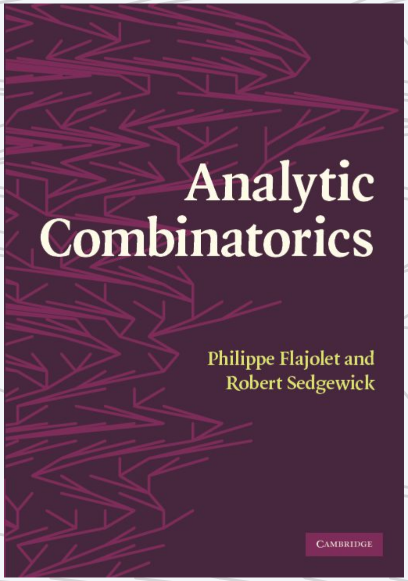
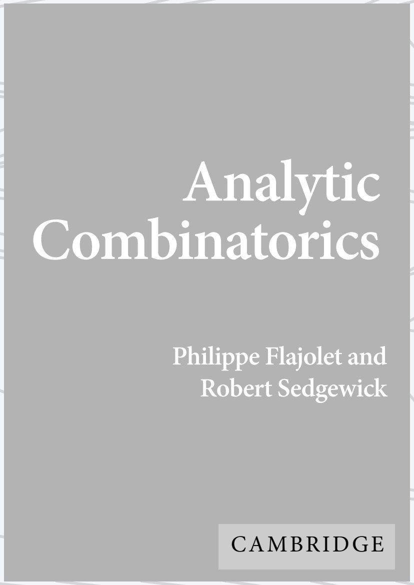


# Combinatorial structures and OGFs



Analytic Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

# Combinatorial structures and OGFs

- **Symbolic method**
- Trees and strings
- Powersets and multisets
- Compositions and partitions
- Substitution

## Analytic combinatorics overview

To analyze properties of a large combinatorial structure:

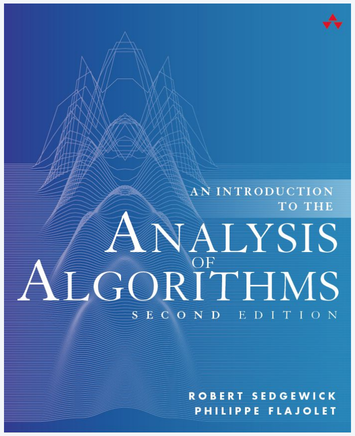
1. Use the symbolic method
   - Define a *class* of combinatorial objects
   - Define a notion of *size* (and associated generating function)
   - Use standard operations to develop a *specification* of the structure

Result: A direct derivation of a GF equation (implicit or explicit)

Classic next steps:

- Extract coefficients
- Use classic asymptotics to estimate coefficients

Result: Asymptotic estimates that quantify the desired properties



http://aofa.cs.princeton.edu

See *An Introduction to the Analysis of Algorithms* for a gentle introduction

## Analytic combinatorics overview

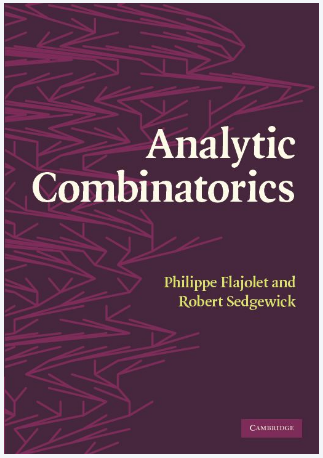
To analyze properties of a large combinatorial structure:

1. Use the symbolic method
   - Define a *class* of combinatorial objects.
   - Define a notion of *size* (and associated generating function)
   - Use standard operations to develop a *specification* of the structure.

Result: A direct derivation of a GF equation (implicit or explicit).

2. Use complex asymptotics to estimate growth of coefficients.
   - [no need for explicit solution]
   - [stay tuned for details]

Result: Asymptotic estimates that quantify the desired properties



http://ac.cs.princeton.edu

See *Analytic Combinatorics* for a rigorous treatment

# Analytic combinatorics overview

**A. SYMBOLIC METHOD**

➡ 1. OGFs

2. EGFs

3. MGFs

**B. COMPLEX ASYMPTOTICS**

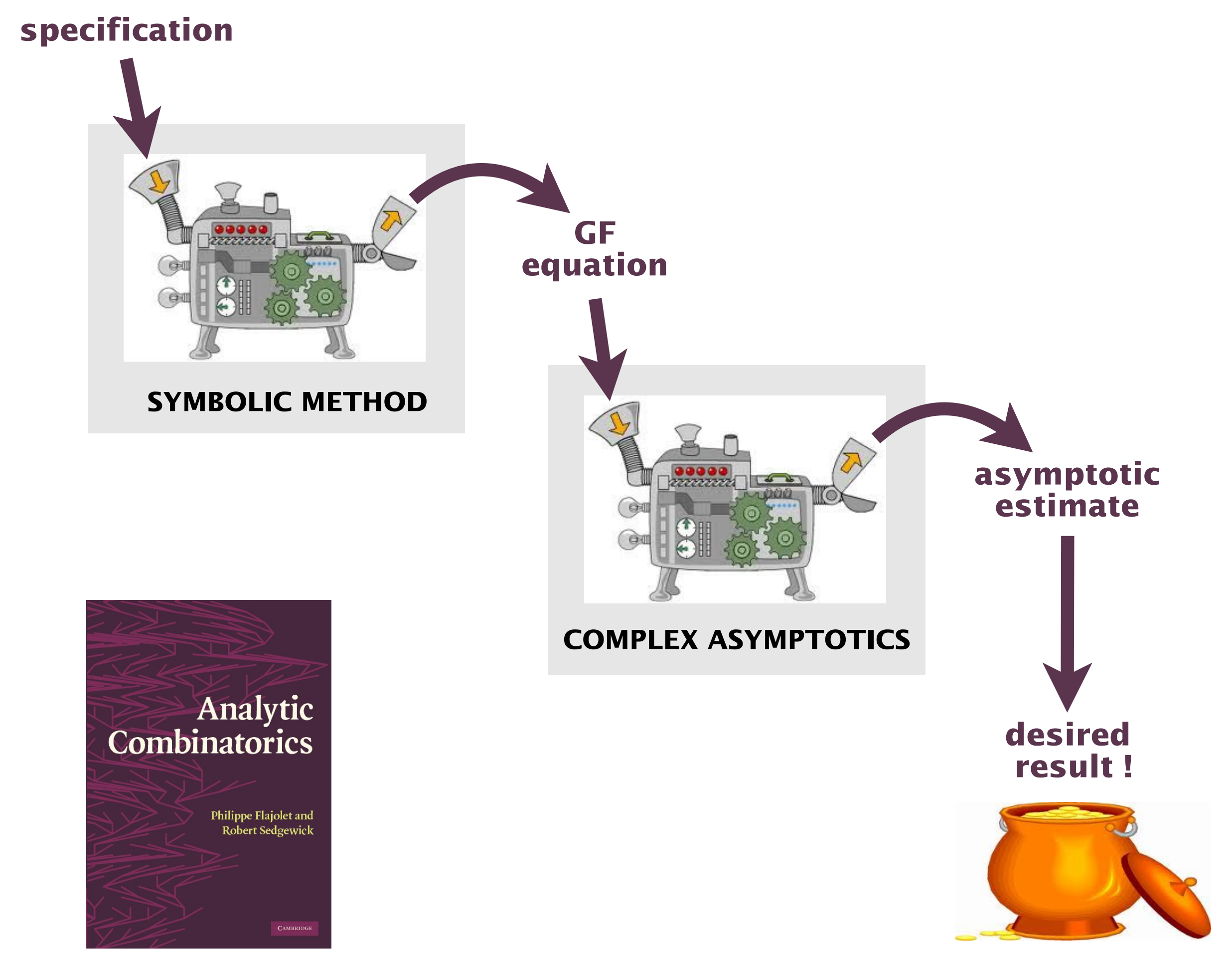
4. Rational & Meromorphic

5. Applications of R&M

6. Singularity Analysis

7. Applications of SA

8. Saddle point



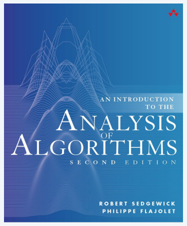
# The symbolic method

An approach for *directly* deriving GF equations.

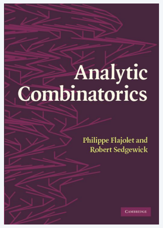
- Define a *class* of combinatorial objects.
- Define a notion of *size* (and associated generating function)
- Define *operations* suitable for constructive definitions of objects.
- Prove *correspondences* between operations and GFs.

Result: A GF equation (implicit or explicit).

See *An Introduction to the Analysis of Algorithms* for a gentle introduction



See *Analytic Combinatorics* for a rigorous treatment



# Basic definitions

**Def.** A *combinatorial class* is a set of combinatorial objects and an associated *size* function.

**Def.** The *ordinary generating function* (OGF) associated with a class is the formal power series $A(z) = \sum_{a \in A} z^{|a|}$ ← size function

object name

class name

Fundamental (elementary) identity

$$A(z) \equiv \sum_{a \in A} z^{|a|} = \sum_{N \geq 0} A_N z^N$$

**Q.** How many objects of size $N$ ?

**A.** $A_N = [z^N]A(z)$

*Fantasy*:
Different letter for each class

*Reality*:
Only 26 letters!

*Usual conventions*

| | | |
|---|---|---|
| class name | roman | $A$ |
| OGF name | roman with arg | $A(z)$ |
| object variable | lowercase | a |
| coefficient | subscripted | $A_N$ |
| size | $N$ or $n$ | |

With the symbolic method, we specify the class *and at the same time* characterize the OGF

# Unlabeled classes: cast of characters

TREES
*Recursive structures*
$T_N$ = [*Catalan #s*]

STRINGS
*Sequences of characters*
$S_N = N^M$

INTEGERS
*N objects*
$I_N = 1$

COMPOSITIONS
*Positive integers sum to N*
$C_N = 2^{N-1}$

LANGUAGES
*Sets of strings*
[REs and CFGs]

PARTITIONS
*Unordered compositions*
[enumeration not elementary]

# The symbolic method (basic constructs)

Suppose that *A* and *B* are classes of unlabeled objects with enumerating OGFs *A*(*z*) and *B*(*z*).

| operation | notation | semantics | OGF |
|---|---|---|---|
| *disjoint union* | $A + B$ | disjoint copies of objects from *A* and *B* | $A(z) + B(z)$ |
| *Cartesian product* | $A \times B$ | ordered pairs of copies of objects, one from *A* and one from *B* | $A(z)B(z)$ |
| *sequence* | $SEQ(A)$ | sequences of objects from *A* | $\frac{1}{1 - A(z)}$ |

# Proofs of correspondences

*A* + *B*

$$\sum_{c\in A+B} z^{|c|} = \sum_{a\in A} z^{|a|} + \sum_{b\in\ B} z^{|b|} = A(z) + B(z)$$

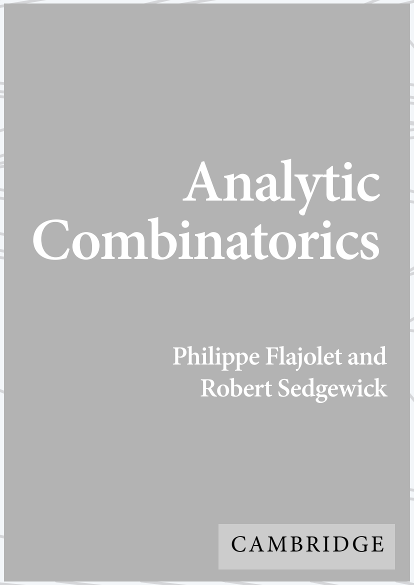
*A* × *B*

$$\sum_{c\in a\times b} z^{|c|} = \sum_{a\in A}\sum_{b\in B} z^{|a|+|b|} = \Big(\sum_{a\in A} z^{|a|}\Big)\Big(\sum_{b\in B} z^{|b|}\Big) = A(z)B(z)$$

Text

*SEQ*( *A* )

| construction | OGF |
|---|---|
| $SEQ_k(A) \equiv A^k$ | $A(z)^k$ |
| $SEQ_T(A) \equiv A^{t_1} + A^{t_2} + A^{t_3} + \ldots$<br>where $T \equiv t_1, t_2, t_3, \ldots$ is a subset of the integers | $A(z)^{t_1} + A(z)^{t_2} + A(z)^{t_3} + \ldots$ |
| $SEQ(A) \equiv \epsilon + A + A^2 + A^3 + \ldots$ | $1 + A(z) + A(z)^2 + A(z)^3 + \ldots = \dfrac{1}{1 - A(z)}$ |

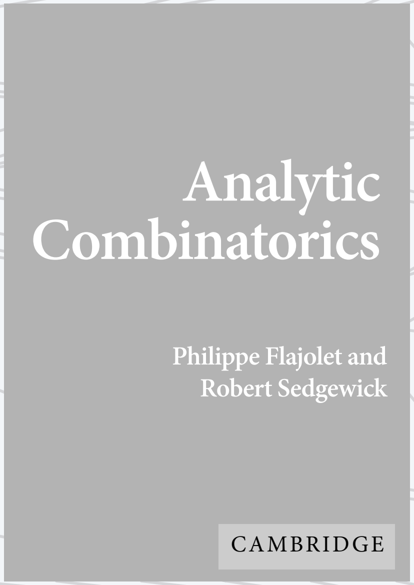

Analytic Combinatorics

Philippe Flajolet and Robert Sedgewick

CAMBRIDGE

# Combinatorial structures and OGFs

- **Symbolic method**
- Trees and strings
- Powersets and multisets
- Compositions and partitions
- Substitution



Analytic Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

# Combinatorial structures and OGFs

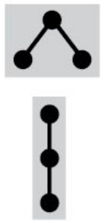
- Symbolic method
- **Trees and strings**
- Powersets and multisets
- Compositions and partitions
- Substitution

# Classic example of the symbolic method

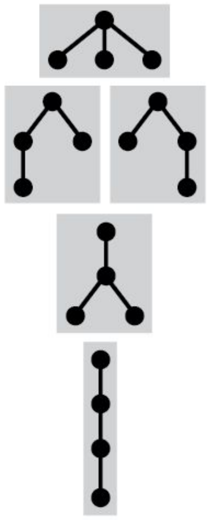
Q. How many trees with $N$ nodes?

$G_1 = 1$

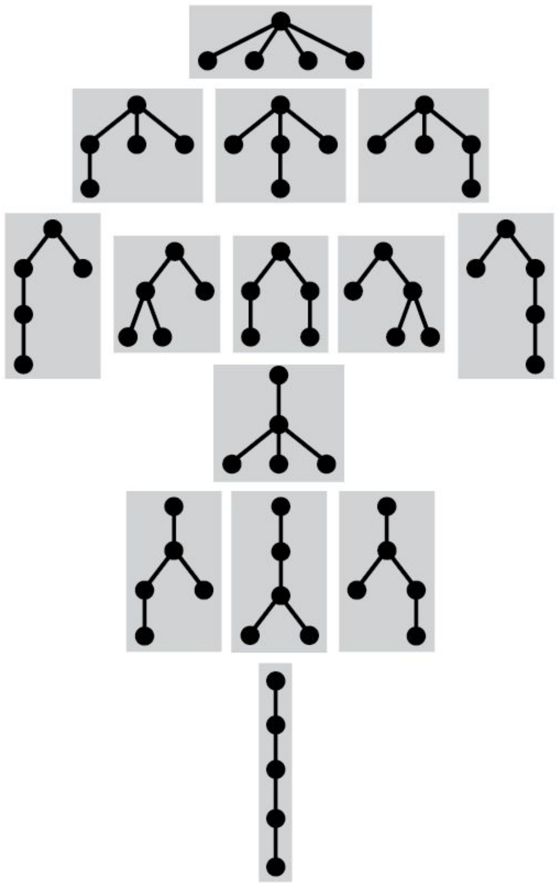
$G_2 = 1$



$G_3 = 2$



$G_4 = 5$



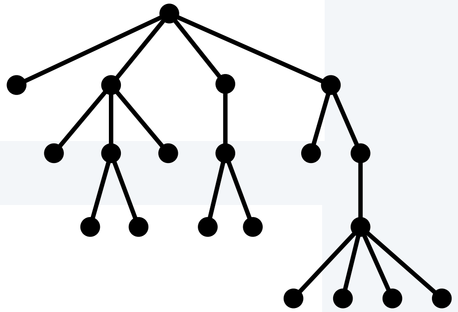
$G_5=14$

# Analytic combinatorics: How many trees with *N* nodes?

## Symbolic method

| | |
|---|---|
| Combinatorial class | *G*, the class of all trees |
| Construction | $G = \bullet \times SEQ(G)$ |
| OGF equation | $G(z) = z(1 + G(z) + G(z)^2 + G(z)^3 + \ldots) = \dfrac{z}{1 - G(z)}$ |
| | $G(z) - G(z)^2 = z$ |
| Quadratic equation | $G(z) = \dfrac{1 + \sqrt{1 - 4z}}{2}$ |

"a tree is a node and a sequence of trees"



## Classic next steps

| | |
|---|---|
| Binomial theorem | $G(z) = -\dfrac{1}{2} \sum_{N \geq 1} \binom{\frac{1}{2}}{N} (-4z)^N$ |
| Extract coefficients | $G_N = -\dfrac{1}{2} \binom{\frac{1}{2}}{N} (-4)^N = \dfrac{1}{N} \binom{2N-2}{N-1} \quad = \dfrac{1}{4N-2} \binom{2N}{N}$ |
| Stirling's approximation | $\sim \dfrac{1}{4N} \exp\big(2N \ln(2N) - 2N + \ln\sqrt{4\pi N} - 2(N \ln(N) - N + \ln\sqrt{2\pi N})\big)$ |
| Simplify | $G_N \sim \dfrac{4^{N-1}}{\sqrt{\pi N^3}}$ |

detailed calculations omitted

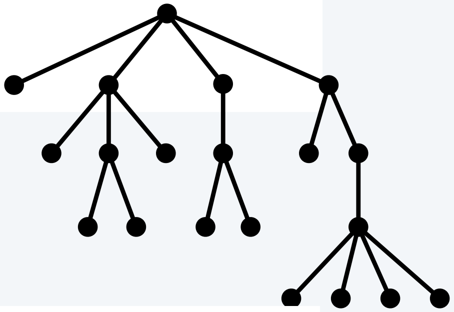
# Analytic combinatorics: How many trees with *N* nodes?

## Symbolic method

| | |
|---|---|
| Combinatorial class | *G*, the class of all trees |
| Construction | $G = \bullet \times SEQ(G)$ |
| OGF equation | $G(z) = z(1 + G(z) + G(z)^2 + G(z)^3 + \ldots) = \dfrac{z}{1 - G(z)}$ |

$$G(z) - G(z)^2 = z$$

"a tree is a node and a sequence of trees"



## Complex asymptotics

Singularity analysis

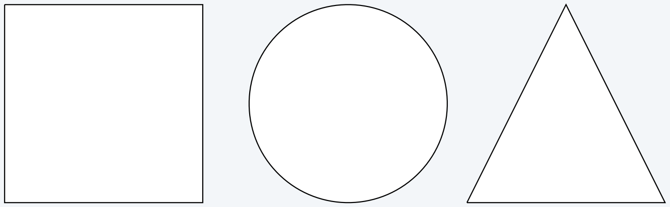
$$G_N = [z^N]G(z) \sim \frac{4^N}{\Gamma(1/2)\sqrt{N}} = \frac{4^N}{\sqrt{\pi N}}$$

GF equation *directly* implies asymptotics

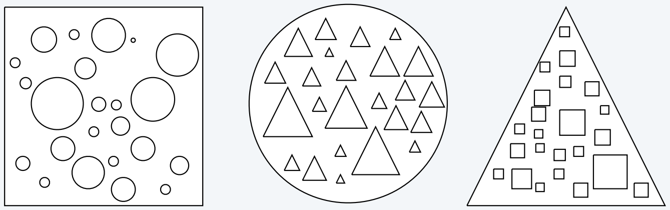
**This lecture**: Focus on symbolic method for deriving OGF equations (stay tuned for asymptotics).

# A standard paradigm for the symbolic method

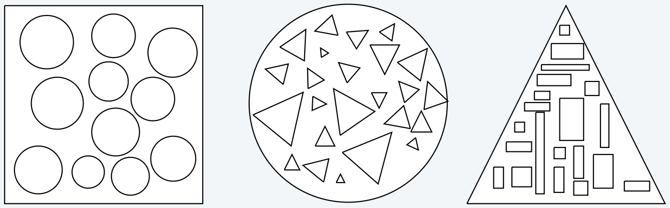
Fundamental constructs
- elementary or trivial
- confirm intuition



Compound constructs
- many possibilities
- classical combinatorial objects
- expose underlying structure
- one of many paths to known results



Variations
- unlimited possibilities
- *not* easily analyzed otherwise

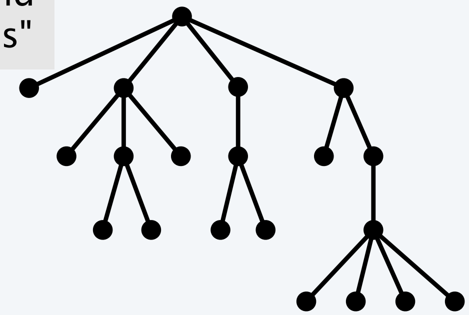


# Variations on a theme 1: Trees

Fundamental construct

| | |
|---|---|
| Combinatorial class | $G$, the class of all trees |
| Construction | $G = \bullet \times SEQ(G)$ |
| OGF equation | $G(z) = z(1 + G(z) + G(z)^2 + G(z)^3 + \ldots) = \dfrac{z}{1 - G(z)}$ |
| | $G(z) - G(z)^2 = z$ |

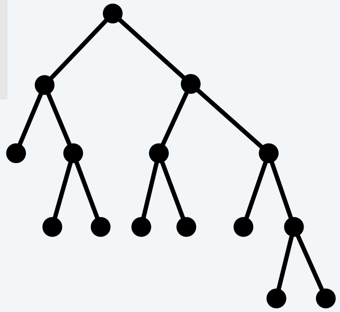
"a tree is a node and a sequence of trees"



Variation on the theme: *restrict each node to* 0 *or* 2 *children*

| | |
|---|---|
| Combinatorial class | $T$, the class of binary trees |
| Construction | $T = \bullet \times SEQ_{0,2}(T)$ |
| OGF equation | $T(z) = z(1 + T(z)^2)$ |

"a binary tree is a node and a sequence of 0 or 2 binary trees"



# Variations on a theme 1: Trees (continued)

Variation on the theme: multiple node types

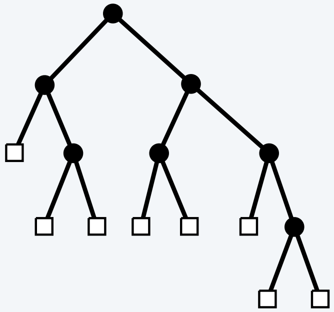
Combinatorial class $T^\bullet$, binary trees, *enumerated by internal nodes*

Atoms

| *type* | *class* | *size* | *GF* |
|---|---|---|---|
| external node | □ | 0 | 1 |
| internal node | ● | 1 | $z$ |

Construction $T = \square + T \times \bullet \times T$

OGF equation $T^\bullet(z) = 1 + zT^\bullet(z)^2$



Combinatorial class $T^\bullet$, binary trees, *enumerated by external nodes*

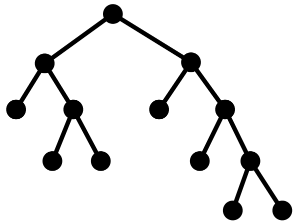
OGF equation $T^\square(z) = z + T^\square(z)^2$

More variations: unary-binary trees, ternary trees, ...

Still more variations: gambler's ruin sequences, context-free languages, triangulations, ...

# Some variations on ordered (rooted plane) trees

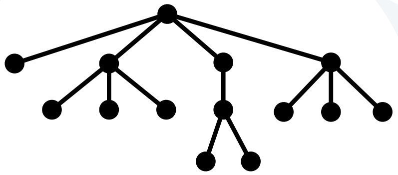
**Binary**



$$T = \bullet \times SEQ_{0,2}(T)$$

$$T(z) = z(1 + T(z)^2)$$

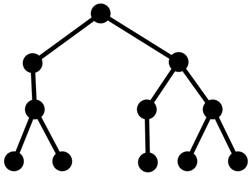
**Ordered**



$$G = \bullet \times SEQ(G)$$

$$G(z) = \frac{z}{1 - G(z)}$$

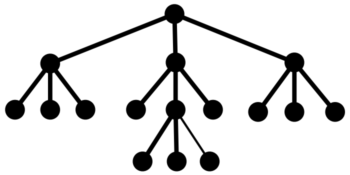
**Unary-binary**



$$M = \bullet \times SEQ_{\le 2}(M)$$

$$M(z) = z(1 + M(z) + M(z)^2)$$

**Ternary**



$$T = \bullet \times SEQ_{0,3}(T)$$

$$T(z) = z(1 + T(z)^3)$$

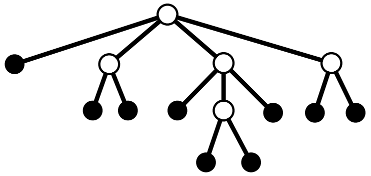
**Arbitrary restrictions**

$$T = \bullet \times SEQ_{\Omega}(T)$$

$$T^{\Omega}(z) = z\phi(T^{\Omega}(z))$$

$$\phi(u) \equiv \sum_{\omega \in \Omega} u^{\omega}$$

**Bracketings**



$$S = \bullet + SEQ_{\ge 2}(S)$$

$$S(z) = z + \frac{S(z)^2}{1 - S(z)}$$

# Variation on a theme 2: Strings

Fundamental construct

| | |
|---|---|
| Combinatorial class | $B$, the class of all binary strings |
| Construction | $B = E + (Z_0 + Z_1) \times B$ |
| OGF equation | $B(z) = 1 + 2zB(z)$ |

"a binary string is empty or a bit followed by a binary string"

Variation on the theme: *disallow sequences of P or more 0s*

| | |
|---|---|
| Combinatorial class | $B_P$, the class of all binary strings with no $0^P$ |
| Construction | $B_P = Z_{<P}(E + Z_1 B_P)$ |
| OGF equation | $B_P(z) = (1 + z + \ldots + z^P)(1 + zB_P(z))$ |

"a string with no $0^P$ is a string of 0s of length <P followed by an empty string or a 1 followed by a string with no $0^P$"

More variations: disallow any pattern (autocorrelation), REs, CFGs ...

# Some variations on strings

*M*-ary

$$B = SEQ(Z_0 + \ldots + Z_{M-1})$$

$$B(z) = \frac{1}{1 - Mz}$$

Binary

$$B = E + (Z_0 + Z_1) \times B$$
$$B = SEQ(Z_0 + Z_1)$$

$$B(z) = \frac{1}{1 - 2z}$$

Exclude $0^P$

$$B_P = Z_{<P}(E + Z_1 \times B_P)$$

$$B_P(z) = \frac{1 - z^P}{1 - 2z + z^{P+1}}$$

Regular languages

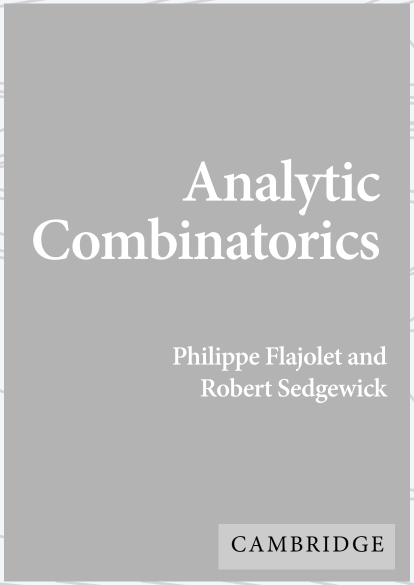
[Rational OGFs]

Context-free languages

[Algebraic OGFs]

Exclude pattern $p$

$$S_p(z) = \frac{c_p(z)}{z^P + (1 - 2z)c_p(z)}$$



Analytic
Combinatorics

Philippe Flajolet and
Robert Sedgewick

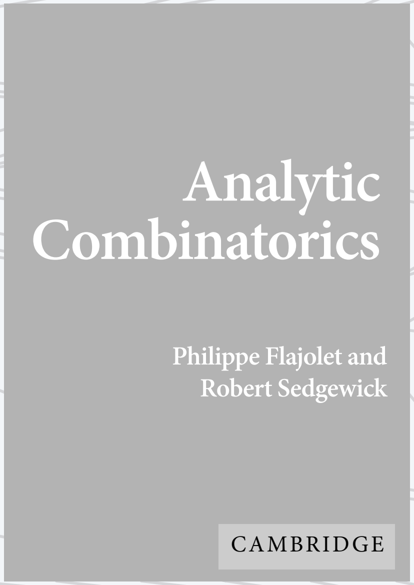
CAMBRIDGE

# Combinatorial structures and OGFs

- Symbolic method
- **Trees and strings**
- Powersets and multisets
- Compositions and partitions
- Substitution

ANALYTIC COMBINATORICS

PART TWO



Analytic
Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

http://ac.cs.princeton.edu

# 1. Combinatorial structures and OGFs

- Symbolic method
- Trees and strings
- **Powersets and multisets**
- Compositions and partitions
- Substitution

# The symbolic method (two additional constructs)

Suppose that $A$ is a class of unlabeled objects with enumerating OGF $A(z)$.

| operation | notation | semantics | OGF |
|---|---|---|---|
| *powerset* | $PSET(A)$ | finite sets of objects from $A$<br>(no repetitions) | [stay tuned] |
| *multiset* | $MSET(A)$ | finite sets of objects from $A$<br>(with repetitions) | [stay tuned] |

# Powersets

Def. The *powerset* of a class A is the class consisting of all subsets of A.

PSET {*a*}

{}
{*a*}

$P_1 = 2$

PSET {*a*, *b* }

{}
{*a*}
{*b*}
{*a*, *b*}

$P_2 = 4$

PSET {*a*, *b*, *c* }

{}
{*a*}
{*b*}
{*a*, *b*}
{*c*}
{*a*, *c*}
{*b*, *c*}
{*a*, *b*, *c*}

$P_3 = 8$

PSET {*a*, *b*, *c*, *d* }

| | |
|---|---|
| {} | {*d*} |
| {*a*} | {*a*, *d*} |
| {*b*} | {*b*, *d*} |
| {*a*, *b*} | {*a*, *b*, *d*} |
| {*c*} | {*c*, *d*} |
| {*a*, *c*} | {*a*, *c*, *d*} |
| {*b*, *c*} | {*b*, *c*, *d*} |
| {*a*, *b*, *c*} | {*a*, *b*, *c*, *d*} |
| ↑ subsets without *d* | ↑ same subsets with *d* |

$P_4 = 16$

Lemma: $\text{PSET}\ \{a_1,\ a_2,\ \ldots\ a_M\} = \text{PSET}\ \{a_1,\ a_2,\ \ldots\ a_{M-1}\} \times (\ \{\}\ +\ \{a_M\}\ )$

# Powersets

Atoms

| *notation* | *size* | *GF* |
|---|---|---|
| $a_k$ | 1 | $z$ |

**Combinatorial class** $P_M$, the powerset class for $M$ atoms

**Example** `{a, c, f, g, h}`

**OGF**

$$P_M(z) = \sum_{p \in P_M} z^{|p|} = \sum_{N \geq 0} P_{MN} z^N \longleftarrow$$

$P_{MN}$ is the # of subsets of size $N$
(no repetitions)

**Construction**

$$P_M = ( \{\} + \{a_1\} ) \times ( \{\} + \{a_2\} ) \times \ldots \times ( \{\} + \{a_M\} )$$

**OGF equation**

$$P_M(z) = (1 + z)^M$$

**Expansion**

$$P_{MN} = \binom{M}{N} \quad ✓$$

$$P_M(1) = 2^M \quad ✓$$

$\uparrow$

total # subsets
of $M$ atoms

# Multisets

**Def.** The *multiset* of a class A is the class consisting of all subsets of A *with repetitions allowed*.

MSET $\{a\}$

```
{}
{a}
{a, a}
{a, a, a}
...
```

MSET $\{a, b\}$

```
{}
{a}
{a, a}
{a, a, a}

{b}
{a, b}
{a, a, b}
{a, a, a, b}

{b, b}
{a, b, b}
{a, a, b, b}
{a, a, a, b, b}
```

MSET $\{a, b, c\}$

```
{}                {c}                   {c, c}
{a}               {a, c}                {a, c, c}
{a, a}            {a, a, c}             {a, a, c, c}
{a, a, a}         {a, a, a, c}          {a, a, a, c, c}

{b}               {b, c}                {b, c, c}
{a, b}            {a, b, c}             {a, b, c, c}
{a, a, b}         {a, a, b, c}          {a, a, b, c, c}
{a, a, a, b}      {a, a, a, b, c}       {a, a, a, b, c, c}

{b, b}            {b, b, b, c}          {b, b, c, c}
{a, b, b}         {a, b, b, b, c}       {a, b, b, c, c}
{a, a, b, b}      {a, a, b, b, b, c}    {a, a, b, b, c, c}
{a, a, a, b, b}   {a, a, a, b, b, b, c} {a, a, a, b, b, c, c}
```

**Lemma:** MSET $\{a_1, a_2, \ldots a_M\}$ = MSET $\{a_1, a_2, \ldots a_{M-1}\}$ × SEQ $\{a_M\}$

# Multisets

Atoms

| *notation* | *size* | *GF* |
|---|---|---|
| $a_k$ | 1 | $z$ |

**Combinatorial class** $S_M$, the multiset class for $M$ atoms

**Example** `{a, a, a, b, b, b, c}`

**OGF**

$$S_M(z) = \sum_{s \in S_M} z^{|s|} = \sum_{N \geq 0} S_{MN} z^N \longleftarrow$$

$S_{MN}$ is the # of subsets of size $N$ (with repetitions)

**Construction**

$$S_M = SEQ(a_1) \times SEQ(a_2) \times \ldots \times SEQ(a_M)$$

**OGF equation**

$$S_M(z) = \frac{1}{(1-z)^M}$$

**Expansion**

$$S_{MN} = \binom{N+M-1}{M-1} \checkmark$$

# The symbolic method (two additional constructs)

Suppose that $A$ is a class of unlabeled objects with enumerating OGF $A(z)$.

| operation | notation | semantics | OGF |
|---|---|---|---|
| *powerset* | $PSET(A)$ | finite sets of objects from $A$ (no repetitions) | $\prod_{n\geq 1}(1+z^n)^{A_n} = \exp\Big(-\sum_{k\geq 1}\frac{(-1)^k A(z^k)}{k}\Big)$ |
| *multiset* | $MSET(A)$ | finite sets of objects from $A$ (with repetitions) | $\prod_{n\geq 1}\frac{1}{(1-z^n)^{A_n}} = \exp\Big(\sum_{k\geq 1}\frac{A(z^k)}{k}\Big)$ |

# Proof of correspondences for powersets

$PSET(A)$

| construction | OGF |
|---|---|
| $PSET(\{a, b\}) = (\{\} + \{a\}) \times (\{\} + \{b\})$ | $(1 + z^{\lvert a \rvert})(1 + z^{\lvert b \rvert})$ |
| $PSET(A) \equiv \prod_{a \in A} (\{\} + \{a\})$ | $\prod_{a \in \mathcal{A}} (1 + z^{\lvert a \rvert}) = \prod_{N \geq 0} (1 + z^N)^{A_N}$ |

exp-log version

$$\begin{aligned}\prod_{N\geq 0}(1+z^N)^{A_N} &= \exp\Big(\sum_{N\geq 0} A_N \ln(1+z^N)\Big)\\ &= \exp\Big(-\sum_{N\geq 0} A_N \sum_{k\geq 1}(-1)^k \frac{z^{Nk}}{k}\Big)\\ &= \exp\Big(-\sum_{k\geq 1}(-1)^k \frac{A(z^k)}{k}\Big)\\ &= \exp\Big(A(z) - \frac{A(z^2)}{2} + \frac{A(z^3)}{3} - \ldots\Big)\end{aligned}$$

# Proof of correspondences for multisets

$MSET(A)$

| construction | OGF |
|---|---|
| $MSET(\{a,b\}) = SEQ(\{a\}) \times SEQ(\{b\})$ | $\dfrac{1}{(1-z^{\lvert a \rvert})(1-z^{\lvert b \rvert})}$ |
| $MSET(A) \equiv \prod_{a \in A} SEQ(\{a\})$ | $\prod_{a \in A} \dfrac{1}{(1-z^{\lvert a \rvert})} = \prod_{N \geq 0} \dfrac{1}{(1-z^{N})^{A_N}}$ |

exp-log version

$$\prod_{N \geq 0} \frac{1}{(1-z^N)^{A_N}} = \exp\Big(\sum_{N \geq 0} A_N \ln \frac{1}{1-z^N}\Big)$$

$$= \exp\Big(\sum_{N \geq 0} A_N \sum_{k \geq 1} \frac{z^{Nk}}{k}\Big)$$

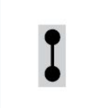
$$= \exp\Big(\sum_{k \geq 1} \frac{A(z^k)}{k}\Big)$$

$$= \exp\Big(A(z) + \frac{A(z^2)}{2} + \frac{A(z^3)}{3} + \ldots\Big)$$

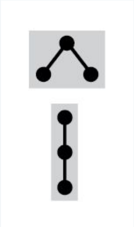
# Multiset application example

Q. How many unordered trees with $N$ nodes?

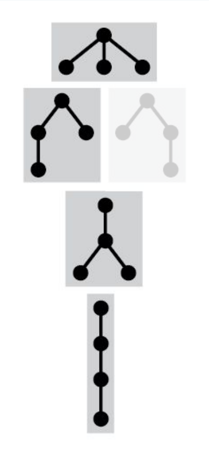
$H_1 = 1$



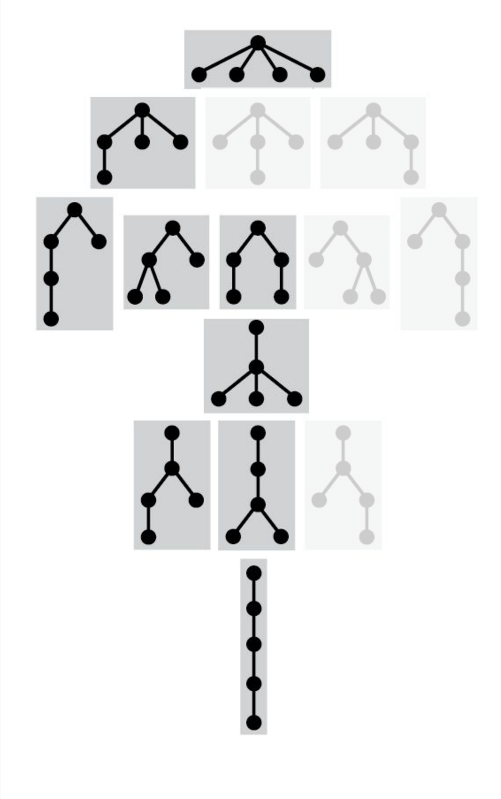
$H_2 = 1$



$H_3 = 2$



$H_4 = 4$



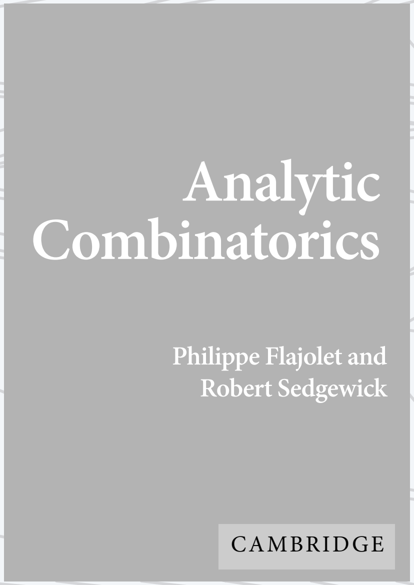
$H_5=9$

| | |
|---|---|
| Combinatorial class | $H$, the class of all unordered trees |
| Construction | $H = \bullet \times MSET(H) \longleftarrow$ |
| OGF equation | $H(z) = z\exp\bigl(H(z) + H(z^2)/2 + H(z^3)/3 + \ldots\bigr)$ |

"a tree is a node and
a multiset of trees"

ANALYTIC COMBINATORICS

PART TWO


Analytic
Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

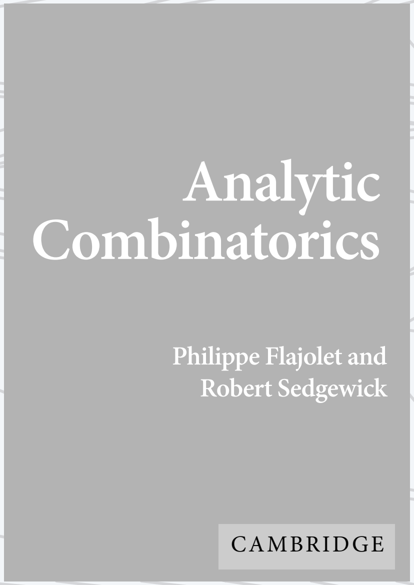
http://ac.cs.princeton.edu

# 1. Combinatorial structures and OGFs

- Symbolic method
- Trees and strings
- **Powersets and multisets**
- Compositions and partitions
- Substitution

ANALYTIC COMBINATORICS

PART TWO



Analytic
Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

http://ac.cs.princeton.edu

# 1. Combinatorial structures and OGFs

- Symbolic method
- Trees and strings
- Powersets and multisets
- **Compositions and partitions**
- Substitution

`II.1d.OGFs.Compositions`

## Compositions

Q. How many ways to express $N$ as a sum of positive integers?

1

$I_1 = 1$

1 + 1
2

$I_2 = 2$

1 + 1 + 1
1 + 2
2 + 1
3

$I_3 = 4$

1 + 1 + 1 + 1
1 + 1 + 2
1 + 2 + 1
1 + 3
2 + 1 + 1
2 + 2
3 + 1
4

$I_4 = 8$

1 + 1 + 1 + 1 + 1
1 + 1 + 1 + 2
1 + 1 + 2 + 1
1 + 1 + 3
1 + 2 + 1 + 1
1 + 2 + 2
1 + 3 + 1
1 + 4
2 + 1 + 1 + 1
2 + 1 + 2
2 + 2 + 1
2 + 3
3 + 1 + 1
3 + 2
4 + 1
5

$I_5 = 16$

A. $I_N = 2^{N-1}$

# Integers as a combinatorial class

| Atom | *notation* | *size* | *GF* |
|---|---|---|---|
| | ● | 1 | $z$ |

Combinatorial class $I$, the class of all positive integers

Example ● ● ● ● ● ● ● ← unary notation for 7

OGF
$$I(z) = \sum_{i \in I} z^{|i|} = \sum_{N \geq 0} I_N z^N$$

Construction
$$I = SEQ_{>0}(\bullet)$$

OGF equation
$$I(z) = \frac{z}{1 - z}$$

Expansion
$$I_N = 1 \text{ for } N > 0$$
✓

# Compositions

**Combinatorial class** $C$, the class of all compositions

**Example** ● | ●●● | ● | ●●●●● | ●● = ●●●●●●●●●●●●●●●●

unary notation for
1+3+1+5+2=12

**OGF**

$$C(z) = \sum_{c \in C} z^{|c|} = \sum_{N \geq 0} C_N z^N$$

**Construction** $C = SEQ(I)$

"a composition is a sequence of positive integers"

**OGF equation**

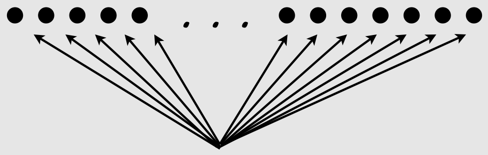
$$C(z) = \frac{1}{1 - I(z)}$$

$$= \frac{1}{1 - \frac{z}{1-z}} = \frac{1 - z}{1 - 2z}$$

**Expansion**

$$C_N = 2^N - 2^{N-1} = \boxed{2^{N-1}} \text{ for } N > 0$$

✓



$N$–1 spaces between dots
each could have a bar or not
$=2^{N-1}$ possibilities ✓

# Partitions

Q. How many ways to express *N* as a sum of *unordered* positive integers?

1

$P_1 = 1$

1 + 1
2

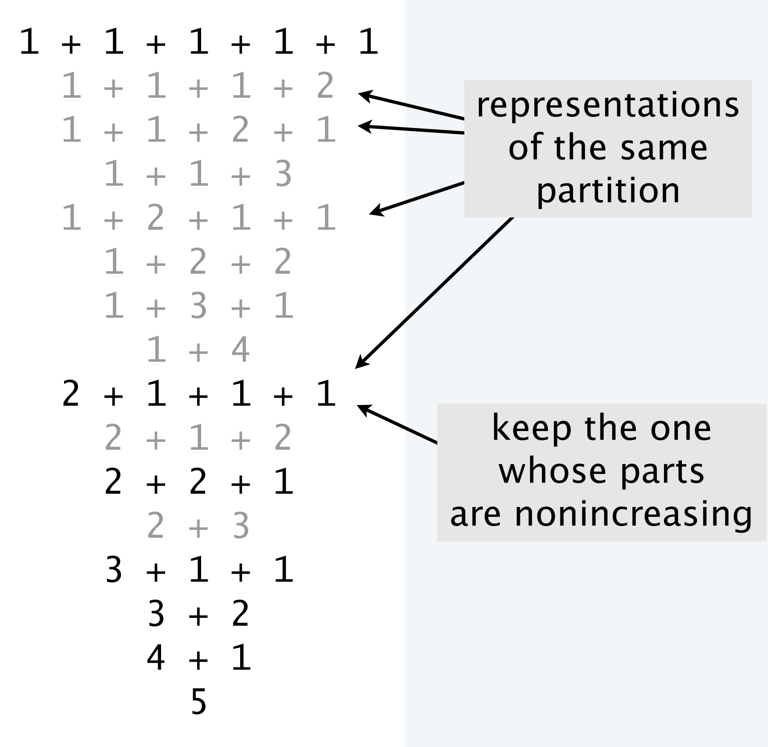
$P_2 = 2$

1 + 1 + 1
1 + 2
2 + 1
3

$P_3 = 3$

1 + 1 + 1 + 1
1 + 1 + 2
1 + 2 + 1
1 + 3
2 + 1 + 1
2 + 2
3 + 1
4

$P_4 = 5$



1 + 1 + 1 + 1 + 1
1 + 1 + 1 + 2
1 + 1 + 2 + 1
1 + 1 + 3
1 + 2 + 1 + 1
1 + 2 + 2
1 + 3 + 1
1 + 4
2 + 1 + 1 + 1
2 + 1 + 2
2 + 2 + 1
2 + 3
3 + 1 + 1
3 + 2
4 + 1
5

representations of the same partition

keep the one whose parts are nonincreasing

$P_5 = 7$

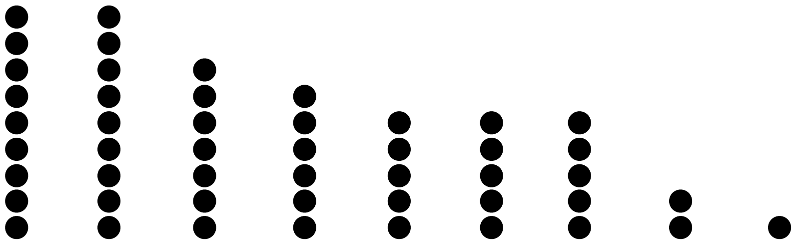
A. *Not so obvious* !

# Ferrers diagrams

Def. A *Ferrers diagram* is a 2D representation of a partition: one column of dots per part.

partition $8 + 8 + 6 + 5 + 4 + 4 + 4 + 2 + 1 = 42$

Ferrers diagram



← *42 dots*

Q. How many Ferrers diagrams with *N* dots?

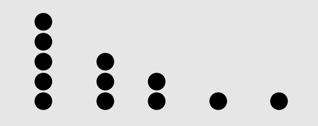
A. *Not so obvious* [need symbolic method plus saddle-point asymptotics—stay tuned]

Applications. AofA, representation theory, Lie algebras, particle physics, . . .

# Partitions

Combinatorial class $P$, the class of all partitions

Example



Ferrers diagram for
5+3+2+1+1=12

OGF

$$P(z) = \sum_{p \in P} z^{|p|} = \sum_{N \geq 0} P_N z^N$$

Construction

$$P \ = MSET\,(I\,)$$

"a partition is a *multiset* of positive integers"

$$MSET(A) \equiv \prod_{a \in A} SEQ(\{a\})$$

$$\prod_{a \in A} \frac{1}{(1 - z^{|a|})} = \prod_{N \geq 0} \frac{1}{(1 - z^N)^{A_N}}$$

OGF equation

$$P(z) = \frac{1}{(1-z)(1-z^2)(1-z^3)\ldots}$$

Expansion

$$P_N \sim \boxed{\frac{e^{\pi\sqrt{2N/3}}}{4N\sqrt{3}}}$$

Classic result of Hardy and Ramanujan
(need saddle-point asymptotics)

# Some variations on compositions and partitions

Restricted compositions

$T = \{$ any subset of $I\}$

$C^T = SEQ(SEQ_T(Z))$

$$C^T(z) = \frac{1}{1 - T(z)}$$

Compositions

$C = SEQ(I)$

$$C(z) = \frac{1-z}{1-2z}$$

Compositions into $M$ parts

$C_M = SEQ_M(I)$

$$C_M(z) = \frac{z^M}{1 - z^M}$$

Partitions into distinct parts

$Q = PSET(I)$

$$Q(z) = (1+z)(1+z^2)(1+z^3)\ldots$$

Partitions

$P = MSET(I)$

$$P_N \sim \frac{e^{\pi\sqrt{2N/3}}}{4N\sqrt{3}}$$

Restricted partitions

$T = \{$ any subset of $I\}$

$P^T = MSET(SEQ_T(Z))$

$$P^T(z) = \prod_{N \in T} \frac{1}{1 - z^N}$$

## In-class exercises

Q. OGF for compositions into parts less than or equal to *R* ?

Q. How many partitions into parts that are powers of 2?

A. **1**

$$\prod_{j\geq 0}(1+z^{2^j}) = (1+z)(1+z^2)(1+z^4)(1+z^8)\ldots$$

$$= (1+z+z^2+z^3)(1+z^4)(1+z^8)\ldots$$

$$= (1+z+z^2+z^3+z^4+z^5+z^6+z^7)(1+z^8)\ldots$$

$$= 1+z+z^2+z^3+z^4+z^5+z^6+z^7+z^8+z^9+z^{10}+\ldots$$

Q. How many ways to represent an integer as a sum of powers of 2?

A. **1**

$$\prod_{j\geq 0}(1+z^{2^j}) = \frac{1}{1-z}$$

# How many ways to change a dollar?

Q. How many ways to change a dollar with quarters ?

A. **1**

$$[z^{100}]\frac{1}{1-z^{25}} = [z^{100}](1+z^{25}+z^{50}+\ldots) = 1$$

Q. How many ways to change a dollar with quarters *and dimes*?

A. **3**

$$[z^{100}]\frac{1}{1-z^{25}}\frac{1}{1-z^{10}} = [z^{100}](1+z^{25}+z^{50}+\ldots)(1+z^{10}+z^{20}+\ldots)$$

$$= [z^{100}](1+z^{50}+z^{100})(1+z^{50}+z^{100})$$

## How many ways to change a dollar?

Q. How many ways to change a dollar with quarters ?

A. **1** $$[z^{100}]\frac{1}{1-z^{25}} = [z^{100}](1+z^{25}+z^{50}+\ldots) = 1$$

Q. How many ways to change a dollar with quarters *and dimes* ?

A. **3** $$[z^{100}]\frac{1}{1-z^{25}}\frac{1}{1-z^{10}} = [z^{100}](1+z^{25}+z^{50}+\ldots)(1+z^{10}+z^{20}+\ldots)$$

Q. How many ways to change a dollar with quarters, dimes *and nickels* ?

A. ? $$[z^{100}]\frac{1}{1-z^{25}}\frac{1}{1-z^{10}}\frac{1}{1-z^{5}}$$ ← need a computer?

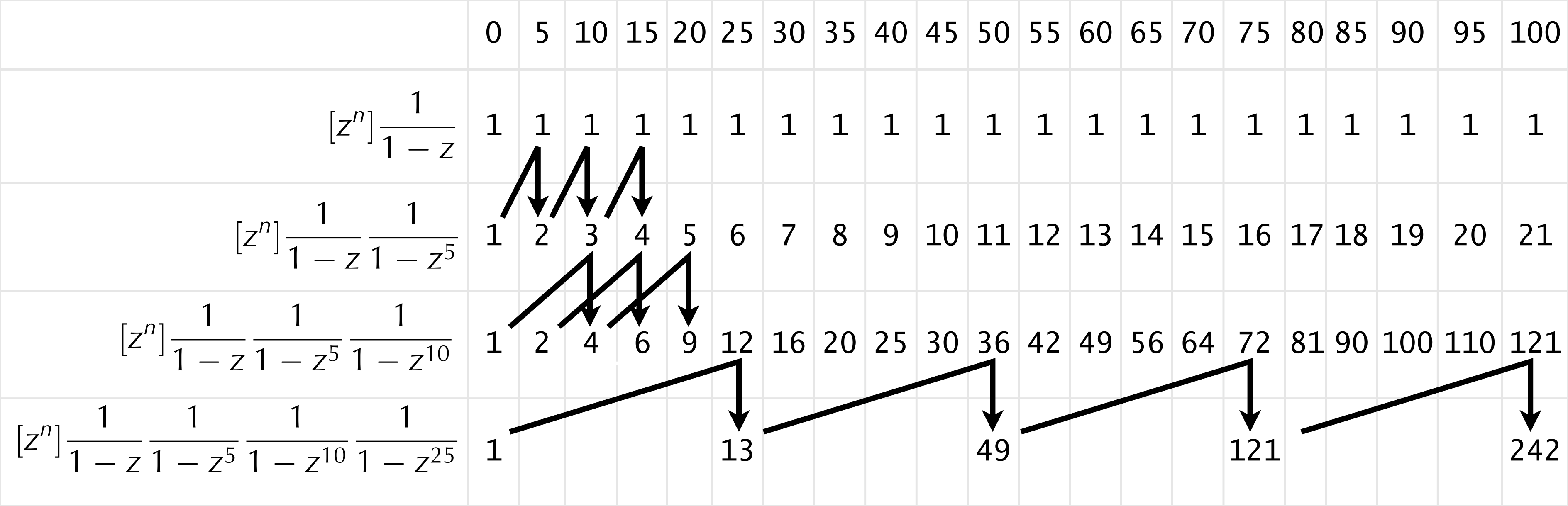
Q. How many ways to change a dollar with quarters, dimes, nickels *and pennies* ?

A. ? $$[z^{100}]\frac{1}{1-z^{25}}\frac{1}{1-z^{10}}\frac{1}{1-z^{5}}\frac{1}{1-z}$$ ← need a computer?

# How many ways to change a dollar?

Key insight (Pólya): If $b(z) = a(z)\dfrac{1}{1-z^M}$ then $b(z)(1-z^M) = a(z)$ and therefore $b_n = b_{n-M} + a_n$

Gives an easy way to compute small values by hand.



| | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $[z^n]\dfrac{1}{1-z}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $[z^n]\dfrac{1}{1-z}\dfrac{1}{1-z^5}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| $[z^n]\dfrac{1}{1-z}\dfrac{1}{1-z^5}\dfrac{1}{1-z^{10}}$ | 1 | 2 | 4 | 6 | 9 | 12 | 16 | 20 | 25 | 30 | 36 | 42 | 49 | 56 | 64 | 72 | 81 | 90 | 100 | 110 | 121 |
| $[z^n]\dfrac{1}{1-z}\dfrac{1}{1-z^5}\dfrac{1}{1-z^{10}}\dfrac{1}{1-z^{25}}$ | 1 | | | | | 13 | | | | | 49 | | | | | 121 | | | | | 242 |

## In-class exercise

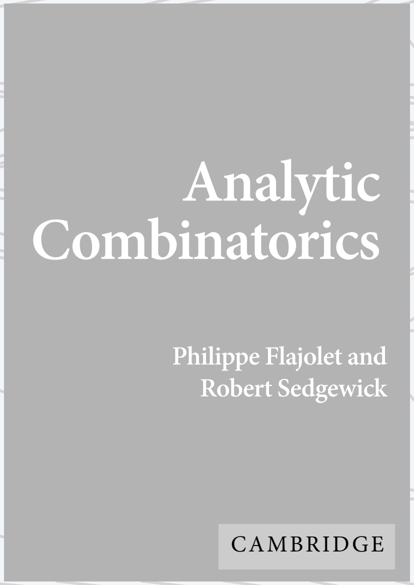
For whatever reason, the government switches to 20-cent pieces instead of dimes.

How many ways to change a dollar?

| | 0 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 95 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $[z^n]\frac{1}{1-z}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $[z^n]\frac{1}{1-z}\frac{1}{1-z^5}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| $[z^n]\frac{1}{1-z}\frac{1}{1-z^5}\frac{1}{1-z^{10}}$ | 1 | 2 | 3 | 4 | 6 | 8 | 10 | 12 | 15 | 18 | 21 | 24 | 28 | 32 | 36 | 40 | 45 | 50 | 55 | 60 | 66 |
| $[z^n]\frac{1}{1-z}\frac{1}{1-z^5}\frac{1}{1-z^{10}}\frac{1}{1-z^{25}}$ | 1 | | | | | 9 | | | | | 30 | | | | | 70 | | | | | 136 |

ANALYTIC COMBINATORICS

PART TWO



http://ac.cs.princeton.edu

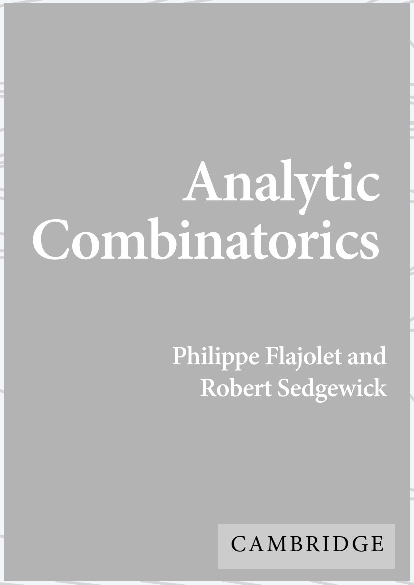
# 1. Combinatorial structures and OGFs

- Symbolic method
- Trees and strings
- Powersets and multisets
- **Compositions and partitions**
- Substitution

`II.1d.OGFs.Compositions`

ANALYTIC COMBINATORICS

PART TWO



Analytic Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

http://ac.cs.princeton.edu

# 1. Combinatorial structures and OGFs

- Symbolic method
- Trees and strings
- Powersets and multisets
- Compositions and partitions
- **Substitution**

`II.1e.OGFs.Substitution`

# The symbolic method for unlabeled objects (summary)

| operation | notation | semantics | OGF |
|---|---|---|---|
| *disjoint union* | $A + B$ | disjoint copies of objects from $A$ and $B$ | $A(z) + B(z)$ |
| *Cartesian product* | $A \times B$ | ordered pairs of copies of objects, one from $A$ and one from $B$ | $A(z)B(z)$ |
| *sequence* | $SEQ(A)$ | sequences of objects from $A$ | $\dfrac{1}{1 - A(z)}$ |
| *powerset* | $PSET(A)$ | finite sets of objects from $A$ (no repetitions) | $\prod_{n\geq 1}(1+z^n)^{A_n} = \exp\Big(-\sum_{k\geq 1}\frac{(-1)^k A(z^k)}{k}\Big)$ |
| *multiset* | $MSET(A)$ | finite sets of objects from $A$ (with repetitions) | $\prod_{n\geq 1}\frac{1}{(1-z^n)^{A_n}} = \exp\Big(\sum_{k\geq 1}\frac{A(z^k)}{k}\Big)$ |

Additional constructs are available (and still being invented)—one example to follow

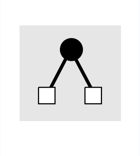
# Another construct for the symbolic method: substitution

Suppose that $A$ and $B$ are classes of unlabeled objects with enumerating OGFs $A(z)$ and $B(z)$.

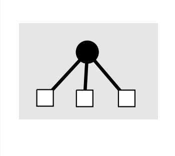
| operation | notation | semantics | OGF |
|---|---|---|---|
| *substitution* | $A \circ [\, B \,]$ | replace each object in an instance of A with an object from B | $A(B(z))$ |

# Substitution application example

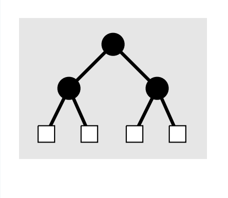
Q. How many 2-3 trees with $N$ nodes?



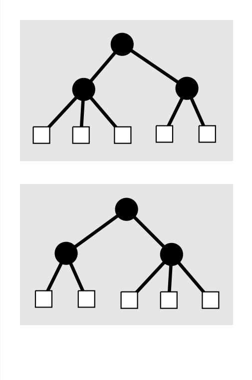
$W_2 = 1$



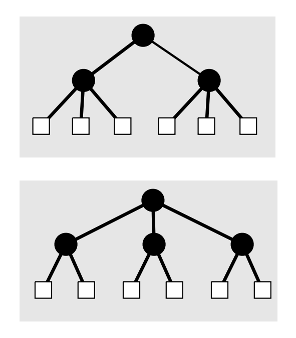
$W_3 = 1$



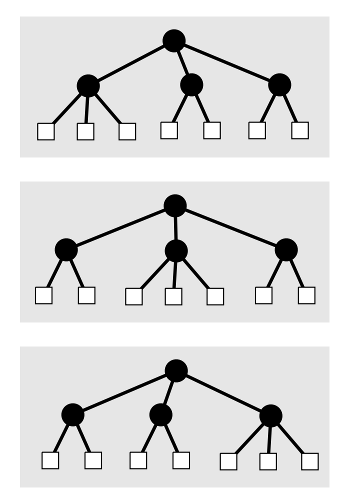
$W_4 = 1$



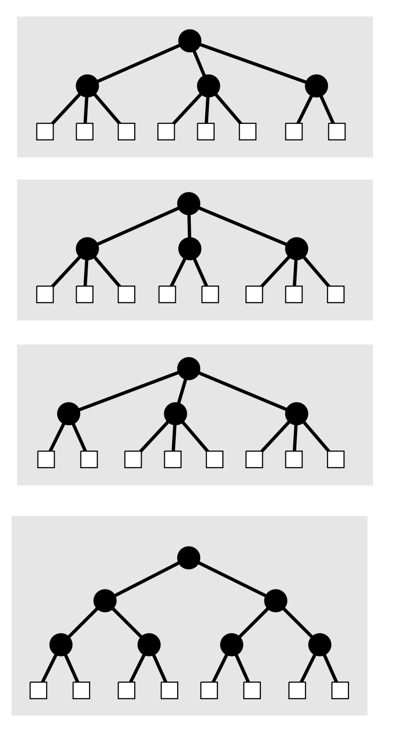
$W_5 = 2$



$W_6 = 2$



$W_7 = 3$



$W_8 = 4$

# Substitution application example

Q. How many 2-3 trees with $N$ nodes?

Combinatorial class $W$, the class of all 2-3 trees

Construction $W = Z + W \circ [\, (Z \times Z) + (Z \times Z \times Z)\, ]$ ⟵ "a 2-3 tree is a 2-3 tree with each external node replaced by a 2-node or a 3-node"

OGF equation $W(z) = z + W(z^2 + z^3)$

$$W(z) = z^2 + z^3 + z^4 + 2z^5 + 2z^6 + 3z^7 + 4z^8 + \ldots$$

$$W(z^2 + z^3) = z^2 + z^3 + (z^2 + z^3)^2 + (z^2 + z^3)^3 + (z^2 + z^3)^4 + \ldots$$

$$= z^2 + z^3 + (z^4 + 2z^5 + z^6) + (z^6 + 3z^7 + 3z^8 + z^9) + z^8 + \ldots \checkmark$$

Coefficient asymptotics are complicated (oscillations in the leading term).

See A. Odlyzko, *Periodic oscillations of coefficients of power series that satisfy functional equations*, Adv. in Mathematics (1982).

# Two French mathematicians on the utility of GFs

*"A property... is understood better, when one constructs a bijection... than when one calculates the coefficients of a polynomial whose variables have no particular meaning. The method of generating functions, which has had devastating effects for a century, has fallen into obsolescence, for this reason.*

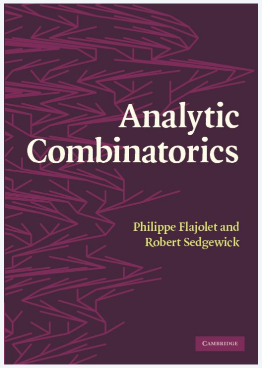
*— Claude Bergé, 1968*

*" Generating functions are the central objects of the theory, rather than a mere artifact to solve recurrences, as it is still often believed. "*

*— Philippe Flajolet, 2007*

# Analytic combinatorics overview

To analyze properties of a large combinatorial structure:



1. Use the symbolic method
   - Define a *class* of combinatorial objects.
   - Define a notion of *size* (and associated generating function)
   - Use standard operations to develop a *specification* of the structure.

Result: A direct derivation of a GF equation (implicit or explicit).

*Important note: GF equations vary widely in nature*

$$P(z) = \frac{1}{(1-z)(1-z^2)(1-z^3)\ldots}$$

$$C(z) = \frac{1}{1-I(z)}$$

$$T(z) = z + T(z^2 + z^3)$$

$$B(z) = \frac{1}{1-2z}$$

$$S_M(z) = \frac{1}{(1-z)^M}$$

$$H(z) = z\exp\bigl(H(z) + H(z^2)/2 + H(z^3/3 + \ldots\bigr)$$

$$B_P(z) = \frac{1-z^P}{1-2z+z^{P+1}}$$

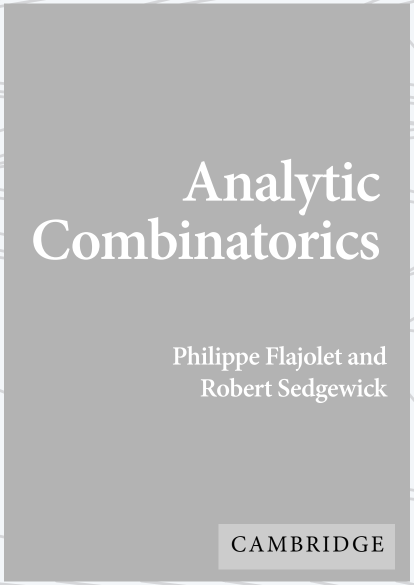
$$G(z)^2 - G(z) + z = 0$$

$$Q(z) = (1+z)(1+z^2)(1+z^3)\ldots$$

2. Use complex asymptotics to estimate growth of coefficients (stay tuned).

ANALYTIC COMBINATORICS

PART TWO


Analytic
Combinatorics

Philippe Flajolet and
Robert Sedgewick

CAMBRIDGE

http://ac.cs.princeton.edu

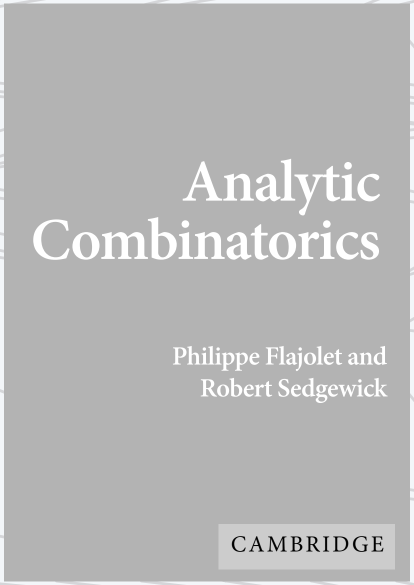
# 1. Combinatorial structures and OGFs

- Symbolic method
- Trees and strings
- Powersets and multisets
- Compositions and partitions
- **Substitution**

II.1e.OGFs.Substitution

ANALYTIC COMBINATORICS

PART TWO



Analytic Combinatorics

Philippe Flajolet and Robert Sedgewick

CAMBRIDGE

http://ac.cs.princeton.edu

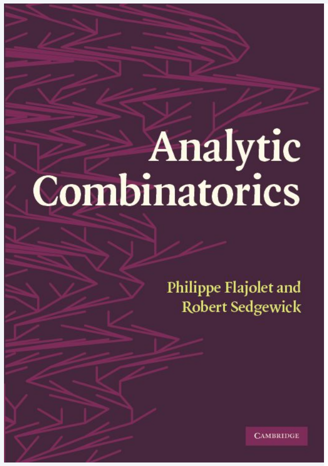
# 1. Combinatorial structures and OGFs

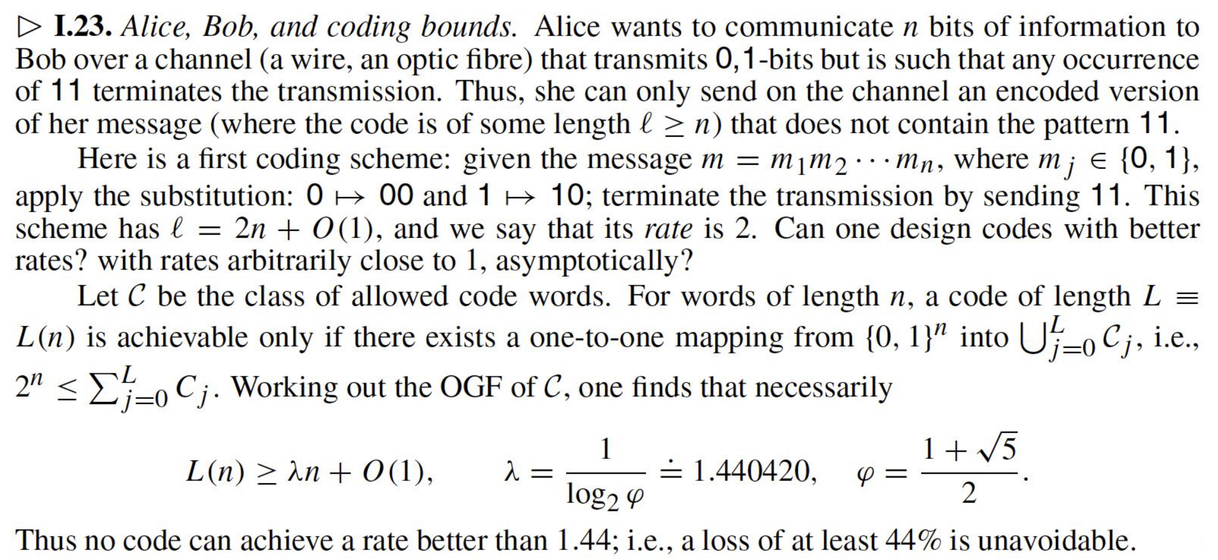
- Symbolic method
- Trees and strings
- Powersets and multisets
- Compositions and partitions
- Substitution
- **Exercises**

II.1f.OGFs.Exercises

# Note 1.23

Alice, Bob, and coding bounds





▷ **I.23.** *Alice, Bob, and coding bounds.* Alice wants to communicate $n$ bits of information to Bob over a channel (a wire, an optic fibre) that transmits 0,1-bits but is such that any occurrence of 11 terminates the transmission. Thus, she can only send on the channel an encoded version of her message (where the code is of some length $\ell \geq n$) that does not contain the pattern 11.

Here is a first coding scheme: given the message $m = m_1 m_2 \cdots m_n$, where $m_j \in \{0, 1\}$, apply the substitution: $0 \mapsto 00$ and $1 \mapsto 10$; terminate the transmission by sending 11. This scheme has $\ell = 2n + O(1)$, and we say that its *rate* is 2. Can one design codes with better rates? with rates arbitrarily close to 1, asymptotically?

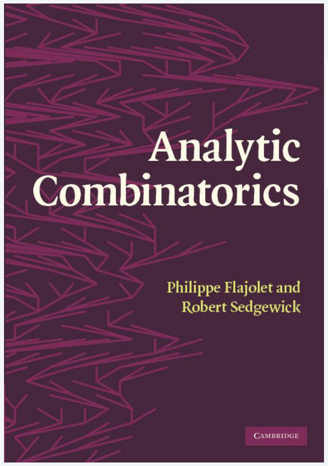
Let $\mathcal{C}$ be the class of allowed code words. For words of length $n$, a code of length $L \equiv L(n)$ is achievable only if there exists a one-to-one mapping from $\{0,1\}^n$ into $\bigcup_{j=0}^{L} \mathcal{C}_j$, i.e., $2^n \leq \sum_{j=0}^{L} C_j$. Working out the OGF of $\mathcal{C}$, one finds that necessarily

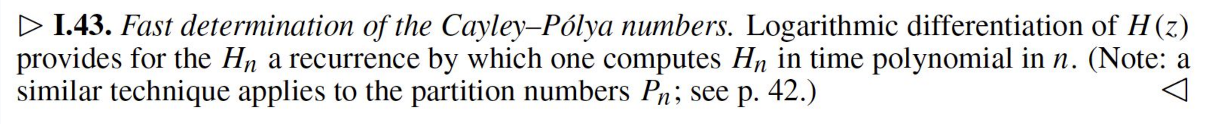
$$L(n) \geq \lambda n + O(1), \qquad \lambda = \frac{1}{\log_2 \varphi} \doteq 1.440420, \quad \varphi = \frac{1+\sqrt{5}}{2}.$$

Thus no code can achieve a rate better than 1.44; i.e., a loss of at least 44% is unavoidable.

# Note 1.43

Calculating Cayley numbers and partition numbers

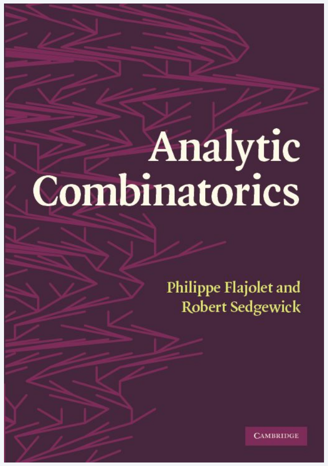




▷ **I.43.** *Fast determination of the Cayley–Pólya numbers.* Logarithmic differentiation of $H(z)$ provides for the $H_n$ a recurrence by which one computes $H_n$ in time polynomial in $n$. (Note: a similar technique applies to the partition numbers $P_n$; see p. 42.) ◁

# Assignments

1. Read pages 15-94 in text.



2. Write up solutions to Notes 1.23 and 1.43.

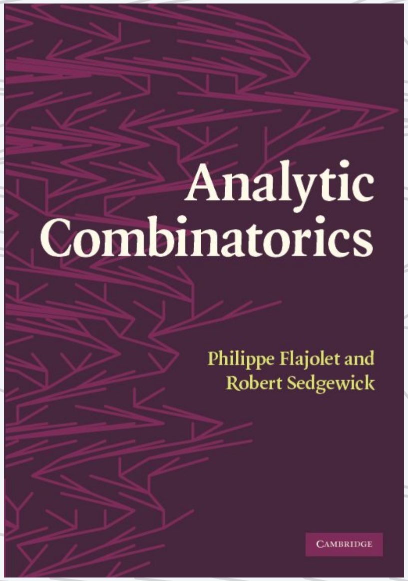
3. Programming exercises.

**Program I.1.** Determine the choice of four coins that maximizes the number of ways to change a dollar.

**Program I.2.** Write programs that estimate the rate of growth of the Cayley numbers and the partition numbers ($H_n/H_{n-1}$ and $P_n/P_{n-1}$). See Note I.43.

ANALYTIC COMBINATORICS

PART TWO



http://ac.cs.princeton.edu

# 1. Combinatorial structures and OGFs