Google DeepMind's AlphaGo vs. world Go champion Lee Sedol

Review of Nature paper: Mastering the game of Go with Deep Neural Networks & Tree Search

Tapani Raiko

Aalto University

THE CURIOUS AI COMPANY

Thanks to Antti Tarvainen for some slides

Slide from my lecture on 26 March 2015: Future of Go AI

- In 2015, a computer vision based approach (deep convolutional neural network) was trained to mimick human player moves
- This resulted in surprisingly strong AI without any look-ahead (almost competitive with Monte-Carlo)
- Two approaches are very different in nature
- Perhaps a good combination will beat humans in near future

Aalto University

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

At *last* — a computer program that can beat a champion Go player PAGE 484

ALL SYSTEMS GO

CONSERVATION

SONGBIRDS À LA CARTE

Illegal harvest of millions of Mediterranean birds

PAGE 452

RESEARCH ETHICS

SAFEGUARD TRANSPARENCY

Don't let openness backfire on individuals

PAGE 459

POPULAR SCIENCE

WHEN GENES GOT 'SELFISH'

Dawkins's calling card 40 years on

PAGE 462

NATUREASIA.COM

28 January 2016

Vol. 529. No. 7587

Lee Sedol vs. AlphaGo

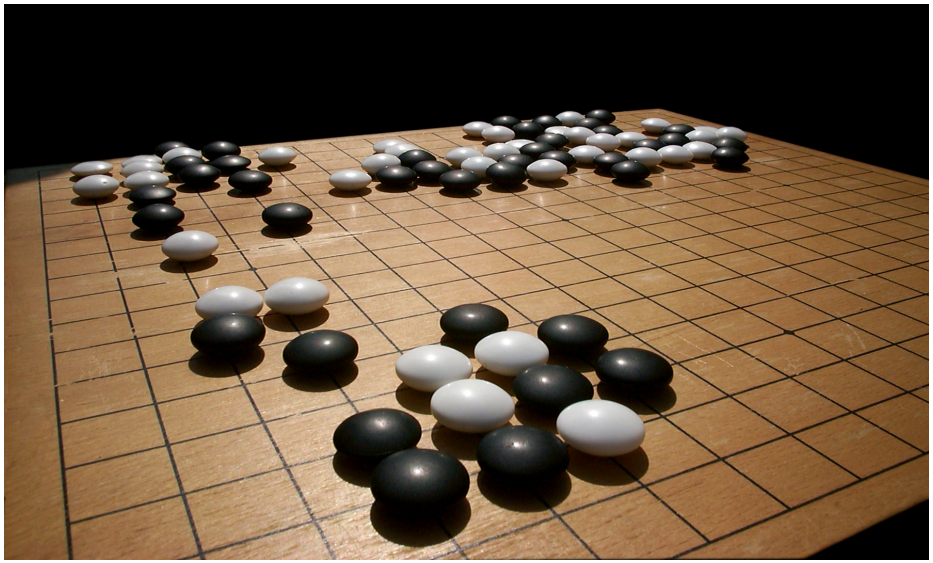
1 brain
80 watts

1202 CPUs, 176 GPUs
$>$100 000 watts

Table of Contents

Aalto University

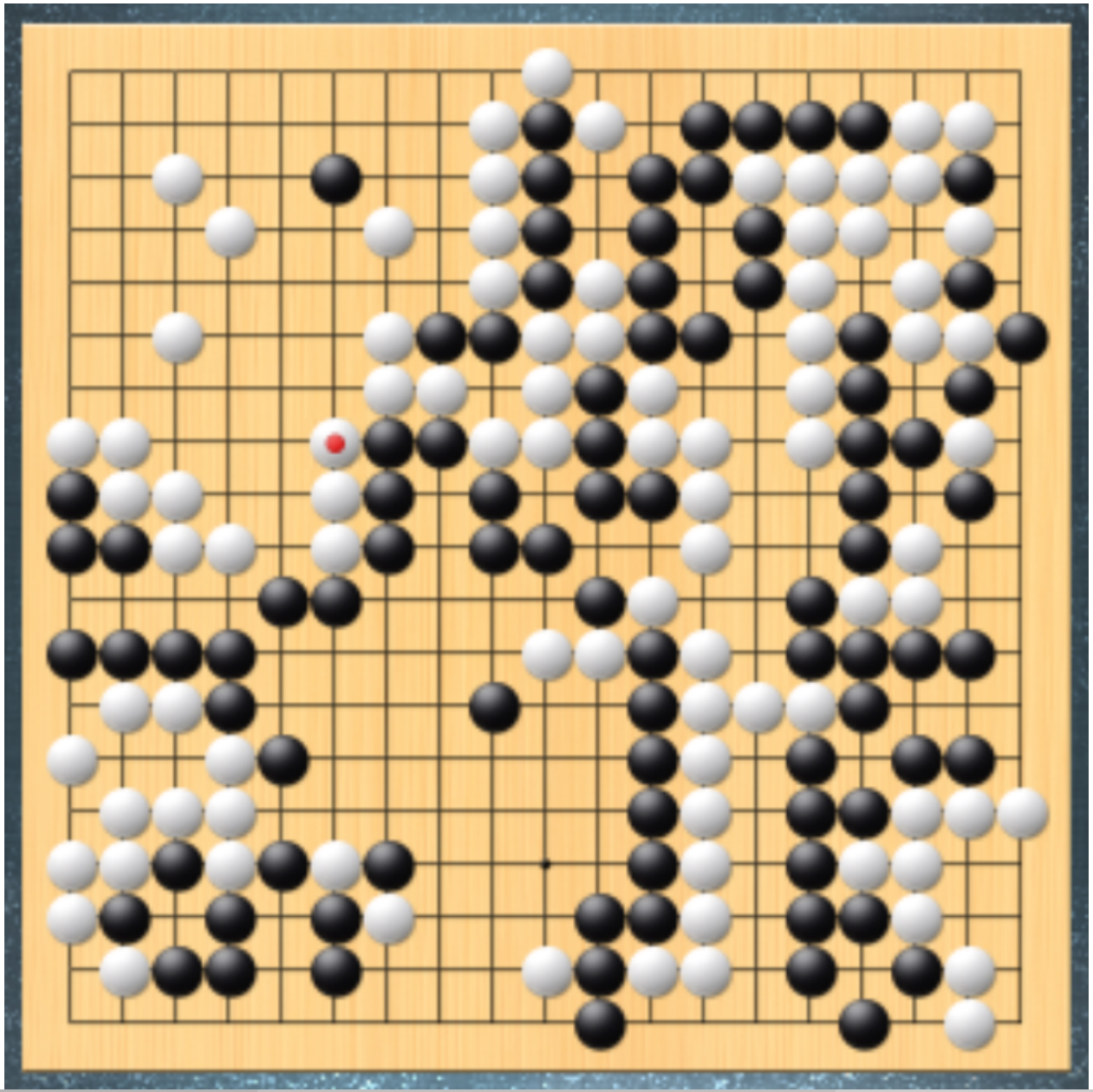
Go (or Baduk or Weiqi)



- Two-player fully-observable deterministic zero-sum board game
- Oldest game still played with same simple rules
- Most popular board game in the world
- AI won human champion in March 2016

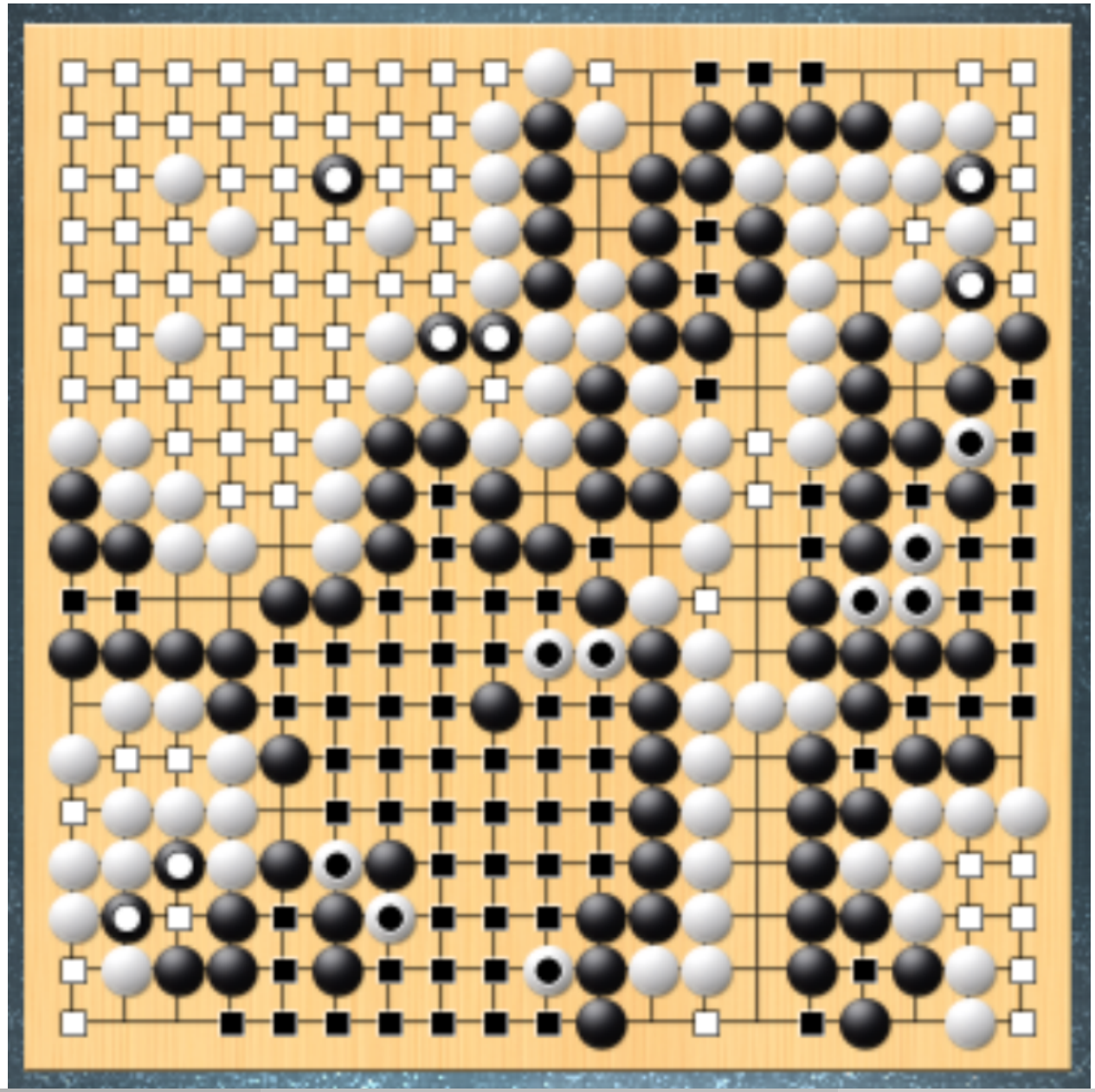
Aalto University

Game 1: Lee Sedol (black) resigned here



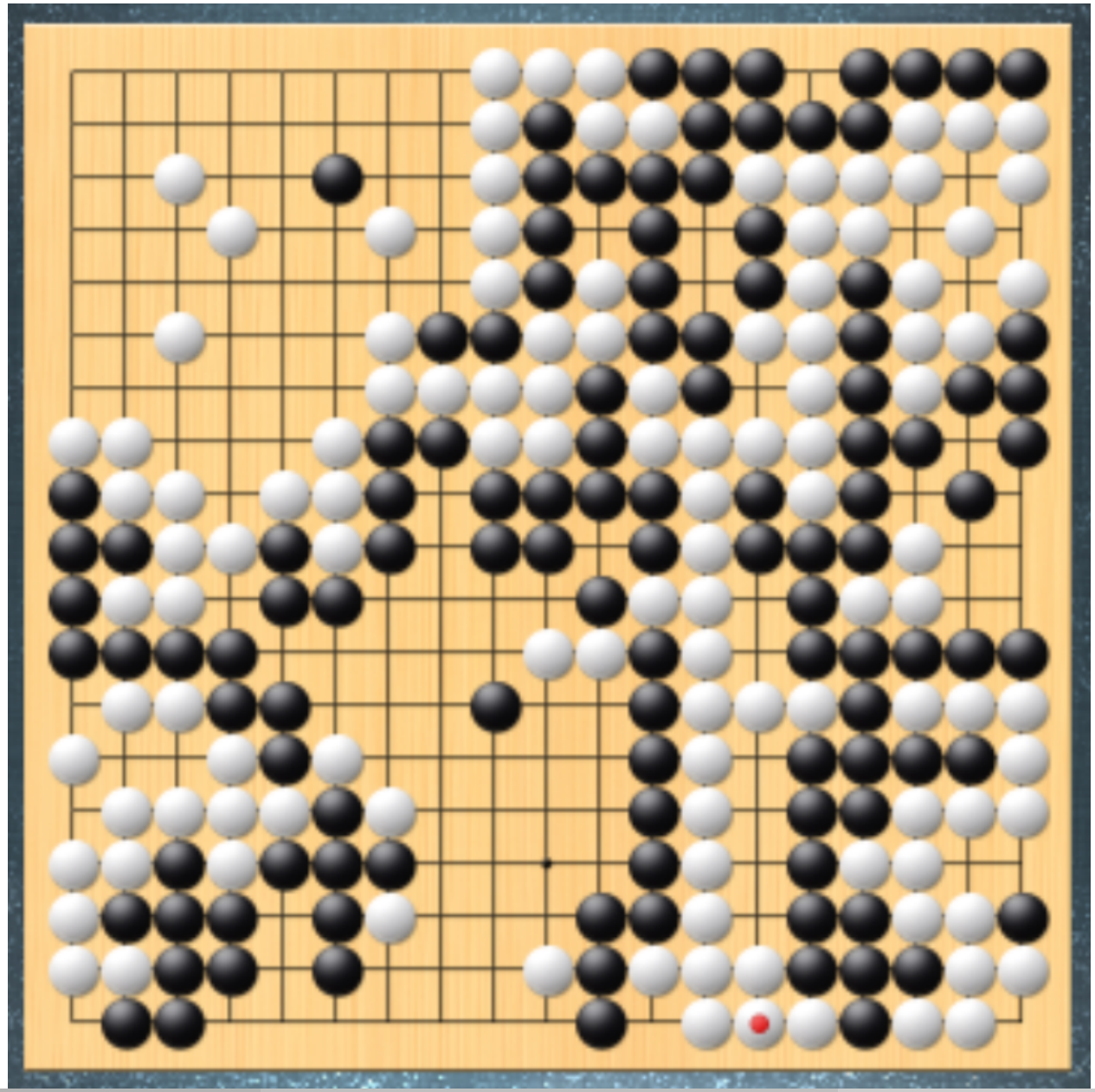
Aalto University

White is leading a bit



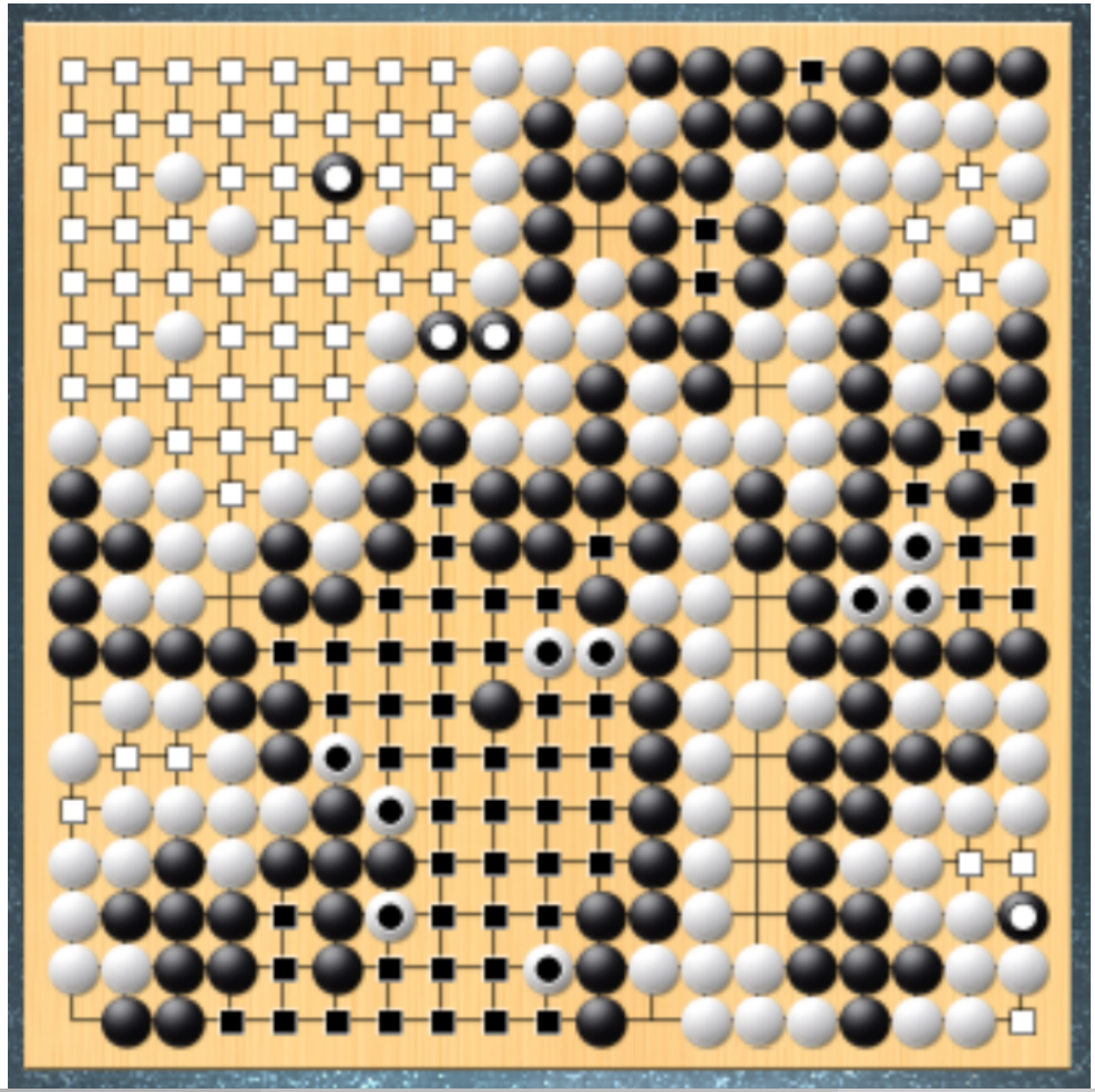
Aalto University

Amateurs would have finished e.g. like this



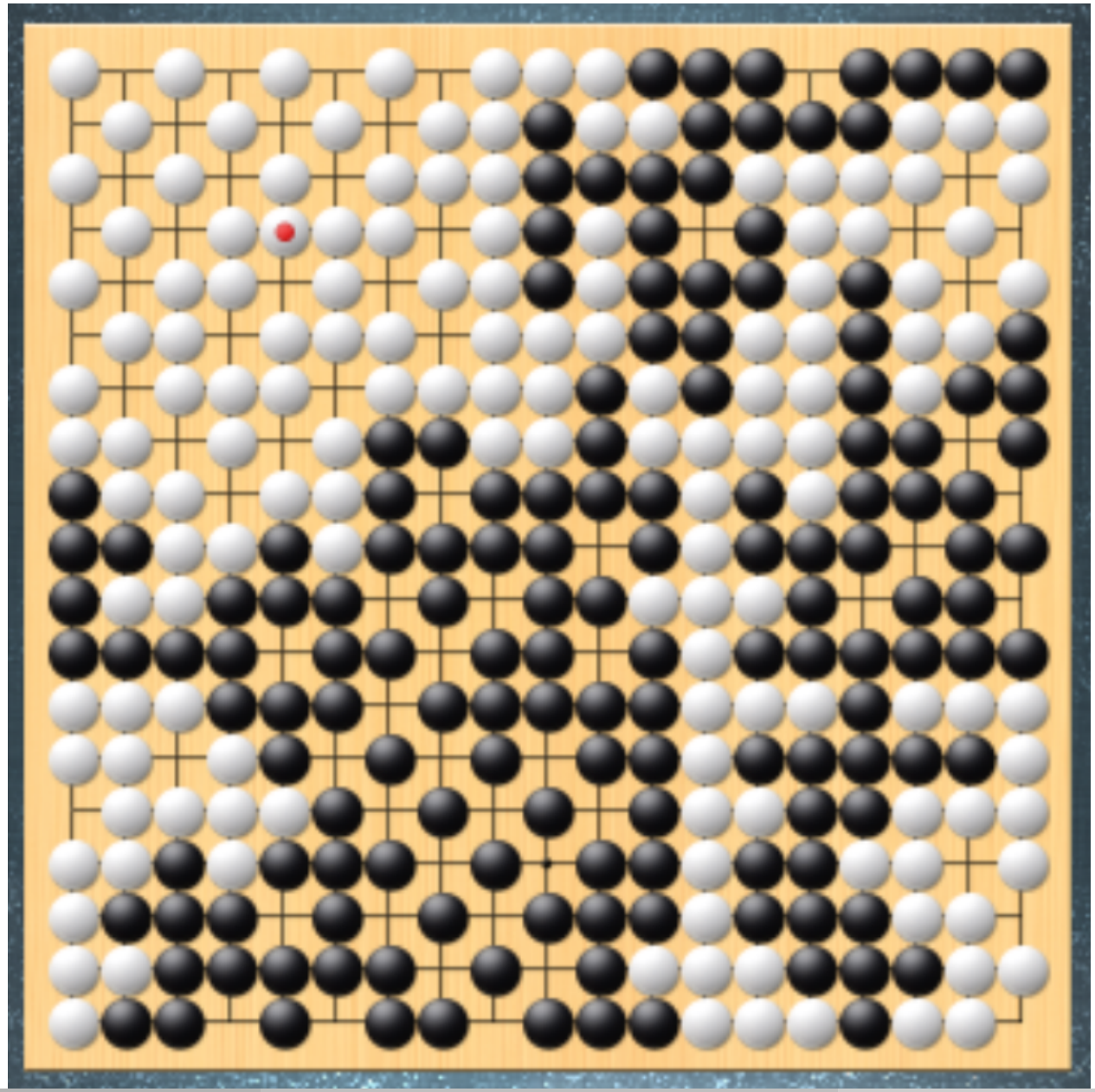
Aalto University

...and count that white won by 4.5 points



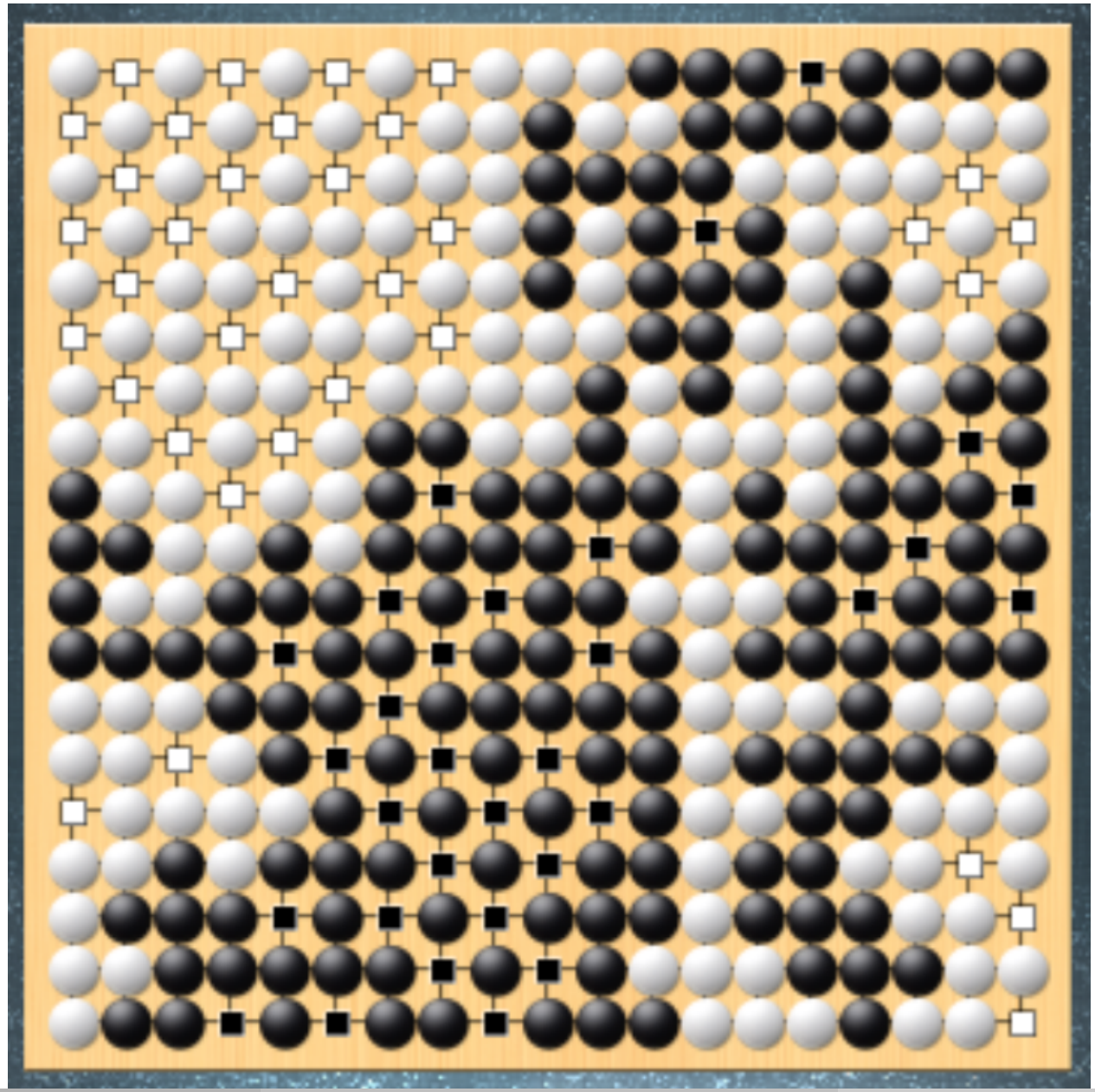
Aalto University

Beginners (or simple AI rollouts) continue



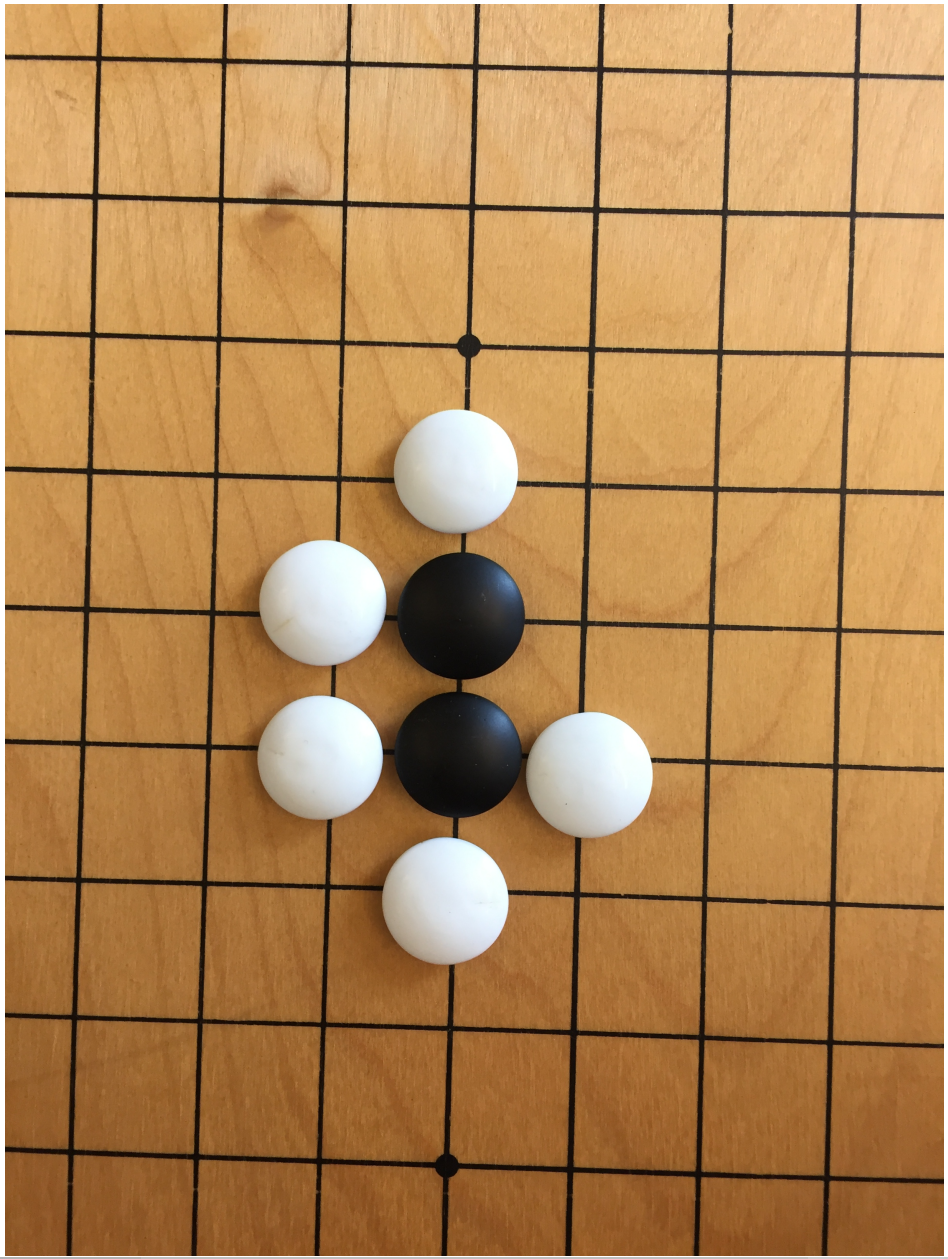
Aalto University

...until scoring is really obvious



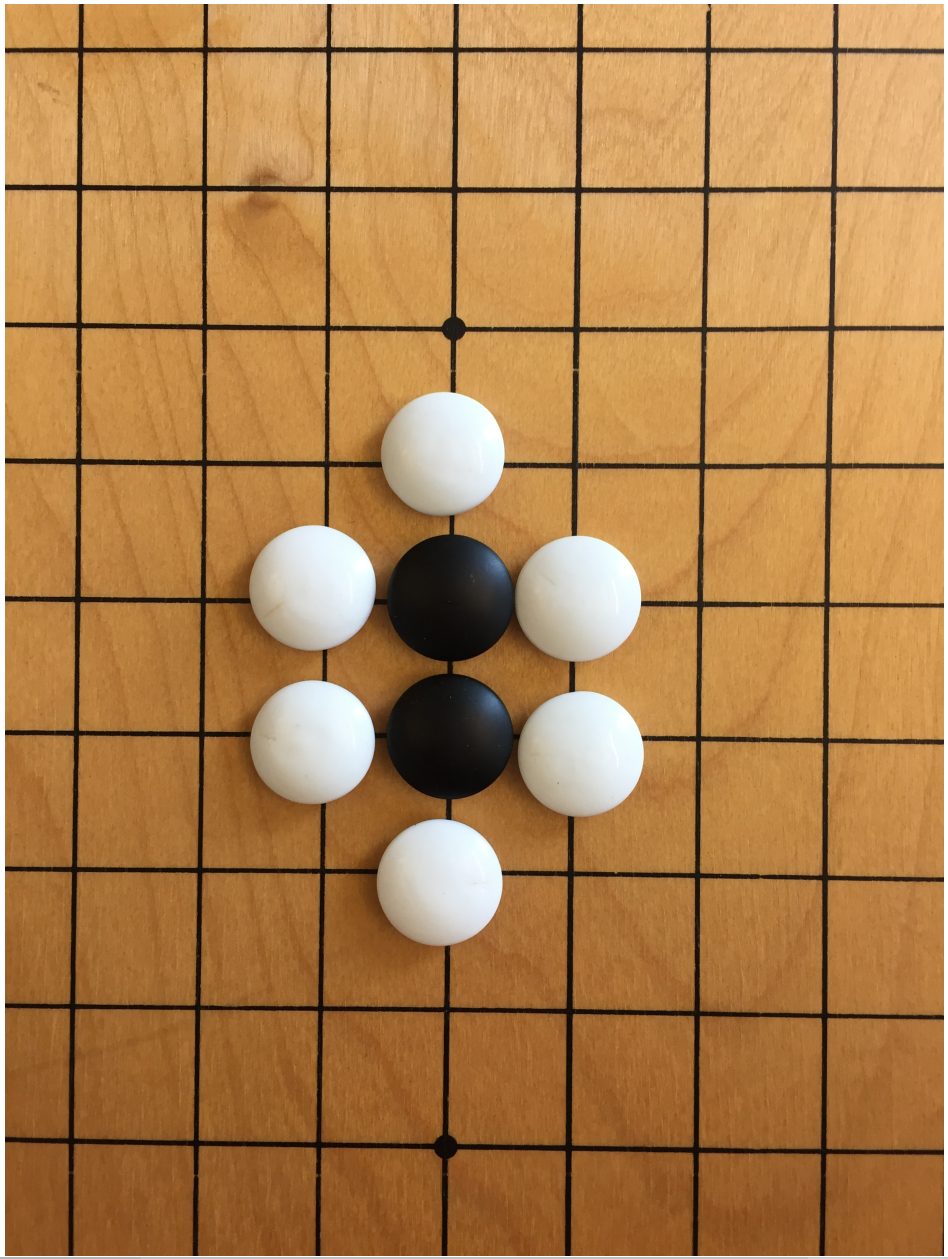
Aalto University

Capturing



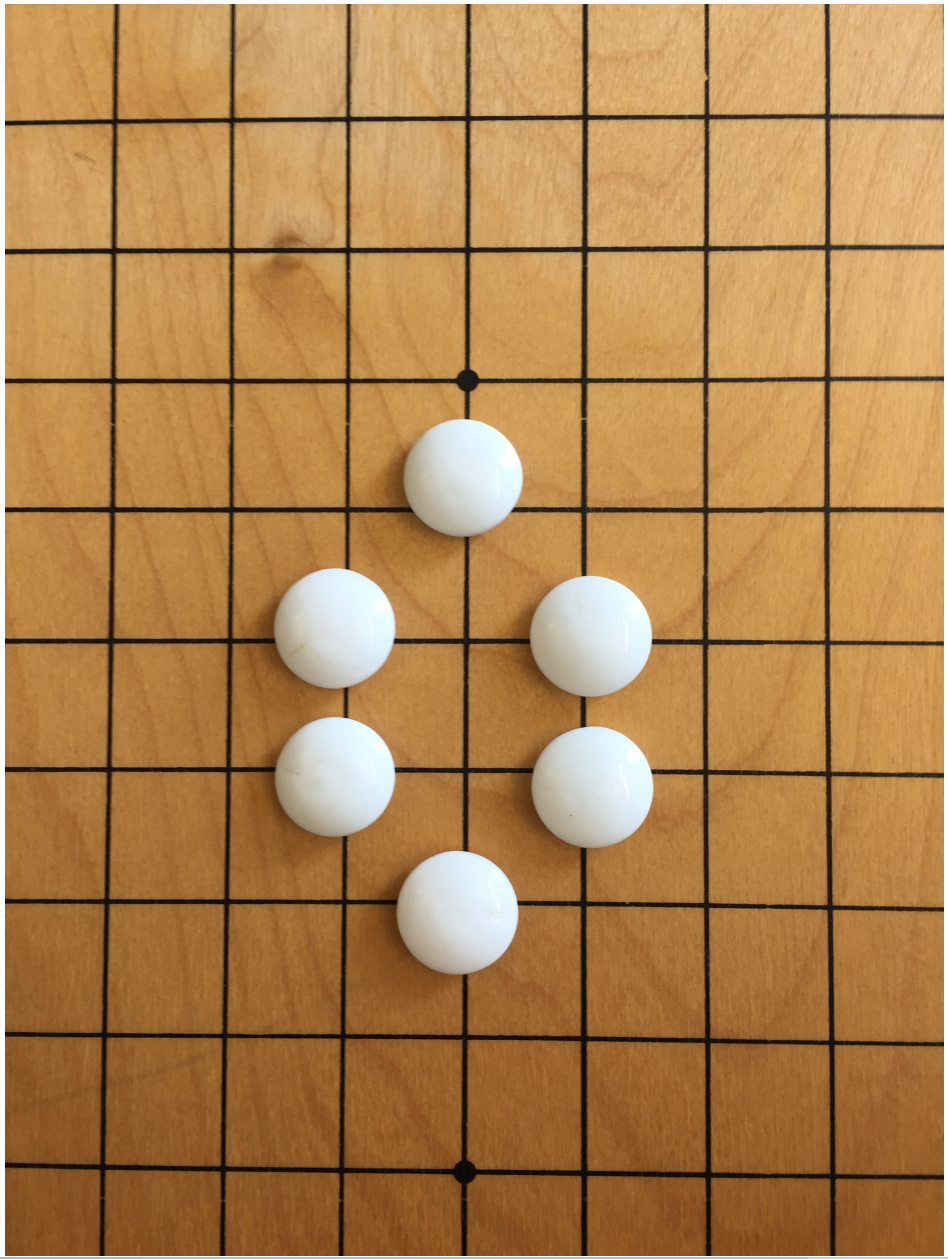
Aalto University

Capturing



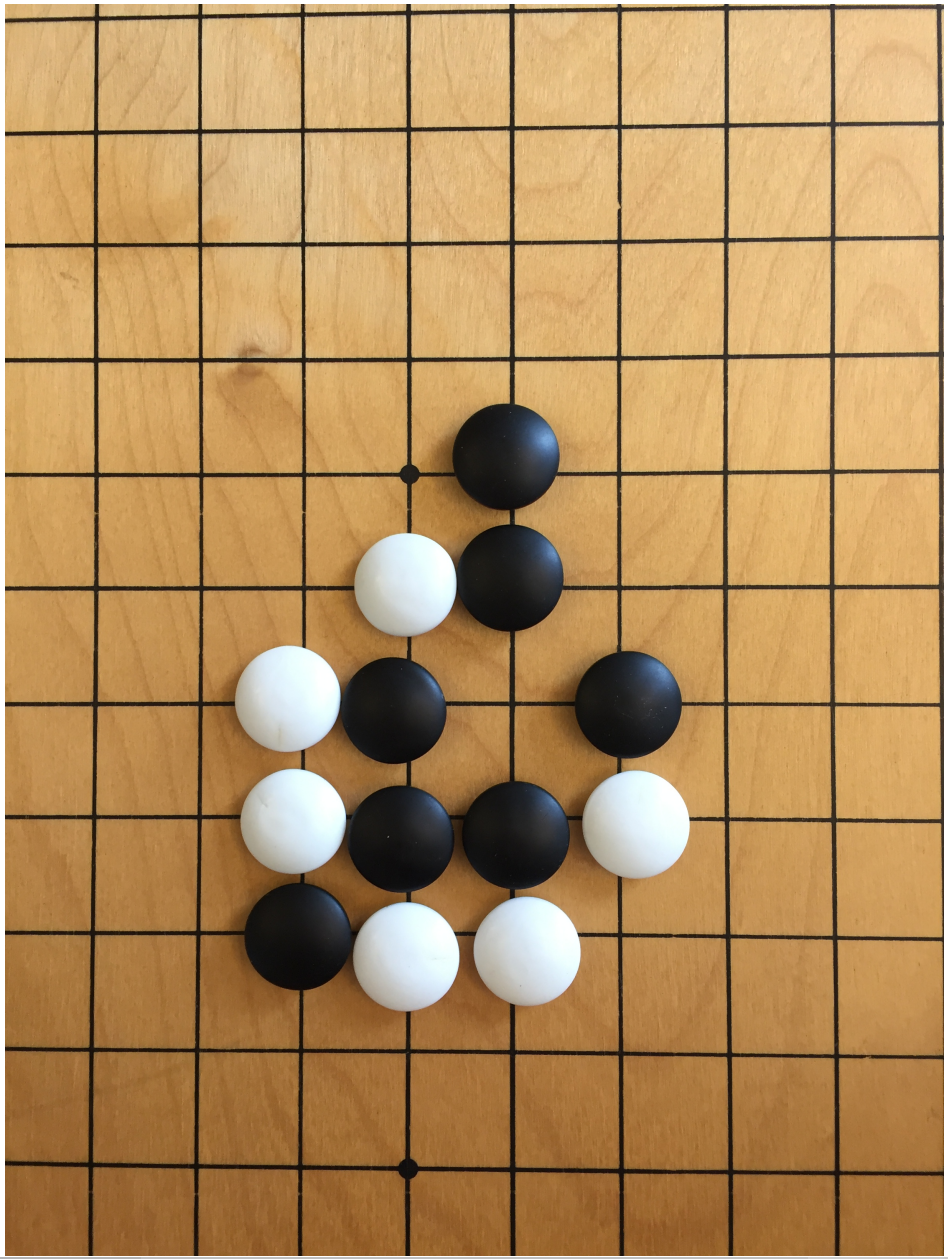
Aalto University

Capturing



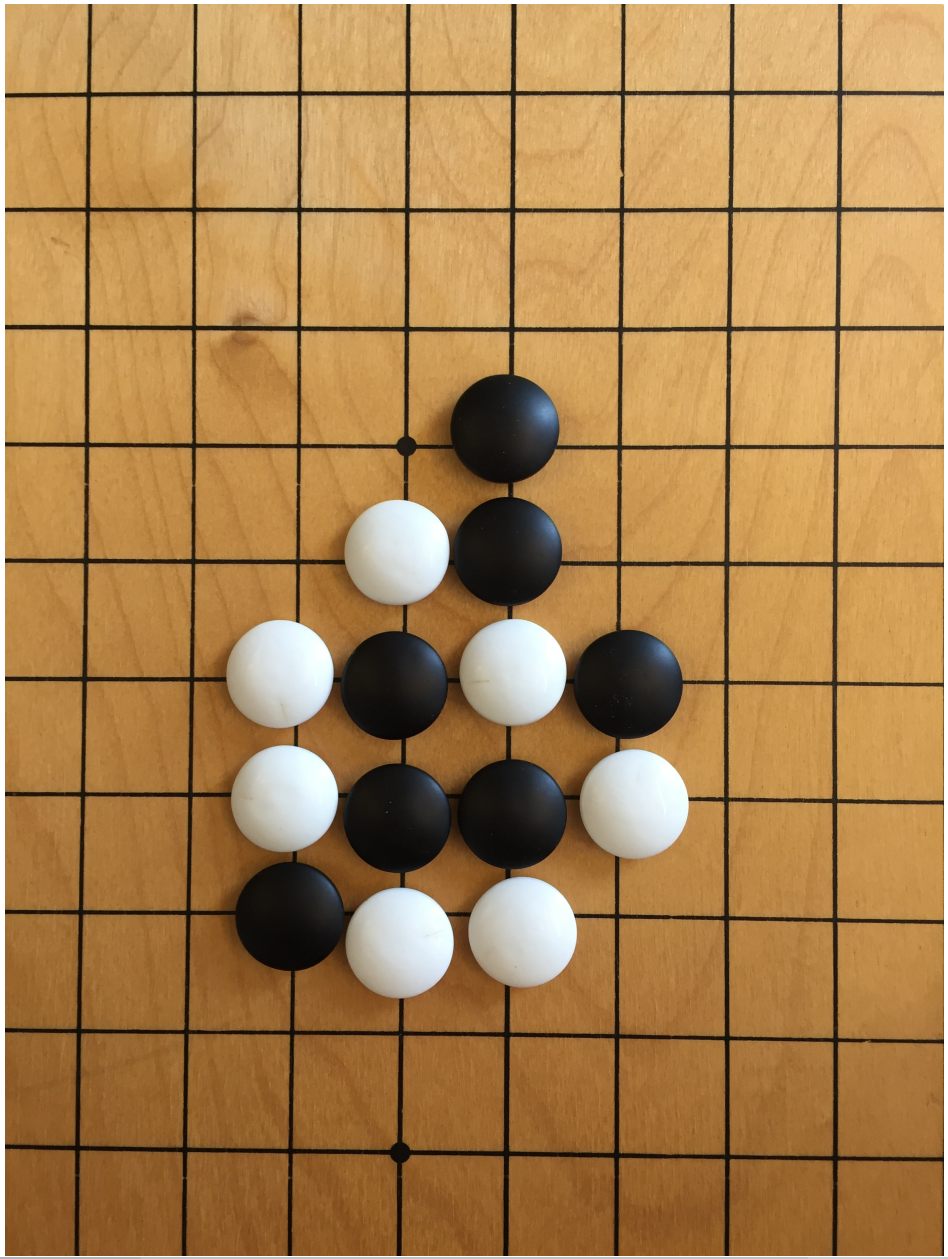
Aalto University

Capturing



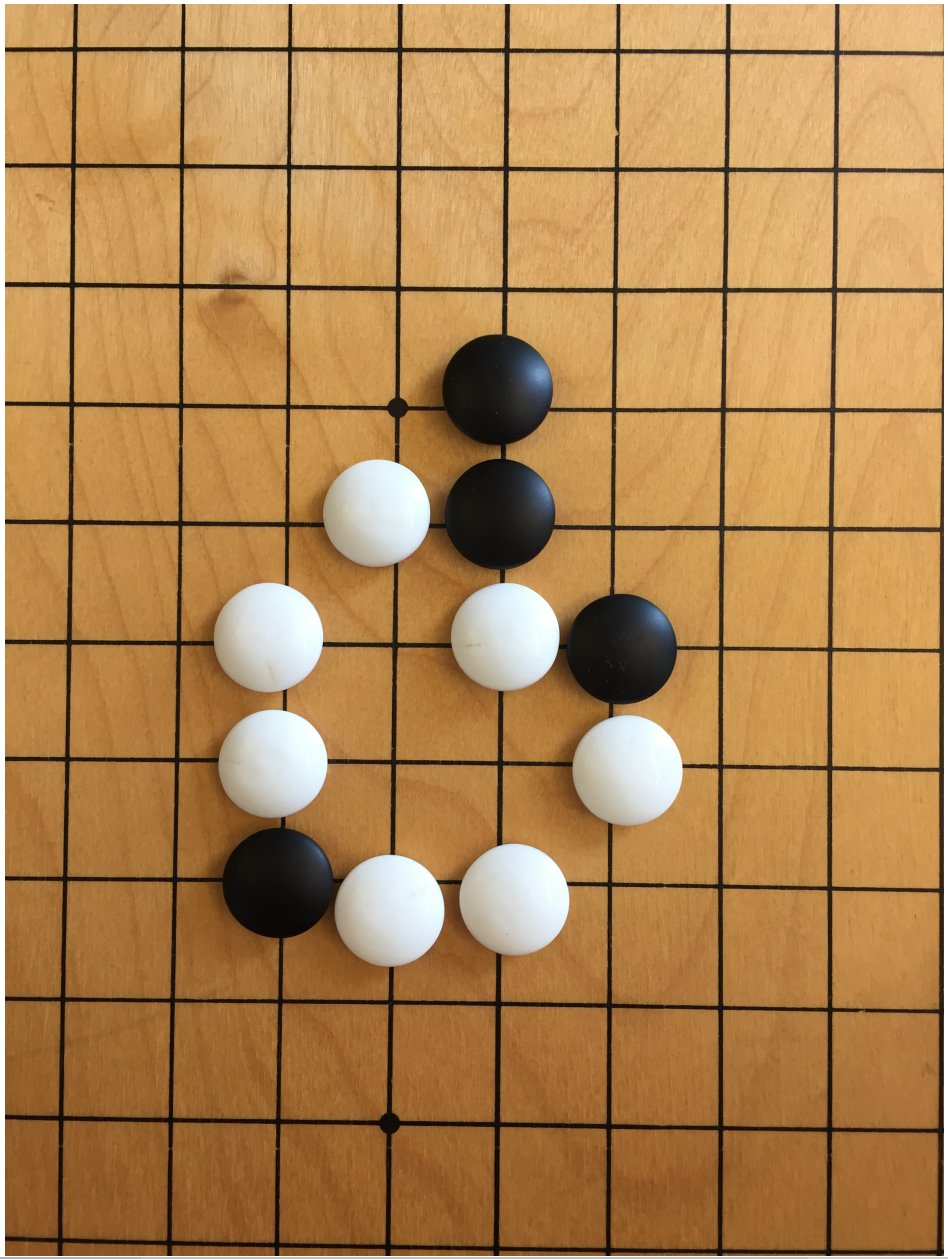
Aalto University

Capturing

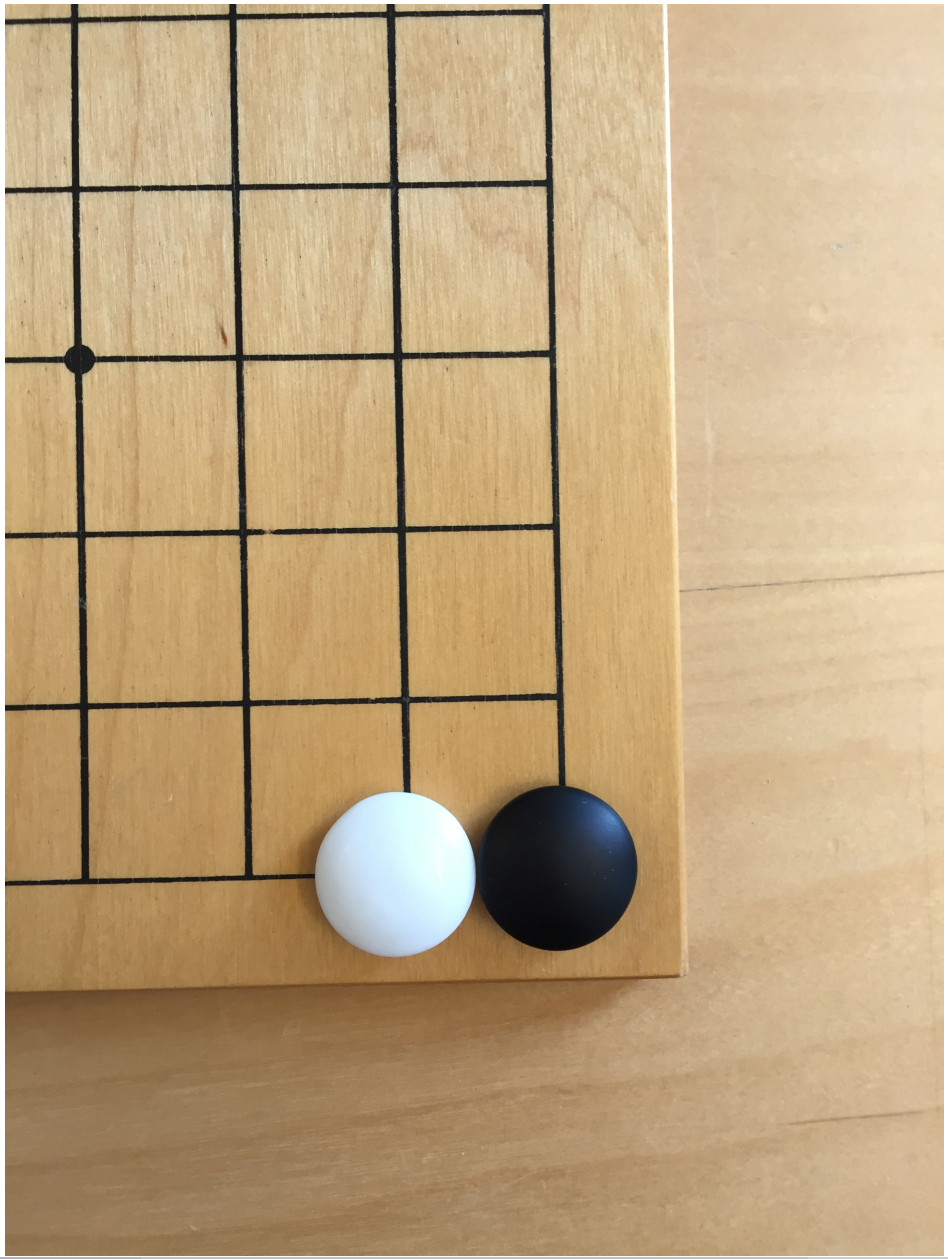


Aalto University

Capturing

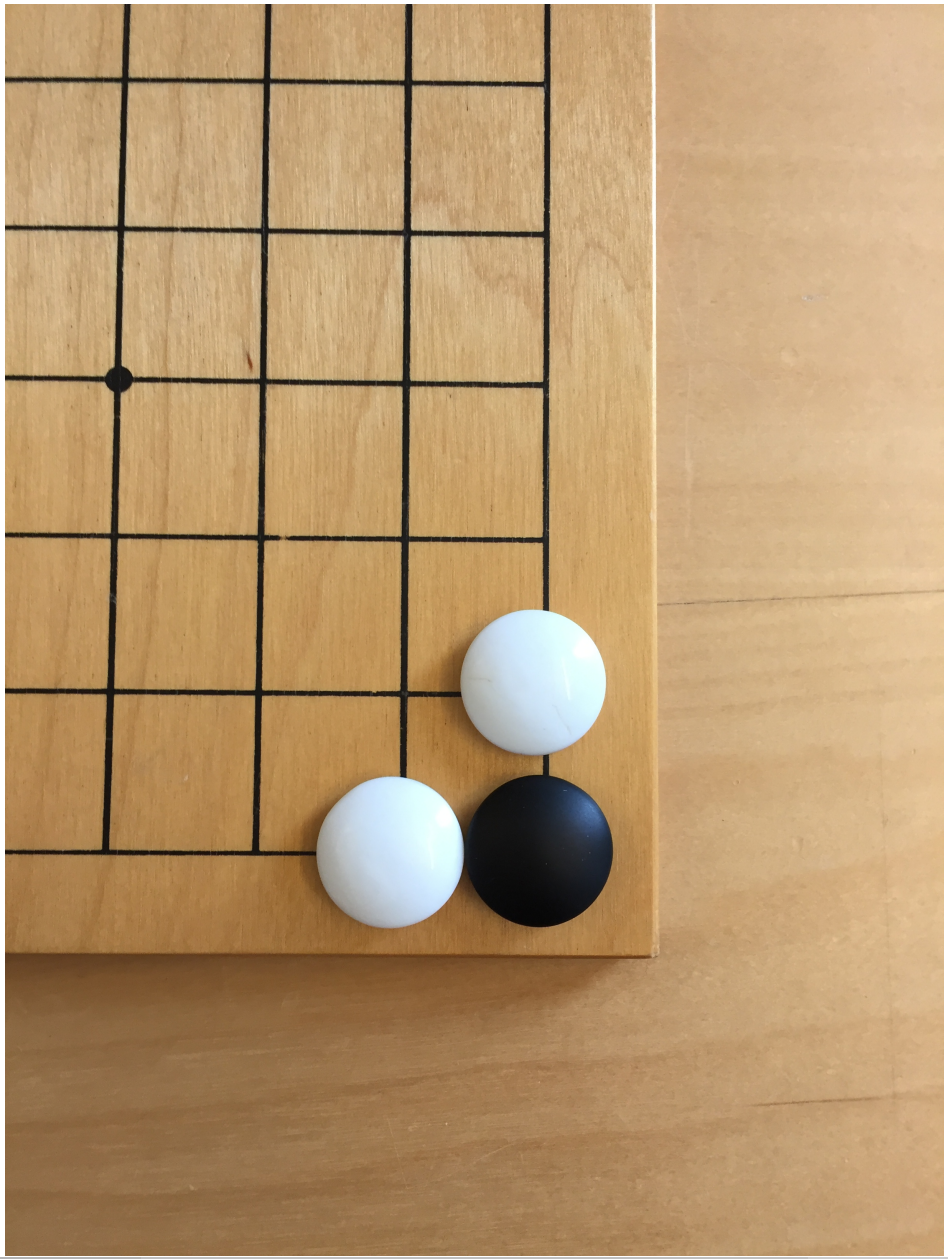


Capturing



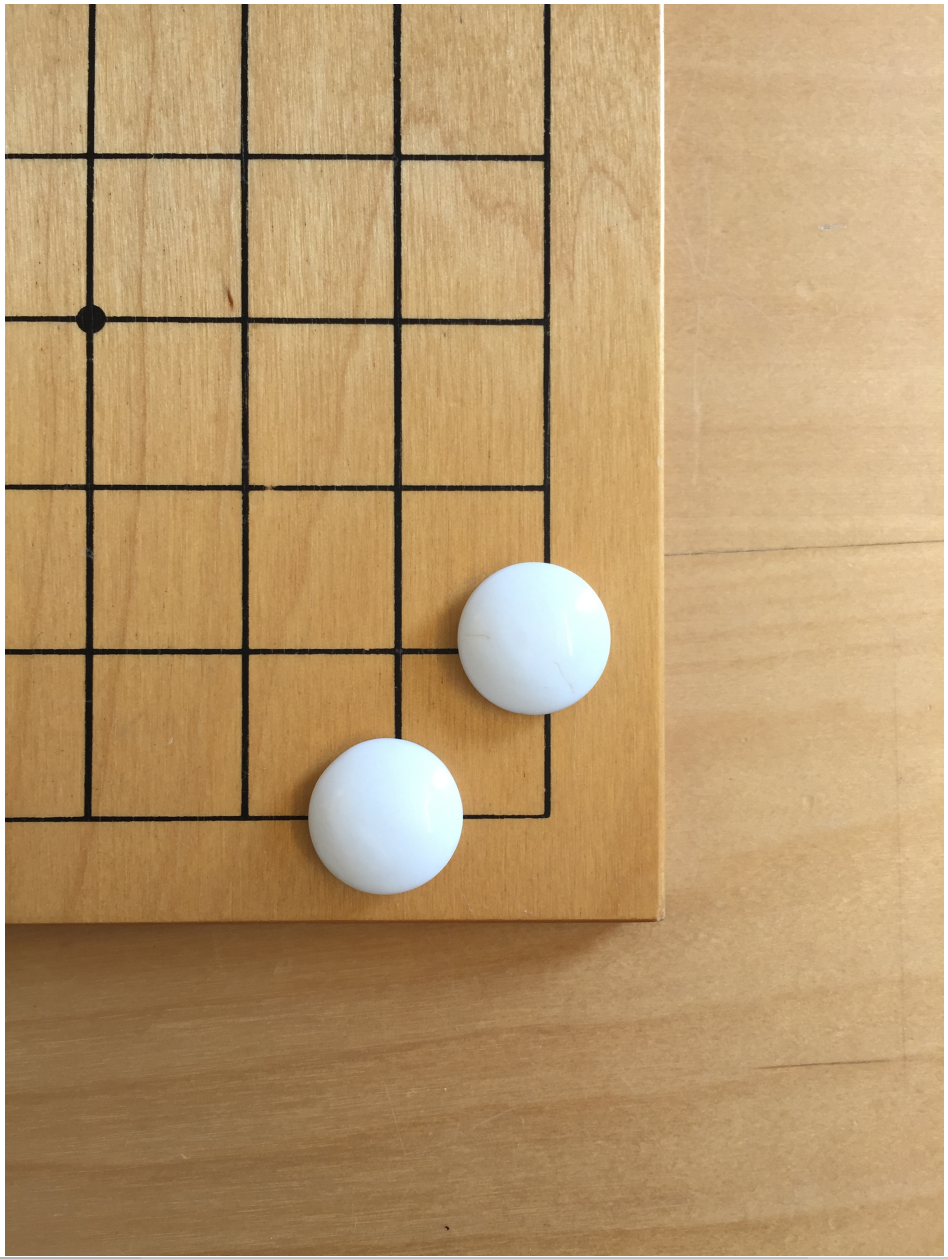
Aalto University

Capturing



Aalto University

Capturing



Aalto University

Rules of Chess (1/2)

Chess is played on a chessboard, a square board divided into 64 squares (eight-by-eight) of alternating color. No matter what the actual colors of the board, the lighter-colored squares are called "light" or "white", and the darker-colored squares are called "dark" or "black". Sixteen "white" and sixteen "black" pieces are placed on the board at the beginning of the game. The board is placed so that a white square is in each player's near-right corner. Horizontal rows are called ranks and vertical rows are called files. At the beginning of the game, the pieces are arranged as follows: The rooks are placed on the outside corners, right and left edge. The knights are placed immediately inside of the rooks. The bishops are placed immediately inside of the knights. The queen is placed on the central square of the same color of that of the player: white queen on the white square and black queen on the black square. The king takes the vacant spot next to the queen. The pawns are placed one square in front of all of the other pieces.
White moves first, then players alternate moves. Making a move is required; it is not legal to skip a move. Play continues until a king is checkmated, a player resigns, or a draw is declared. Each type of chess piece has its own method of movement. A piece moves to a vacant square except when capturing an opponent's piece. Except for any move of the knight and castling, pieces cannot jump over other pieces. A piece is captured (or taken) when an attacking enemy piece replaces it on its square (en passant is the only exception). The captured piece is thereby permanently removed from the game. The king can be put in check but cannot be captured (see below). The king moves exactly one square horizontally, vertically, or diagonally. A special move with the king known as castling is allowed only once per player, per game (see below). A rook moves any number of vacant squares in a horizontal or vertical direction. It also is moved when castling. A bishop moves any number of vacant squares in any diagonal direction. The queen moves any number of vacant squares in a horizontal, vertical, or diagonal direction. A knight moves to the nearest square not on the same rank, file, or diagonal. The knight is not blocked by other pieces: it jumps to the new location. Pawns have the most complex rules of movement: A pawn moves straight forward one square, if that square is vacant. If it has not yet moved, a pawn also has the option of moving two squares straight forward, provided both squares are vacant. Pawns cannot move backwards. Pawns are the only pieces that capture differently from how they move. A pawn can capture an enemy piece on either of the two squares diagonally in front of the pawn (but cannot move to those squares if they are vacant). The pawn is also involved in the two special moves en passant and promotion. Castling consists of moving the king two squares towards a rook, then placing the rook on the other side of the king, adjacent to it. Castling is only permissible if all of the following conditions hold: The king and rook involved in castling must not have previously moved; There must be no pieces between the king and the rook; The king may not currently be in check, nor may the king pass through or end up in a square that is under attack by an enemy piece (though the rook is permitted to be under attack and to pass over an attacked square); The king and the rook must be on the same rank.

.

Aalto University

Rules of Chess (2/2)

En passant. When a pawn advances two squares from its original square and ends the turn adjacent to a pawn of the opponent's on the same rank, it may be captured by that pawn of the opponent's, as if it had moved only one square forward. This capture is only legal on the opponent's next move immediately following the first pawn's advance.
Pawn promotion. If a player advances a pawn to its eighth rank, the pawn is then promoted (converted) to a queen, rook, bishop, or knight of the same color at the choice of the player (a queen is usually chosen). The choice is not limited to previously captured pieces.
Check. A king is in check when it is under attack by at least one enemy piece. A piece unable to move because it would place its own king in check (it is pinned against its own king) may still deliver check to the opposing player. A player may not make any move which places or leaves his king in check. The possible ways to get out of check are: Move the king to a square where it is not threatened. Capture the threatening piece (possibly with the king). Block the check by placing a piece between the king and the opponent's threatening piece. If it is not possible to get out of check, the king is checkmated and the game is over.
Draw. The game ends in a draw if any of these conditions occur: The game is automatically a draw if the player to move is not in check but has no legal move. This situation is called a stalemate. An example of such a position is shown in the diagram to the right. The game is immediately drawn when there is no possibility of checkmate for either side with any series of legal moves. This draw is often due to insufficient material, including the endgames king against king; king against king and bishop; king against king and knight; king and bishop against king and bishop, with both bishops on squares of the same color. Both players agree to a draw after one of the players makes such an offer. The player having the move may claim a draw by declaring that one of the following conditions exists, or by declaring an intention to make a move which will bring about one of these conditions: Fifty-move rule: There has been no capture or pawn move in the last fifty moves by each player. Threefold repetition: The same board position has occurred three times with the same player to move and all pieces having the same rights to move, including the right to castle or capture en passant.

Aalto University

Rules of Go

Go is played on 19×19 square grid of points, by players called Black and White. Each point on the grid may be colored black, white or empty.

A point P, is said to reach color C, if there is a path of (vertically or horizontally) adjacent points of P's color from P to a point of color C.

Clearing a color means emptying all points of that color that don't reach empty.

Starting with an empty grid, the players alternate turns, starting with Black.

A turn is either a pass; or a move that doesn't repeat an earlier grid coloring.

A move consists of coloring an empty point one's own color; then clearing the opponent color, and then clearing one's own color.

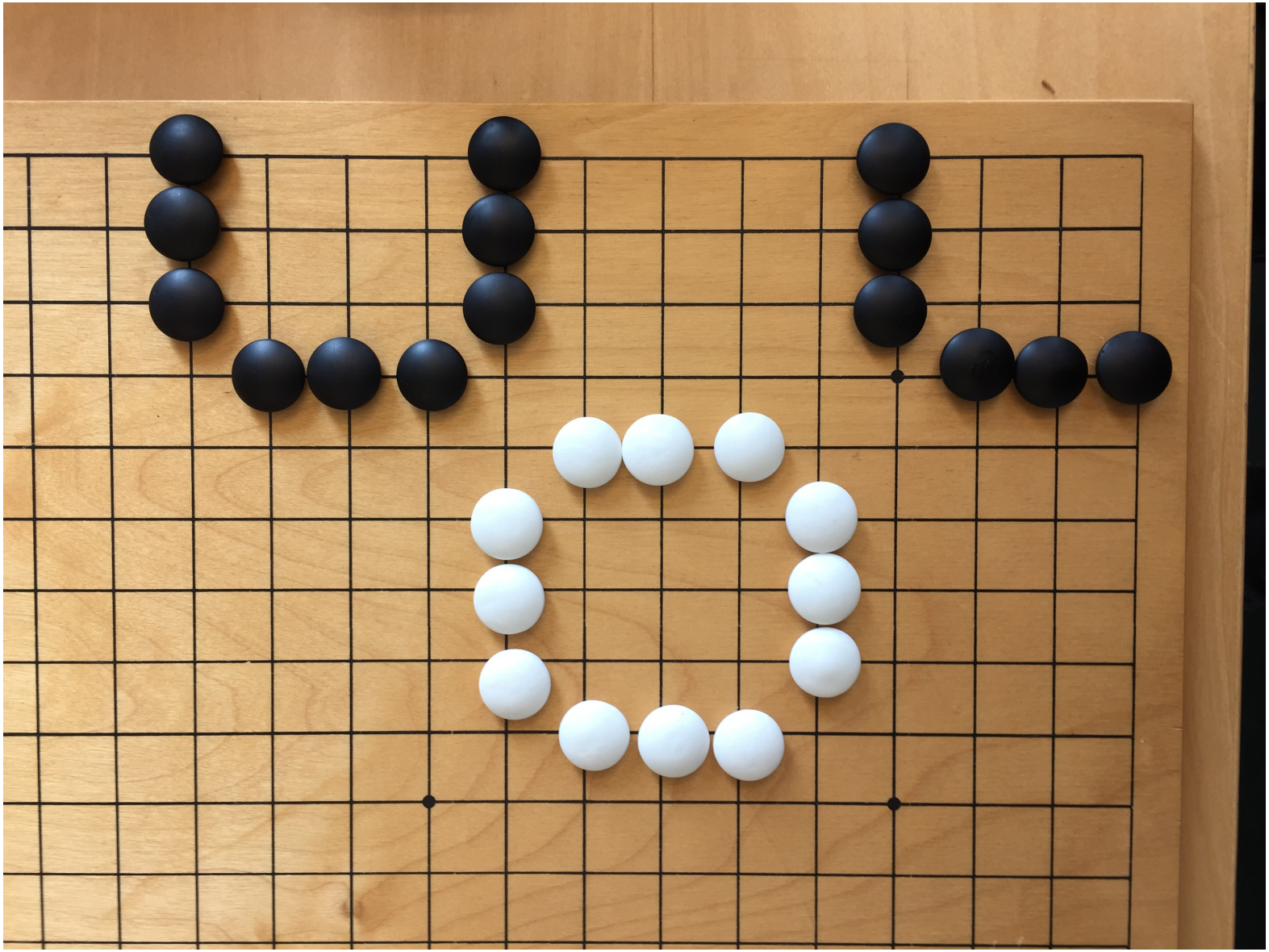
The game ends after two consecutive passes.

A player's score is the number of points of her color, plus the number of empty points that reach only her color. White gets 7.5 points extra.

The player with the higher score at the end of the game is the winner.

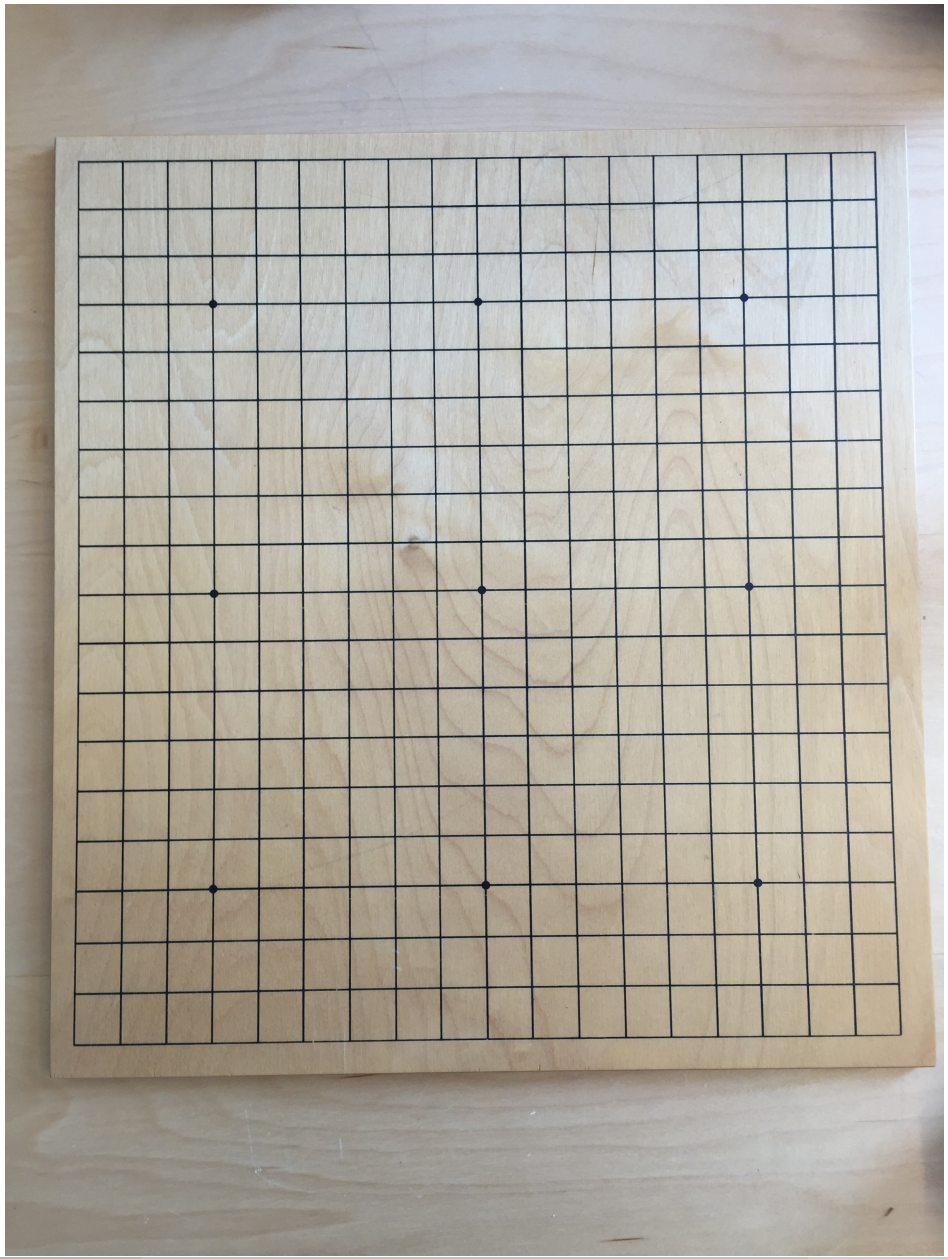
Aalto University

Easier to make territory in the corner



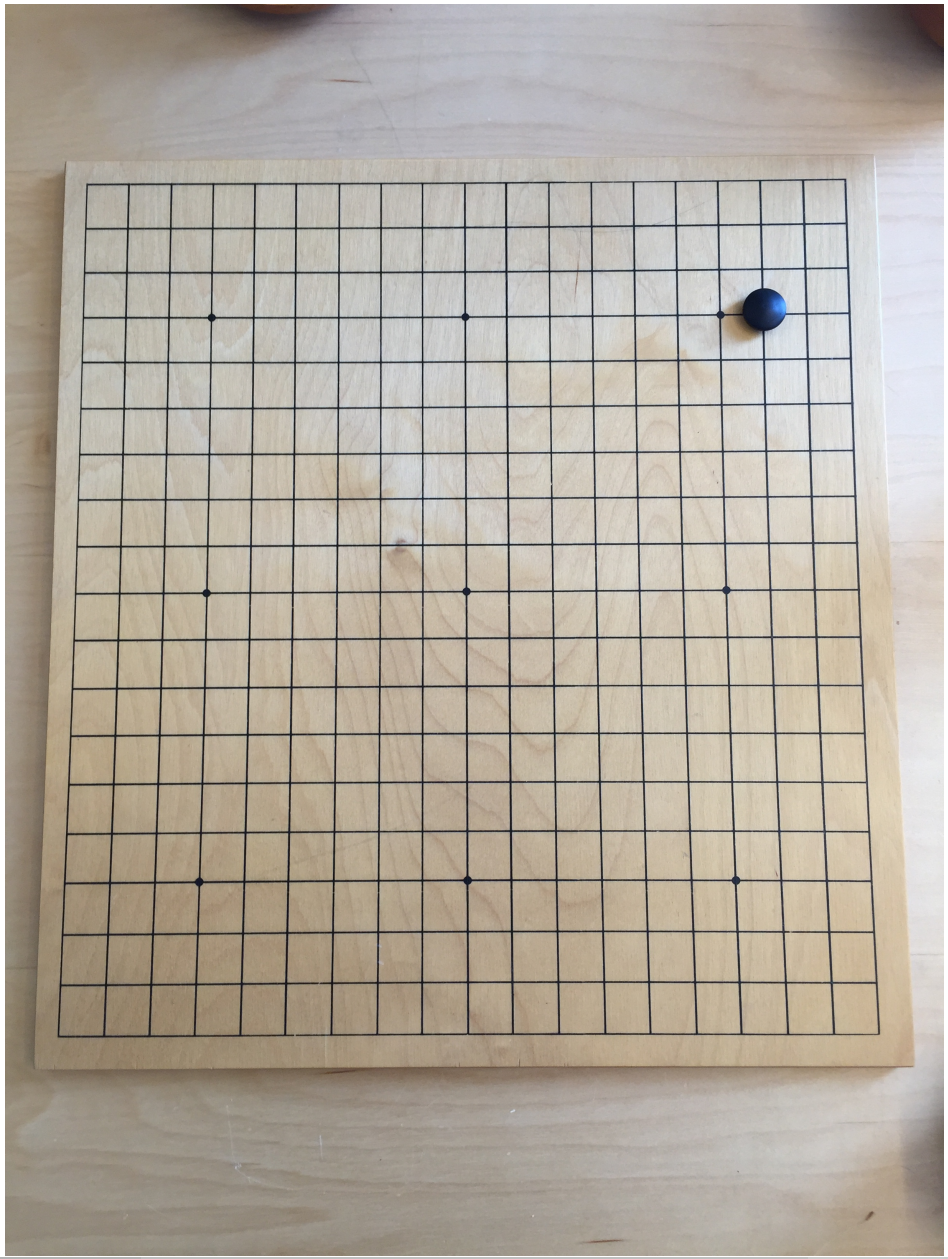
Aalto University

Opening



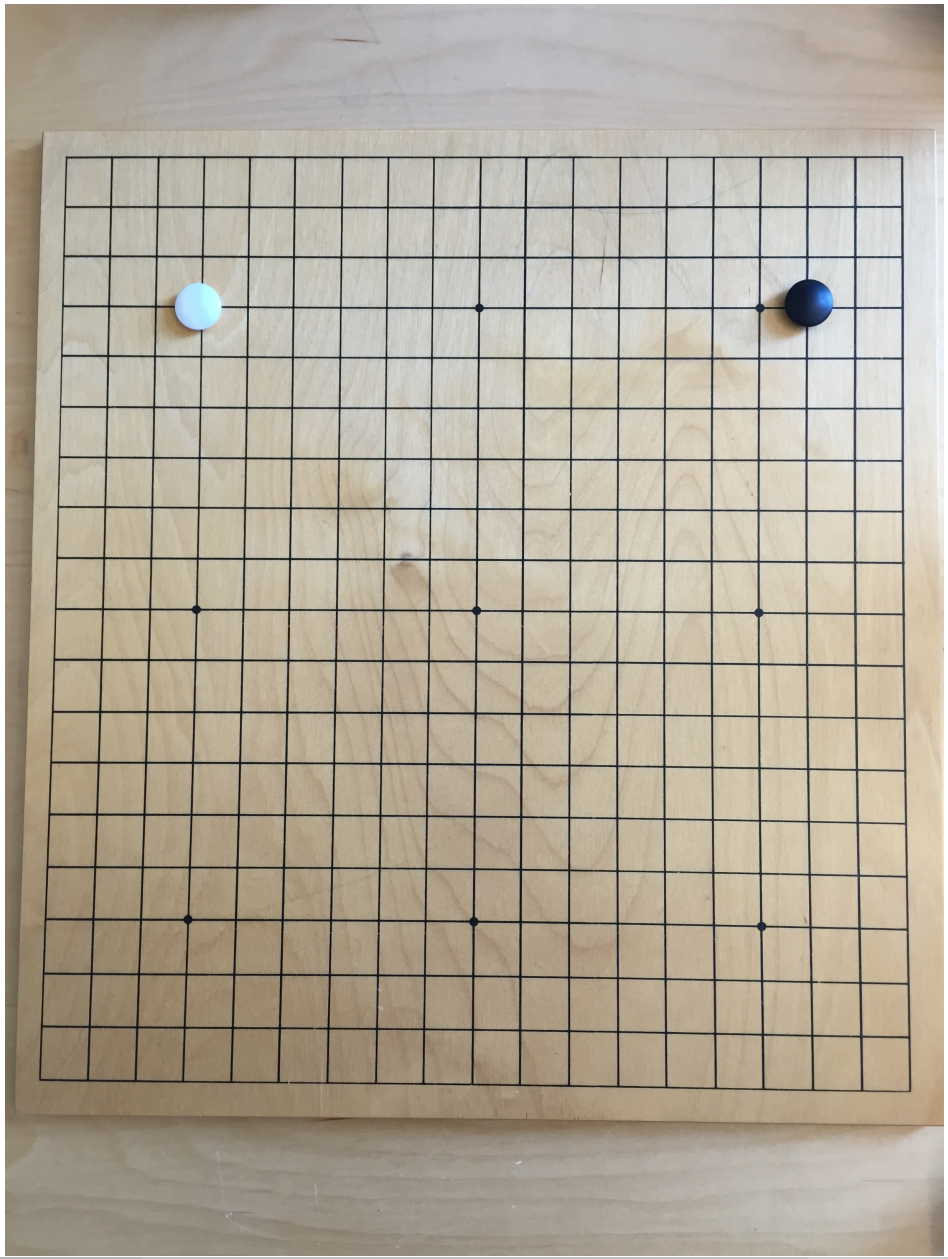
Aalto University

Opening



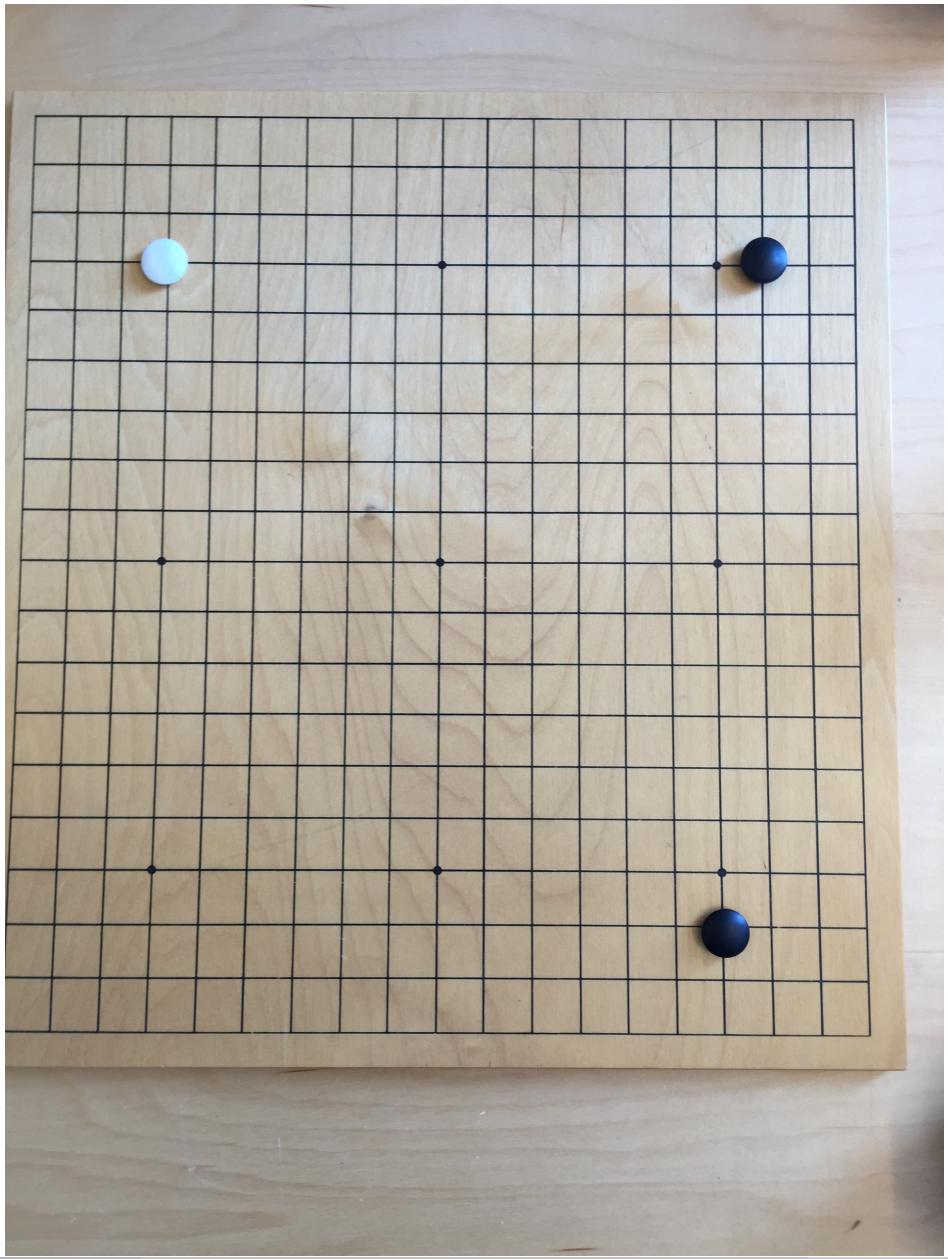
Aalto University

Opening



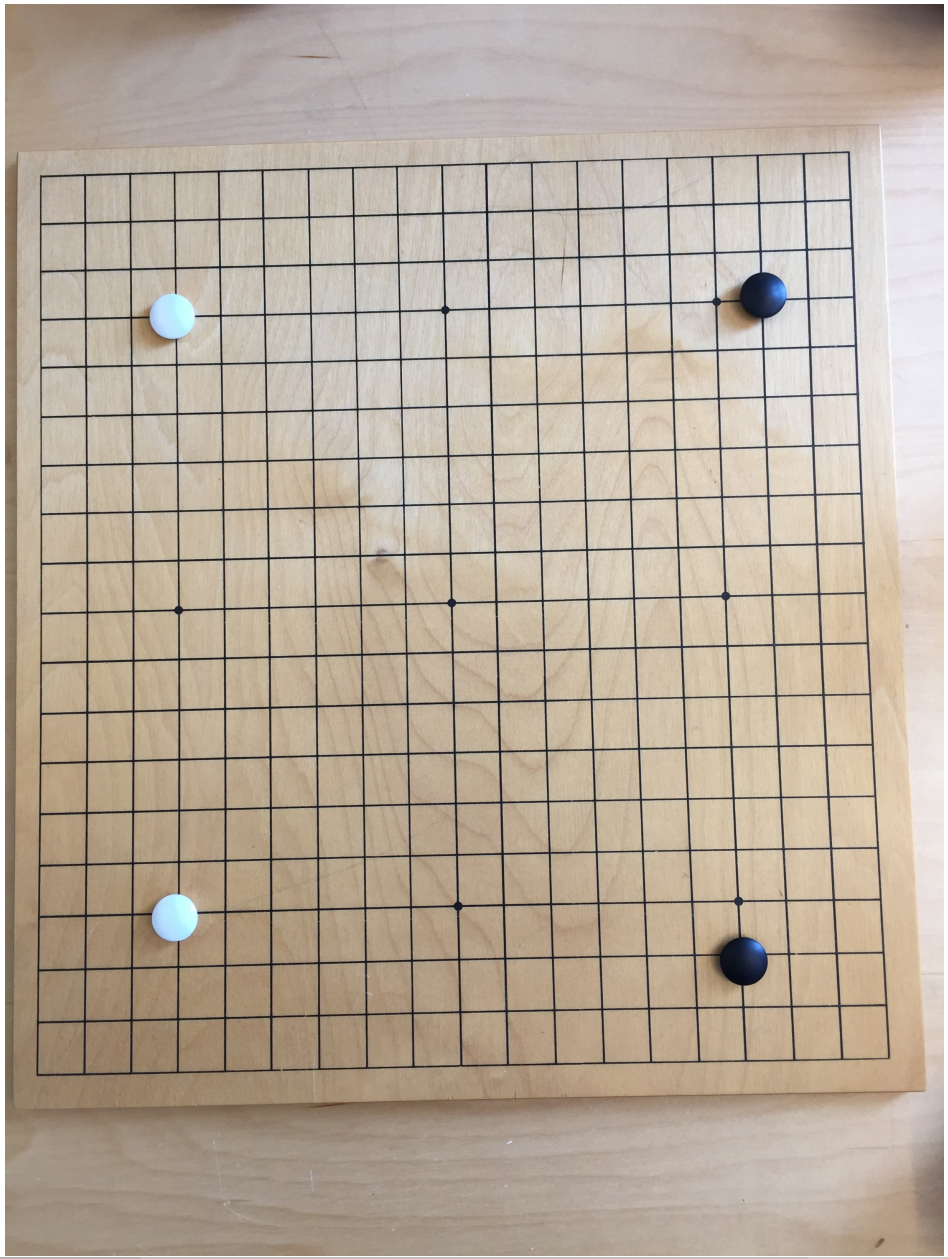
Aalto University

Opening



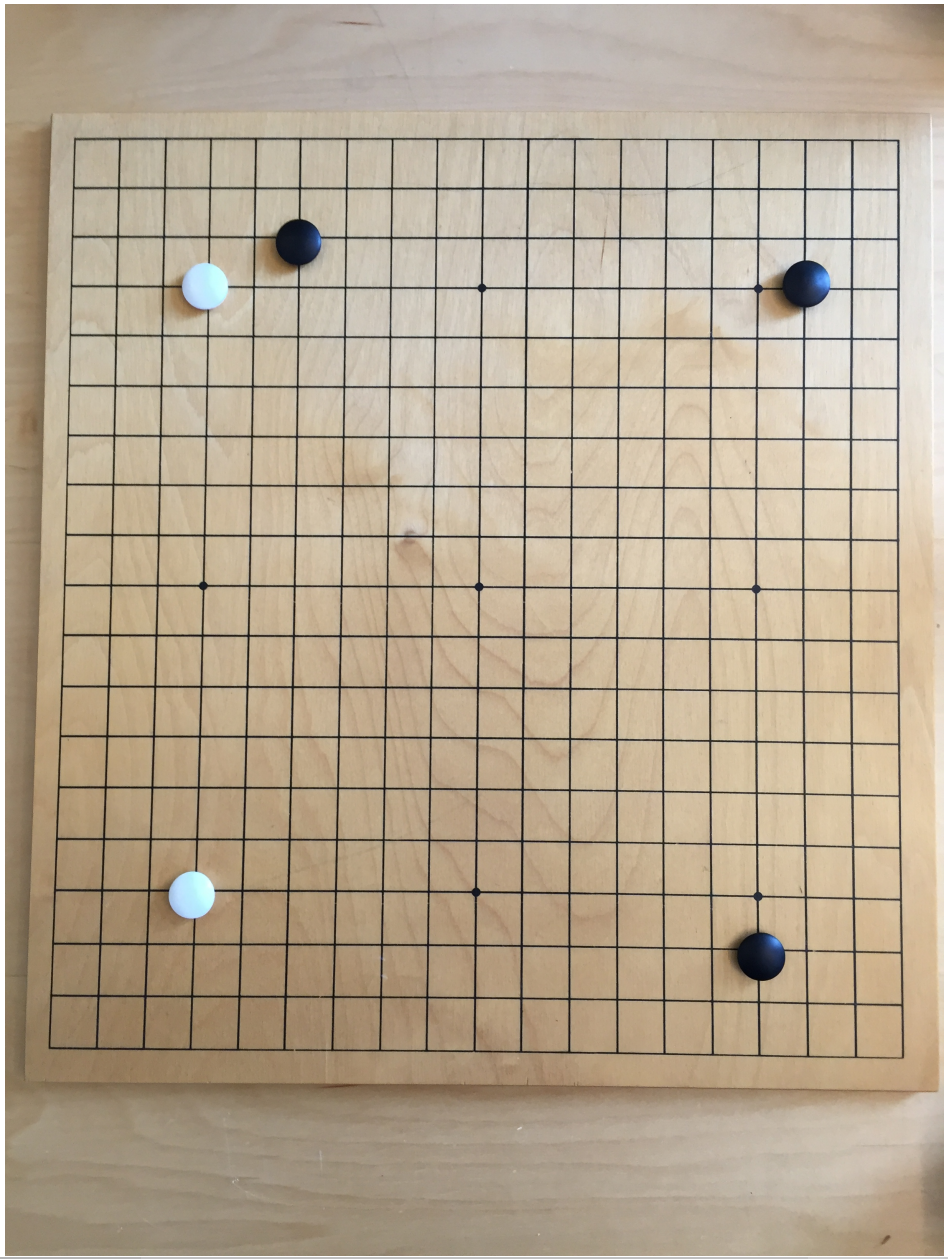
Aalto University

Opening



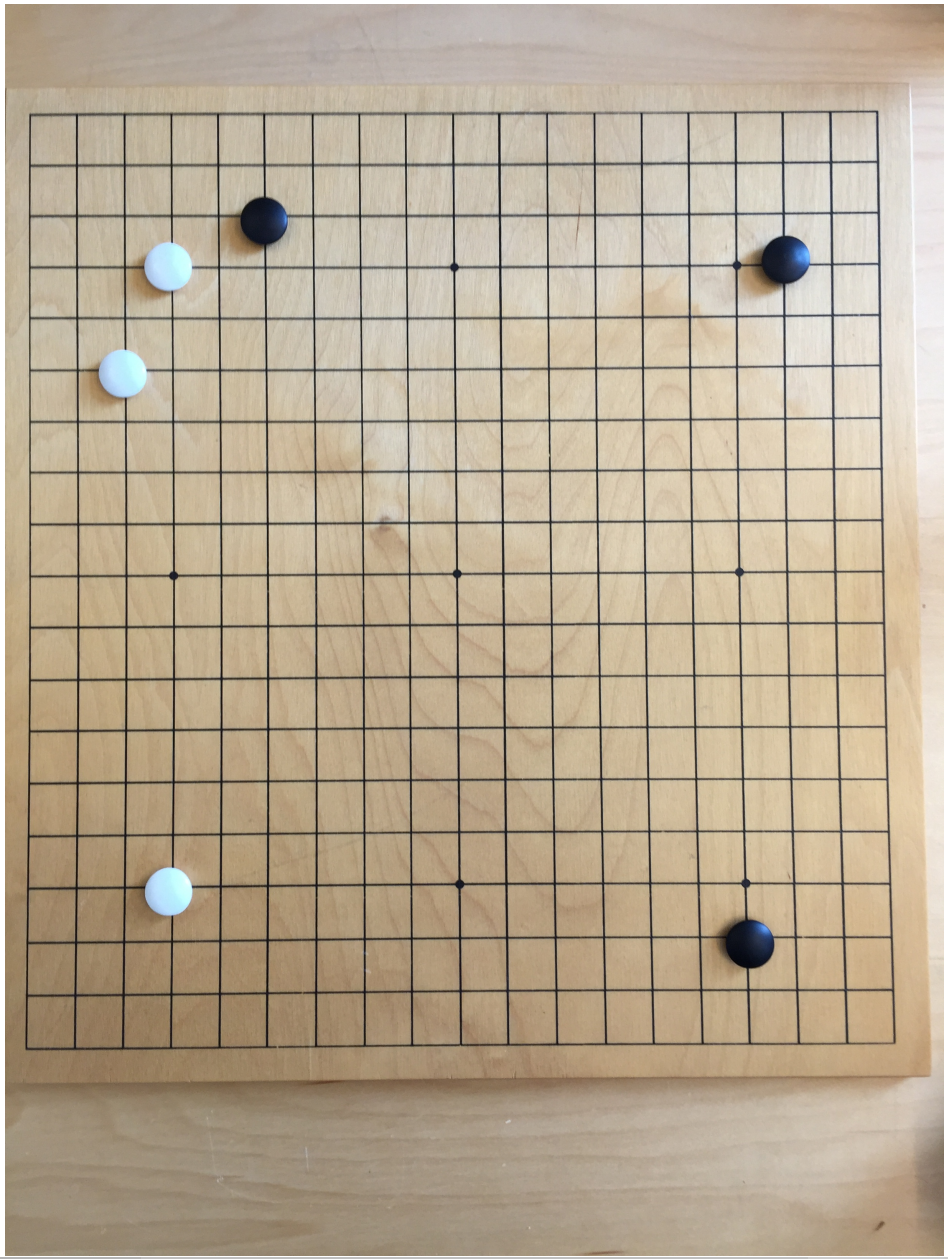
Aalto University

Opening



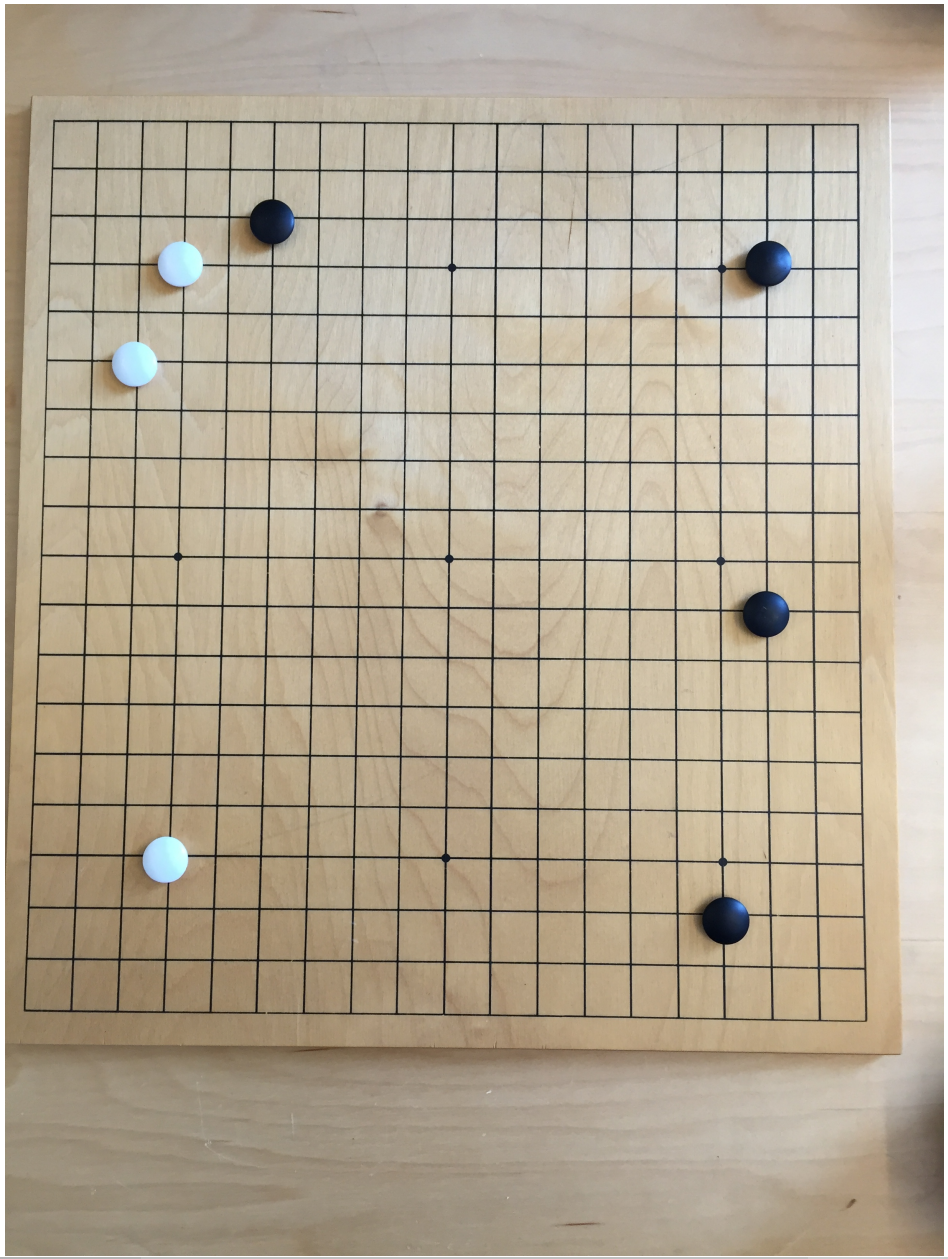
Aalto University

Opening



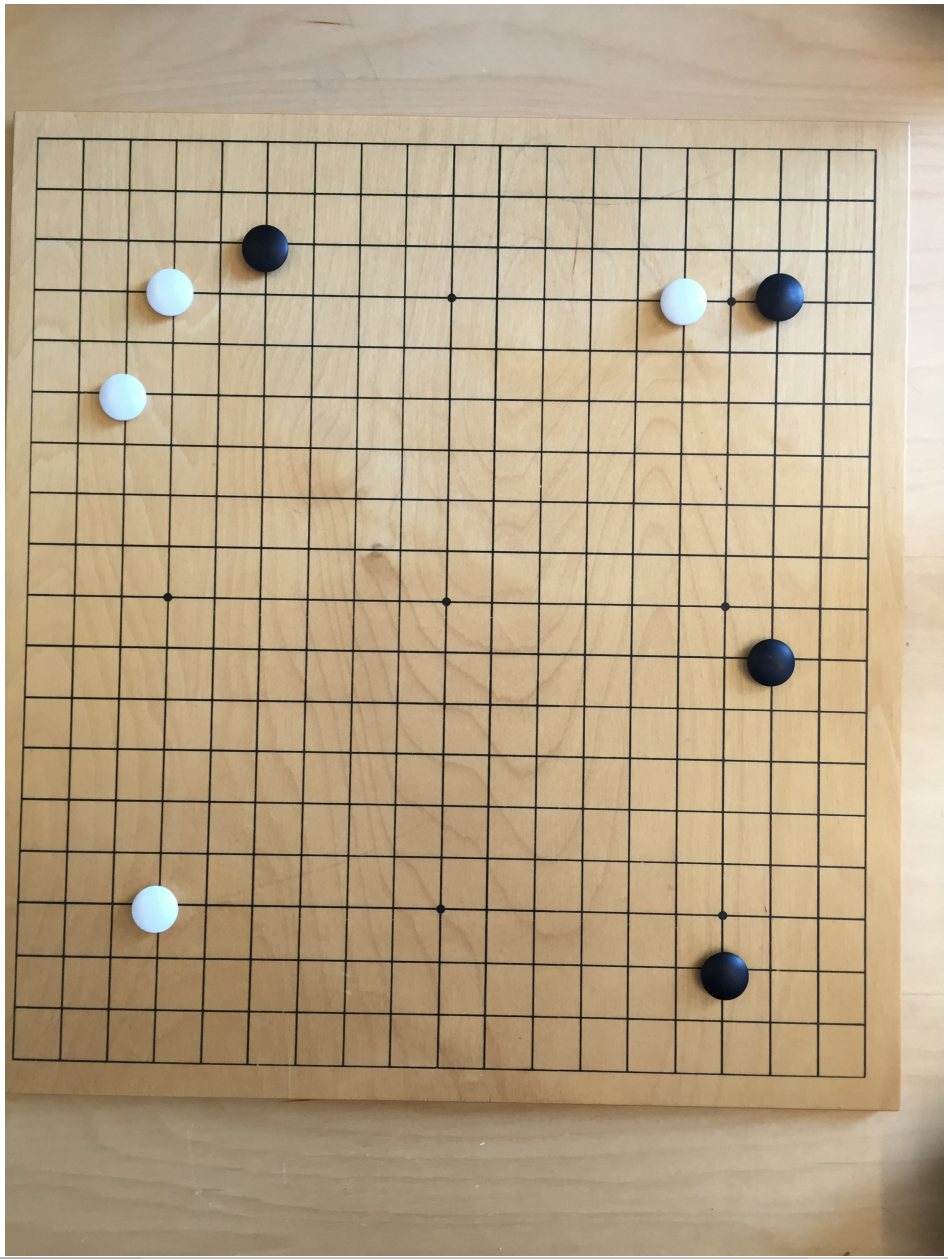
Aalto University

Opening



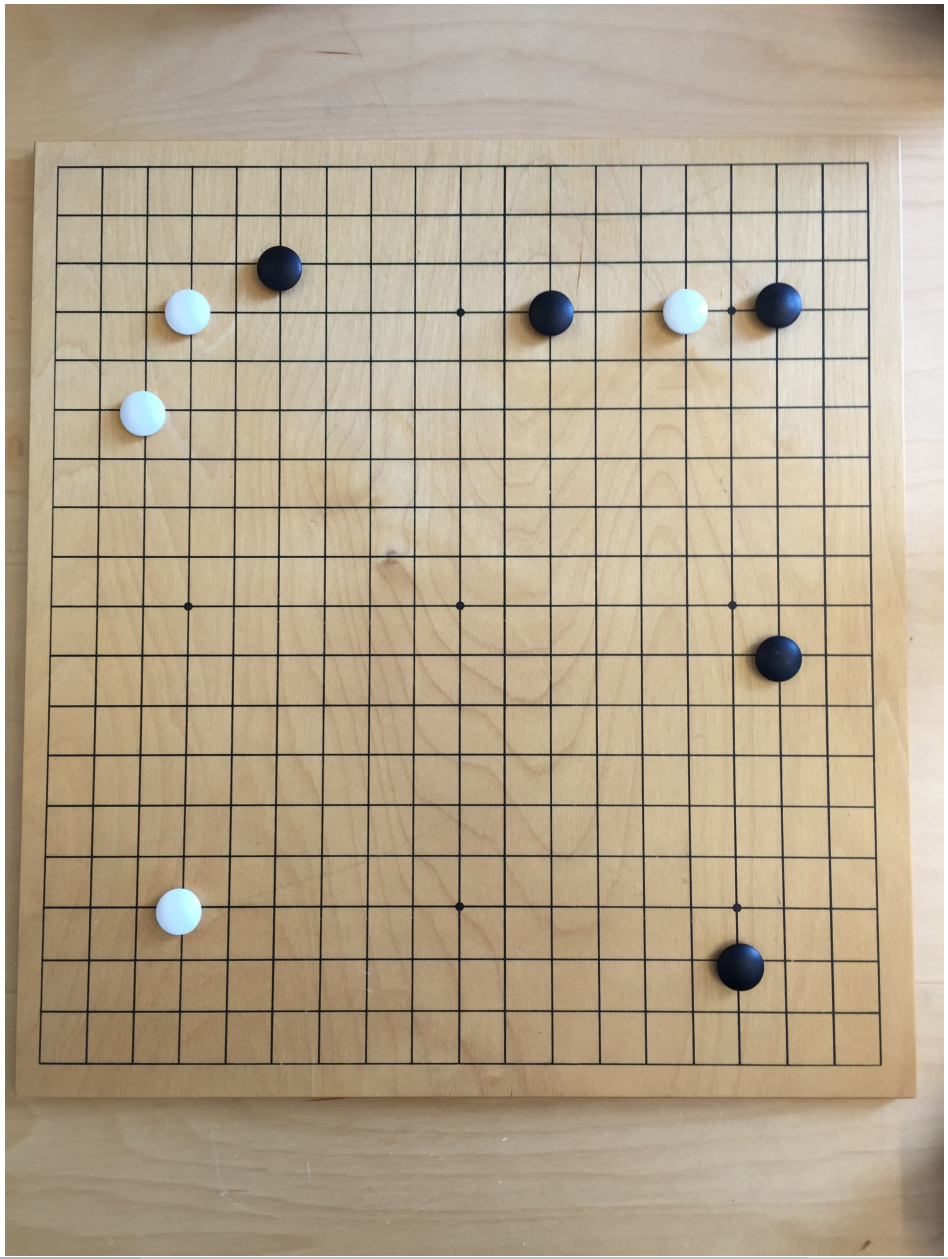
Aalto University

Opening



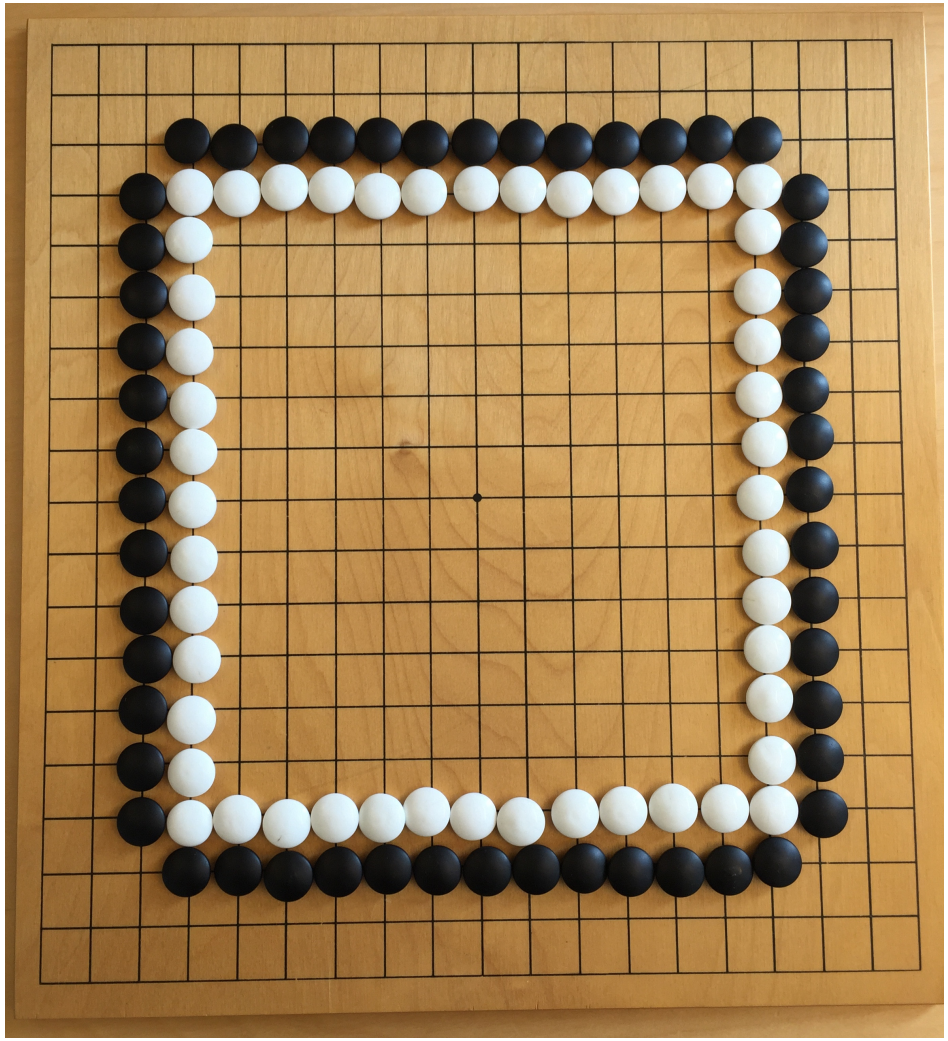
Aalto University

Opening



Aalto University

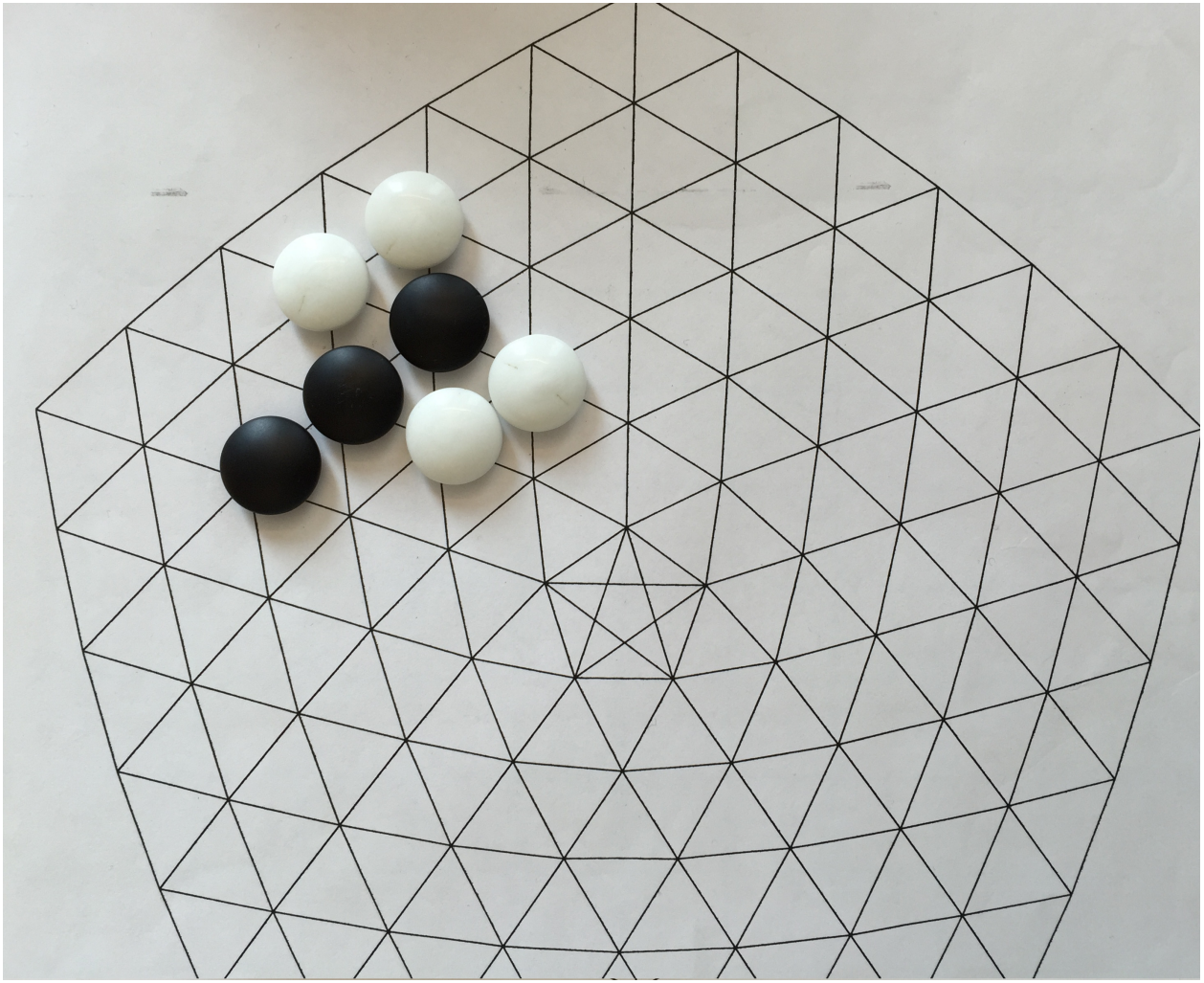
Why 19x19 grid?



Third and forth lines about equally important.

Aalto University

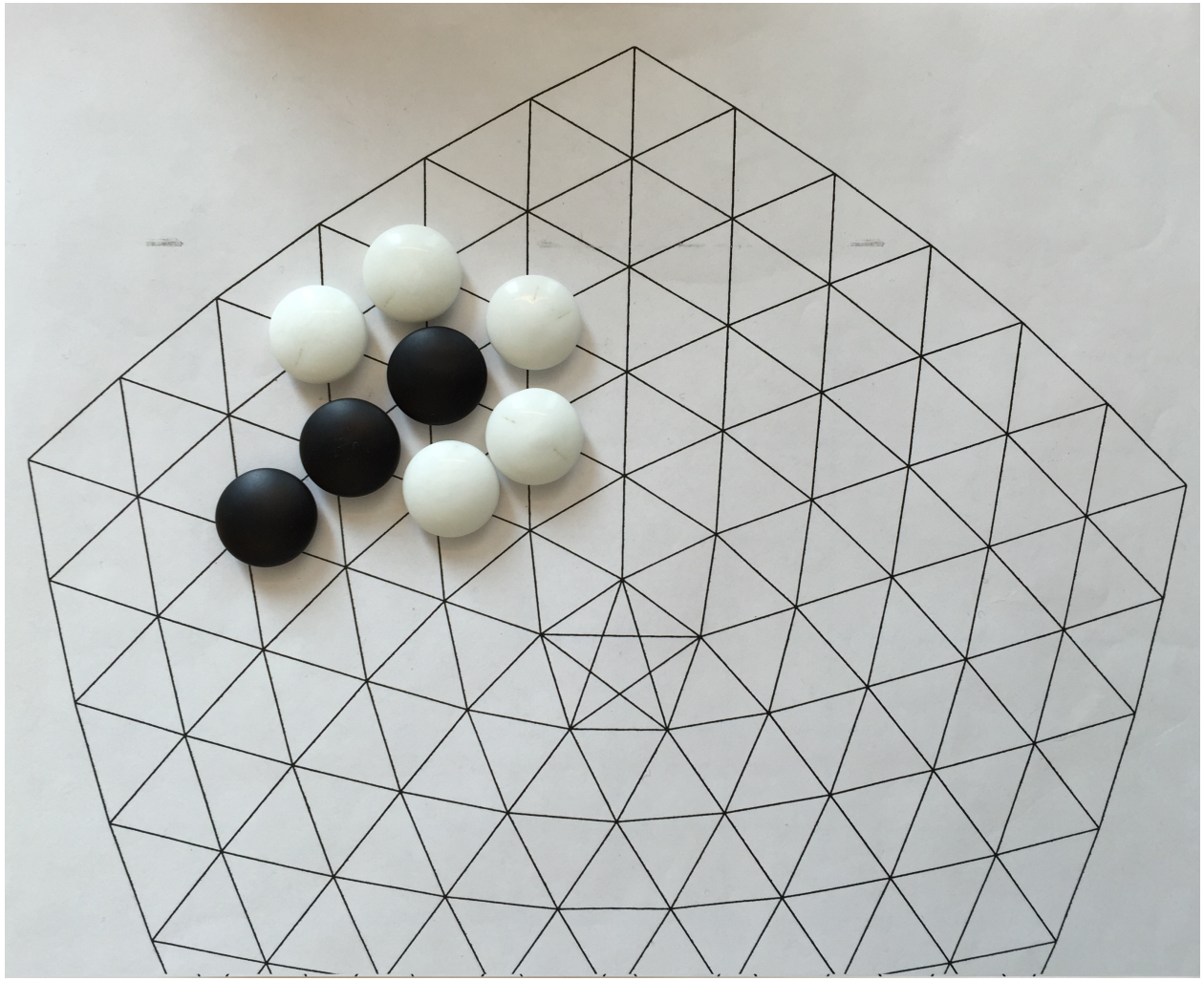
Why 19x19 grid?



If white blocks black...

Aalto University

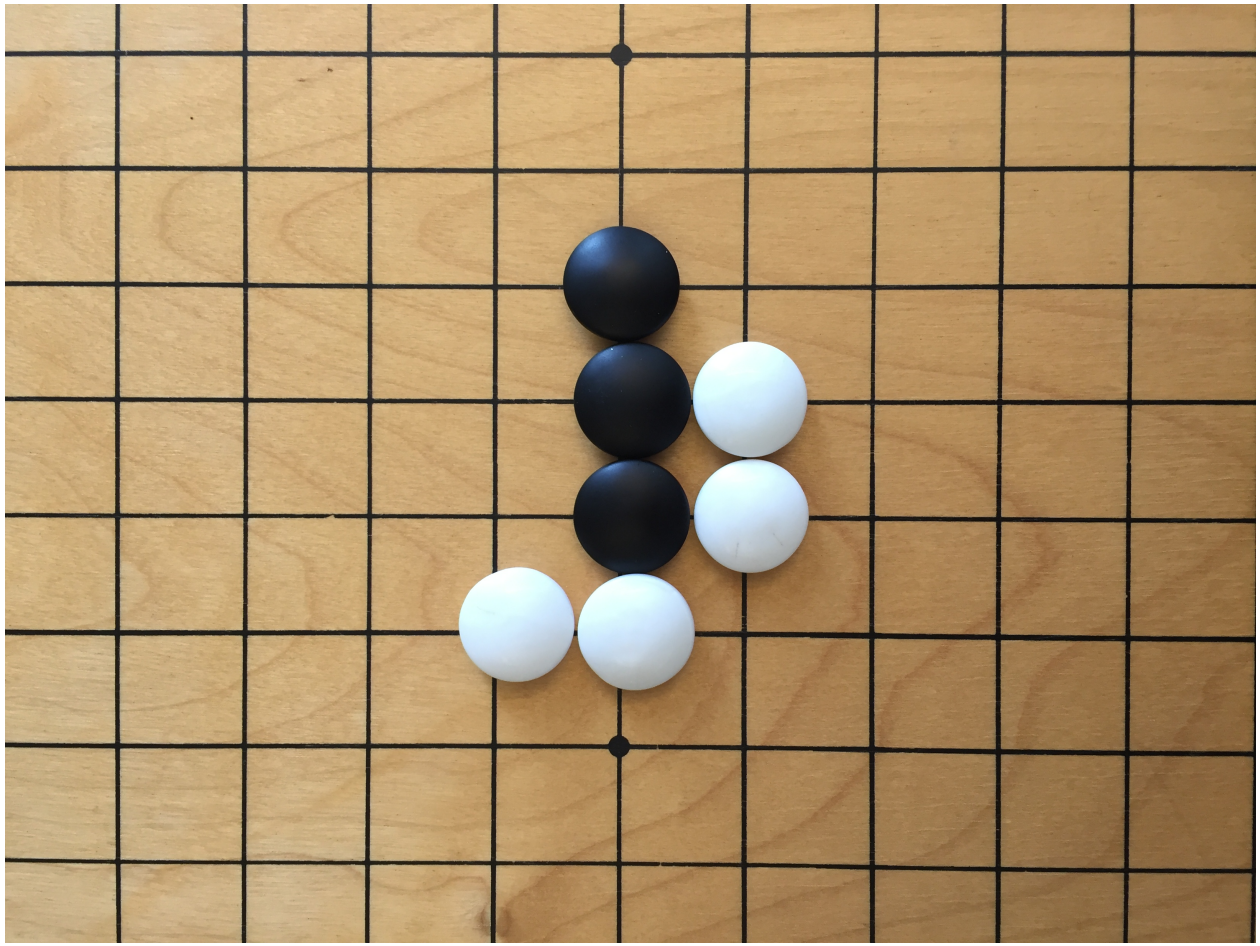
Why 19x19 grid?



...white is always connected.

Aalto University

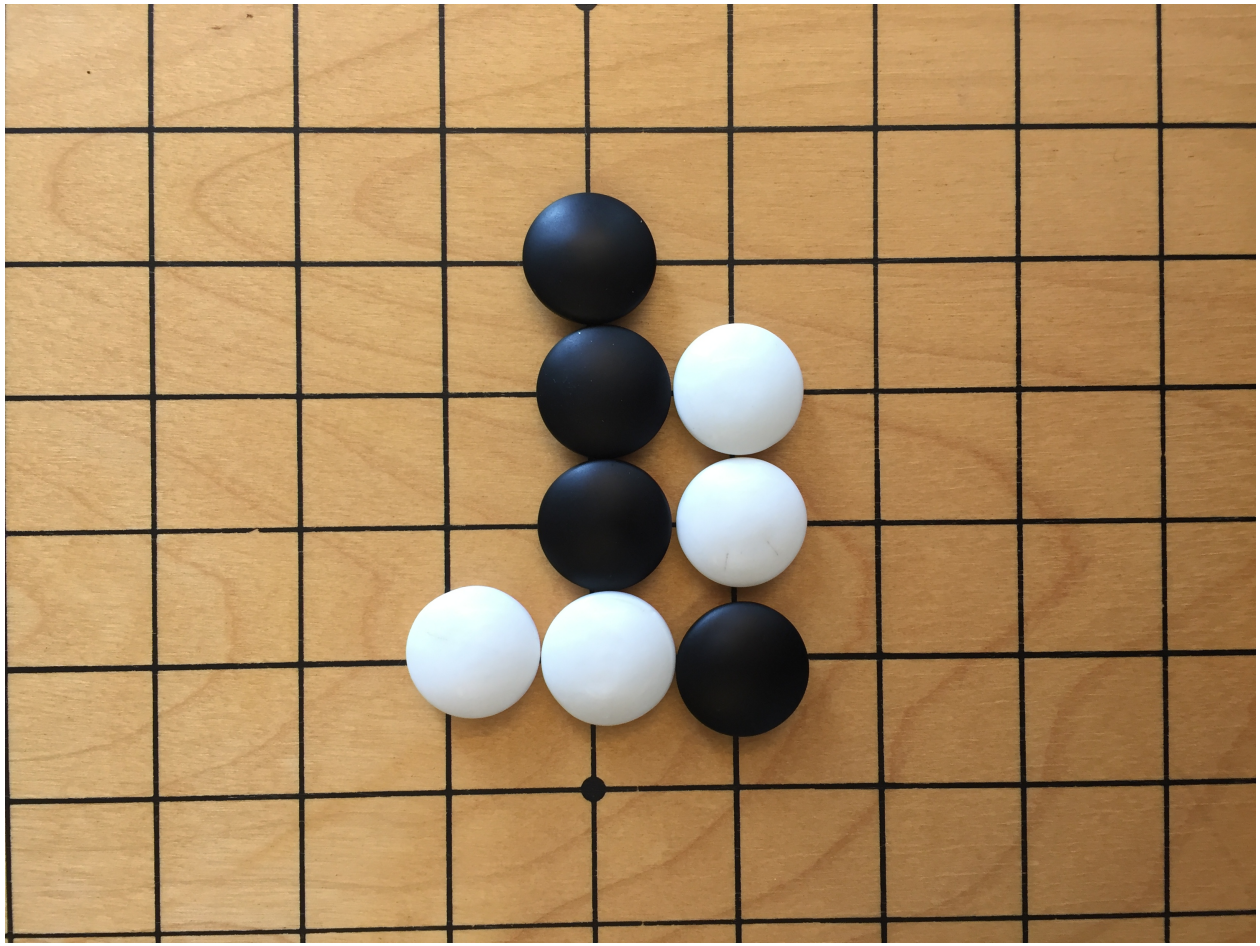
Why 19x19 grid?



White has blocked black.

Aalto University

Why 19x19 grid?



But black can cut. A fight will ensue.

Aalto University

Go should be in "easy" category of AI problems

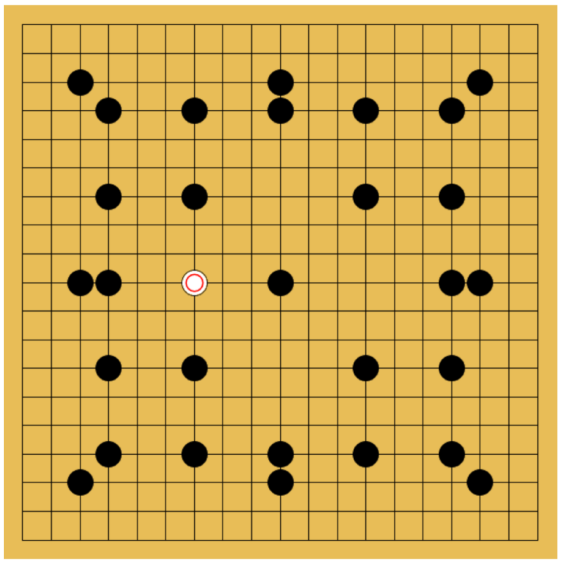
- **Fully observable** vs. partially observable
- Single agent vs. **multiagent**
- **Deterministic** vs. stochastic
- Episodic vs. **sequential**
- **Static** vs. dynamic
- **Discrete** vs. continuous
- **Known** vs. unknown

Aalto University

Go AI progress

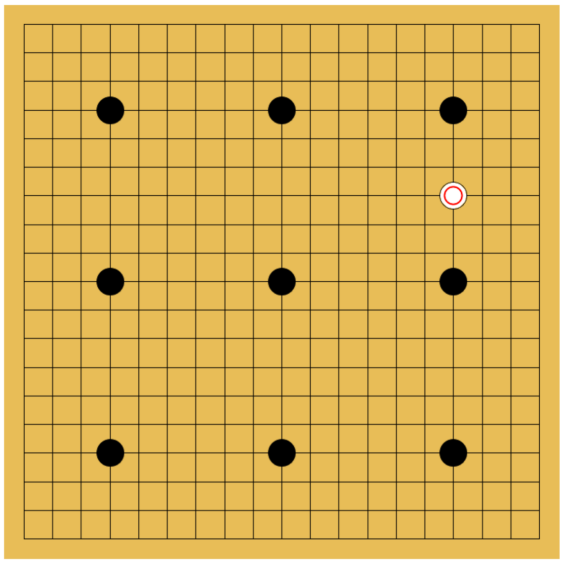
1998

29 stones handicap, human wins



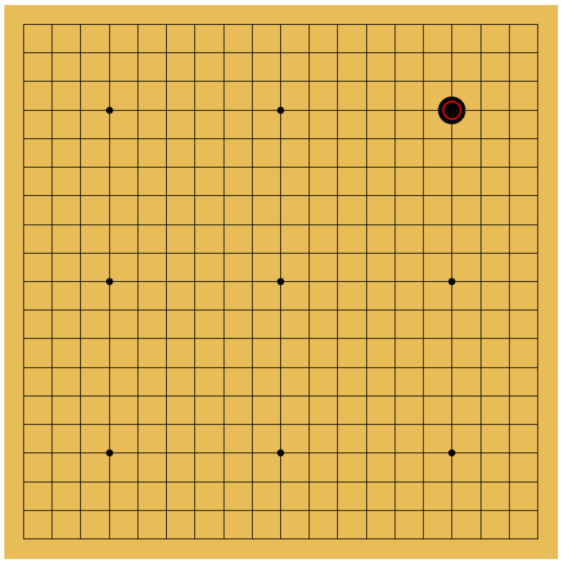
2008

9 stones handicap, computer wins



2016

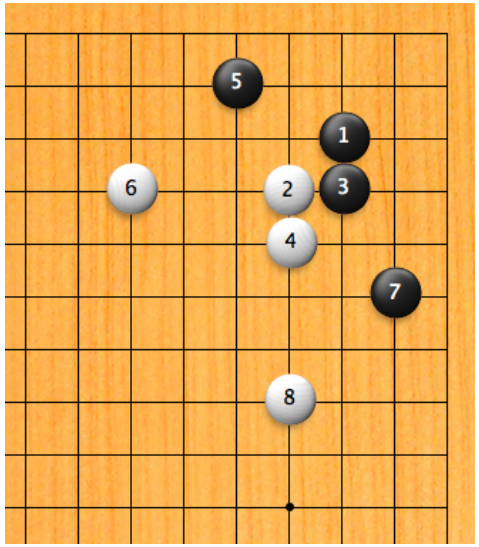
no handicap, computer wins



Aalto University

Why Go was difficult for AI

- Go is visual and thus easy for people



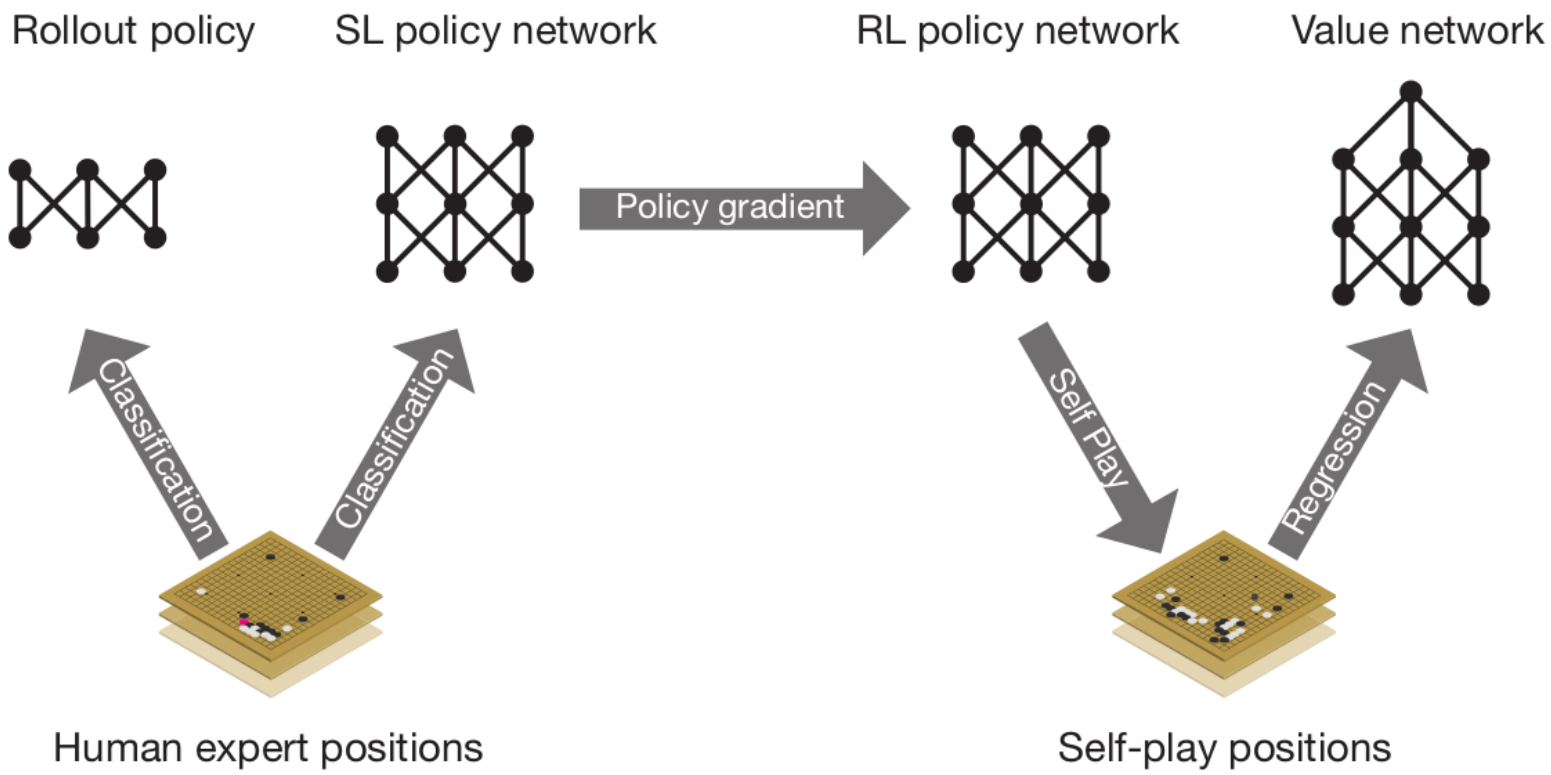
  Could not show 8 Chess moves in one image
- Branching factor is large (no brute force)
- Heuristic evaluation is difficult
- Horizon effect is strong (easy to delay capture)

Aalto University

Table of Contents

Aalto University

AlphaGo uses four networks



- Three of them output a probability distribution $p(a|s)$ over the next move a
- Last one evaluates the value $v(s)$ of a position

Aalto University

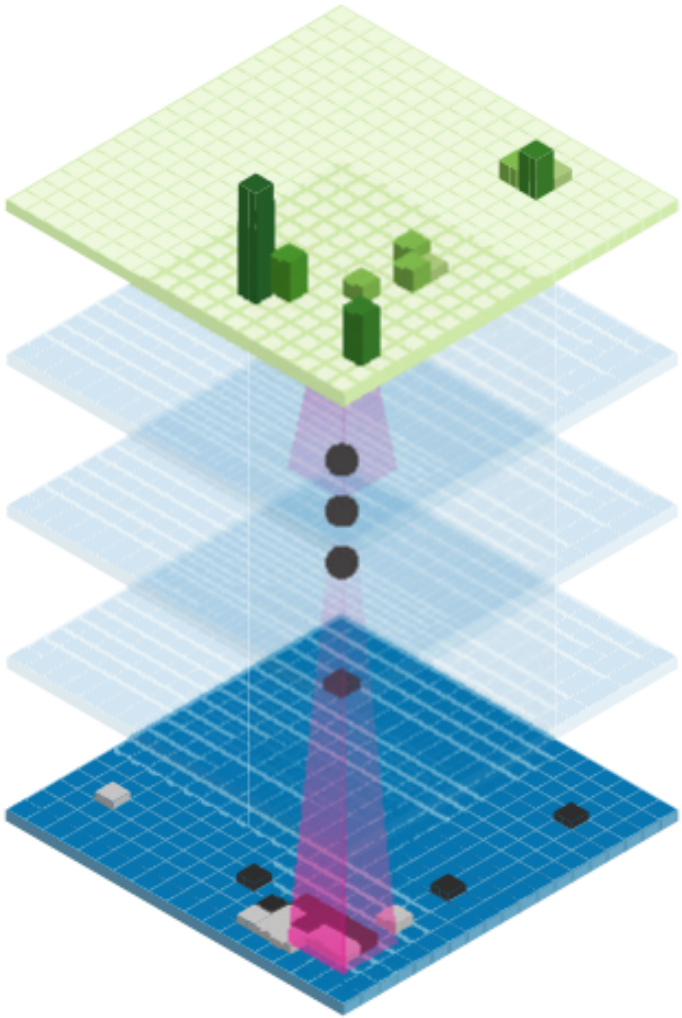
Supervised learning (nets 1&2)

- Predict human moves $p(a|s)$
- 160 000 game records from 6-9 dan amateurs
- Gradient-based learning: $\Delta\boldsymbol{\theta} \propto \frac{\partial \log p(a|s)}{\partial \boldsymbol{\theta}}$
- *Rollout policy* uses handcrafted features and no hidden units (2μs to evaluate)
- *SL policy network* uses a convolutional network (4.8ms to evaluate, 57% prediction accuracy)

(Sutskever & Nair, 2008) (Maddison et al., 2015) (Clark & Storkey, 2015)

Aalto University

Convolutional architecture for policy nets (2&3)



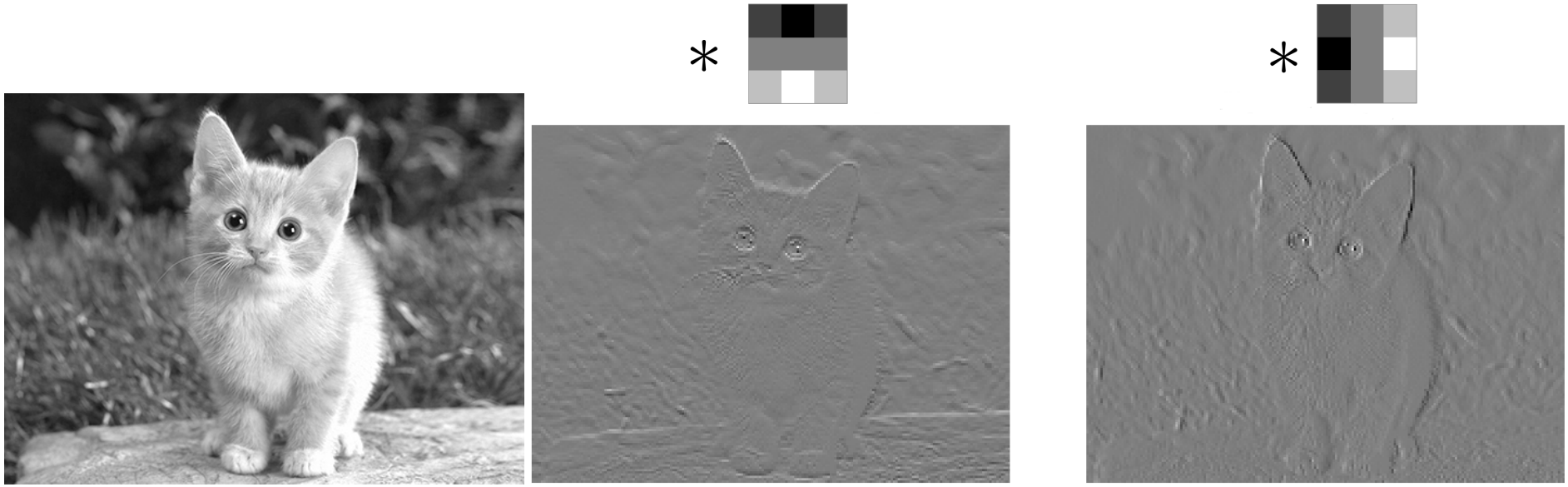
- Output: probability of next move a in each location
- Input: board state s as $19 \times 19 \times 48$, 5×5 filters
- 12 layers of $19 \times 19 \times 192$, 3×3 filters

  zero padding, stride 1
- ReLU activations

  softmax at the end
- About 4 million params θ

Aalto University

Convolution $*$



Convolutional networks are used in computer vision
(LeCun et al., 1989)

Reinforcement learning of policy net (3)

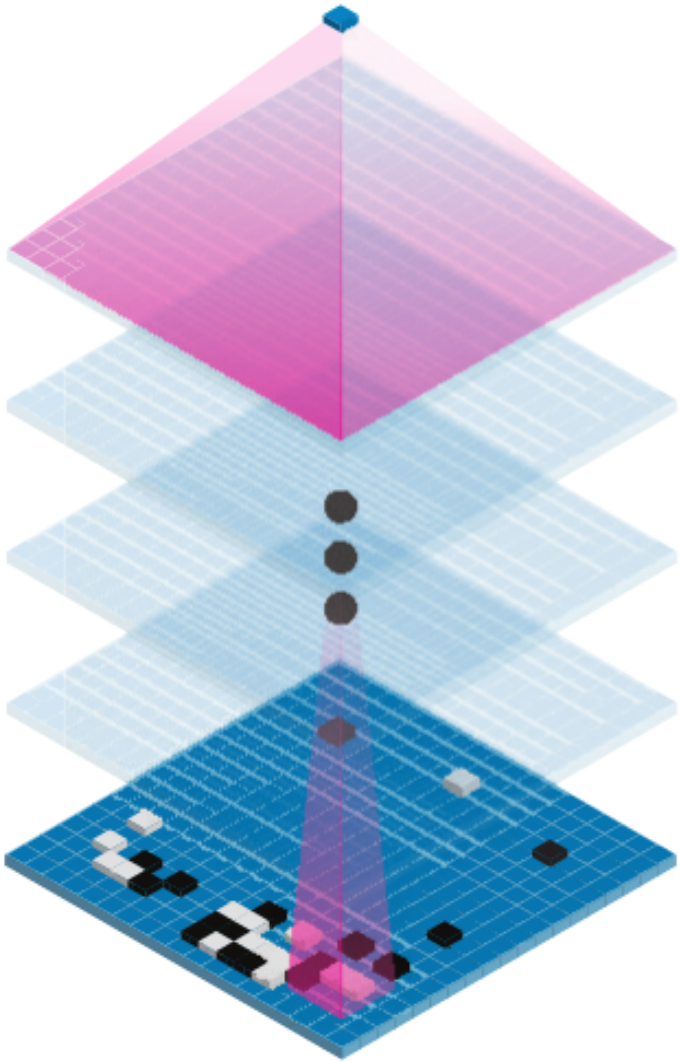
- Initialize by net 2, learn from self-play
  $\geq$ 30 million games
- Policy gradient (Williams, 1992, Sutton et al., 2000):

$$\Delta\boldsymbol{\theta} \propto \frac{\partial \log p(a_t|s_t)}{\partial\boldsymbol{\theta}} z_t$$

  where $z_t \in \{-1, 1\}$ is the outcome of the game
- This reaches 3 dan amateur strength
  (without any look-ahead, 4.8ms per move)

Aalto University

Convolutional architecture for value net (4)



- Output: expected outcome $v_t = \mathbb{E}[z_t]$ of the game from state s with RL policy
- Similar architecture, but a fully connected layer with 256 ReLUs and single output with tanh at the top

Aalto University

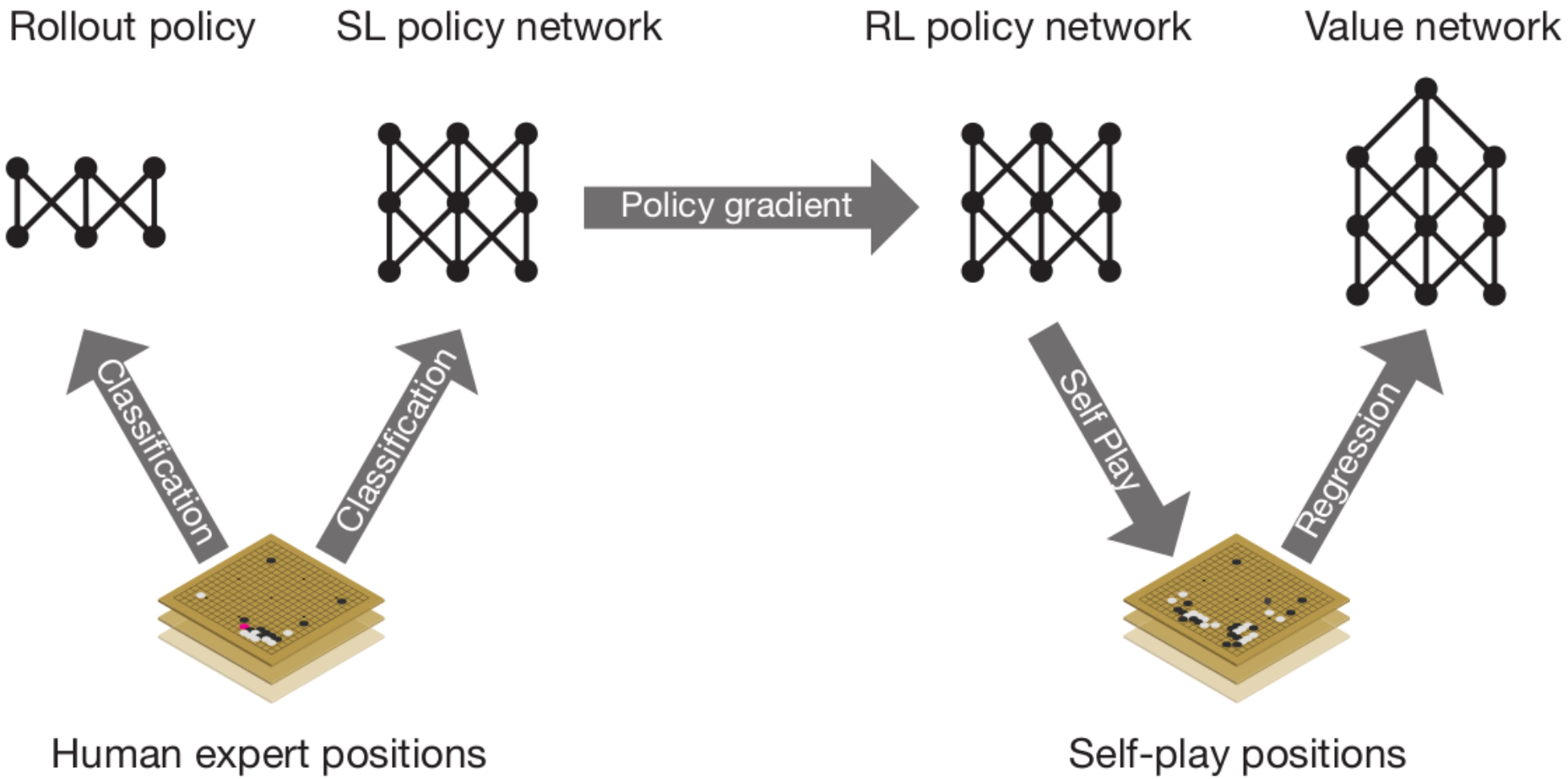
Reinforcement learning of value net (4)

- Self-play (net 3) to generate ≥ 30 million positions, each from different game
- Regression:

$$\Delta\boldsymbol{\theta} \propto \frac{\partial v(s_t)}{\partial \boldsymbol{\theta}}(z_t - v(s_t))$$

where $z_t \in \{-1, 1\}$ is the outcome of the game, $v(s_t)$ is the predicted outcome

Aalto University



Aalto University

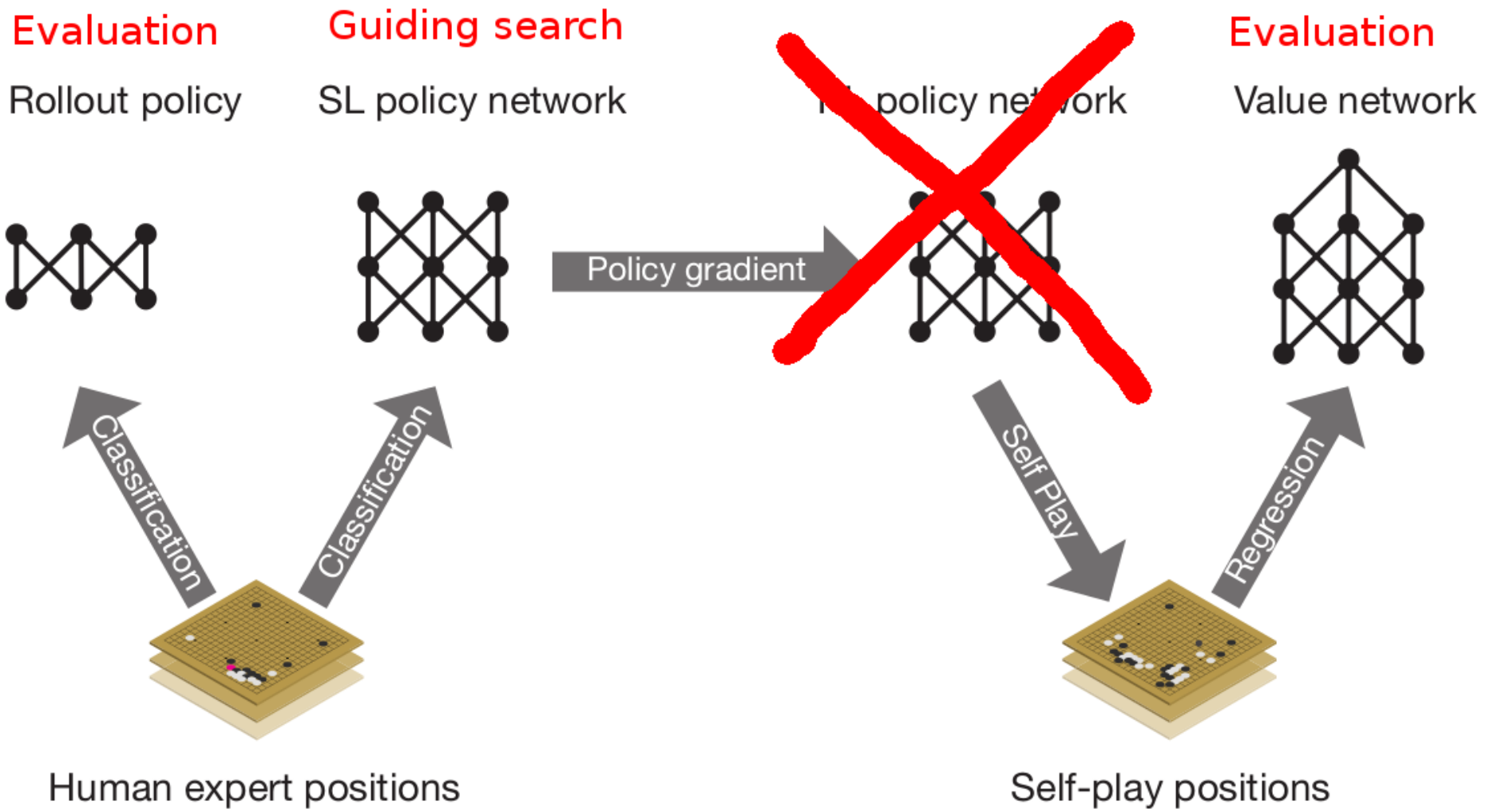

Evaluation
Rollout policy
Guiding search
SL policy network
Policy gradient
Evaluation
Value network
Classification
Classification
Self Play
Regression
Human expert positions
Self-play positions

Table of Contents

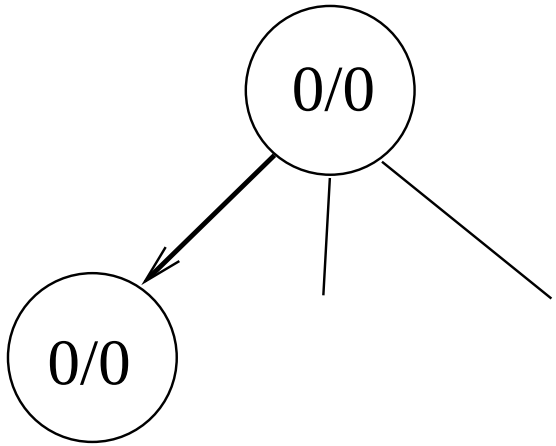
Aalto University

Monte Carlo tree search (Coulom, 2006)

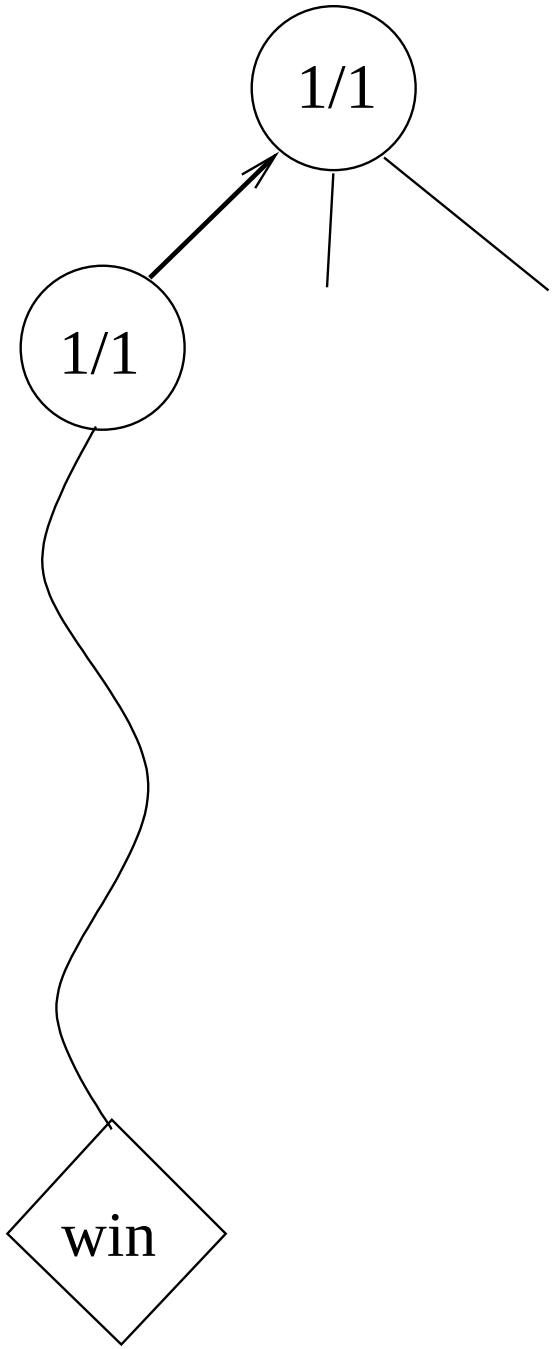
- Self-play from current state s
- $\sim$17 000 simulations per second
- Each path adds a new leaf node to search tree
- Not full-width search
- Final choice is move that was chosen most times

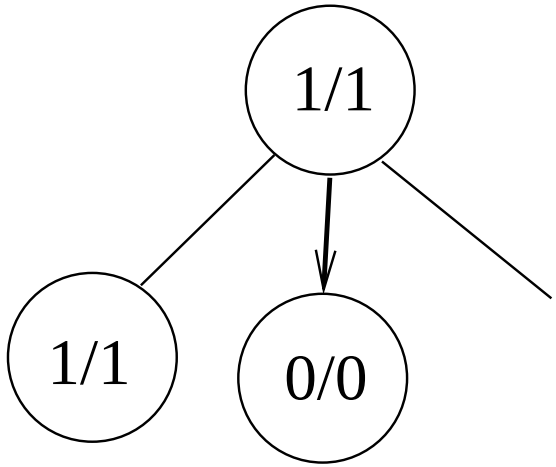
Aalto University

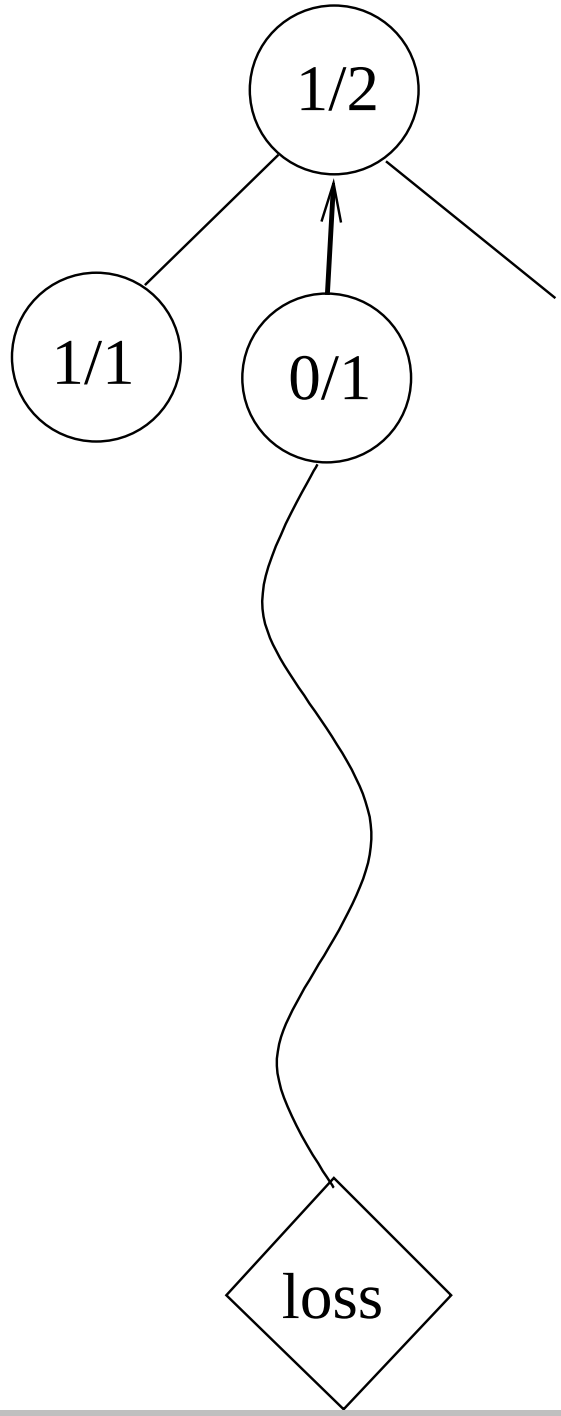
Example (single agent)

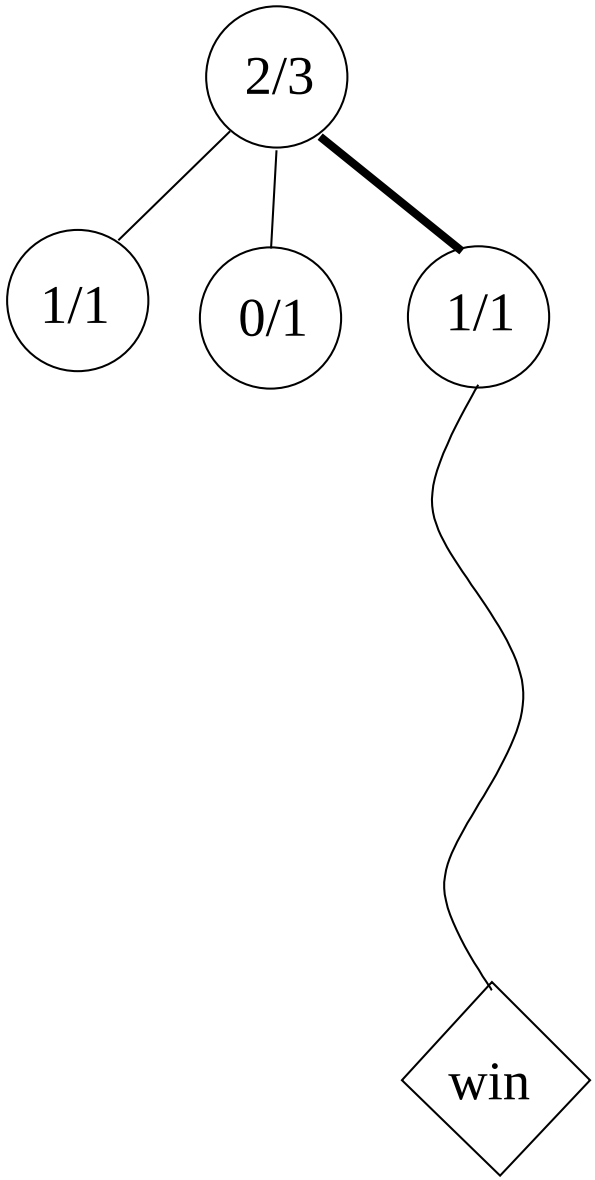


- Show number of wins/trials for each node

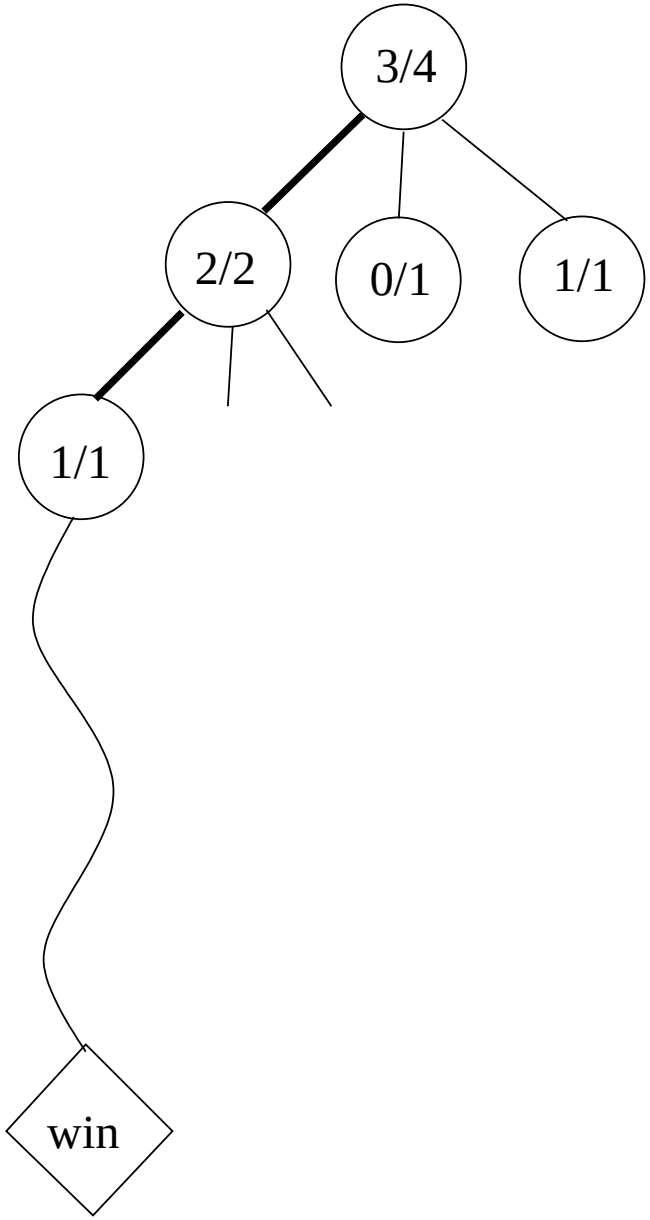

1/1
1/1
win


1/1
1/1
0/0

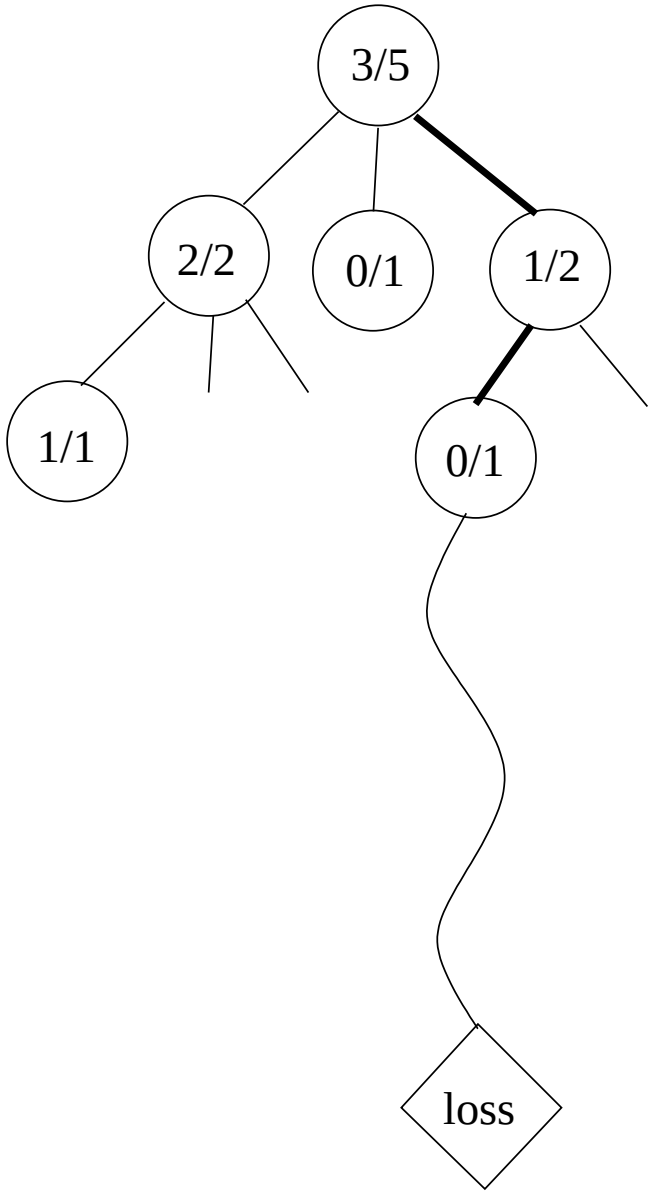

1/2
1/1
0/1
loss

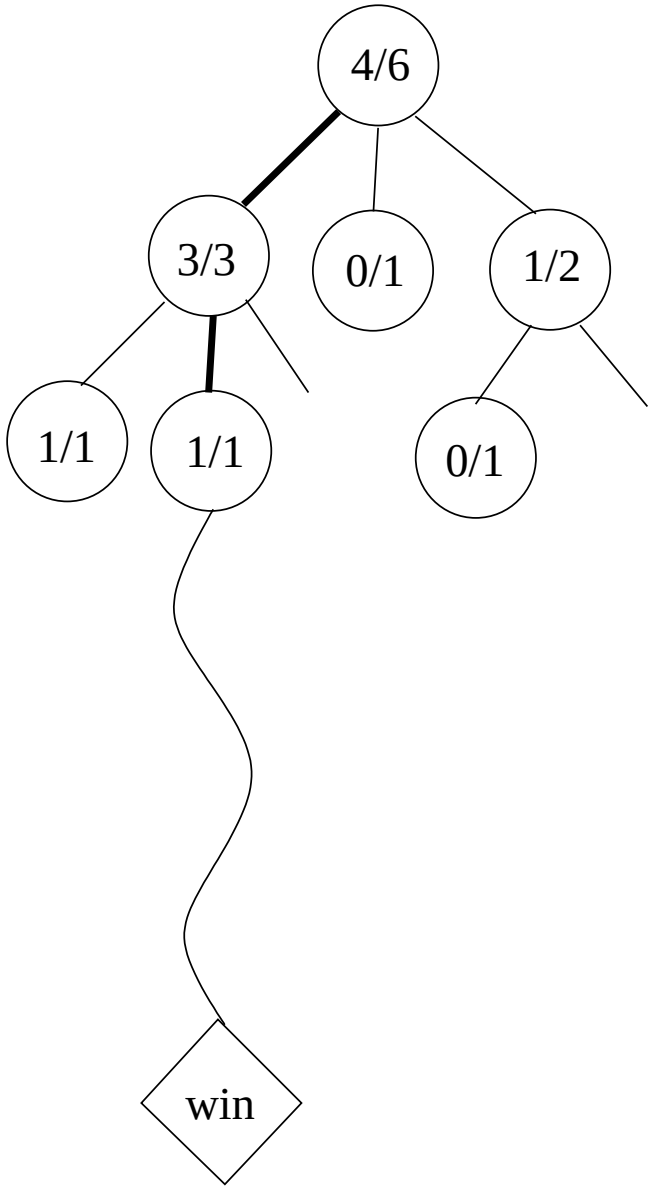


Aalto University

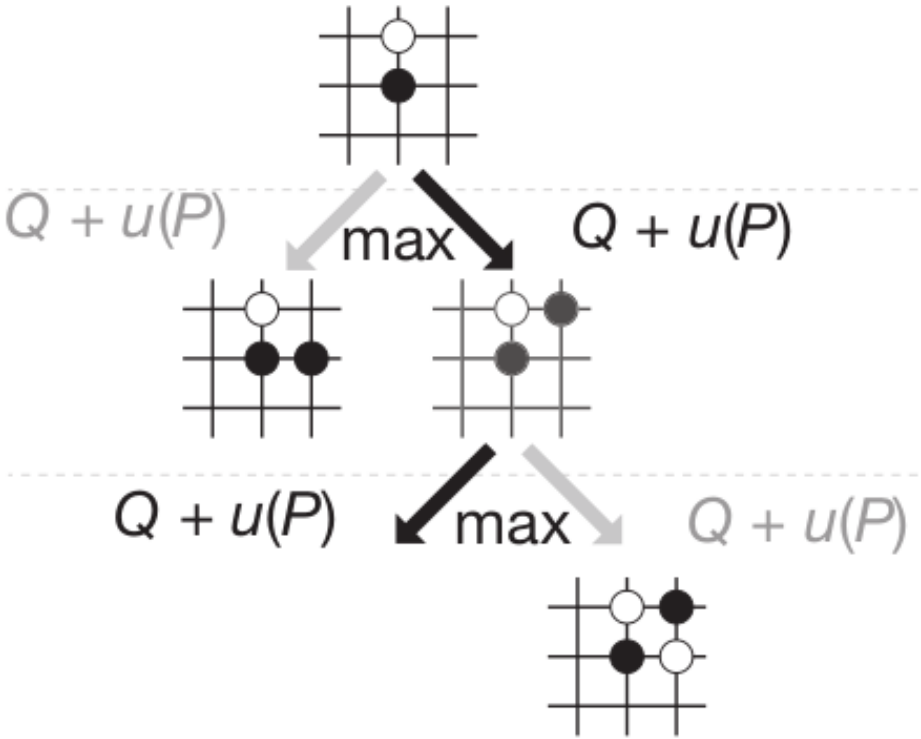


Aalto University


3/5
2/2
0/1
1/2
1/1
0/1
loss


4/6
3/3
0/1
1/2
1/1
1/1
0/1
win

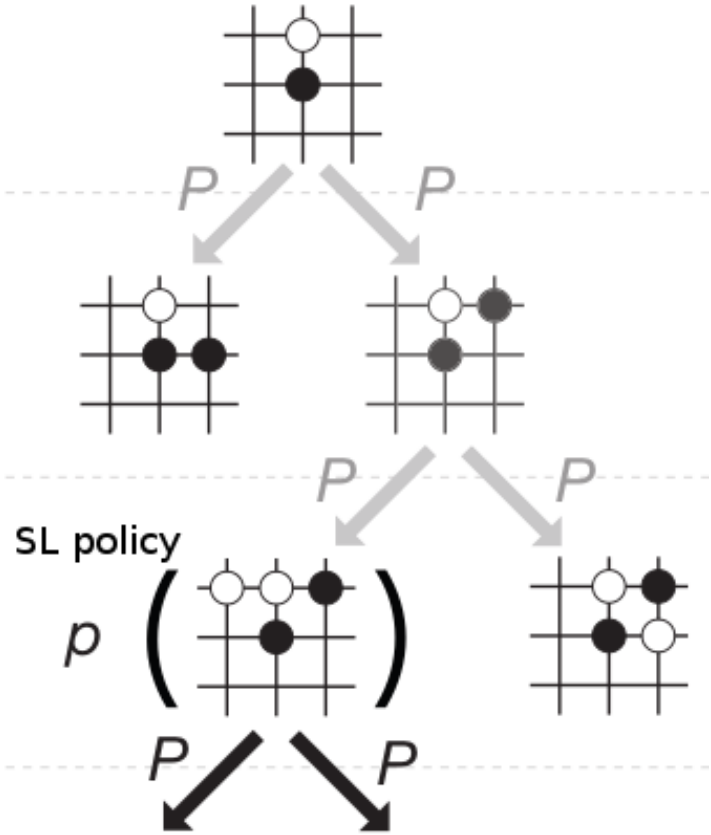
Step 1/4: Selection



Choose action by $\arg\max_a Q(a|s) + \frac{5p(a|s)}{1+\text{count}(a|s)}$
where $p(a|s)$ is from SL policy net (2)

Aalto University

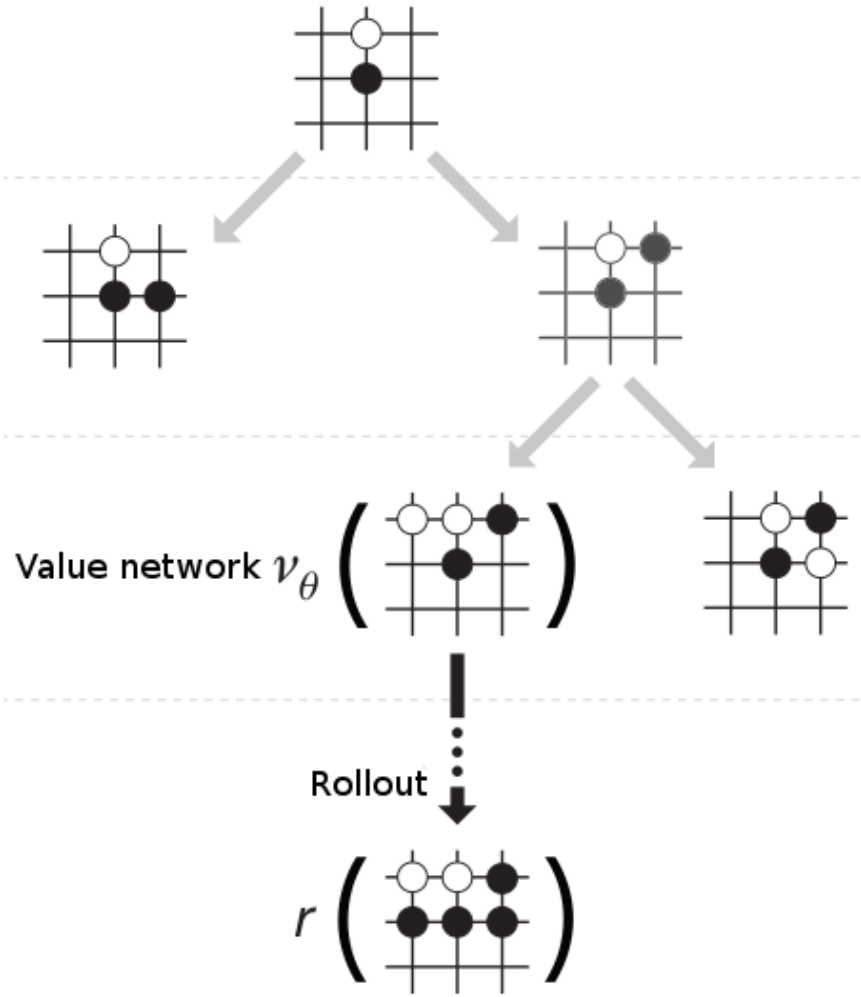
Step 2/4: Expansion



Create one new node
(Run SL policy net only once)

Aalto University

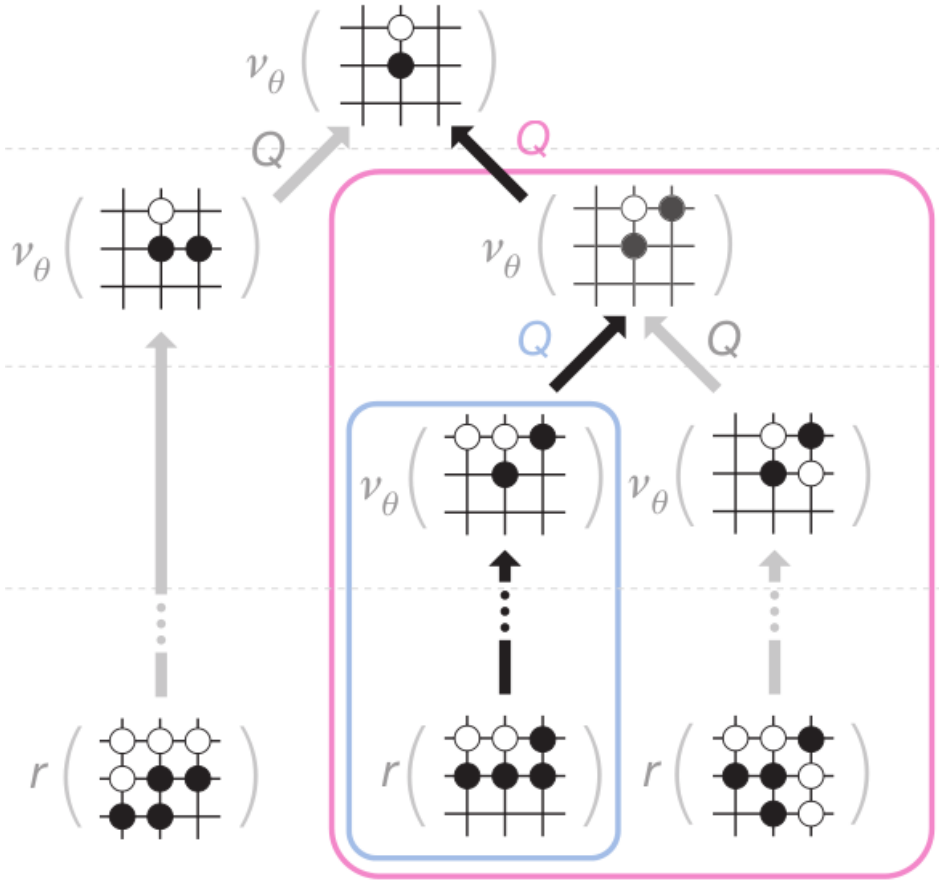
Step 3/4: Evaluation



Evaluation $Q(s)$ is average of Value network (4) and outcome of self-play by Rollout policy (net 1)

Aalto University

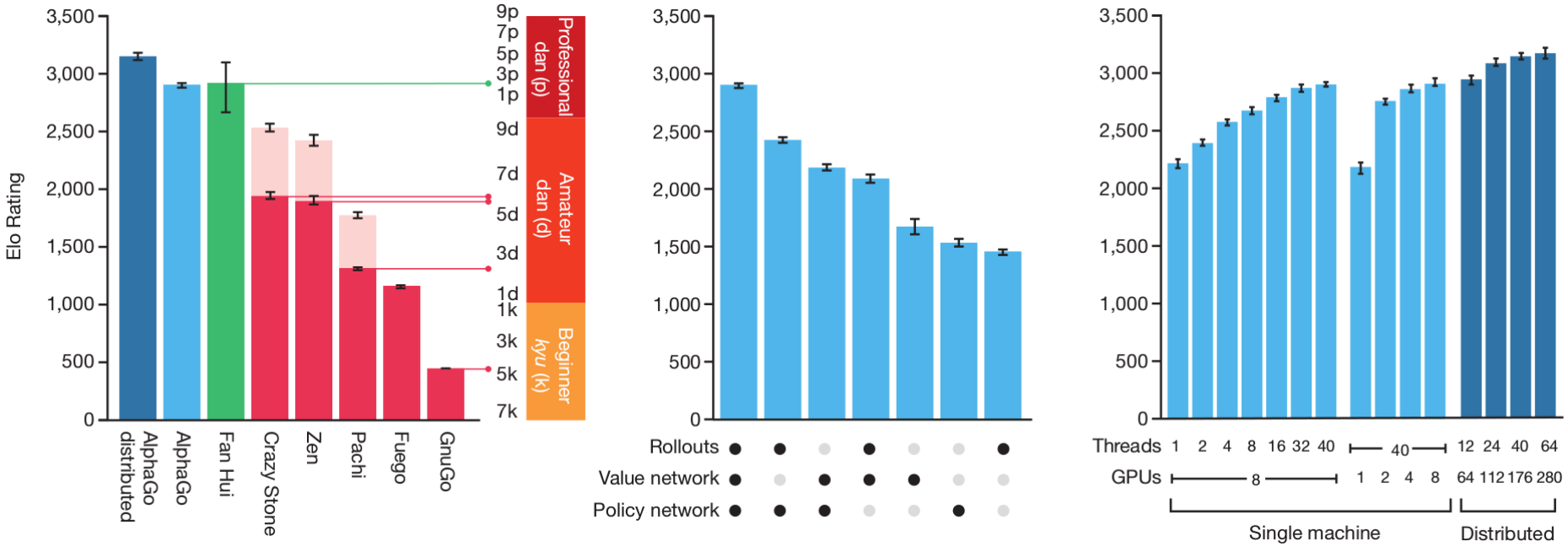
Step 4/4: Backup



$Q(a|s)$ along the path are updated to track the mean of Q in subtree below

Aalto University

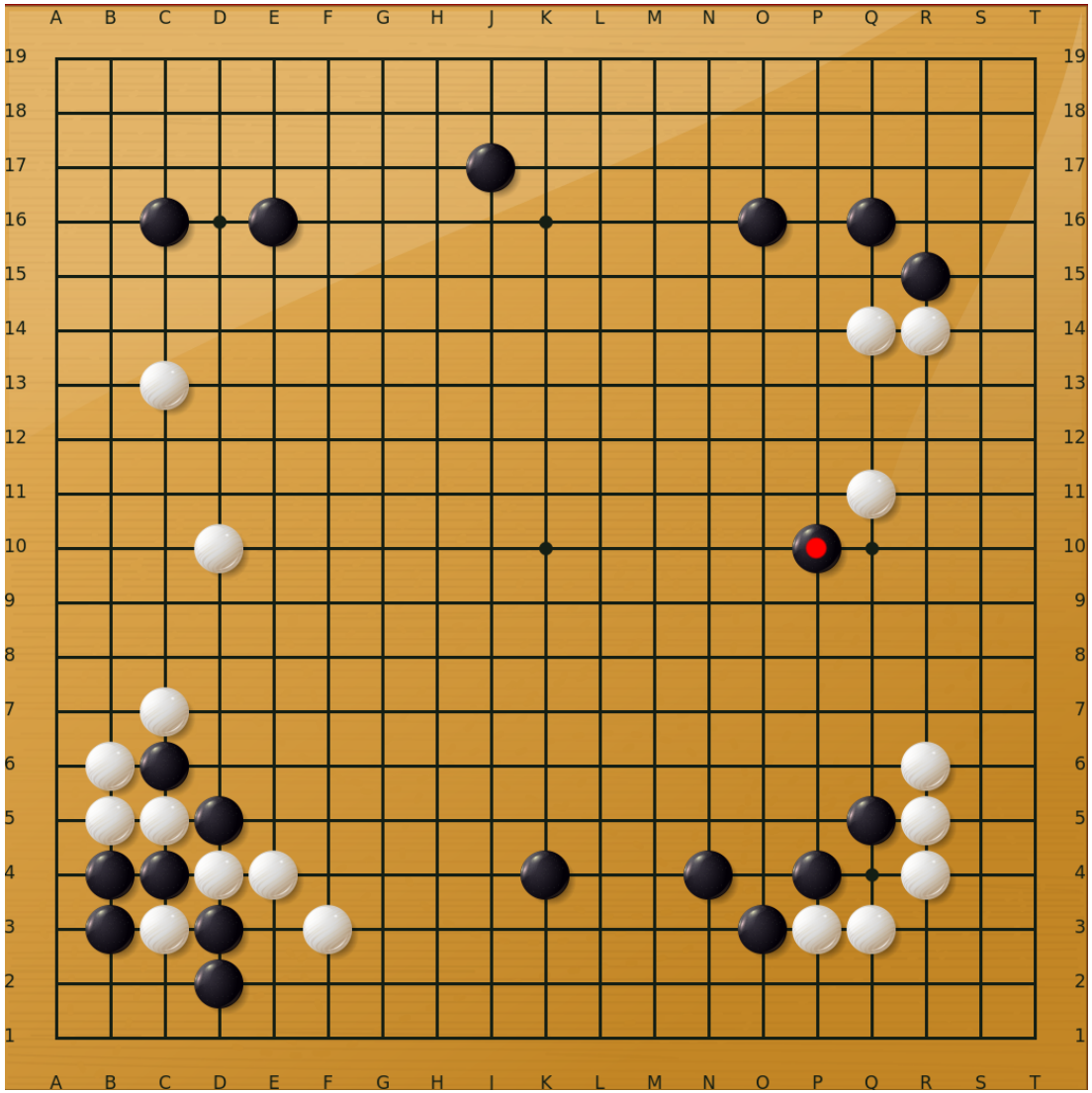
Strength



- AlphaGo won Lee Sedol (4 games out of 5)
- Strongest AI with any 2/3 features enabled
- Single machine version professional level, too

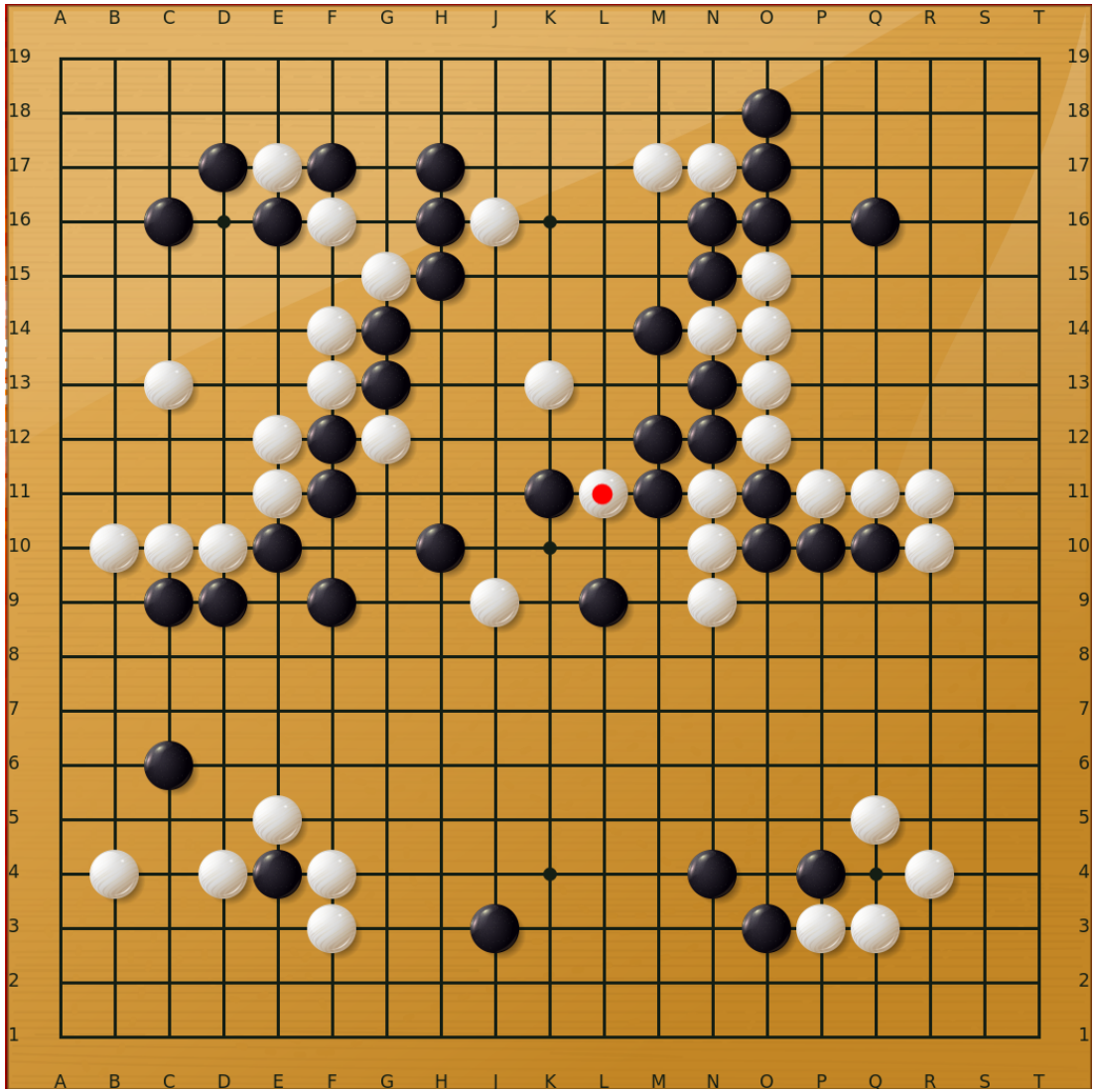
Aalto University

AlphaGo's new move (game 2, move 37)



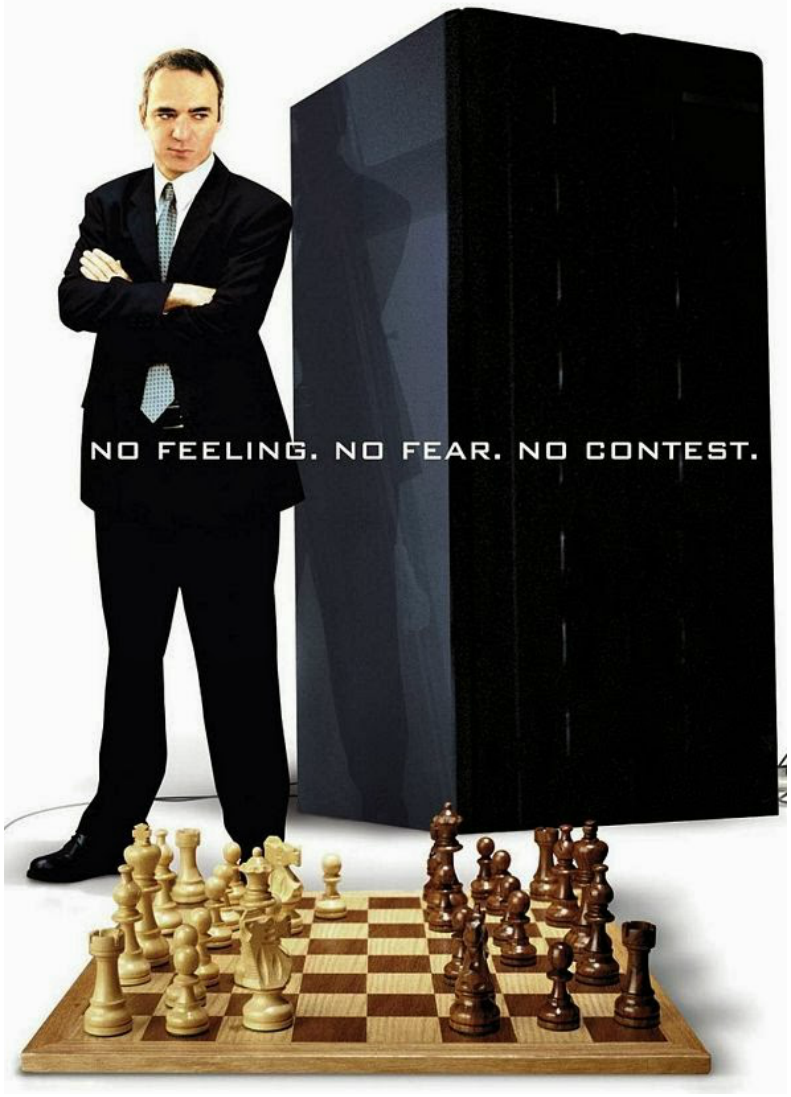
Aalto University

Lee Sedol's hand of god (game 4, move 78)



Aalto University

Comparison to Deep Blue (1997)



AlphaGo

- learns
- uses 100 000 times faster hardware
- evaluates 1000 times fewer positions per second
- is not restricted to deterministic adversarial search
  (Deep Blue was based on alpha-beta pruning)