Combinatorics and Matrix model on RNA secondary structures

SangKwan Choi
Physics department, Sichuan University

1612.07468 w/ Chaiho Rim and Hwajin Um

Integrability in low-dimensional quantum systems
MATRIX, The University of Melbourne
26 June – 21 July 2017
Enumeration example

Number of ways connecting vertices with non-crossing arcs on the upper half plane?
Enumeration example

Number of ways connecting vertices with non-crossing arcs on the upper half plane?
**Enumeration example**

Number of ways connecting vertices with non-crossing arcs on the upper half plane?

- **Catalan number**

  Recursion relation

  \[
  C_0 = 1 \quad C_{\ell+1} = \sum_{k=0}^{\ell} C_k C_{k-\ell}
  \]
Enumeration example

Number of ways connecting vertices with non-crossing arcs on the upper half plane?

- Catalan number

Recursion relation

\[ C_0 = 1 \]
\[ C_{\ell+1} = \sum_{k=0}^{\ell} C_k C_{k-\ell} \]

Generating function

\[ c(z) = \sum_{\ell=0}^{\infty} C_\ell z^\ell \]

\[ c(z) = 1 + zc(z)^2 \]
Enumeration example

Number of ways connecting vertices with non-crossing arcs on the upper half plane?

- Catalan number

Recursion relation

\[ C_0 = 1 \quad C_{\ell+1} = \sum_{k=0}^{\ell} C_k C_{k-\ell} \]

Generating function

\[ c(z) = \sum_{\ell=0}^{\infty} C_\ell z^\ell \]

\[ c(z) = \frac{1 - \sqrt{1 - 4z}}{2z} \]

\[ c(z) = 1 + z c(z)^2 \]
**Enumeration example**

Number of ways connecting vertices with non-crossing arcs on the upper half plane?

- *Catalan number*
  
  Recursion relation
  
  $$C_0 = 1 \quad C_{\ell+1} = \sum_{k=0}^{\ell} C_k C_{k-\ell}$$

  Generating function
  
  $$c(z) = \sum_{\ell=0}^{\infty} C_\ell z^\ell$$
  
  $$c(z) = 1 + z c(z)^2$$
  
  $$c(z) = \frac{1 - \sqrt{1 - 4z}}{2z}$$
  
  $$C_\ell = \frac{1}{\ell + 1} \binom{2\ell}{\ell}$$
**Enumeration in terms of Matrix model**

Gaussian matrix model

\[
\int dM \text{tr} M^4 e^{-\frac{N}{2} \text{tr} M^2} \sim \langle \text{tr} M^4 \rangle = N^2 \left( \begin{array}{c}
\begin{array}{c}
\end{array}
\end{array} \right) + N^0 \begin{array}{c}
\begin{array}{c}
\end{array}
\end{array}
\]

\[
V - E + F = \chi = 2 - 2g
\]
Enumeration in terms of Matrix model

Gaussian matrix model

\[
\int dM \, \text{tr} M^4 \, e^{-\frac{N}{2} \text{tr} M^2} \sim \langle \text{tr} M^4 \rangle = N^2 \left( \begin{array}{c} \text{Diagram 1} \\ + \\ \text{Diagram 2} \end{array} \right) + N^0 \quad \begin{array}{c} \text{Diagram 3} \end{array}
\]

\[ V - E + F = \chi = 2 - 2g \]

\[ N^{2-2g} \]
Enumeration in terms of Matrix model

Gaussian matrix model

\[
\int dM \, \text{tr} M^4 \, e^{-\frac{N}{2} \text{tr} M^2} \sim \langle \text{tr} M^4 \rangle = N^2 \left( \begin{array}{c}
\begin{array}{c}
\text{Fatgraph}
\end{array}
\end{array} \right) + N^0
\]

\[
V - E + F = \chi = 2 - 2g
\]

\[
N^{2-2g}
\]
Enumeration in terms of Matrix model

Gaussian matrix model

\[ \int dM \, \text{tr} M^4 \, e^{-\frac{N}{2} \text{tr} M^2} \sim \langle \text{tr} M^4 \rangle = N^2 \left( \underbrace{\text{Fatgraph}}_{\text{Fatgraph}} + \underbrace{\text{Fatgraph}}_{\text{Fatgraph}} \right) + N^0 \]

\[ V - E + F = \chi = 2 - 2g \]

\[ N^{2-2g} \]
Enumeration in terms of Matrix model

Gaussian matrix model

\[ \int dM \text{tr} M^4 e^{-\frac{N}{2} \text{tr} M^2} \sim \langle \text{tr} M^4 \rangle \]

\[ = N^2 \left( \begin{array}{c}
\text{Fatgraph}
\end{array} \right) + N^0 \]

\[ N^{2-2g} \]

\[ V - E + F = \chi = 2 - 2g \]

Number of ways connecting \( k \) vertices with non-crossing arcs on the upper half plane

\[ \cdots \]

\[ C_6 \]

Number of genus-0 diagrams by connecting \( k \)-legs vertex with propagators

\[ \langle \text{tr} M^{12} \rangle_0 \]
Enumeration in terms of Matrix model

Loop equation possesses an associated recursion relation

• Loop equation

\[ \int dM \, e^{-N \text{tr}V(M)} \]

\[ W(z)^2 - V'(z)W(z) = f(z) \]

\[ V(z) = \frac{z^2}{2} \]

\[ W(z) := \frac{1}{N} \left\langle \text{tr} \frac{1}{z - M} \right\rangle \]

\[ f(z) := \frac{1}{N} \left\langle \text{tr} \frac{V'(z) - V'(M)}{z - M} \right\rangle = 1 \]
Enumeration in terms of Matrix model

Loop equation possesses an associated recursion relation

- Loop equation \[ \int dM \, e^{-N \text{tr} V(M)} \]

\[ W(z)^2 - V'(z)W(z) = f(z) \]

\[ W(z) := \frac{1}{N} \left\langle \text{tr} \frac{1}{z - M} \right\rangle \quad f(z) := \frac{1}{N} \left\langle \text{tr} \frac{V'(z) - V'(M)}{z - M} \right\rangle = 1 \]

Large z expansion

\[ W(z) = \frac{1}{N} \sum_{a \geq 0} \frac{\left\langle \text{tr} M^a \right\rangle}{z^{a+1}} \quad W(z)^2 = \frac{1}{N^2} \sum_{a \geq 0} \sum_{b=0}^{a} \frac{\left\langle \text{tr} M^b \right\rangle \left\langle \text{tr} M^{b-a} \right\rangle}{z^{a+2}} \]

At each order of 1/z,

\[ \frac{\left\langle \text{tr} M^0 \right\rangle}{N} = 1 \quad \frac{\left\langle \text{tr} M^{2(\ell+1)} \right\rangle}{N} = \sum_{k=0}^{\ell} \frac{\left\langle \text{tr} M^{2k} \right\rangle \left\langle \text{tr} M^{2(k-\ell)} \right\rangle}{N} \]

\[ C_0 = 1 \quad C_{\ell+1} = \sum_{k=0}^{\ell} C_k C_{k-\ell} \]
Enumeration in terms of Matrix model

Loop equation possesses an associated recursion relation

- Loop equation

\[ W(z)^2 - V'(z)W(z) = f(z) \]

\[ W(z) = \frac{1}{N} \left\langle \mathrm{tr} \frac{1}{z - M} \right\rangle \quad f(z) = \frac{1}{N} \left\langle \mathrm{tr} \frac{V'(z) - V'(M)}{z - M} \right\rangle = 1 \]

Large \( z \) expansion

\[ W(z) = \frac{1}{N} \sum_{a \geq 0} \frac{\langle \mathrm{tr} M^a \rangle}{z^{a+1}} \]

\[ W(z)^2 = \frac{1}{N^2} \sum_{a \geq 0} \sum_{b=0}^{a} \frac{\langle \mathrm{tr} M^b \rangle \langle \mathrm{tr} M^{b-a} \rangle}{z^{a+2}} \]

At each order of \( 1/z \),

\[ \frac{\langle \mathrm{tr} M^0 \rangle}{N} = 1 \]

\[ \frac{\langle \mathrm{tr} M^{2(\ell+1)} \rangle}{N} = \sum_{k=0}^{\ell} \frac{\langle \mathrm{tr} M^{2k} \rangle}{N} \frac{\langle \mathrm{tr} M^{2(k-\ell)} \rangle}{N} \]

\[ C_0 = 1 \quad C_{\ell+1} = \sum_{k=0}^{\ell} C_k C_{k-\ell} \]

This enumeration is related to RNA secondary structures \textit{without} pseudoknot.
Background to RNA

- RNA (Ribonucleic acid) is a linear polymer with a backbone of ribose sugar rings linked by phosphate groups. Each sugar has one of the four bases, adenine (A), cytosine (C), guanine (G), and uracil (U) attached to the 1’ position.
- A phosphate group is attached to the 3’ position of one ribose and the 5’ position of the next. This imposes directionality on the backbone.

- Primary structure is a sequence of letters (ACGU) that indicate the order of nucleotides from 5’ to 3’. GCGCUGUGUCGA
Background to RNA

• RNA is usually single stranded and base pairs are formed intra-molecularly.

• Base pair: hydrogen bonds are possible between AU and CG pairs (Watson-Crick base pair) and between less stable GU pairs (Wobble base pairs).

AAGUCAGGGAUUGCACCAGUUGUUCG

• RNA secondary structure consists of the base pairs within a molecule.
Background to RNA

- Tertiary structure is its three-dimensional shape.
- The functions of RNA require a precise tertiary structure.
Background to RNA

• Tertiary structure is its three-dimensional shape.

• The functions of RNA require a precise tertiary structure.

• There has been little experimental information on tertiary structure.
Background to RNA

- Tertiary structure is its three-dimensional shape.

- The functions of RNA require a precise tertiary structure.

- There has been little experimental information on tertiary structure.

- Prediction of RNA secondary structure is important.
  - Scaffold of a tertiary structure is its secondary structure.
  - It is thought that stable secondary structures form first and tertiary structures form afterwards.
RNA secondary structure

• RNA secondary structures can be viewed as diagrams in which nucleotides are represented by labeled vertices \(\{1,...,n\}\) in a horizontal line and base pairs are drawn as arcs in the upper half-plane.
RNA secondary structure

- RNA secondary structures can be viewed as diagrams in which nucleotides are represented by labeled vertices \{1,\ldots,n\} in a horizontal line and base pairs are drawn as arcs in the upper half-plane.

- No crossing arcs (crossing arc: \((i,j)\) \((k,l)\) and \(i<k<j<l\))

There are good algorithms to the secondary structure predictions based on the recursion [Nusinov, Waterman ’78]

\[
T_{n+1} = T_n + \sum_{k=\lambda}^{n-1} T_k T_{n-k-1}
\]
**RNA secondary structure**

- RNA secondary structures can be viewed as diagrams in which nucleotides are represented by labeled vertices \( \{1, \ldots, n\} \) in a horizontal line and base pairs are drawn as arcs in the upper half-plane.

- No crossing arcs (crossing arc: \((i,j)\) \((k,l)\) and \(i<k<j<l\))

There are good algorithms to the secondary structure predictions based on the recursion [Nusinvo, Waterman ’78]

\[
T_{n+1} = T_n + \sum_{k=\lambda}^{n-1} T_k T_{n-k-1}
\]
RNA secondary structure

- RNA secondary structures can be viewed as diagrams in which nucleotides are represented by labeled vertices \{1,...,n\} in a horizontal line and base pairs are drawn as arcs in the upper half-plane.

- No crossing arcs (crossing arc: \((i,j)\) \((k,l)\) and \(i<k<j<l\))

There are good algorithms to the secondary structure predictions based on the recursion [Nusinov, Waterman '78]

\[ T_{n+1} = T_n + \sum_{k=\lambda}^{n-1} T_k \cdot T_{n-k-1} \]

- efficiently (Polynomial time) find the structure with minimum free energy.
RNA secondary structure with pseudoknot

- Structures with crossing arcs have been considerably discovered.
RNA pseudoknots
RNA secondary structure with pseudoknot

- Structures with crossing arcs have been considerably discovered.
RNA pseudoknots

- It is now clear that pseudoknots are common in RNA and often have a functional role.
RNA secondary structure with pseudoknot

- Structures with crossing arcs have been considerably discovered. RNA pseudoknots

- It is now clear that pseudoknots are common in RNA and often have a functional role.

- To predict the secondary structures with arbitrary pseudoknots is *NP-complete* [Lyngsø, Pedersen '00] which means that there would be no efficient algorithms.

- We need further filtration.
Combinatorial classifications on pseudoknot

- $k$-noncrossing [Reidys, Jin, Qin ‘08]:
  There exist no $k$-mutually crossing arcs.
Combinatorial classifications on pseudoknot

- $k$-noncrossing [Reidys, Jin, Qin ’08]: There exist no $k$-mutually crossing arcs.

- Bi-secondary [Haslinger, Stadler ’99]: Diagram that can be drawn in the plane without intersections of arcs. 2-page book-embedding.
Combinatorial classifications on pseudoknot

- **k-noncrossing** [Reidys, Jin, Qin ‘08]:
  There exist no k-mutually crossing arcs.

- **Bi-secondary** [Haslinger, Stadler ‘99]:
  Diagram that can be drawn in the plane without intersections of arcs.
  2-page book-embedding.

- **Genus** [Zee, Orland ‘08, Penner, Andersen ‘10, Reidys ‘11]:
  Matrix model

\[ V - E + F = \chi = 2 - 2g \]

- Example:
  \[ V = 1 \quad E = 3 \quad F = 2 \]
  \[ g = 1 \]
Combinatorial classifications on pseudoknot

• k-noncrossing [Reidys, Jin, Qin ‘08]:
  There exist no k-mutually crossing arcs.

• Bi-secondary [Haslinger, Stadler ‘99]:
  Diagram that can be drawn in the plane without intersections of arcs.
  2-page book-embedding.
  No known generating function

• Genus [Zee, Orland ‘08, Penner, Andersen ‘10, Reidys ‘11]: Matrix model

\[ V - E + F = \chi = 2 - 2g \]

\[ V = 1 \quad E = 3 \quad F = 2 \]

\[ g = 1 \]
**Bi-secondary structure**

- Bi-secondary structures
  - Generating function
  - Recursion associated with the Generating function
Bi-secondary structure

- Bi-secondary structures
  - Generating function
  - Recursion associated with the Generating function

Bi-secondary + Genus description

Surface with boundaries

\[ V - E + F = \chi = 2 - 2g - b \]
**Bi-secondary structure**

- Bi-secondary structures
  - Generating function
  - Recursion associated with the Generating function

Bi-secondary + Genus description
Surface with boundaries

\[ V - E + F = \chi = 2 - 2g - b \]

\[ \chi(D_U \cup D_L) = \chi(D_U) + \chi(D_L) - \chi(D_U \cap D_L) \]
Bi-secondary structure

- Bi-secondary structures
  - Generating function
  - Recursion associated with the Generating function

Bi-secondary + Genus description
Surface with boundaries

\[ V - E + F = \chi = 2 - 2g - b \]

\[ \chi(D_U \cup D_L) = \chi(D_U) + \chi(D_L) - \chi(D_U \cap D_L) \]
Island

- Introduce an auxiliary element

Island: a set of consecutive nucleotides all of which participate in the planar diagram
**Island**

- Introduce an auxiliary element

**Island**: a set of consecutive nucleotides all of which participate in the planar diagram

- It is forbidden to have an arc between $i$ and $i + 1$ due to the tension of the backbone.
**Island**

- Introduce an auxiliary element

Island: a set of consecutive nucleotides all of which participate in the planar diagram

- It is forbidden to have an arc between $i$ and $i + 1$ due to the tension of the backbone.

It’s impossible to have an arc inside an island
Matrix model description

Matrix model

Enumeration of genus 0 configurations *without self-interaction*

\[ \int dM \, \text{tr} M^8 \, e^{-N \text{tr} M^2} \sim \langle \text{tr} M^8 \rangle_0 \]
Matrix model description

Matrix model

Enumeration of genus 0 configurations \textit{without self-interaction}

\[
\int dM \ trM^8 \ e^{-NtrM^2} \sim \langle trM^8 \rangle_0
\]

\[
\langle trM^2 M^2 M^3 M \rangle_0 \quad \xrightarrow{\text{\LARGE \Rightarrow}} \quad \langle trP_2(M)P_2(M)P_3(M)P_1(M) \rangle_0
\]
Matrix model description

Matrix model

Enumeration of genus 0 configurations *without self-interaction*

\[
\langle \text{tr} M^2 M^2 M^3 M \rangle_0 \quad \Rightarrow \quad \langle \text{tr} P_2(M) P_2(M) P_3(M) P_1(M) \rangle_0
\]

Recursion relation

\[
P_0(M) = 1 \quad P_1(M) = M \quad P_{n+1}(M) = M P_n(M) - P_{n-1}(M)
\]

Loop equation (MM with vector)

\[
\int dM \: \text{tr} M^8 \: e^{-N\text{tr} M^2} \sim \langle \text{tr} M^8 \rangle_0
\]
Matrix model description

Matrix model
Enumeration of genus 0 configurations *without self-interaction*

\[ \int dM \, \text{tr} M^8 \, e^{-N\text{tr} M^2} \sim \langle \text{tr} M^8 \rangle_0 \]

\[ \langle \text{tr} M^2 M^2 M^3 M \rangle_0 \rightarrow \langle \text{tr} P_2(M) P_2(M) P_3(M) P_1(M) \rangle_0 \]

**Recursion relation**

\[ P_0(M) = 1 \quad P_1(M) = M \quad P_{n+1}(M) = M P_n(M) - P_{n-1}(M) \]

\[ = \quad \times \left( \quad - \quad - \quad - \right) \quad - \quad \times \left( \quad - \quad \right) \]
Matrix model description

Matrix model

Enumeration of genus 0 configurations \textit{without self-interaction}

\[
\int dM \, \text{tr} M^8 \, e^{-N \text{tr} M^2} \sim \langle \text{tr} M^8 \rangle_0
\]

\[
\langle \text{tr} M^2 M^2 M^3 M \rangle_0 \quad \Rightarrow \quad \langle \text{tr} P_2(M) P_2(M) P_3(M) P_1(M) \rangle_0
\]

Recursion relation

\[
P_0(M) = 1 \quad P_1(M) = M \quad P_{n+1}(M) = M \, P_n(M) - P_{n-1}(M)
\]

\[
= \quad \cdots - \quad \hat{\quad} \cdots - \quad \cdots + \quad \hat{\quad} \quad \cdots
\]

Loop equation (MM with vector)

\[
U_0(\xi) = 1 \quad U_1(\xi) = 2\xi \quad U_{n+1}(\xi) = 2\xi \, U_n(\xi) - U_{n-1}(\xi)
\]

Chebyshev polynomial of 2\textsuperscript{nd} kind
Matrix model description

Matrix model
Enumeration of genus 0 configurations without self-interaction

\[ \langle \text{tr} M^2 M^2 M^3 M \rangle_0 \quad \Rightarrow \quad \langle \text{tr} P_2(M)P_2(M)P_3(M)P_1(M) \rangle_0 \quad \Rightarrow \quad \langle \text{tr} U_2 U_2 U_3 U_1 \rangle_0 \]

Recursion relation

\[ P_0(M) = 1 \quad P_1(M) = M \quad P_{n+1}(M) = MP_n(M) - P_{n-1}(M) \]

\[ \begin{align*} &\quad - \quad - \quad - \quad + \quad + \\ = &\quad \times \quad ( \quad - \quad - \quad ) \quad - \quad \times \quad ( \quad - \quad ) \end{align*} \]

Chebyshev polynomial of 2\(^{nd}\) kind

\[ U_0(\xi) = 1 \quad U_1(\xi) = 2\xi \quad U_{n+1}(\xi) = 2\xi U_n(\xi) - U_{n-1}(\xi) \]

\[ \int dM \text{tr} M^8 e^{-N\text{tr} M^2} \sim \langle \text{tr} M^8 \rangle_0 \]
Matrix model description

Matrix model
enumeration of genus 0 configurations \textit{without self-interaction}

\[ \langle \text{tr}U_2 U_3 U_1 \rangle_0 \]

\[ U_m(\xi) U_n(\xi) = \sum_{k=0}^{n} U_{m-n+2k}(\xi) \]
Matrix model description

Matrix model

count of genus 0 configurations \textit{without self-interaction}

\[
\langle \text{tr} U_2 U_2 U_3 U_1 \rangle_0 \\
U_m(\xi)U_n(\xi) = \sum_{k=0}^{n} U_{m-n+2k}(\xi)
\]

\[
\langle \text{tr} U_2 U_2 U_3 U_1 \rangle_0 = \langle \text{tr} (U_8 + 3U_6 + 5U_4 + 5U_2 + 2U_0) \rangle_0
\]
Matrix model description

Matrix model
enumeration of genus 0 configurations \textit{without self-interaction}

\[
\langle \text{tr} U_2 U_2 U_3 U_1 \rangle_0 = \langle \text{tr} (U_8 + 3U_6 + 5U_4 + 5U_2 + 2U_0) \rangle_0
\]

Non-zero contribution = Coefficient of $U_0$
Matrix model description

Matrix model

enumeration of genus 0 configurations *without self-interaction*

\[ \langle \text{tr} U_2 U_2 U_3 U_1 \rangle_0 = \langle \text{tr} (U_8 + 3U_6 + 5U_4 + 5U_2 + 2U_0) \rangle_0 \]

*Non-zero contribution = Coefficient of \( U_0 \)*

orthogonality

\[ \int_{-1}^{1} U_k(\xi) U_\ell(\xi) \sqrt{1 - \xi^2} d\xi = \frac{\pi}{2} \delta_{k,\ell} \]

\[ y^3 z^\ell \sum_{a+b+c=\ell} \langle \text{tr} U_a U_b U_c \rangle_0 = \sum_{a+b+c=\ell} \frac{2}{\pi} \int_{-1}^{1} U_a(\xi) U_b(\xi) U_c(\xi) U_0(\xi) \sqrt{1 - \xi^2} d\xi \]
Generating function

\( g(h, I, \ell) \): number of planar diagrams with \( h \) hairpins, \( I \) islands and \( \ell \) arcs

Generating function

\[
G(x, y, z) = \sum_{h, I, \ell} g(h, I, \ell) x^h y^I z^\ell
\]

\[
G(x, y, z) = \left( \frac{y}{1 + y} \right) \frac{1 - A(1 + B) - \sqrt{1 - 2A(1 + B) + A^2(1 - B)^2}}{2A}
\]

\[
A = z(1 + y)^2, \quad B = \frac{xy}{1 + y}
\]
Generating function

\( g(h, I, \ell) \): number of planar diagrams with \( h \) hairpins, \( I \) islands and \( \ell \) arcs

Generating function

\[
G(x, y, z) = \sum_{h, I, \ell} g(h, I, \ell) x^h y^I z^\ell
\]

\[
G(x, y, z) = \left( \frac{y}{1 + y} \right) \frac{1 - A(1 + B) - \sqrt{1 - 2A(1 + B) + A^2(1 - B)^2}}{2A}
\]

\[
A = z(1 + y)^2, \quad B = \frac{xy}{1 + y}
\]

- One may manipulate the generating function to take into account further structural elements such as stem, bulge, interior loop and multi loop.
Generating function

$g(h, I, \ell)$: number of planar diagrams with $h$ hairpins, $I$ islands and $\ell$ arcs

Generating function

$$G(x, y, z) = \sum_{h,I,\ell} g(h, I, \ell) x^h y^I z^\ell$$

$$G(x, y, z) = \left(\frac{y}{1 + y}\right) \frac{1 - A(1 + B) - \sqrt{1 - 2A(1 + B) + A^2(1 - B)^2}}{2A}$$

$$A = z(1 + y)^2, \quad B = \frac{xy}{1 + y}$$

• One may manipulate the generating function to take into account further structural elements such as *stem, bulge, interior loop and multi loop*.

• Stem: a set of maximally consecutive parallel arcs
Generating function with stem

\[ f(k, h, I, \ell) \]: number of diagrams with \( k \) stems, \( h \) hairpins, \( I \) islands and \( \ell \) arcs

\[ s(h, I, k) \]: number of single-stack diagram(#stem=#arc)
  with \( k \) stems, \( h \) hairpins and \( \ell \) arcs
Generating function with stem

\( f(k, h, I, \ell) \): number of diagrams with \( k \) stems, \( h \) hairpins, \( I \) islands and \( \ell \) arcs

\( s(h, I, k) \): number of single-stack diagram (#stem=#arc)

with \( k \) stems, \( h \) hairpins and \( \ell \) arcs

The number of ways stacking \( \ell - k \) arcs on \( k \) arcs of the single-stack diagram:

\[
\binom{\ell - 1}{k - 1}
\]

\[ f(k, h, I, \ell) = \binom{\ell - 1}{k - 1} s(h, I, k) \]
Generating function with stem

relation

\[ f(k, h, I, \ell) = \binom{\ell - 1}{k - 1} s(h, I, k) \]

\[
\sum_{k, h, I, \ell} f(k, h, I, \ell) u^k x^h y^I z^\ell = \sum_{k, h, I, \ell} \binom{\ell - 1}{k - 1} s(h, I, k) u^k x^h y^I z^\ell
\]
Generating function with stem

relation

\[ f(k, h, I, \ell) = \binom{\ell - 1}{k - 1} s(h, I, k) \]

\[
\sum_{k,h,I,\ell} f(k, h, I, \ell) u^k x^h y^I z^\ell = \sum_{k,h,I,\ell} \binom{\ell - 1}{k - 1} s(h, I, k) u^k x^h y^I z^\ell
\]

\[
= \sum_{k,h,I} s(h, I, k) \left( \frac{u z}{1 - z} \right)^k x^h y^I
\]
Generating function with stem

relation

\[ f(k, h, I, \ell) = \binom{\ell - 1}{k - 1} s(h, I, k) \]

\[ \sum_{k, h, I, \ell} f(k, h, I, \ell) u^k x^h y^I z^\ell = \sum_{k, h, I, \ell} \left( \binom{\ell - 1}{k - 1} \right) s(h, I, k) u^k x^h y^I z^\ell \]

\[ = \sum_{k, h, I} s(h, I, k) \left( \frac{u z}{1 - z} \right)^k x^h y^I \]

In terms of generating function

\[ F(u, x, y, z) = S(x, y, \frac{u z}{1 - z}) \]
Generating function with stem

relation

\[ f(k, h, I, \ell) = \binom{\ell}{k} s(h, I, k) \]

\[
\sum_{k,h,I,\ell} f(k, h, I, \ell) u^k x^h y^I z^{\ell} = \sum_{k,h,I,\ell} \binom{\ell-1}{k-1} s(h, I, k) u^k x^h y^I z^{\ell}
\]

\[
= \sum_{k,h,I} s(h, I, k) \left( \frac{uz}{1-z} \right)^k x^h y^I
\]

In terms of generating function

\[ F(u, x, y, z) = S(x, y, \frac{uz}{1-z}) \]

\[ F(1, x, y, z) = S(x, y, \frac{z}{1-z}) \]
Generating function with stem

relation

\[ f(k, h, I, \ell) = \binom{\ell - 1}{k - 1} s(h, I, k) \]

\[ \sum_{k,h,I,\ell} f(k, h, I, \ell) u^k x^h y^I z^\ell = \sum_{k,h,I,\ell} \binom{\ell - 1}{k - 1} s(h, I, k) u^k x^h y^I z^\ell \]

\[ = \sum_{k,h,I} s(h, I, k) \left( \frac{u z}{1 - z} \right)^k x^h y^I \]

In terms of generating function

\[ F(u, x, y, z) = S\left( x, y, \frac{u z}{1 - z} \right) \]

\[ F(1, x, y, z) = S\left( x, y, \frac{z}{1 - z} \right) = G(x, y, z) \]
Generating function with stem

relation

\[ f(k, h, I, \ell) = \binom{\ell - 1}{k - 1} s(h, I, k) \]

\[
\sum_{k, h, I, \ell} f(k, h, I, \ell) u^k x^h y^I z^\ell = \sum_{k, h, I, \ell} \binom{\ell - 1}{k - 1} s(h, I, k) u^k x^h y^I z^\ell
\]

\[ = \sum_{k, h, I} s(h, I, k) \left( \frac{uz}{1-z} \right)^k x^h y^I \]

In terms of generating function

\[ F(u, x, y, z) = S\left(x, y, \frac{uz}{1-z} \right) \]

\[ F(1, x, y, z) = S\left(x, y, \frac{z}{1-z} \right) = G(x, y, z) \]

Obtain two generating functions \( F \) and \( S \) in terms of known generating function \( G \)

\[ S(x, y, u) = G\left(x, y, \frac{u}{1+u} \right) \quad F(u, x, y, z) = G\left(x, y, \frac{uz}{1+uz-z} \right) \]
Generating function

Similarly, the generating function with all the structure elements in terms of $G(x, y, z)$

$$
G(x, y, z, x_b, x_i, x_m) = G(x, y, z) \bigg| \left\{ \begin{array}{l}
    x \to \frac{x}{x_m} , \\
    z \to \frac{z \cdot x_m}{1 - z \left(1 - \frac{x_m}{y} + 2 \left(\frac{x_m}{y}\right) + \left(x_m - \frac{x}{y}\right)^2\right)}
\end{array} \right\}
$$
Density of state in structure space

\( \Omega_{k, \ell} \): Number of secondary structures with \( k \) stems and \( \ell \) arcs

\[
F(u, x, y, z) = G\left( x, y, \frac{uz}{1 + uz - z} \right) \quad \Rightarrow \quad \Omega_{k, \ell} = \binom{\ell - 1}{k - 1} 5^{k - 1} {}_2F_1 \left( \frac{1 - k}{2}, \frac{2 - k}{2}; 2; \frac{32}{25} \right)
\]
Density of state in structure space

\( \Omega_{k,\ell} \): Number of secondary structures with \( k \) stems and \( \ell \) arcs

\[
F(u, x, y, z) = G \left( x, y, \frac{u z}{1 + u z - z} \right) \quad \Rightarrow \quad \Omega_{k,\ell} = \binom{\ell - 1}{k - 1} 5^{k-1} _2F_1 \left( \frac{1 - k}{2}, \frac{2 - k}{2}; 2; \frac{32}{25} \right)
\]

Figure: Normalized distribution of \( \Omega_{k,\ell} \) as a function of \( k \) when \( \ell = 200 \)
Density of state in structure space

\( \Omega_{k, \ell} \): Number of secondary structures with \( k \) stems and \( \ell \) arcs

\[
F(u, x, y, z) = G\left(x, y, \frac{u z}{1 + u z - z}\right) \quad \Rightarrow \quad \Omega_{k, \ell} = \binom{\ell - 1}{k - 1} 5^{k-1} \, _2F_1 \left( \frac{1 - k}{2}, \frac{2 - k}{2}; 2; \frac{32}{25} \right)
\]

Minimum free energy structures tend to have the average stack length per stem \( \ell/k \) in the range from 2 to 10, mostly near 4 independent of \( \ell \).
Density of state in structure space

Algorithmic point of view

- Exhaustive algorithms would waste time verifying structures which never be a minimum free energy structure.

\[
\frac{\sum_{k=1}^{100} \Omega_{k,200}}{\sum_{k=1}^{200} \Omega_{k,200}} \approx 3.76 \times 10^{-52}
\]
Density of state in structure space

Algorithmic point of view

- Exhaustive algorithms would waste time verifying structures which never be a minimum free energy structure.

\[
\frac{\sum_{k=1}^{100} \Omega_{k,200}}{\sum_{k=1}^{200} \Omega_{k,200}} \approx 3.76 \times 10^{-52}
\]

- \( r \)-canonical structure: Each stem has at least the number \( r \) of arcs.
Density of state in structure space

Algorithmic point of view

- Exhaustive algorithms would waste time verifying structures which never be a minimum free energy structure.

\[ \frac{\sum_{k=1}^{100} \Omega_{k,200}}{\sum_{k=1}^{200} \Omega_{k,200}} \approx 3.76 \times 10^{-52} \]

- \( r \)-canonical structure: Each stem has at least the number \( r \) of arcs.

- RNA can have a stem with the number of arcs less than \( r > 1 \). Possibly fail to predict minimum free energy structures.
Density of state in structure space

Algorithmic point of view

• Exhaustive algorithms would waste time verifying structures which never be a minimum free energy structure.

\[
\frac{\sum_{k=1}^{100} \Omega_{k,200}}{\sum_{k=1}^{200} \Omega_{k,200}} \approx 3.76 \times 10^{-52}
\]

• \(r\)-canonical structure: Each stem has at least the number \(r\) of arcs.

• RNA can have a stem with the number of arcs less than \(r > 1\). possibly fail to predict minimum free energy structures.

• But worth it if the speed of prediction is important
Density of state in structure space

Algorithmic point of view

• Exhaustive algorithms would waste time verifying structures which never be a minimum free energy structure.

$$\sum_{k=1}^{100} \Omega_{k,200} / \sum_{k=1}^{200} \Omega_{k,200} \approx 3.76 \times 10^{-52}$$

• $r$-canonical structure: Each stem has at least the number $r$ of arcs.

• RNA can have a stem with the number of arcs less than $r > 1$. Possibly fail to predict minimum free energy structures.

• But worth it if the speed of prediction is important

$$F(u, x, y, z) = G\left(x, y, \frac{u z}{1 + u z - z}\right) \quad \rightarrow \quad F(u z^{r-1}, x, y, z) = G\left(x, y, \frac{u z^r}{1 + u z^r - z}\right)$$
Density of state in structure space

Introduce a certain weight (fugacity) in creating a stem (equivalently, loop): $e^\mu$

$$F(u, x, y, z) = G\left(x, y, \frac{u z}{1 + u z - z}\right) \quad \Rightarrow \quad F(u e^{\mu}, x, y, z) = G\left(x, y, \frac{u z e^{\mu}}{1 + u z e^{\mu} - z}\right)$$
Density of state in structure space

Introduce a certain weight (fugacity) in creating a stem (equivalently, loop): $e^\mu$

\[
F(u, x, y, z) = G\left(x, y, \frac{uz}{1 + uz - z}\right) \quad \Rightarrow \quad F(u e^\mu, x, y, z) = G\left(x, y, \frac{uz e^\mu}{1 + uz e^\mu - z}\right)
\]

Histogram of RNA for $80 \leq \ell \leq 120$ obtained from database ($\# = 222$) and normalized distribution $e^{\mu k} \Omega_{k,\ell}$ for $\mu = -3.6$, $\ell = 100$ as a function of $k/\ell$. 

Summary

- Prediction of RNA secondary structure with pseudoknot needs further filtration.
- Matrix model is suitable for enumeration when especially genus filtration is required.
- Matrix integral is easily calculated to find the Generating function with the help of the recursion derived from its loop equation.
- In statistical point of view, fugacity was introduced to compare the structure space with minimum free energy space.

Future work

- Generating function of bi-secondary structure
- Algebraic equation or PDE for the GF.
- Comparison structure space with pseudoknot with minimum free energy space.
- 3-page book-embedding, 4-page, etc.
**Summary**

- Prediction of RNA secondary structure with pseudoknot needs further filtration.
- Matrix model is suitable for enumeration when especially genus filtration is required.
- Matrix integral is easily calculated to find the Generating function with the help of the recursion derived from its loop equation.
- In statistical point of view, fugacity was introduced to compare the structure space with minimum free energy space.

**Future work**

- Generating function of bi-secondary structure
- Algebraic equation or PDE for the GF.
- Comparison structure space with pseudoknot with minimum free energy space.
- 3-page book-embedding, 4-page, etc.

*Thank you for your attention!*