 + 2 \times 49}{22^2} = \frac{203}{484} ;
\end{array} \\
b_{12} = \frac{2 \times 45 + 4 \times 21}{22^2} = \frac{174}{484} ;
\end{array} \\
\begin{array}{l}
\begin{array}{l}
b_{13} = \frac{4 \times 15 + 8 \times 7}{22^2} = \frac{116}{484} ;
\end{array} \\
b_{21} = \frac{45 + 2 \times 11}{22^2} = \frac{87}{484} ;
\end{array} \\
\begin{array}{l}
\begin{array}{l}
b_{22} = \frac{57 + 5 \times 9}{22^2} = \frac{102}{484} ;
\end{array} \\
b_{23} = \frac{2 \times 19 + 10 \times 3}{22^2} = \frac{68}{484} ;
\end{array} \\
\begin{array}{l}
\begin{array}{l}
b_{31} = \frac{15 + 2 \times 7}{22^2} = \frac{29}{484} ;
\end{array} \\
b_{32} = \frac{19 + 5 \times 3}{22^2} = \frac{34}{484} ;
\end{array} \\
\begin{array}{l}
\begin{array}{l}
b_{33} = \frac{21 + 11 \times 1}{22^2} = \frac{32}{484} .
\end{array}
\end{array}
\end{align*}
\]

One can straightforwardly check that these values are reproduced by the generic expression for \( b_{ij} \) in (8) for \( k = 3 \).

In figures 4(a), (b) we have plotted the matrix elements of \( B \) for \( k = 5, 30 \) in form of a 3D relief \( B_{ij} \), over the base plane \((i, j)\), where \((i, j) = 1, \ldots, k\). Computing numerically the maximal eigenvalues of the true matrices \( B \) defined in (8), for different \( k \), we get a plot in figure 4(c).

Extrapolating to \( k \to \infty \) the data shown in figure 4, we obtain

\[
\lambda_{\text{max}} \approx 0.957
\]

meaning that the largest eigenvalue, \( \lambda_{\text{max}} \) of the matrix \( B \) and, consequently, of the initial Kramers matrix \( G \), is bounded from above by some numeric value independent on the matrix size (and without any logarithmic corrections).

It is noteworthy that a rough estimate of the eigenvalue \( \lambda_{\text{max}} \) can be done using the following simple analytical argument. All matrix elements \( B_{ij} \) are positive and

\[
B_{ij} = \begin{cases} 
   b_{ij} & j \geq i, \\
   b_{ij}/2^{i-j} & j < i. 
\end{cases}
\]

Taking into account the explicit form of matrix elements \( b_{ij} \) for \( j \geq i \) (see (8)), we can approximate each element \( B_{ij} \) for any \((i, j)\) by \( b_{ij} \). This approximation gives exact value of \( B_{ij} \) for \( j \geq i \), and provides an upper estimate of elements \( B_{ij} \) for \( j < i \).

The maximal eigenvalue, \( \lambda_{\text{max}} \), of the \( k \times k \) matrix \( \hat{B} \) with entries \( \{b_{ij}\} \) can be computed exactly. Equation (8) means that the element \( b_{ij} \) can be factorized as

\[
\ldots
\]
The largest eigenvalue, $\lambda_{\text{max}}$, reads

$$\bar{\lambda}_{\text{max}} = \sum_{m=1}^{k} f_m g_m = \frac{7 \times 2^{2+2k} - 27k \times 2^k - 15 \times 2^k - 13}{3(3 \times 2^k - 2)^2}. \quad (11)$$

For $k \to \infty$ on gets from (11)

$$\bar{\lambda}_{\text{max}} \mid_{k \to \infty} = \frac{28}{27} \approx 1.037 \quad (12)$$

which is a pretty good approximation for numerically found true limiting value $\lambda_{\text{max}} \approx 0.957$.

Returning to the main question of this paper, which is about conformational entropy, one may want to ask if it is possible to compute this entropy for a regular tree directly, with reference to the generalized Kramers theorem. Surprisingly, we found this exercise non-trivial, and will not do it here.

### 2.2. Sparse regular trees

We can generalize our approach to ‘sparse’ regular trees, in which branchings are connected by linear subchains of $s$ links. By definition, for $s = 1$ we return to the former case of branching at every node of the tree. Particular example of a dendrimer of $N = 19$ nodes with three-branching vertices separated by two-link subchains (i.e. $s = 2$), is shown in figure 2(b).

To construct the Kramers adjacency matrix, $G^{(s)}$, for such a tree, one proceeds as above, using the descent diagram representation, where we first enumerate sequentially links in each linear subchain, and then pass to the next subchain in the same level (compare figures 2(a) and (b). The matrix $G^{(s)}$ constructed in such a way, highly resembles the former one, $G$, where each matrix element $G_{i,j}$ is replaced by a $s \times s$-block (a $2 \times 2$ block for a particular example of figure 2(b).

The largest eigenvalue, $\lambda_{\text{max}}^{(s)}$, of the composite adjacency matrix, $G^{(s)}$, can be estimated as

$$\lambda_{\text{max}}^{(s)} \approx \lambda_{\text{max}}^{(1)} \lambda_{\text{max}}^{(s)} \quad (13)$$

where $\lambda_{\text{max}}^{(1)}$ is the maximal eigenvalue of a Kramers matrix for a linear chain of $s$ links and $\lambda_{\text{max}}^{(s)}$ is given by (9). The value of $\lambda_{\text{max}}^{(1)}$ reads:

$$\lambda_{\text{max}}^{(1)} (s) = \frac{s + 1}{4s^2} \sin^{-2} \frac{\pi}{2(s + 1)}; \quad \lambda_{\text{max}}^{(1)} (s) = \begin{cases} \lambda_{\text{max}}^{(1)} (s = 1) = 1, \\ \lambda_{\text{max}}^{(1)} (s \gg 1) = \frac{s}{\pi^2}. \end{cases} \quad (14)$$

As it follows from (9), the value $\lambda_{\text{max}}^{(s)}$ can be approximated by some constant, $c$, of order of unity. Thus, we can estimate $\lambda_{\text{max}}^{(s)} (s)$ for all regular sparse trees as

$$\lambda_{\text{max}}^{(s)} (s) \approx c \lambda_{\text{max}}^{(1)} (s) = \varepsilon s \quad (15)$$

where $\varepsilon$ absorbs all numerical constants.
3. Discussion

Let us now summarize what we learned about trees. To begin with, recall that for a linear polymer the whole spectrum of the Kramers matrix is known [13] (see also small corrections in [14] and also [15]); in particular
\[ \lambda_{\text{max}} = \frac{N}{\pi}. \]
Since spanning distance in this case is just
\[ L = N, \]
we have
\[ \lambda_{\text{max}} \approx \frac{L}{\pi} \sim L. \]

Next, for the perfect three-branching dendrimer we have established above that
\[ \lambda_{\text{max}} \approx 1. \]
This has to be compared with average spanning distance of this tree. For dendrimer with \( k \) generations and, accordingly, \( 3 \times 2^k \) ends, the sum of distances along a backbone from one end to all other ends is equal to \( 2 + 2^{k+1} (3k + 1) \). Hence,
\[ L = \frac{1 + 2^{k+1}}{3 \times 2^k}, \]
and at large \( k \) this asymptotically tends to
\[ L \approx 2k. \]
Since
\[ N = 3 \times 2^k + 2, \]
(or, equivalently, \( k \approx \log_2 \frac{N}{3} - 1 \approx \log_2 N \)), we have
\[ L \approx 2 \log_2 N. \]
Therefore, we can say that
\[ \lambda_{\text{max}} \approx L/(2 \log_2 N). \]

Thus, for a regular tree, \( \lambda_{\text{max}} \) is related to the average spanning distance \( L \) via a factor of order \( \ln N \). Let us show now that this is the worst case, and for all other trees the estimate
\[ \lambda_{\text{max}} \sim L \]
is even more accurate.

Sparse regular trees provide in this sense a good insight, as they smoothly interpolate between regular trees and linear chains. As we have seen, in this case
\[ \lambda_{\text{max}} \approx s. \]
On the other hand, in this case
\[ N = 3s(2^k - 1) + 1, \]
or
\[ k = \log_2 \left[ \frac{N - s - 1}{3s} \right]; \]
at large \( N \) this is asymptotic to
\[ k \approx \log_2 \left[ \frac{N}{s} \right]. \]
At the same time,
\[ L \approx 2sk \text{ or } L \approx 2s \log_2 \left[ \frac{N}{s} \right]. \]
Thus
\[ \lambda_{\text{max}} \approx \frac{L}{2 \log_2 (N/s)}. \]

We see that the factor relating \( \lambda_{\text{max}} \) to mean spanning distance \( L \) smoothly changes between about \( \ln N \) and about 1 as \( s \) changes from 1 to \( N \).

In fact, it is clear physically that a regularly branching tree is the most compact and most difficult to swell of all quenched branched polymers, while linear polymer is the least compact and the easiest to swell. Since we have rigorously analyzed both of these extremes, we can conclude with the following: first, the generalized Kramers theorem (4) relates conformational entropy to \( \lambda_{\text{max}} \) up to logarithmic corrections (because it results from a saddle point integration, see equation (A.6)); second, \( \lambda_{\text{max}} \) is related to the average spanning distance of the tree, \( L \), by another logarithmic factor. Together, this implies also the validity of the scaling estimate of entropy (1) in terms of the average spanning distance.

The usefulness of our analysis is due to the fact that scaling estimate of entropy finds an increasingly active application, particularly in the physics of viral RNA. Self-assembly of small RNA viruses must involve the work to compress RNA to place it inside the capsid. It is noted that viral RNA are unusually compact in terms of their secondary structure [1], which is likely to be an evolutionary adaption to decrease the packing work [16]. Nevertheless, some amount of work remains [2], and the most natural tool to estimate it [3] is formula (1). In fact, given the detailed knowledge of the viral RNA secondary structure, we can expect it possible to explicitly construct the corresponding adjacency matrix and to numerically compute its largest eigenvalue, which, according to our work, is a good predictor of the relative ease of packing the given RNA into a virus shell [17].
Acknowledgments

Both authors are very grateful to R Everaers, Y Fyodorov, and J Smrek for useful discussions. This work has been started in Paris, where AYG was on a long term visit; AYG acknowledges the hospitality of both the Curie Institute and ESPCI. The work of AYG on this project was supported in part by the National Science Foundation under Grant No. PHY-1066293 and the hospitality of the Aspen Center for Physics. The work of SKN was partially supported by DIONICOS European project.

Appendix A. Generalized Kramers theorem

Here, we compute the gyration radius probability distribution for a Gaussian tree. The approach follows the works of Fixman [13–15], and [18], where it was generalized for Gaussian rings.

Suppose for simplicity that the branching number is $p = 3$, and there are only branchings and ends, namely, $n$ three-valent vertices and $n + 2$ ends, totally $N = 2n + 2$ monomers. If $r_i$ are position vectors of these ‘monomers,’ then the gyration radius reads

$$R^2 = \frac{1}{2N^2} \sum_{i,j} (r_i - r_j)^2.$$  \hspace{1cm} (A.1)

On the tree, every $r_i - r_j$ is uniquely represented as the sum of the set of bond vectors $a_k = r_i - r_j$, where $i$ and $i'$ are monomers connected by bond $k$. Accordingly, gyration radius can be represented as

$$R^2 = \sum_{x,y} G_{x,y} a_x a_y,$$  \hspace{1cm} (A.2)

where the indices $x$ and $y$ label bonds (unlike $i$, $i'$ and $j$ above, which label vertices or monomers), and $G_{x,y}$ is a Kronecker $(N - 1) \times (N - 1)$ matrix, illustrated in figure 1.

For a Gaussian tree we can do more and find the probability distribution of $R^2$, assuming each bond has the probability distribution $e^{-\lambda_s^2/2 \lambda_s^2}$. The characteristic function of $R^2$ reads

$$\Phi(s) = \left\langle e^{-i s R^2} \right\rangle = A \int d^3 \{a\} e^{-\sum \lambda_s^2 / (2 \lambda_s^2) + i s \sum x,y G_{x,y} a_x a_y},$$ \hspace{1cm} (A.3)

where the explicit expression for normalization factor $A$ is dropped for brevity. Rotating now the coordinate system in this $N - 1$ dimensional space of $\{a\}$ to the basis of eigenvectors $\{b\}$ of matrix $G$, we obtain

$$\Phi(s) = A \int d^3 \{b\} e^{-\sum \lambda_p^2 \left( \frac{3}{2 \lambda_p^2} - i s \lambda_p \right) \left( 1 - \frac{2 \lambda_p^2 i s}{3} \right)^{-3/2}},$$ \hspace{1cm} (A.4)

where $\lambda_p$ are eigenvalues, and normalization factor $A$ is reconstructed from the condition $\Phi(s)|_{s=0} = 1$. From here, finding the probability distribution of $R^2$ is a matter of inverse Fourier transform of $\Phi(s)$

$$P(R^2) = \frac{1}{2\pi} \int ds \exp \left( -i s R^2 - \frac{3}{2 \lambda_p^2} \sum_p \ln \left( 1 - \frac{2 \lambda_p^2 i s}{3} \right) \left( 1 - \frac{2 \lambda_p^2 i s}{3} \right)^{-3/2} \sum_p \ln \left( 1 - \frac{2 \lambda_p^2 i s}{3} \right) \left( 1 - \frac{2 \lambda_p^2 i s}{3} \right)^{-3/2} \right).$$ \hspace{1cm} (A.5)

Since we are interested in the behavior of $P(R^2)$ at large $R$, the asymptotics is controlled by the singularity of $\Phi(s)$ closest to the origin in the complex $s$-plane, that corresponds to the largest eigenvalue $\lambda_{max} \approx \lambda_{max}$. In the vicinity of this singularity there is a saddle point
which dominates the integral, and we can evaluate inverse Fourier transform integral by steepest descent. The equation for saddle point location reads

\[ R^2 \approx \frac{a^2 \lambda_{\text{max}}}{1 - i s (2a^2 / 3) \lambda_{\text{max}}} \]

or

\[ s = \left( \frac{3}{2R^2} - \frac{3}{2a^2 \lambda_{\text{max}}} \right). \]

Thus, the result of saddle point integration (up to logarithmic corrections) is

\[ P(R^2)_{R \to \infty} \sim \exp \left( -\frac{3R^2}{2a^2 \lambda_{\text{max}}} + \frac{3}{2} \ln \frac{R^2}{\lambda_{\text{max}}} \right) \sim \exp \left( -\frac{3R^2}{2a^2 \lambda_{\text{max}}} \right), \quad (A.6) \]

yielding the expected generalization of Kramers theorem (4).

References

[1] Gopal A, Egecioglu D E, Yoffe A M, Ben-Shaul A, Rao A L N, Knobler C M and Gelbart W M 2014 PLoS One 9 105875
[2] Gopal A, Zhou Z, Knobler C M and Gelbart W M 2012 RNA 18 284
[3] Bruinsma R F, Comas-Garcia M, Garmann R F and Grosberg A Y 2015 Quasi-equilibrium self-assembly of small RNA viruses arXiv:cond-mat.soft/1505.01224
[4] Zimm B H and Stockmayer W H 1949 J. Chem. Phys. 17 1301
[5] Daoud M, Pincus P, Stockmayer W H and Witten T 1983 Macromolecules 16 1833
[6] Gutzin A M, Grosberg A Y and Shakhnovich E I 1993 Macromolecules 26 1293
[7] Rosa A and Everaers R 2014 Phys. Rev. Lett. 112 118302
[8] Grosberg A Y 2014 Soft Matter 10 560
[9] Smrek J and Grosberg A Y 2015 J. Phys.: Condens. Matter 27 064117
[10] Kramers H A 1946 J. Chem. Phys. 14 415
[11] Rubinstein M and Colby R 2005 Polymer Physics (Oxford: Oxford University Press)
[12] Chui C 1992 An Introduction to Wavelets (San Diego, CA: Academic)
[13] Fixman M 1962 J. Chem. Phys. 36 306
[14] Forsman W C 1965 J. Chem. Phys. 42 2829
[15] Fujita H and Norisuye T 1970 J. Chem. Phys. 52 1115
[16] Tubiana L, Božič A L, Micheletti C and Podgornik R 2015 Biophys. J. 108 194
[17] Bruinsma R 2015 private communication
[18] Moore N T, Lua R C and Grosberg A Y 2005 Physical and Numerical Models in Knot Theory, on Knots and Everything ed J A Calvo, K C Millet, E J Rawdon and A Stasiak (Singapore: World Scientific) ch 19, pp 363–84