Combinatorics of Go

John Tromp

joint work with Gunnar Farnebäck

full paper at http://tromp.github.io/go/legal.html

Overview

- History of Go
- Computational Complexity
- Sample Game and Rules Summary
- Number of Legal Positions
- Number of Games
- Open Problems

Quotes

- Go uses the most elemental materials and concepts – line and circle, wood and stone, black and white – combining them with simple rules to generate subtle strategies and complex tactics that stagger the imagination.
  Iwamoto Kaoru, 9-dan professional Go player and former Honinbo title holder
- While the Baroque rules of chess could only have been created by humans, the rules of go are so elegant, organic, and rigorously logical that if intelligent life forms exist elsewhere in the universe, they almost certainly play go.
  Edward Lasker, Chess International Master

History of Go

- Originating in China between 2000BC and 600BC (wei'qi)
- spread to Japan in 7th century
- gained popularity at imperial court in 8th century
- played in general public in 13th century
- founding of Go academy in early 17th century
- Leibniz (1646-1716) published an article about go
- from 25 to 50 million go players in the Far East; known in Korea as baduk
- central in Japanese manga and anime series

Hikaru no Go

Hikaru no Go

Computational Complexity

- 1980: Lichtenstein and Sipser proved Go PSPACE-hard.
- 1983: Robson showed Go with the basic ko rule to be EXPTIME-complete.
- 2000: Tromp and Crâşmaru showed ladders to be PSPACE-complete.
- 2002: Wolfe showed endgames to be PSPACE-complete.

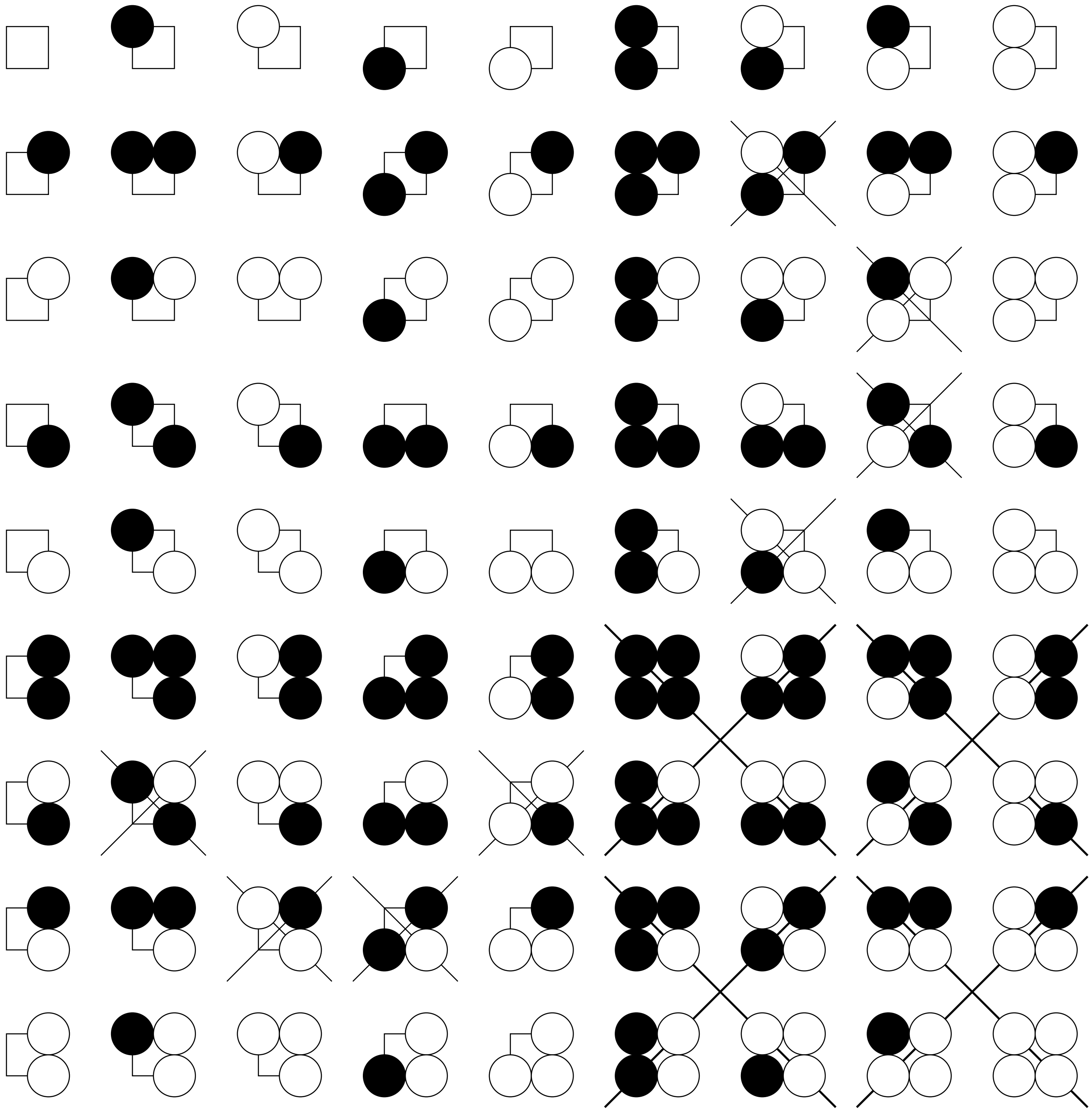
Concepts of Go

- a grid of *points*: $\{0, \ldots, m-1\} \times \{0, \ldots, n-1\}$
- 3 *colors*: $\{\text{empty}, \text{black}, \text{white}\}$
- position: a mapping from points to colors
- stone: a point colored black or white
- string: a connected component of adjacent stones of the same color
- liberty: empty point adjacent to a string

Rules of Go

- play starts on an empty board
- on his turn, a player passes
  or makes a move that doesn't repeat an earlier position
- move: placing a stone and removing libertyless strings,
  removal of opponent string taking precedence
- consecutive passes end the game
- a player's score is number of points he controls

all 2×2 positions

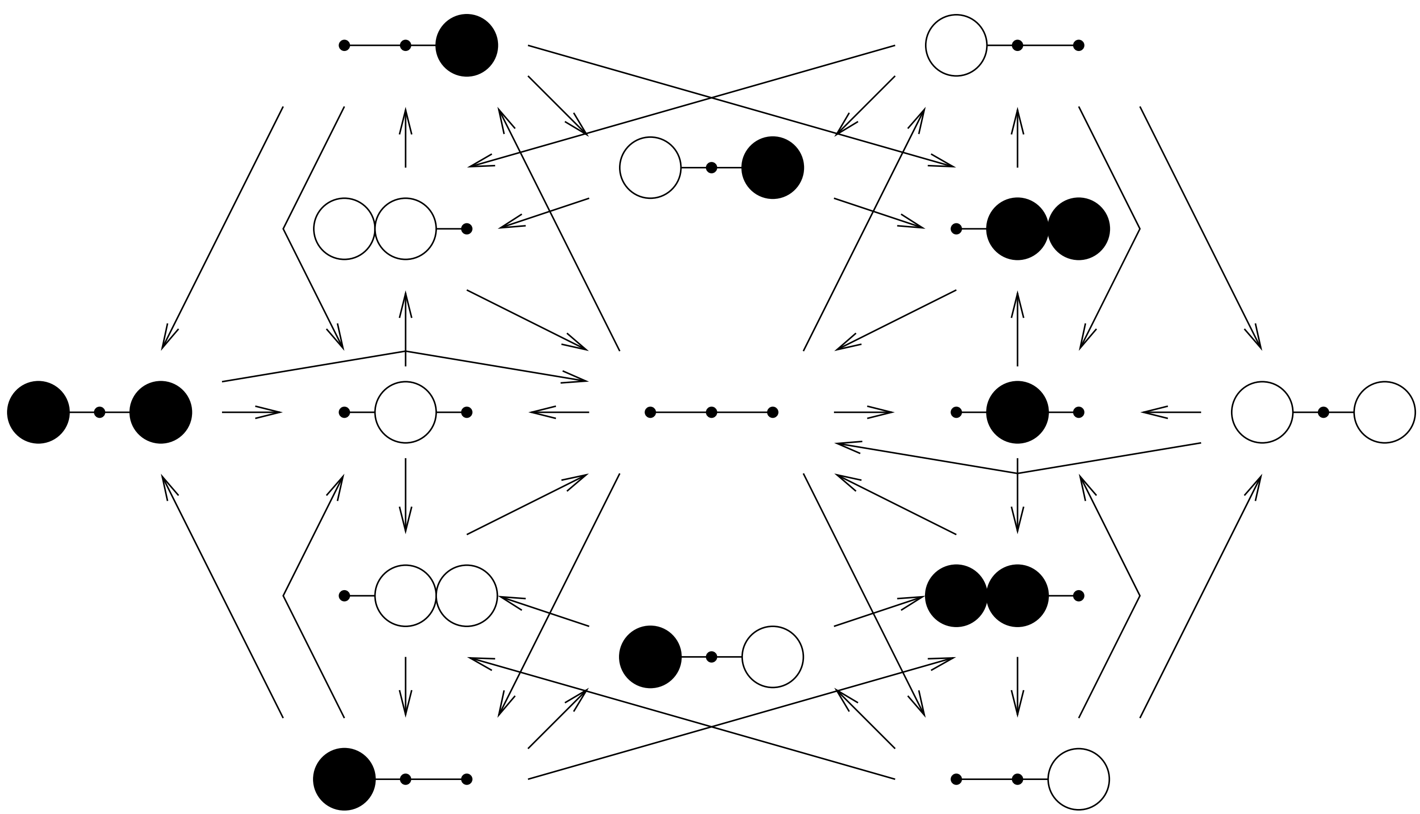


Number of legal positions: $L(m, n)$

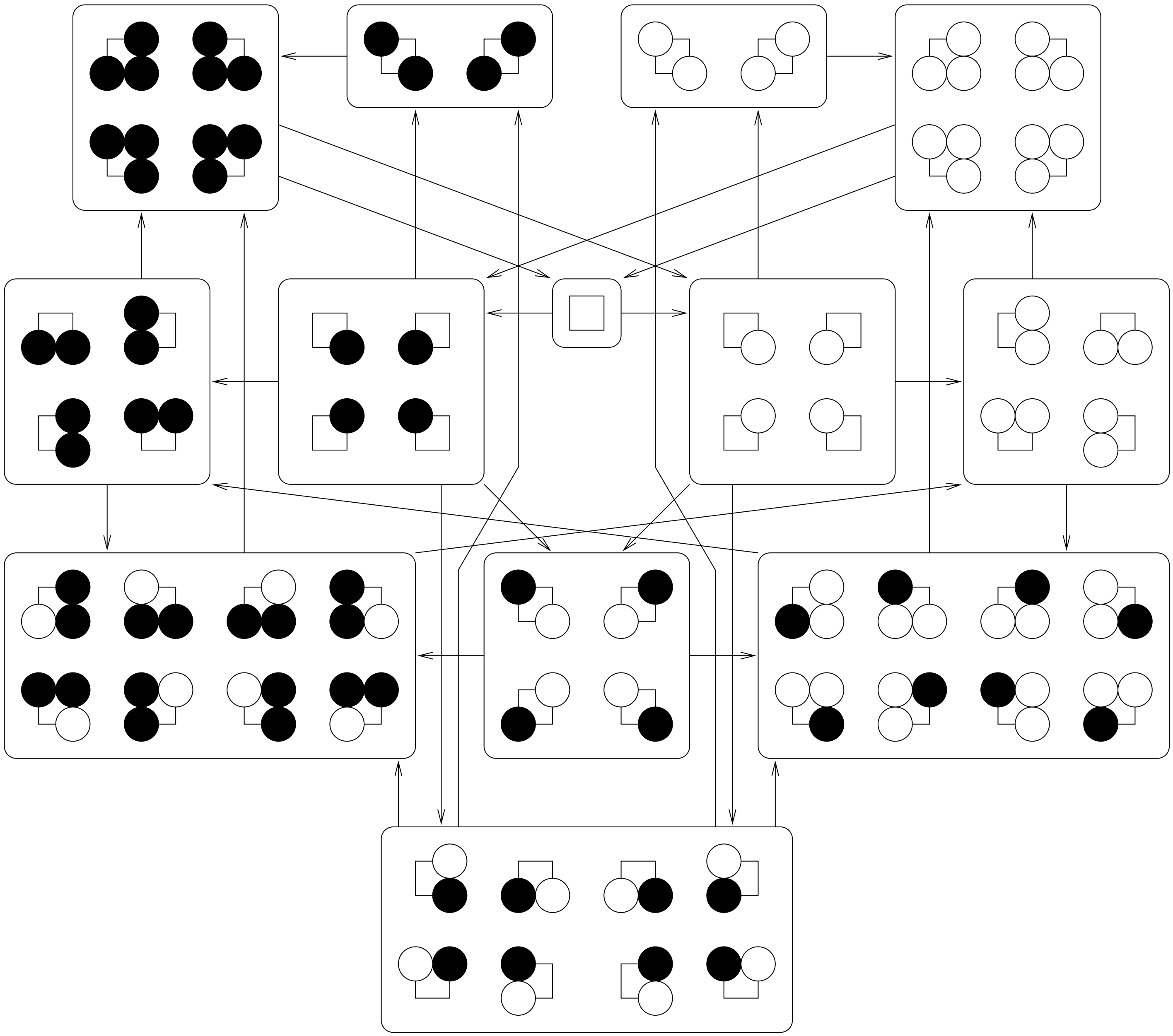
- $L(2, 2) = 3^4$ minus the number of illegal positions.
- all 2^4 positions with 4 stones are illegal.
- all 8 positions with a stone of one color bordered by two stones of the opposite color are illegal.
- all positions with 2 or fewer stones are legal.
- Hence $L(2, 2) = 81 - 16 - 8 = 57$.

Game graph: $G(m, n)$

- vertices: the legal $m \times n$ positions
- edges: moves between different positions
- $G(3, 1)$:



Game graph $G(2,2)$



Basic Lemmas

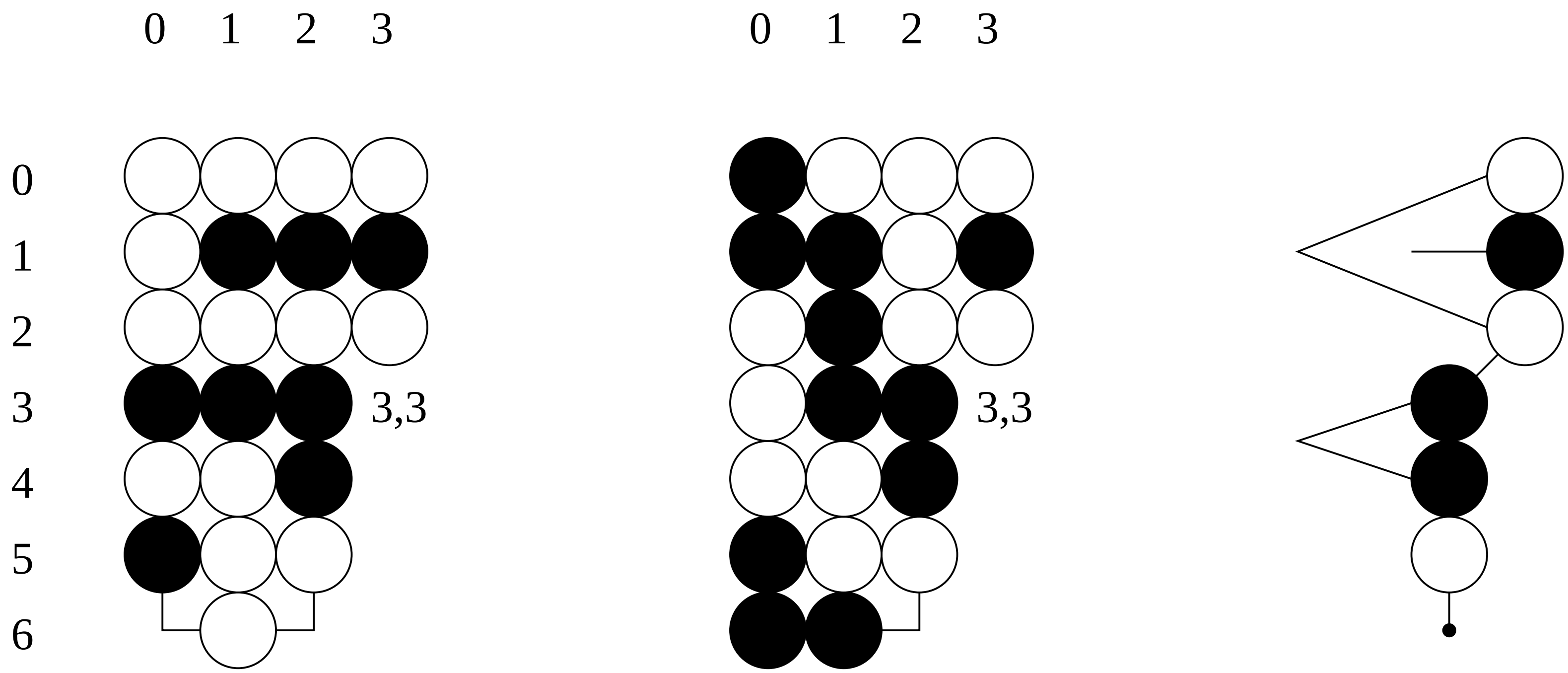
- Go games are in 1-1 correspondence with simple paths starting at the all-empty node in the game graph.
- The game graph is strongly connected.
- For rulesets forbidding suicide: The game graph, minus the empty position and suicide edges, is strongly connected.

Counting legal positions

- Brute force; test each of 3^{mn} positions for legality
- limited to counting up to 5×5.
- Establish correspondence between legal positions and paths in a certain graph of manageable size.
- Path counting can be done efficiently with Dynamic Programming.

Partial Boards

- partial go board up to column x and row y consists of all the points to the left of and above (x, y)
- partial 7×7 positions up to $(3, 3)$:



Border States

- the size $0 \le y < m$ of the partial column,
- the m colors of border points $(x,0),\ldots,(x,y-1),(x-1,y),\ldots,(x-1,m-1)$,
- for each stone on the border, whether it has liberties,
- connections among libertyless stones.

A state with partial column size y is called a y-state.

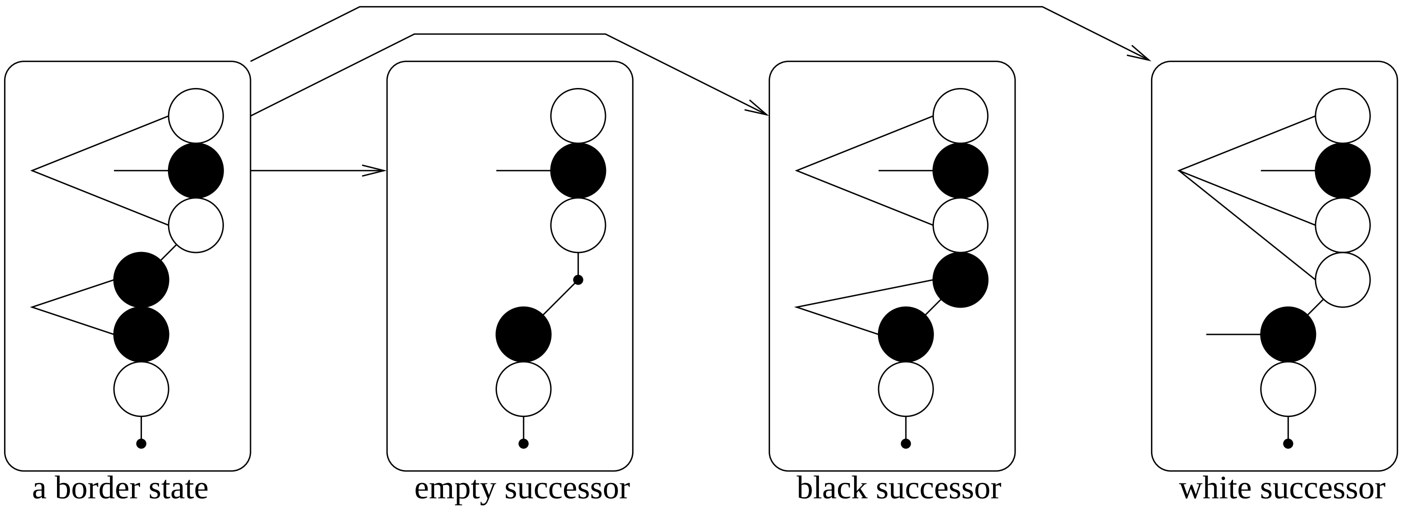
- Pseudo legal partial positions
- Constructible states

Number of State Classes

m	#valid 0-classes	#valid $(m-1)$-classes
1	3	3
2	9	13
3	32	46
5	444	642
7	6742	9808
9	109736	160286
11	1894494	2772774
13	34320647	50258461
15	645949499	945567689
17	12526125303	18320946269
19	248661924718	363324268018

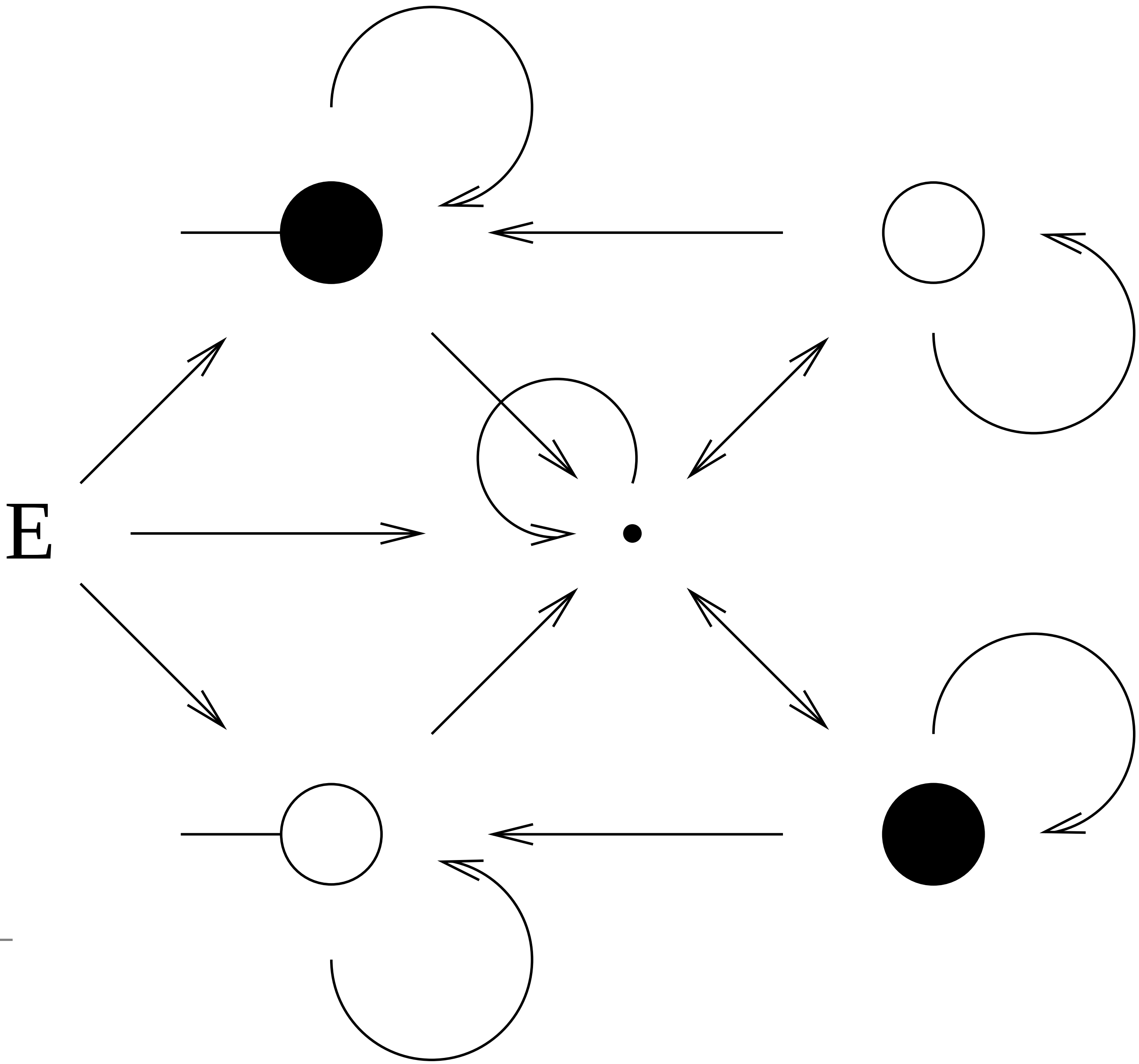
The border state graph $B(m)$

- successor states:

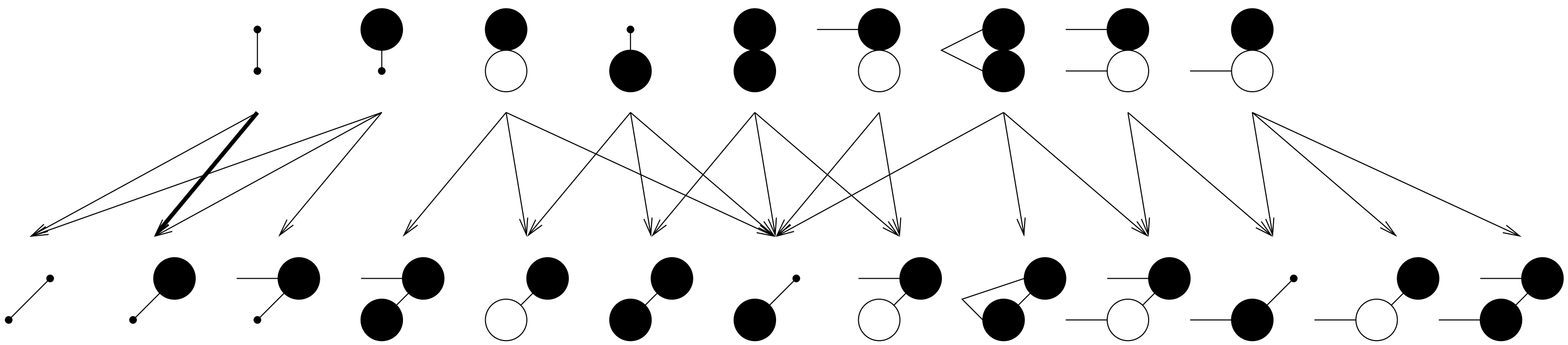


- vertices: the constructible states of height m
- edges: from each y-state to its 2 or 3 successor $((y+1) \bmod m)$-states.

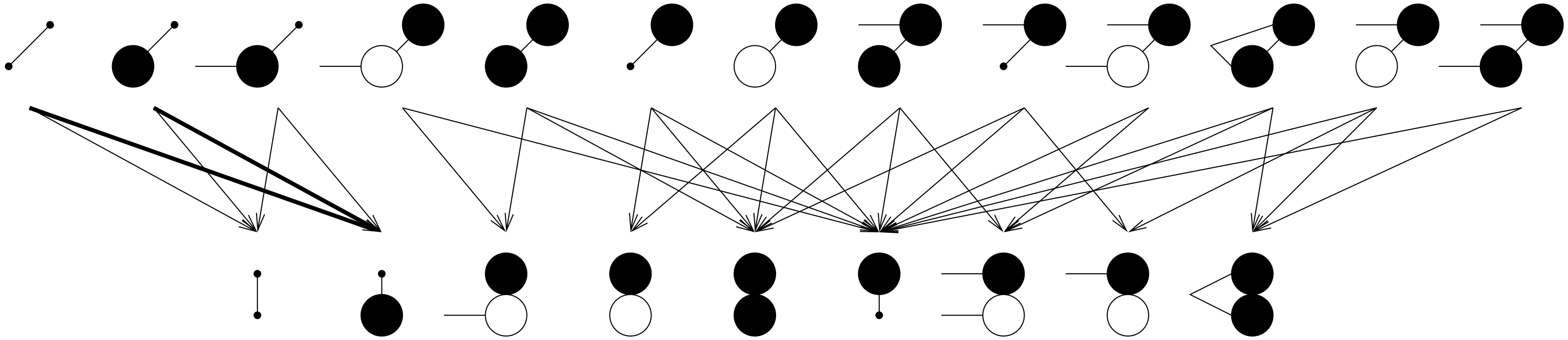
$AB(1)$



$B(2)$



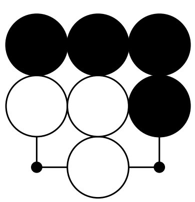
Edges from 0−states to 1−states

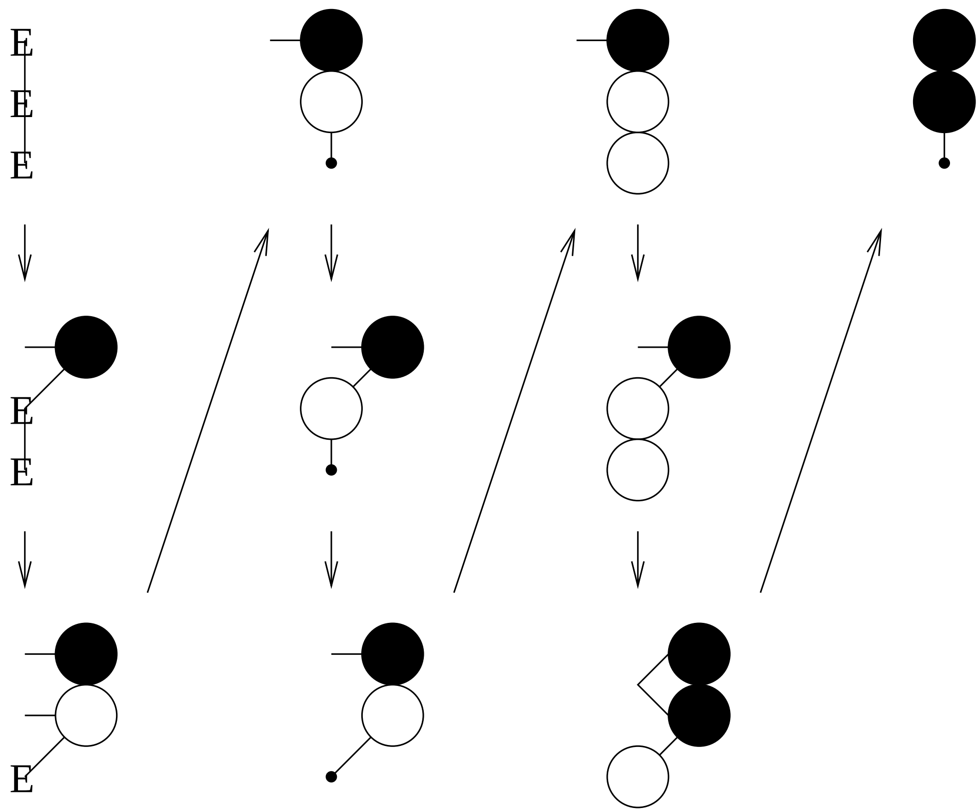


Edges from 1−states to 0−states

Positions are Paths

Legal $m \times n$ positions are in 1-1 correspondence with paths of length mn through $AB(m)$ starting at the all-Edge 0-state and ending at a 0-state with no libertyless stones.





State count vectors

- consider the board height m fixed.
- $\mathbf{L}(n, y)$: vector containing partial board count for all possible border states s
- $$\begin{array}{llll} \mathbf{L}(0,0) & \mathbf{L}(1,0) & \mathbf{L}(2,0) & \mathbf{L}(3,0) \\ \mathbf{L}(0,1) & \mathbf{L}(1,1) & \mathbf{L}(2,1) & \\ \mathbf{L}(0,2) & \mathbf{L}(1,2) & \mathbf{L}(2,2) & \end{array}$$
- border state graph yields linear transformations $\mathbf{T}_y$
- such that $\mathbf{L}(n, y+1) = \mathbf{T}_y \mathbf{L}(n, y)$
- hence $\mathbf{L}(n+1, 0) = \mathbf{T}_{m-1} \mathbf{T}_{m-2} \ldots \mathbf{T}_1 \mathbf{T}_0 \mathbf{L}(n, 0)$

Recurrences

- gives a matrix power expression for $L(m,n)$:
  $L(m,n) = \mathbf{l}^T \mathbf{T}^n \mathbf{L}(n,0)$
- where $\mathbf{T} = \mathbf{T}_{m-1} \ldots \mathbf{T}_0$,
- $\mathbf{L}(n,0)$ is a unit vector for the all-Edge state,
- and $\mathbf{l}$ is the characteristic vector of legal states.

Small dimensional boards

- $L(1, k+3) = 3L(1, k+2) - L(1, k+1) + L(1, k)$ $(\lambda_1 = 2.769)$
- $L(2, n+7) = 10L(2, n+6) - 16L(2, n+5) + 31L(2, n+4) - 13L(2, n+3) + 20L(2, n+2) + 2L(2, n+1) - L(2, n)$ $(\lambda_2 = 8.534)$
- $L(3, n+19) = 33L(3, n+18) - 233L(3, n+17) + 1171L(3, n+16) - 3750L(3, n+15) + 9426L(3, n+14) - 16646L(3, n+13) + 22072L(3, n+12) - 19993L(3, n+11) + 9083L(3, n+10) + 1766L(3, n+9) - 4020L(3, n+8) + 6018L(3, n+7) - 2490L(3, n+6) - 5352L(3, n+5) + 1014L(3, n+4) - 1402L(3, n+3) + 100L(3, n+2) + 73L(3, n+1) - 5L(3, n)$ $(\lambda_3 = 25.45)$

$$L(m,n) = a_m \lambda_m^n (1 + o(1))$$

size	order	a_m	$\sqrt[m]{\lambda_m}$
$1 \times n$	3	0.69412340909080771809	2.769292354238631
$2 \times n$	7	0.77605920648443217564	2.921241604535948
$3 \times n$	19	0.76692462372625158688	2.941265544348697
$4 \times n$	57	0.73972591465609392167	2.949764649676889
$5 \times n$	217	0.71384057986002504205	2.954933728838206
$6 \times n$	791	0.68921150040083474629	2.958390334214090
$7 \times n$	3107	0.66545979340188479816	2.960861834904016
$8 \times n$	12110	0.64252516474515096185	2.962716807025240
$9 \times n$	49361	0.62038058380200867949	2.9641603664723

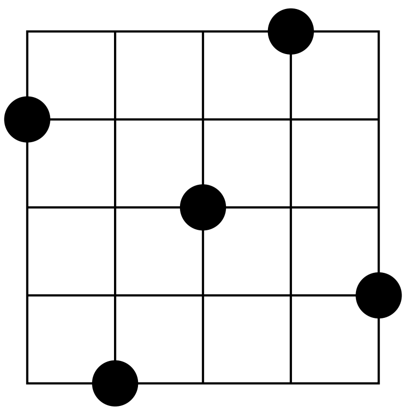
The Dynamic Programming algorithm

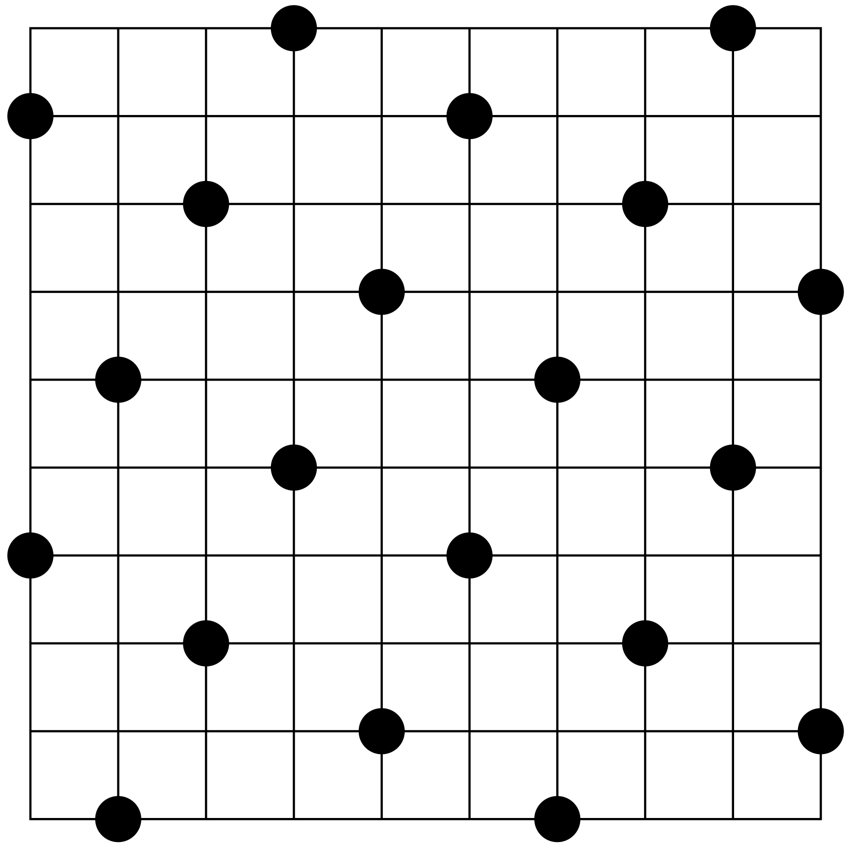
- count modulo some number $M \sim 2^{64}$
- gives equations $L(m,n) = a_i \bmod M_i$, solvable with CRT
- represents border state classes with 3 bits per point
- computes $\mathbf{L}(n, y+1) = \mathbf{T}_y \mathbf{L}(n, y)$
- states are partitioned over multiple cpus so as to exhaust available I/O bandwidth.
- state-count pairs are stored in hundreds of individually sorted files, which are read in parallel and merged.
- new state-counts are stored in blocked hash-table facilitating radix sort (using radix of 2^{12} for 19 count).
- whenever memory is full, one file is written for each cpu.

Results

n	#digits	$L(n,n)$
1	1	1
2	2	57
3	5	12675
4	8	24318165
5	12	414295148741
6	17	62567386502084877
7	23	83677847847984287628595
8	30	990966953618170260281935463385
9	39	103919148791293834318983090438798793469
10	47	96498428501909654589630887978835098088148177857
11	57	793474866816582266820936671790189132321673383112185151899
18	152	66972311428882921289274018884170654350993778064017873281031833769694562442854721810521432601277437139718484889097011183628347046881282790714992650234763 3

Asymptotics





$$3^{\frac{4}{5}n^2}(1-2/81)^{\frac{4}{5}n} \leq L(n,n) \leq 3^{n^2}(1-2/81)^{\frac{4}{5}n}(1-2/243)^{\frac{1}{5}n^2-\frac{4}{5}n}$$

Asymptotics

- Both $\sqrt[m]{\lambda_m}$ and $\sqrt[n^2]{L(n,n)}$ converge to the same value L, the *base of liberties*
- 2-dimensional analogue of the 1-dimensional growth rate $\lambda_1 \sim 2.769$.
- Since $\frac{L(m,n+1)}{L(m,n)}$ converges to λ_m,
  $\frac{L(m,n)L(m+1,n+1)}{L(m,n+1)L(m+1,n)}$ converges to λ_{m+1}/λ_m,
  which we expect to converge to $L^{m+1}/L^m = L$.

Base of Liberties

n	$L(n,n)L(n+1,n+1)/L(n,n+1)^2$
3	2.979
4	2.9756
5	2.975732
6	2.9757343
7	2.9757341927
8	2.9757341918
9	2.9757341920444
10	2.9757341920441
11	2.97573419204335O
12	2.975734192043355
13	2.9757341920433572 7
14	2.9757341920433572 55
15	2.975734192043357249 32
16	2.975734192043357249 3662

A formula for $L(m,n)$

- $L(m,n) \approx \alpha\beta^{m+n}L^{mn}$
- Using the value $L = 2.975734192043357249381$, we can solve for α and β with the computed values of $L(17,17), L(17,18)$ and $L(18,18)$, yielding
- $\alpha \approx 0.8506399258457145387, \beta \approx 0.96553505933837387$
- Achieves relative accuracy 0.99993 at $n = 5$, 0.99999999 at $n = 9$, and 1.00000000000023 at $n = 13$.
- $L(19,19) \approx 2.08168199381982 \cdot 10^{170}$.

Number of Games

$m \setminus n$	1	2	3	4	5	6
1	1	9	907	2098407841	$\sim 10^{31}$	~ 10
2		386356909593	$\sim 10^{86}$	$10^{\sim 5.3 \cdot 10^2}$		
3			$10^{\sim 1.1 \cdot 10^3}$			

Approximate values by Heuristic Sampling.

Upper bounds

- On boards larger than 1×1, every node in the game graph has outdegree at least 2.
- #games $\leq \prod_v$ outdeg$(v) \quad (mn > 1)$
- average outdegree close to $2mn/3$ in the limit
- #games $\leq (mn)^{L(m,n)}$

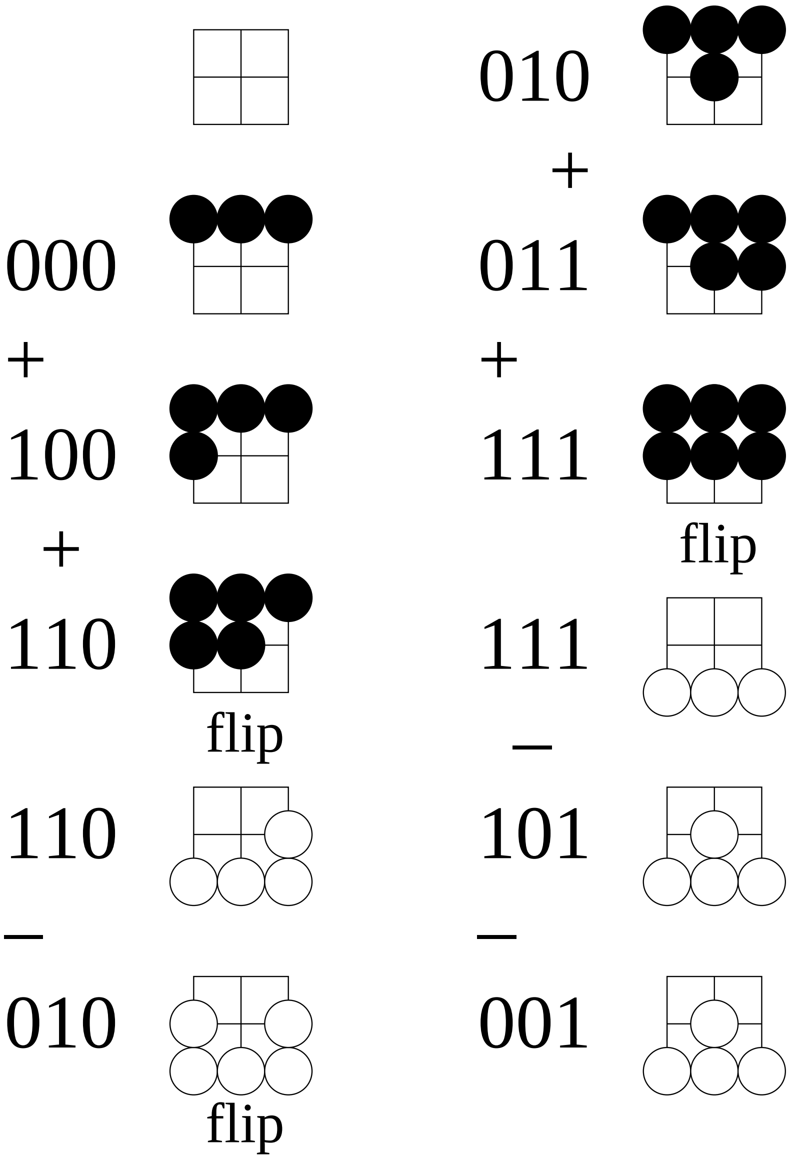
Lower bounds

Suppose the mn points on the board can be partitioned into 3 sets B, W, E such that

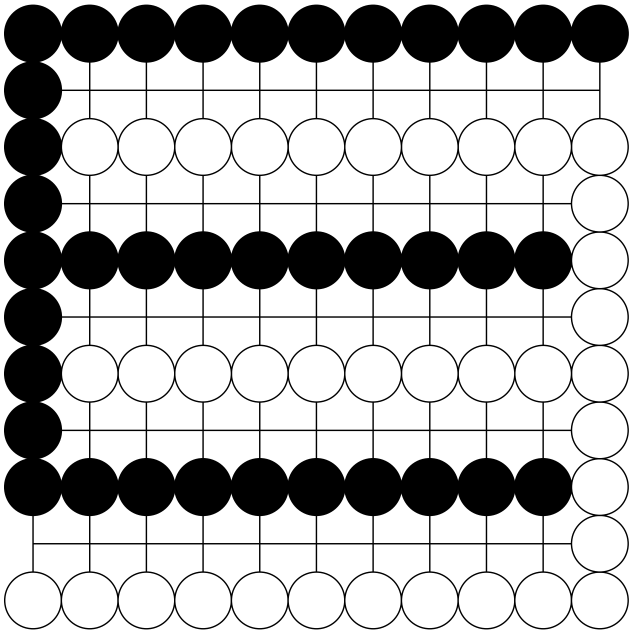
- $|B| = |W| = k, |E| = l = mn - 2k$,
- B and W are connected,
- each point in E is adjacent to both B and W

Then there are at least $(k!)^{2^{l-1}}$ possible games, all lasting over $k2^{l-1}$ moves.

Proof



Bounds

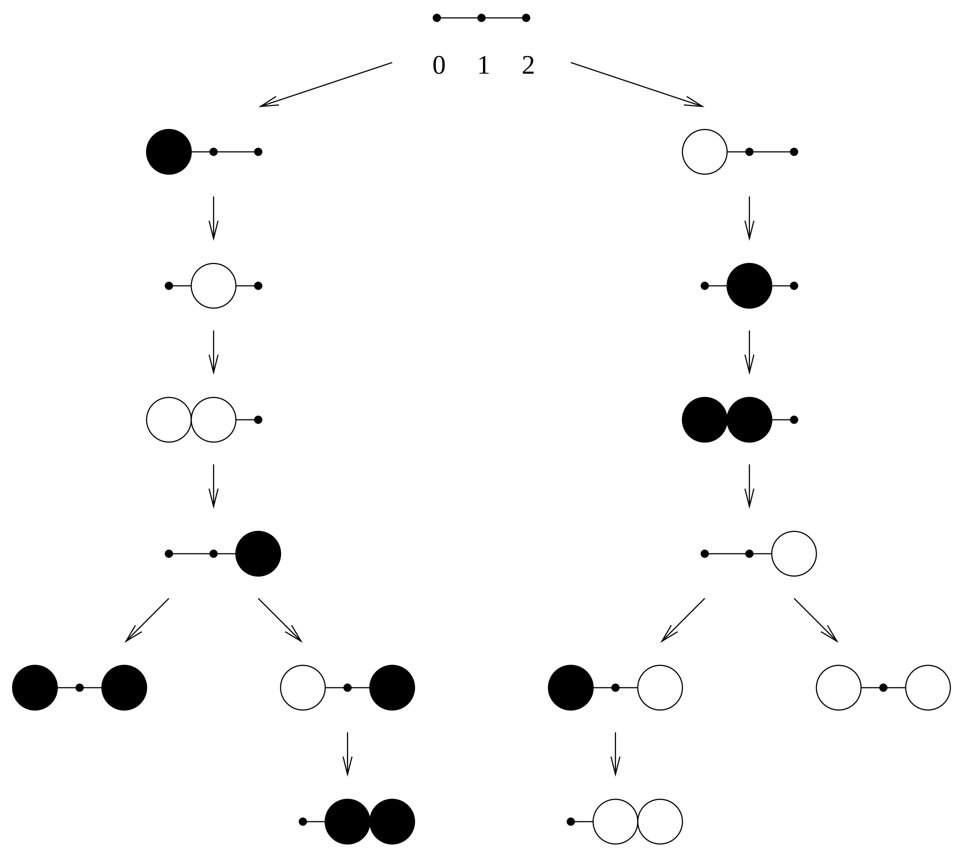
- 
- $k = |B| = |W| = n - 1 + (n-2)(n+1)/4$ and $l = 2 + (n-2)(n-1)/2$
- $2^{2^{n^2/2 - O(n)}} \leq$ #games $\leq 2^{2^{n^2 \log 3 + \log\log n + O(1)}}$

number N of 19x19 games

- $(103!)^{2^{154}} \leq N \leq 361^{0.012 \cdot 3^{361}}$
- in binary: $2^{2^{163}} < N < 2^{2^{569}}$
- in decimal: $10^{10^{48}} < N < 10^{10^{171}}$

In 1 dimension

- stone at i has weight 2^i
- predecessor: replace leftmost stone at i by opposite stones at $0, 1, \ldots, i-1$
- 2^{n-2} skippable positions on each max path
- #games $\geq 2^{2^{n-1}}$



Hamiltonian Games

- Games in which every legal position occurs.
- Only one-dimensional boards can be Hamiltonian.
- Equivalently, $G(1,n)$ must have a directed Hamiltonian path starting at the empty position.
- True for $n = 1, 3, 4, 5, 6, 7$.
- Conjecture: true for all larger n as well.

Open problems

- Finish computing $L(19, 19)$, the number of legal positions on a standard size Go board.
- Prove $L(m, n) \approx \alpha\beta^{m+n}L^{mn}$.
- Prove Hamiltonicity of all $G(1, n), n \geq 3$.
- Find more efficient algorithm for computing $L(m, n)$
- write a Go program to challenge human pros:-)