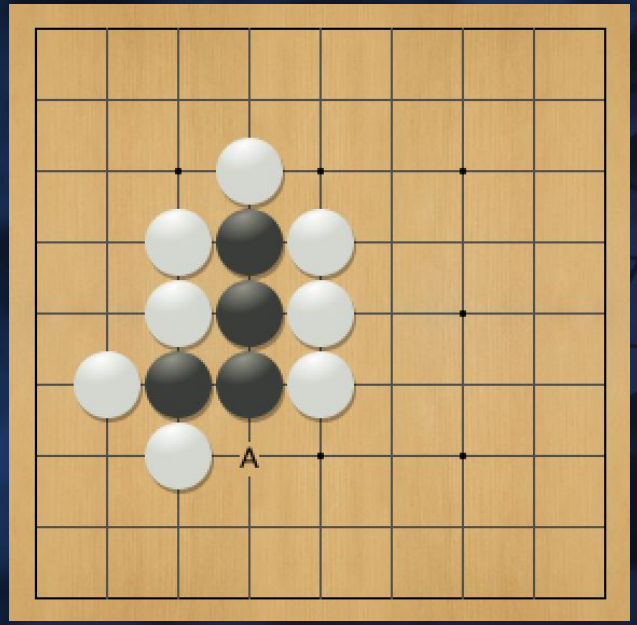
AlphaGo

Go in numbers

The Rules of Go



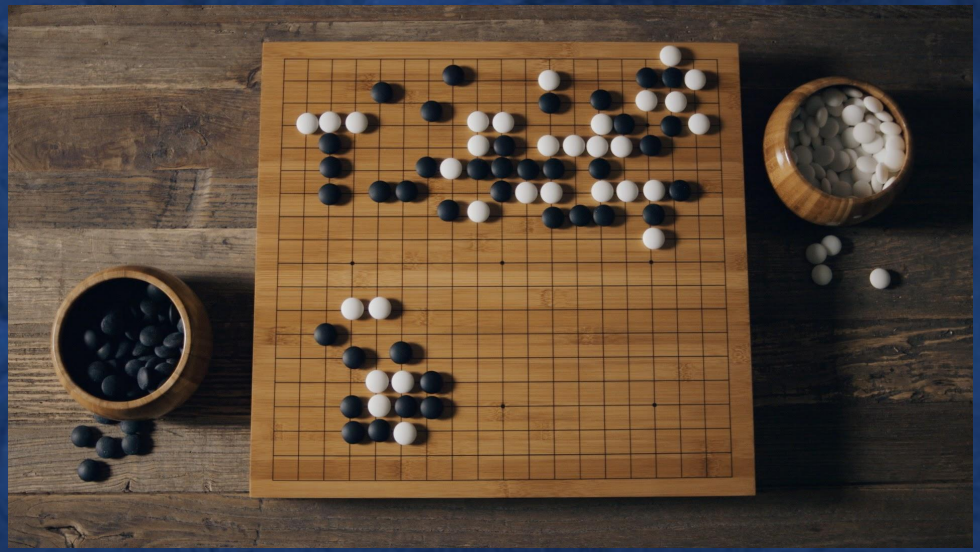
Capture

Territory

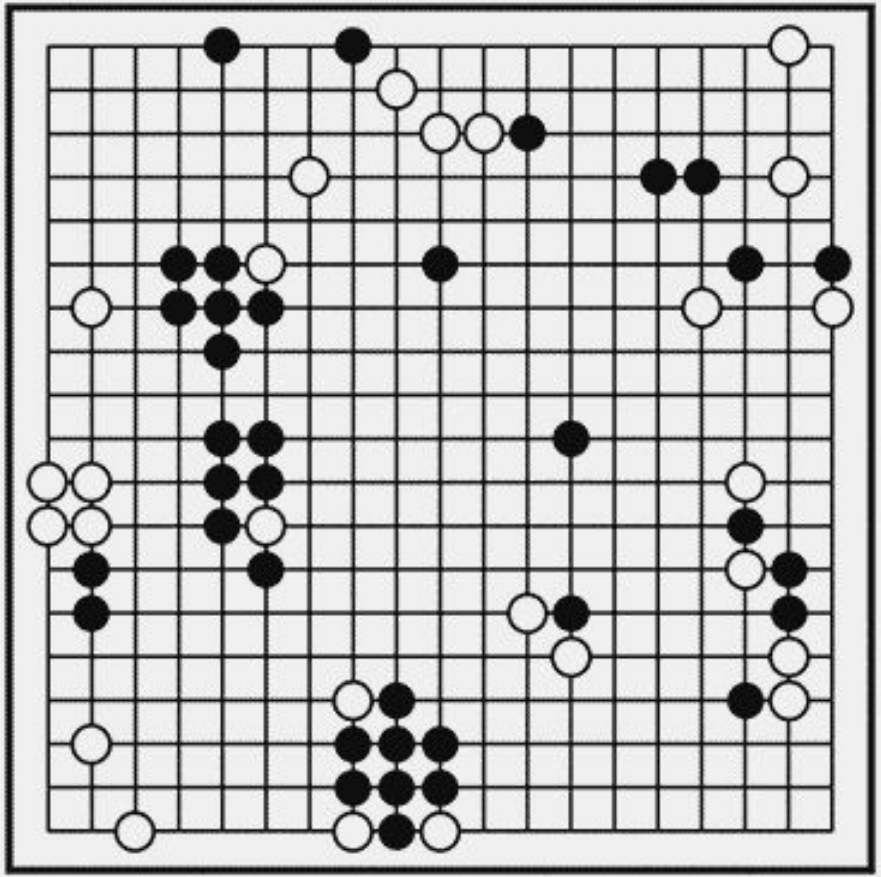
Why is Go hard for computers to play?

Brute force search intractable:

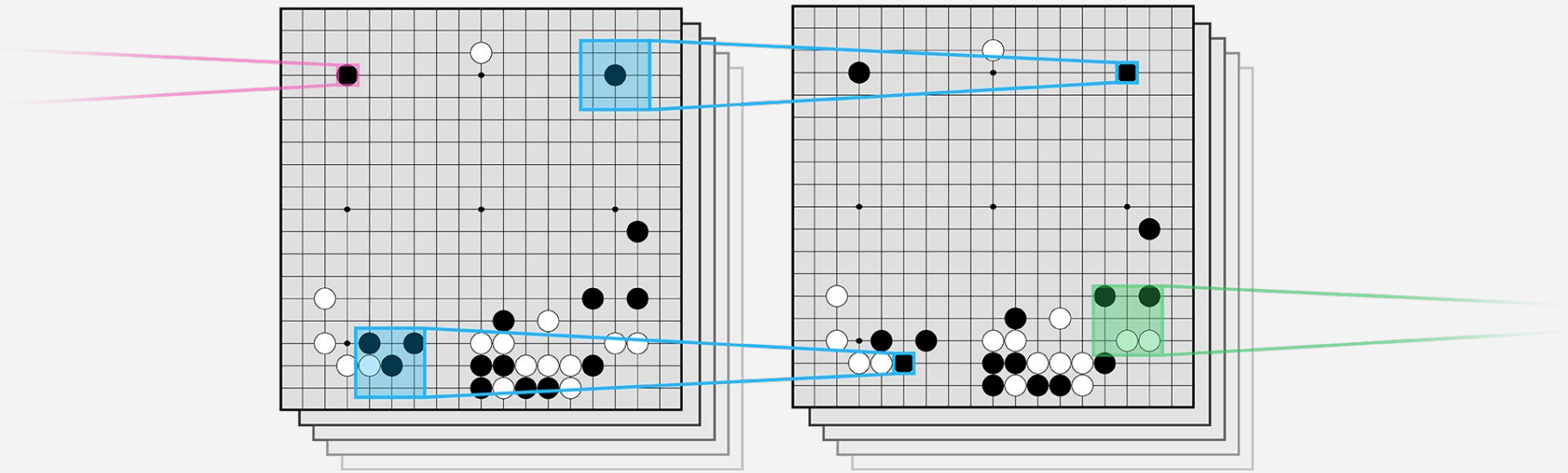
1. Search space is huge
2. “Impossible” for computers to evaluate who is winning



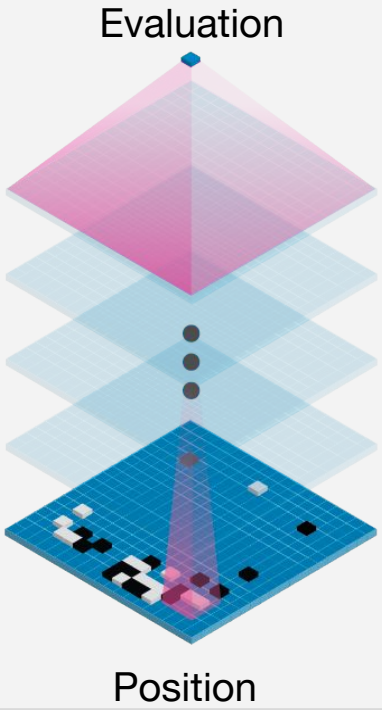
Game tree complexity = b^d

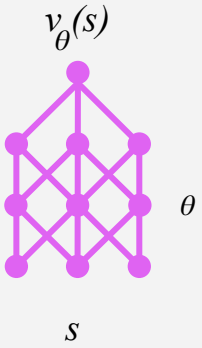


Convolutional neural network



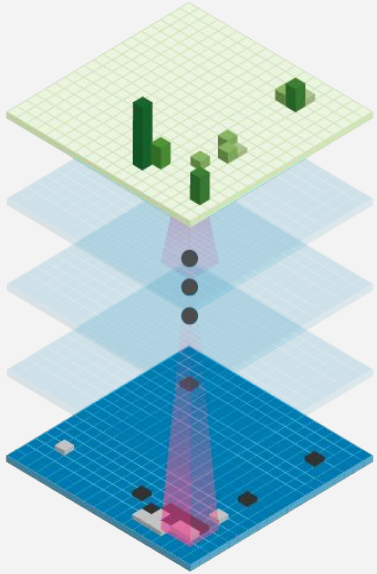
Value network





Policy network

Move probabilities



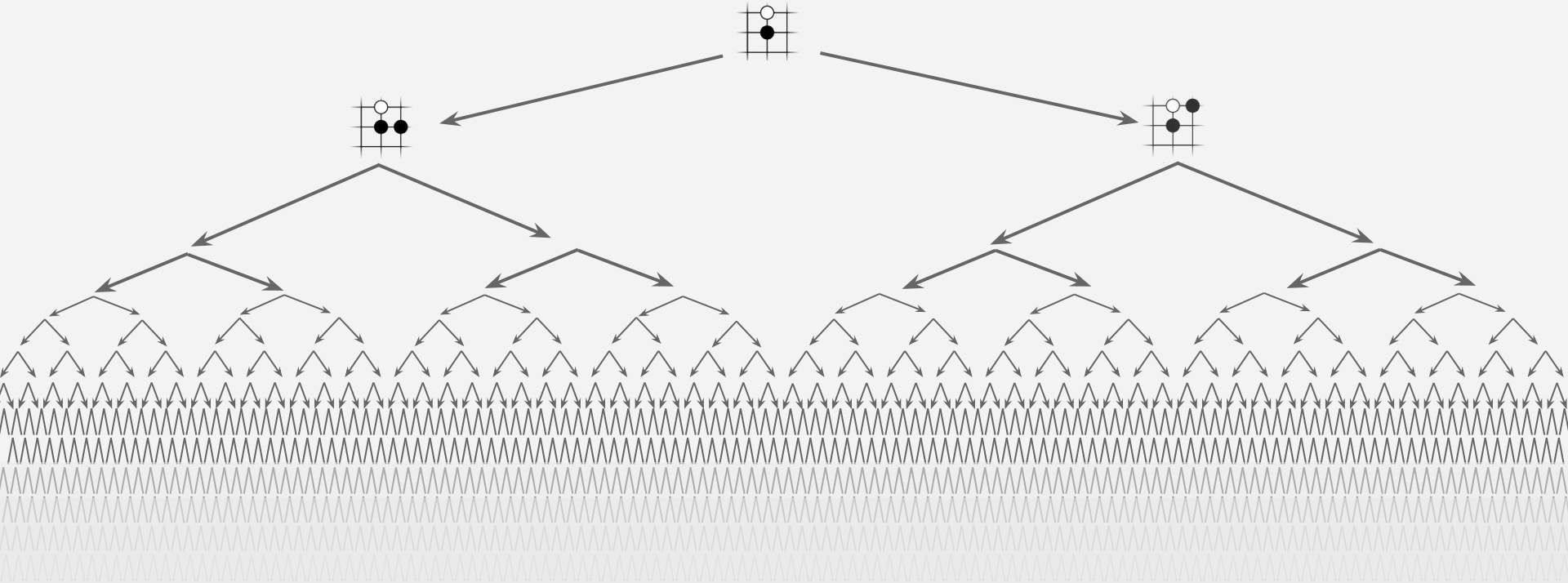
Position

$p_\sigma(a|s)$

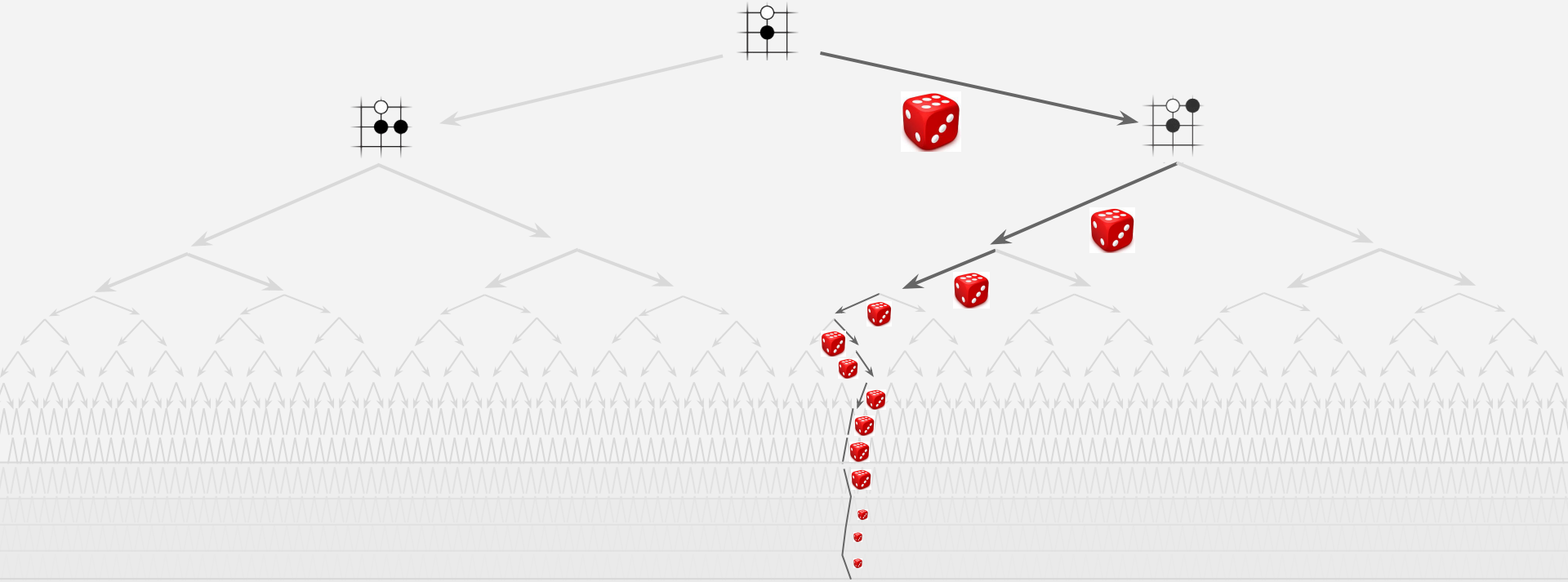
σ

s

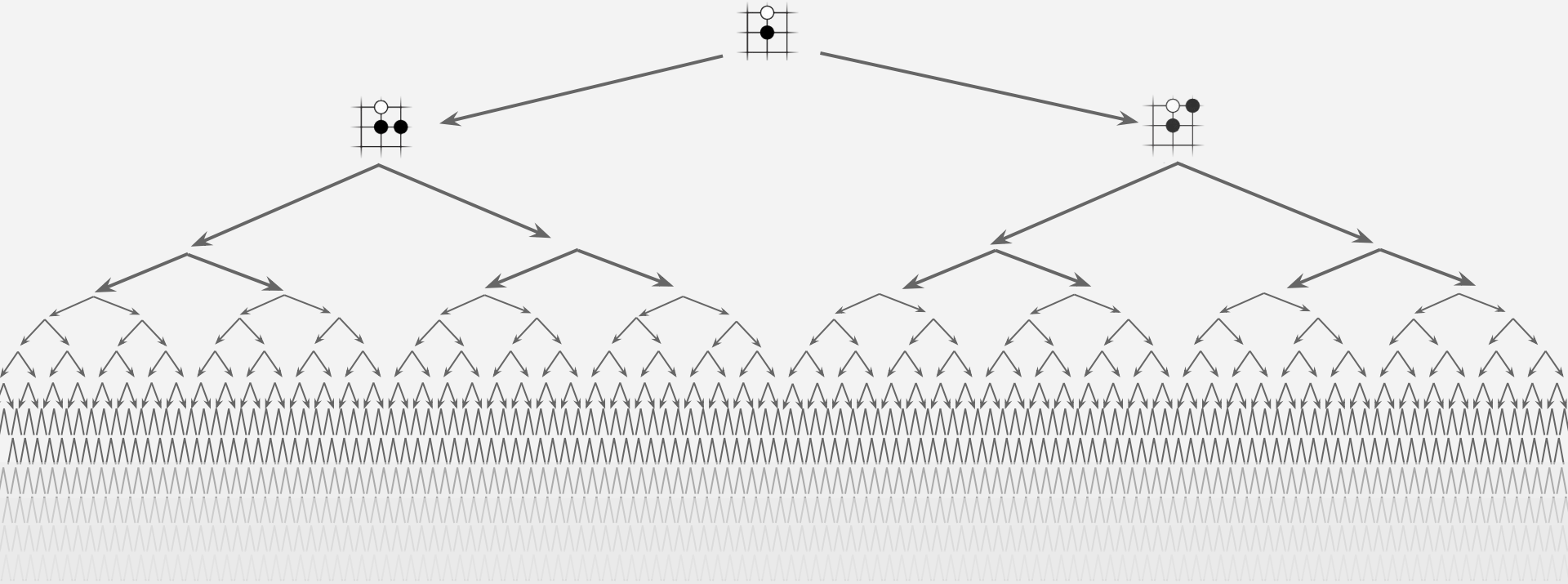
Exhaustive search



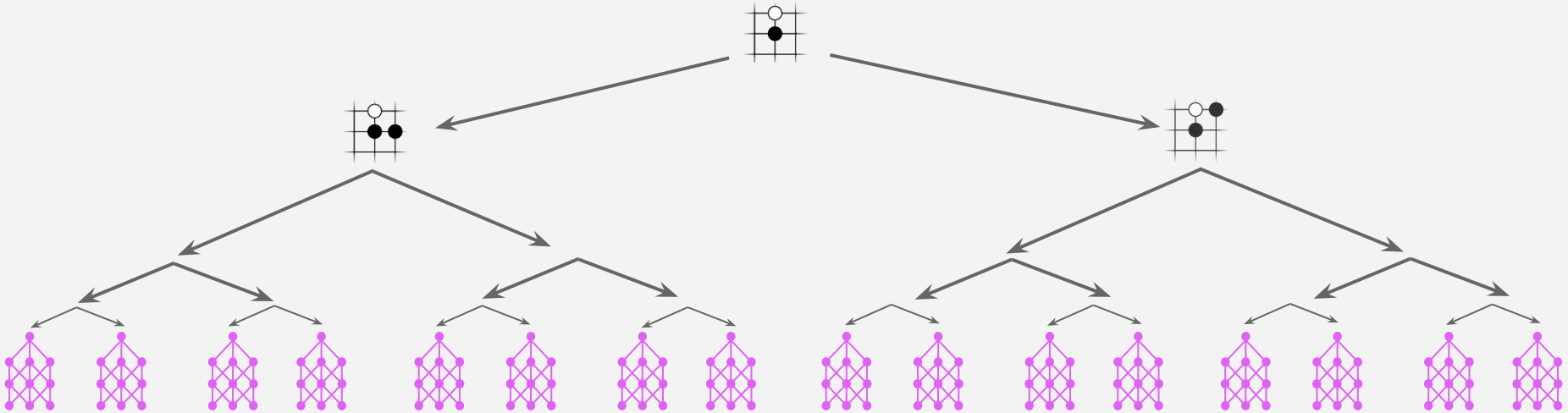
Monte-Carlo rollouts



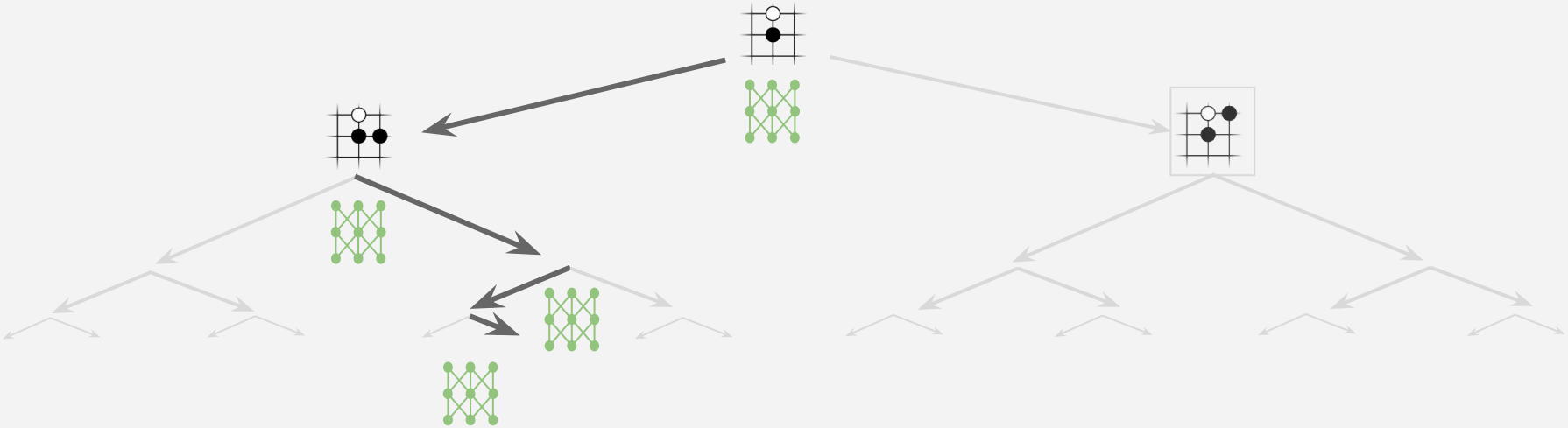
Reducing depth with value network



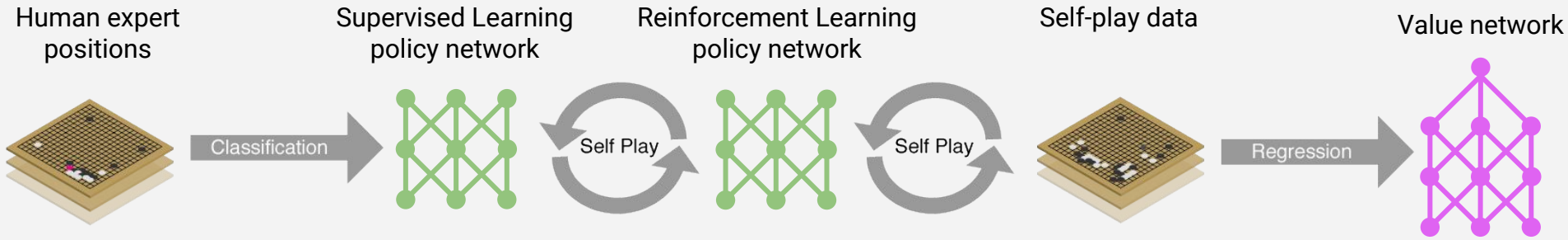
Reducing depth with value network



Reducing breadth with policy network



Deep reinforcement learning in AlphaGo



Supervised learning of policy networks

Policy network: 12 layer convolutional neural network

Training data: 30M positions from human expert games (KGS 5+ dan)

Training algorithm: maximise likelihood by stochastic gradient descent

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial\sigma}$$

Training time: 4 weeks on 50 GPUs using Google Cloud

Results: 57% accuracy on held out test data (state-of-the art was 44%)

Reinforcement learning of policy networks

Policy network: 12 layer convolutional neural network

Training data: games of self-play between policy network

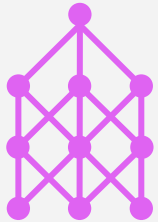
Training algorithm: maximise wins z by policy gradient reinforcement learning

$$\Delta\sigma \propto \frac{\partial \log p_\sigma(a|s)}{\partial \sigma} z$$

Training time: 1 week on 50 GPUs using Google Cloud

Results: 80% vs supervised learning. Raw network ~3 amateur dan.

Reinforcement learning of value networks



Value network: 12 layer convolutional neural network

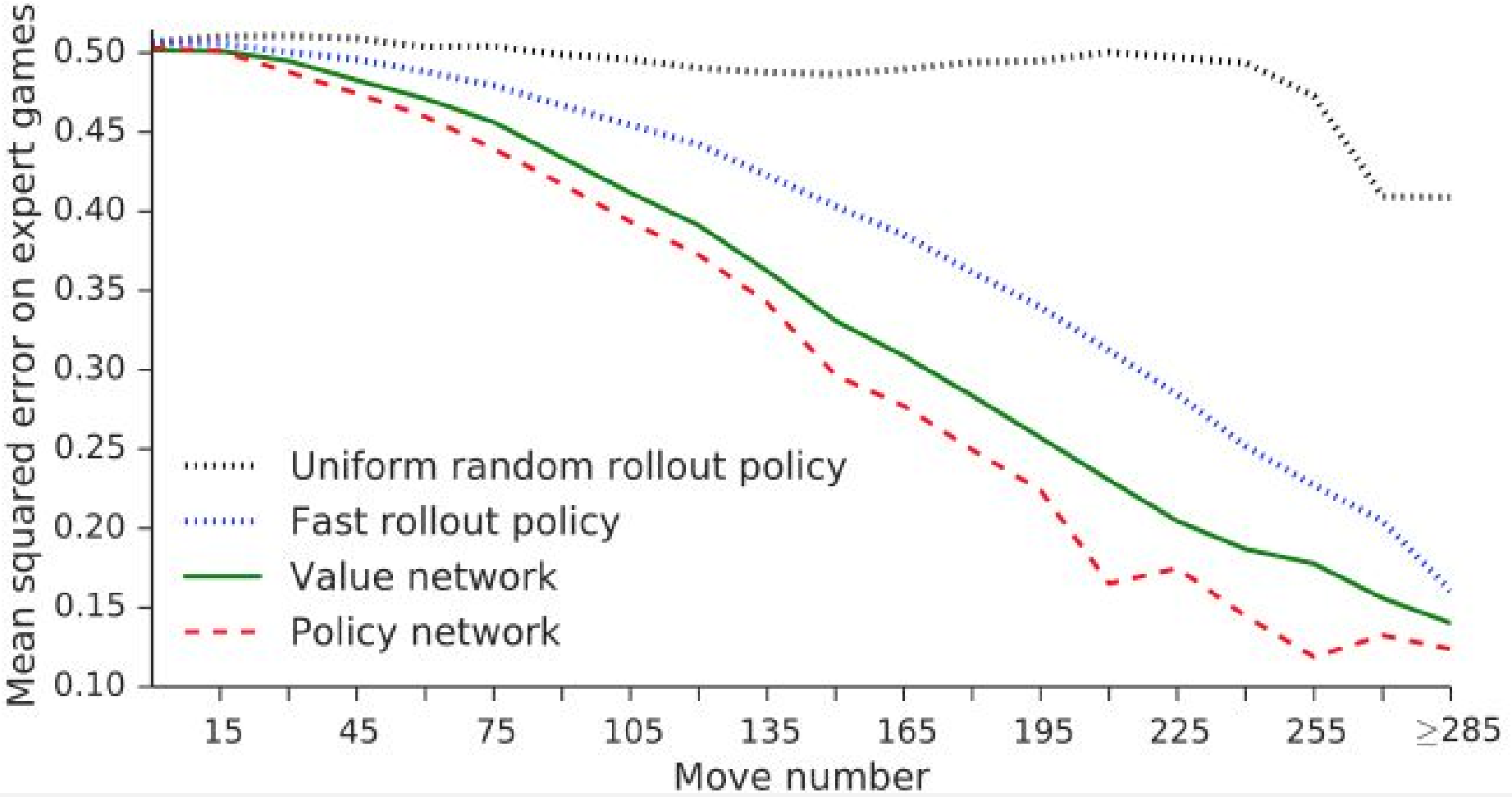
Training data: 30 million games of self-play

Training algorithm: minimise MSE by stochastic gradient descent

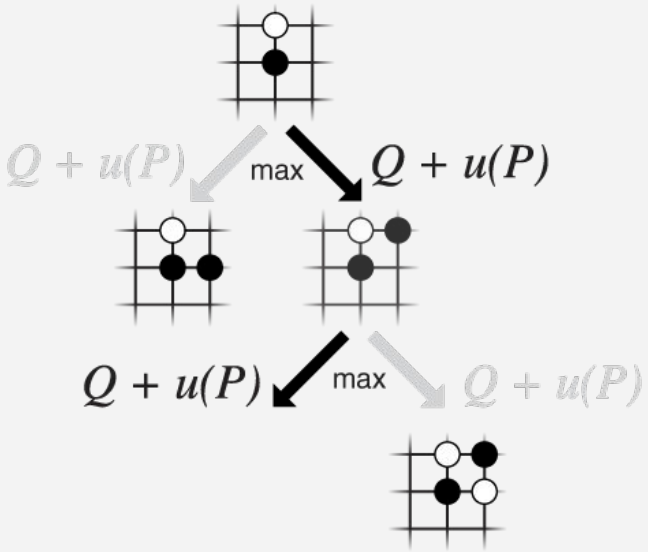
$$\Delta\theta \propto \frac{\partial v_\theta(s)}{\partial\theta}(z - v_\theta(s))$$

Training time: 1 week on 50 GPUs using Google Cloud

Results: First strong position evaluation function - previously thought impossible


Mean squared error on expert games
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
Uniform random rollout policy
Fast rollout policy
Value network
Policy network
15
45
75
105
135
165
195
225
255
≥285
Move number

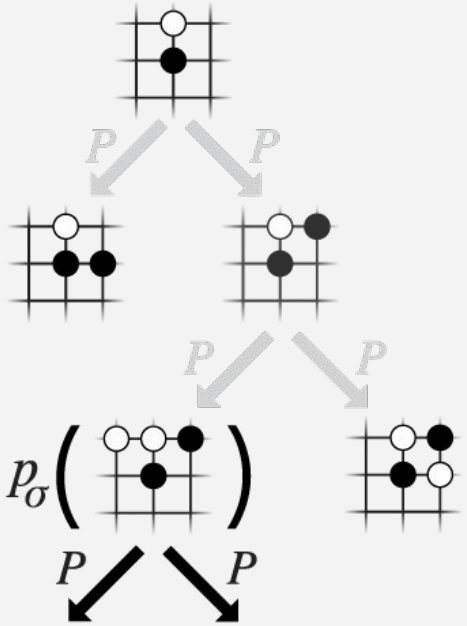
Monte-Carlo tree search in AlphaGo: **selection**



P prior probability
Q action value

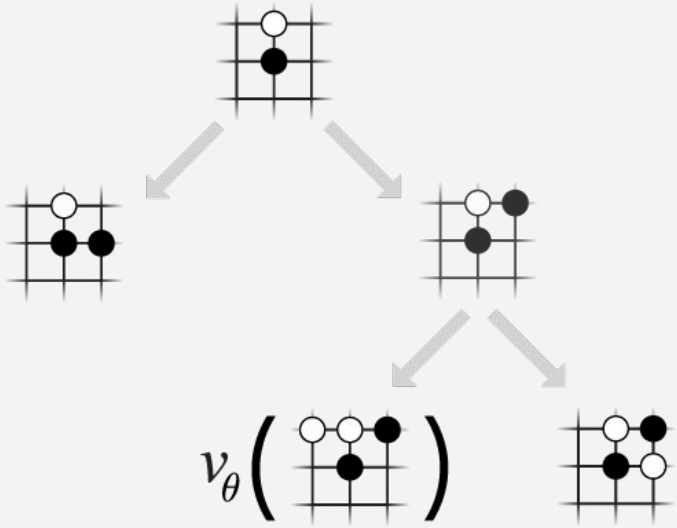
$$u(P) \propto P/N$$

Monte-Carlo tree search in AlphaGo: **expansion**



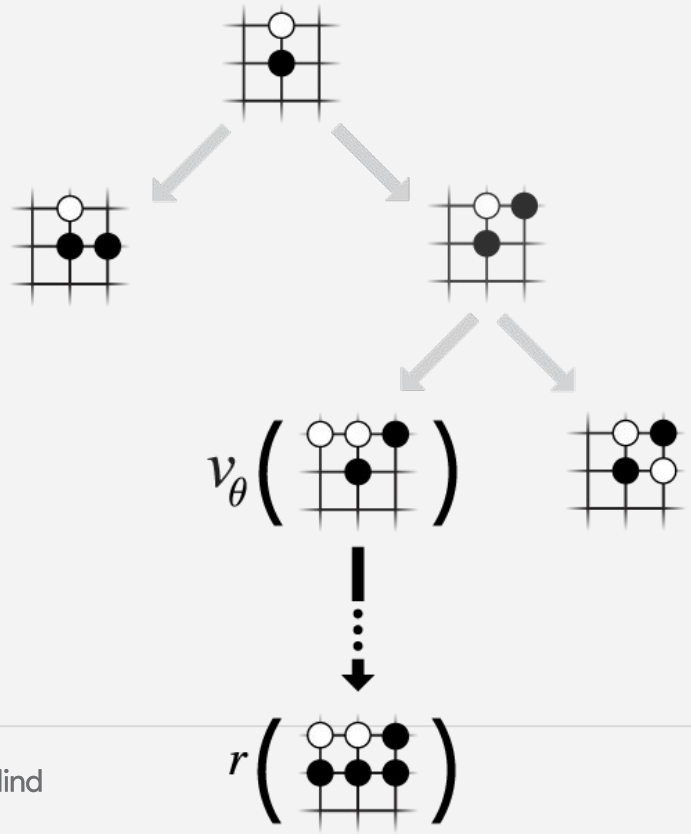
p_σ Policy network
P prior probability

Monte-Carlo tree search in AlphaGo: **evaluation**



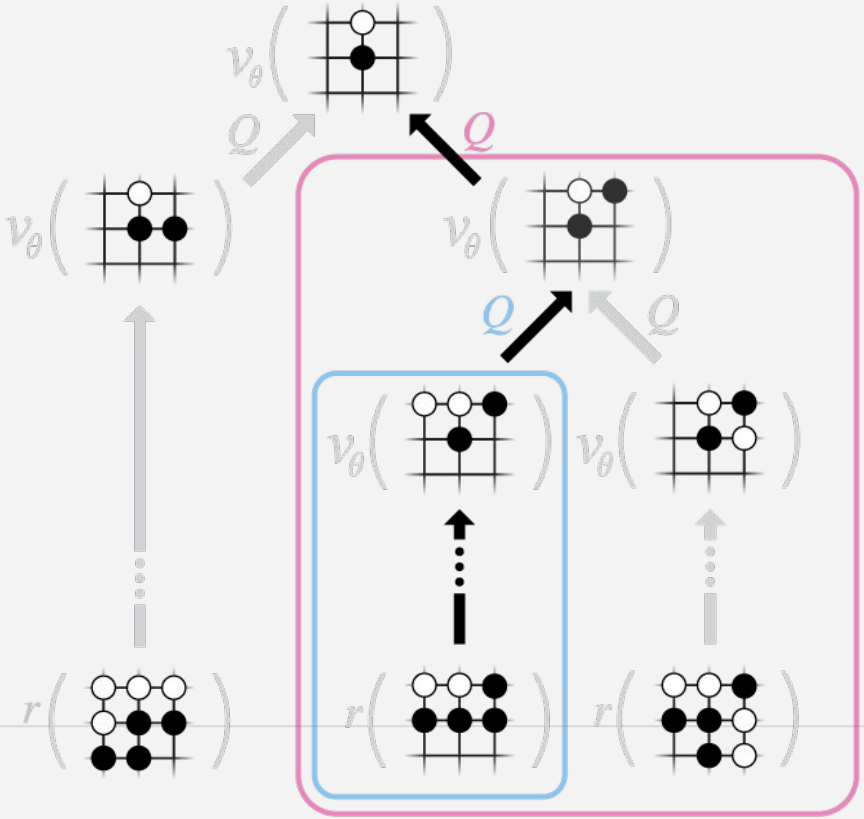
v_θ Value network

Monte-Carlo tree search in AlphaGo: **rollout**



v_θ Value network
r Game scorer

Monte-Carlo tree search in AlphaGo: **backup**



Q Action value
v_θ Value network
r Game scorer

Google DeepMind

Deep Blue	AlphaGo
Handcrafted chess knowledge	Knowledge learned from expert games and self-play
Alpha-beta search guided by heuristic evaluation function	Monte-Carlo search guided by policy and value networks
200 million positions / second	60,000 positions / second

Nature AlphaGo

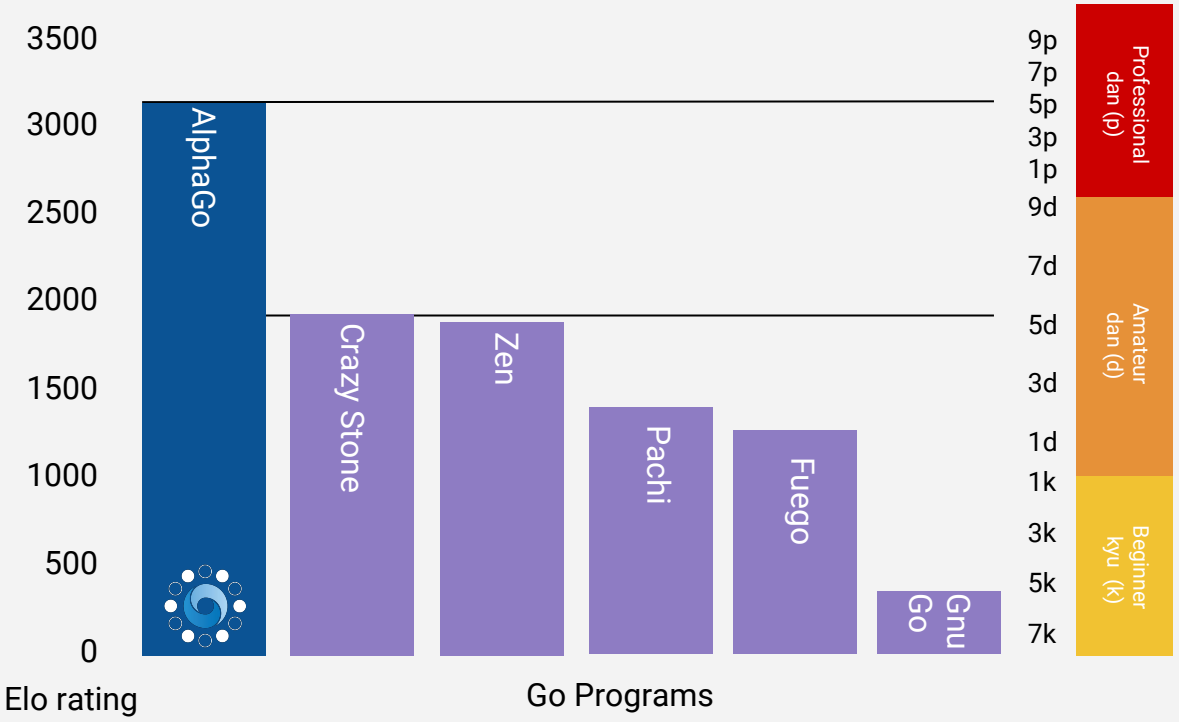
Seoul AlphaGo

Evaluating Nature AlphaGo against computers

494/495 against computer opponents

>75% winning rate with 4 stone handicap

Even stronger using distributed machines



Evaluating Nature AlphaGo against humans

Fan Hui (2p): European Champion 2013 - 2016

Match was played in October 2015

AlphaGo won the match 5-0

First program ever to beat a professional

on a full size 19x19 in an even game

Google DeepMind

Seoul AlphaGo

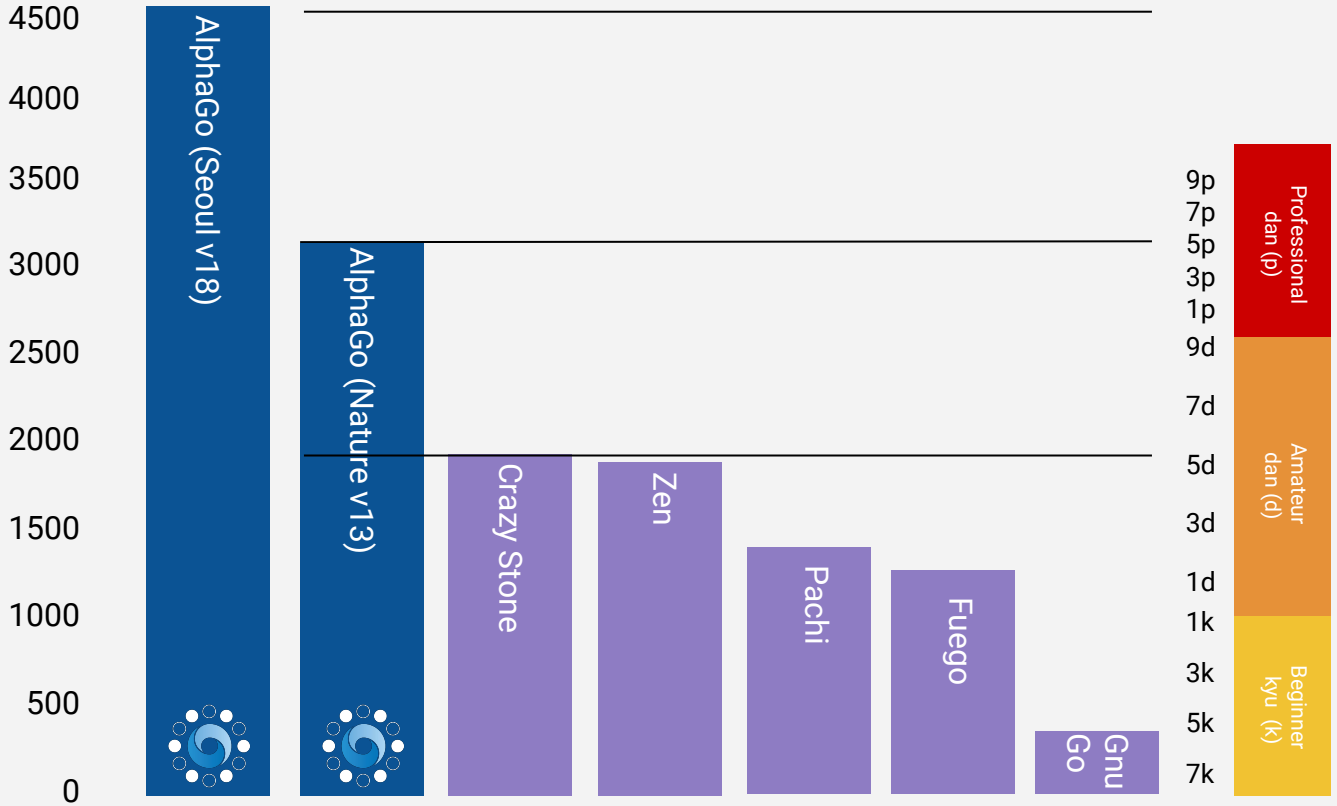
Deep Reinforcement Learning (as Nature AlphaGo)

- Improved value network
- Improved policy network
- Improved search
- Improved hardware (TPU vs GPU)

Evaluating Seoul AlphaGo against computers

>50% against Nature AlphaGo with 3 to 4 stone handicap

CAUTION: ratings based on self-play results



Evaluating Seoul AlphaGo against humans

Lee Sedol (9p): winner of 18 world titles

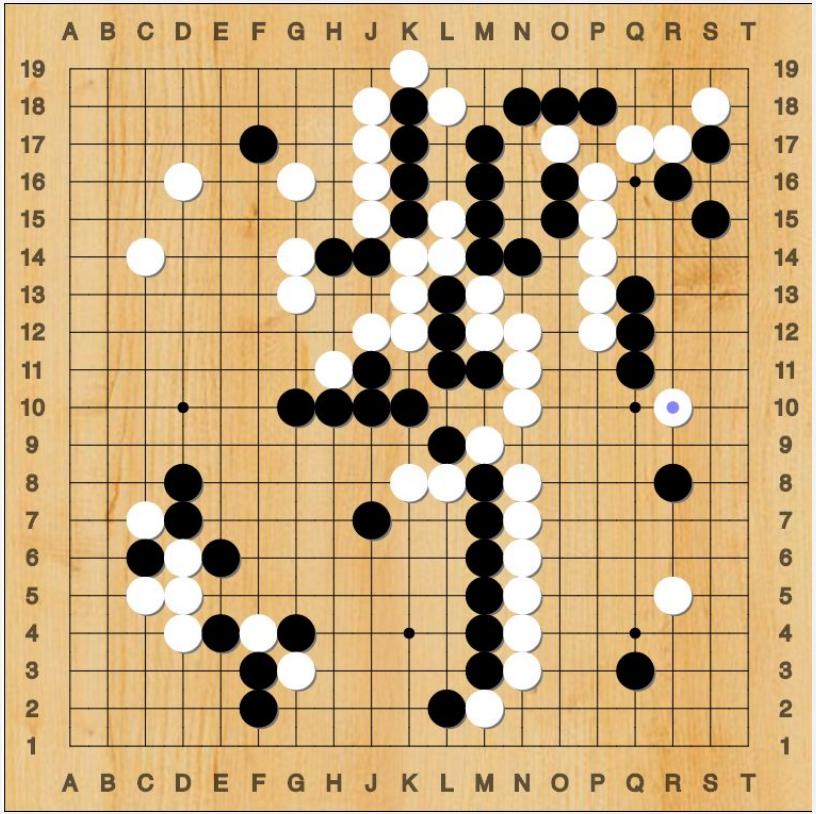
Match was played in Seoul, March 2016

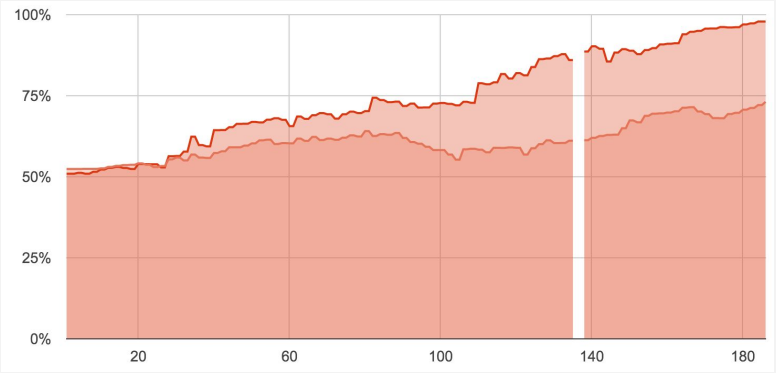
AlphaGo won the match 4-1


KBA
Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo
Google DeepMind
Challenge Match

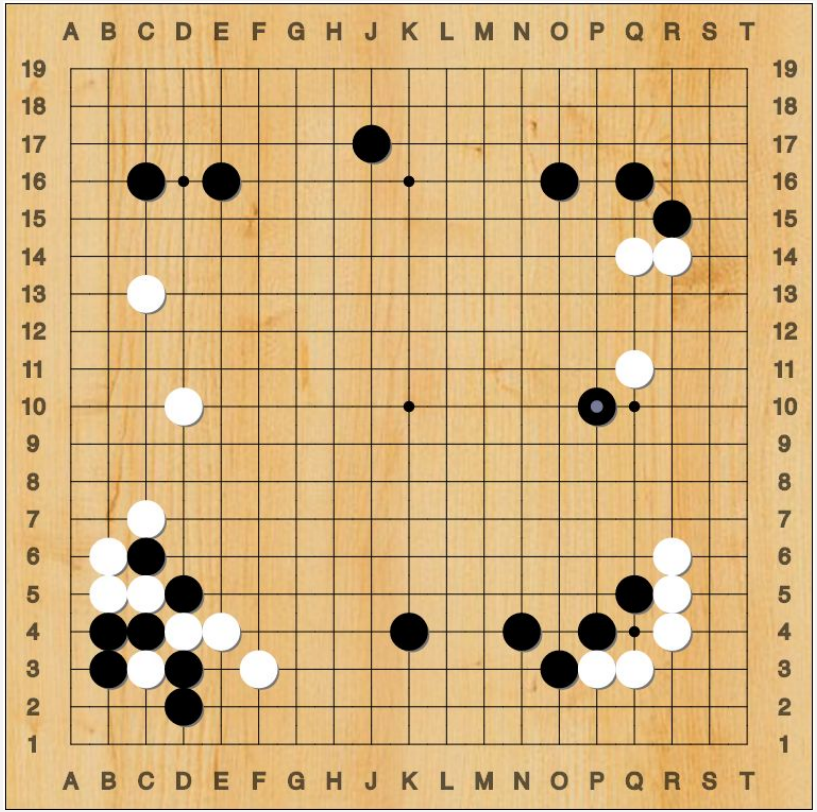

KBA
Google DeepMind
Challenge Match
8 - 15 March 2016

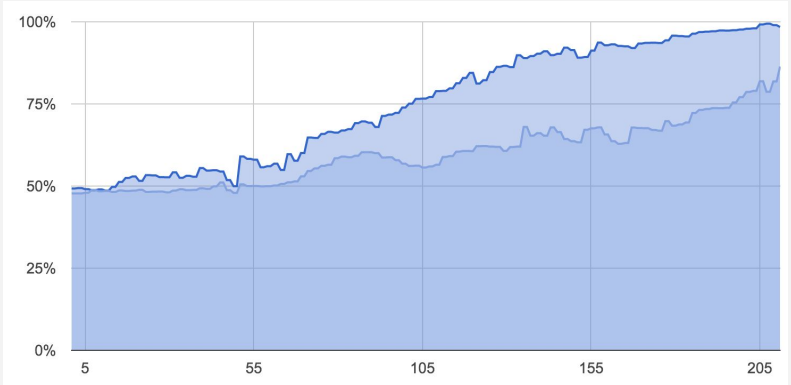
AlphaGo vs Lee Sedol: Game 1



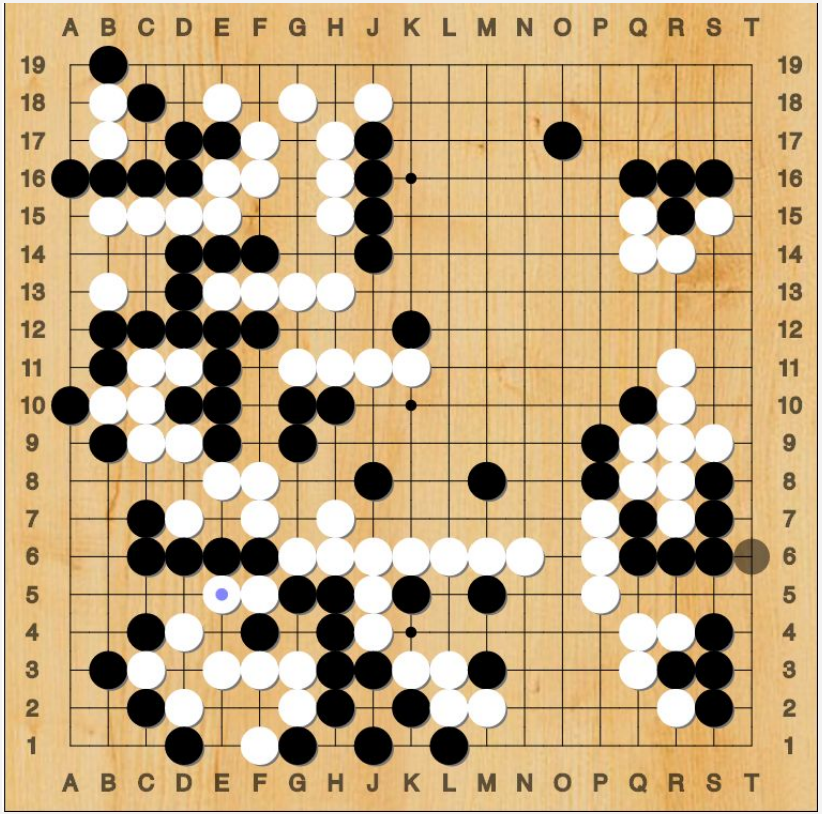


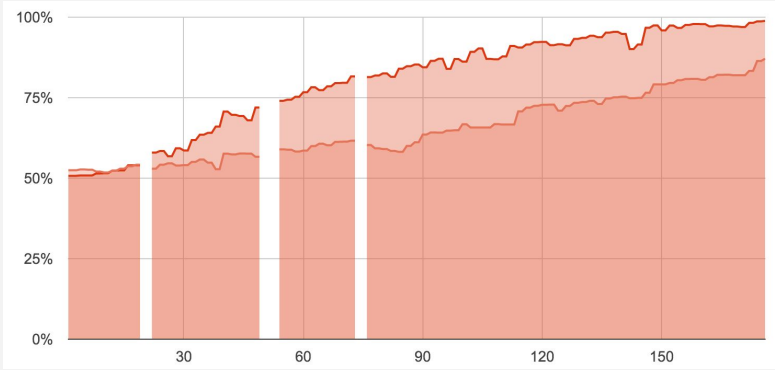
AlphaGo vs Lee Sedol: Game 2



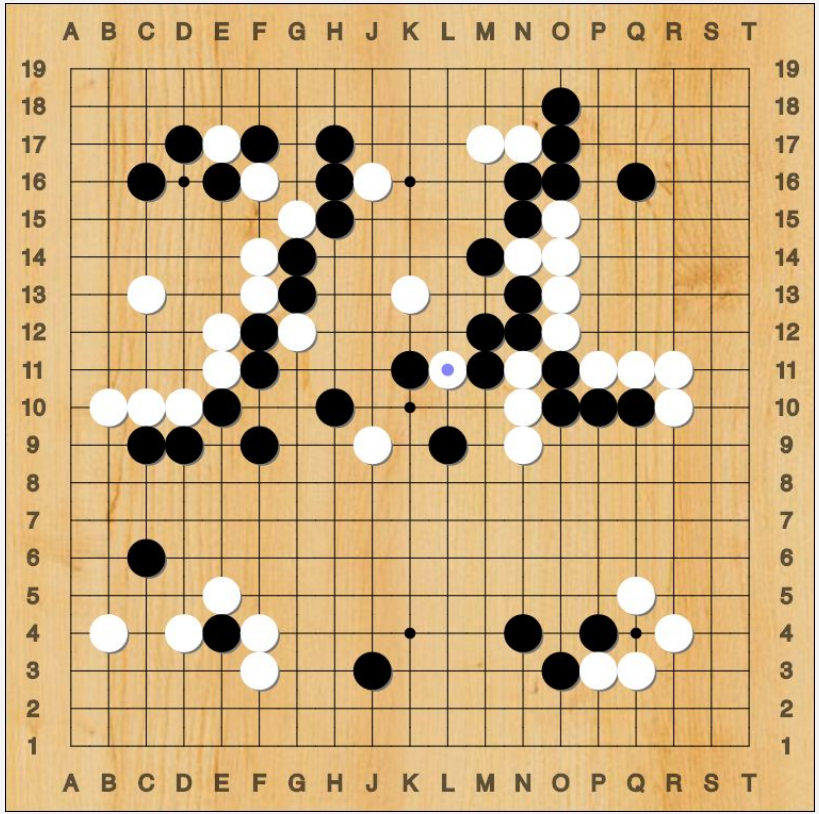


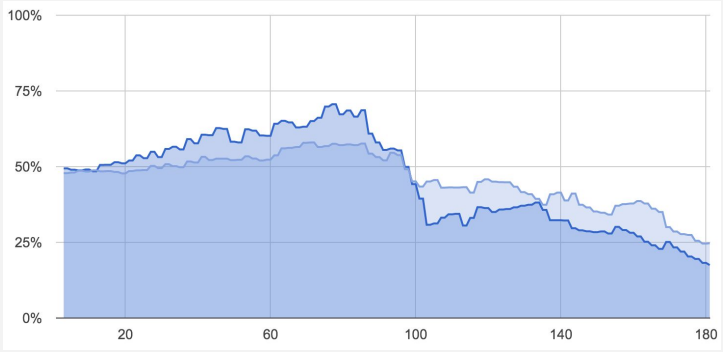
AlphaGo vs Lee Sedol: Game 3



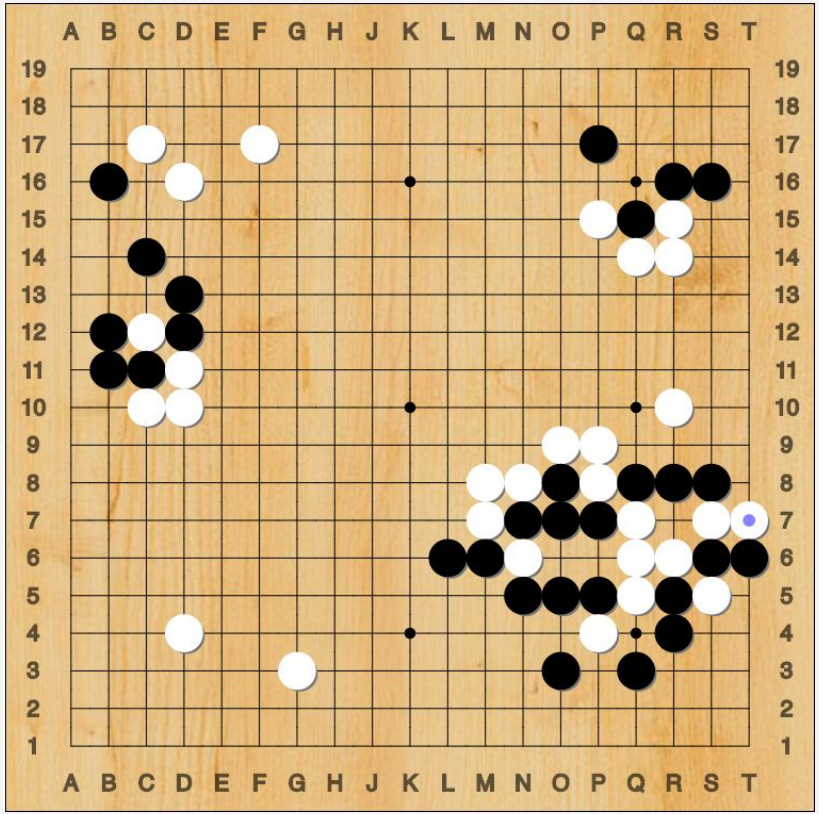


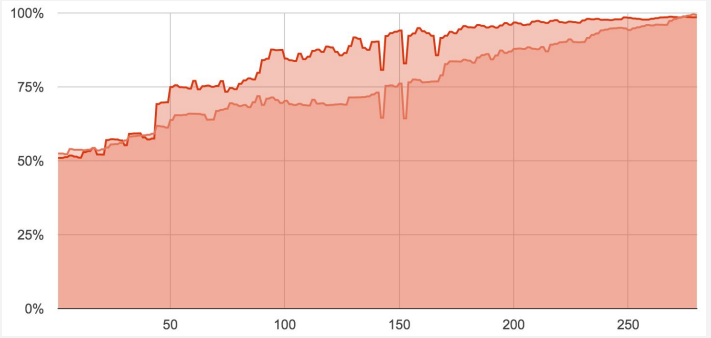
AlphaGo vs Lee Sedol: Game 4





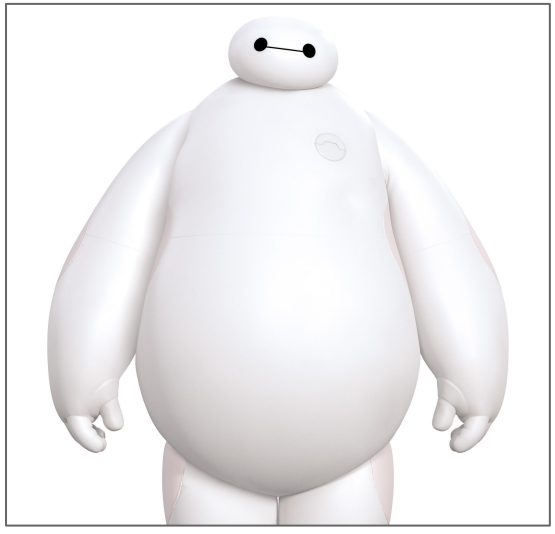
AlphaGo vs Lee Sedol: Game 5





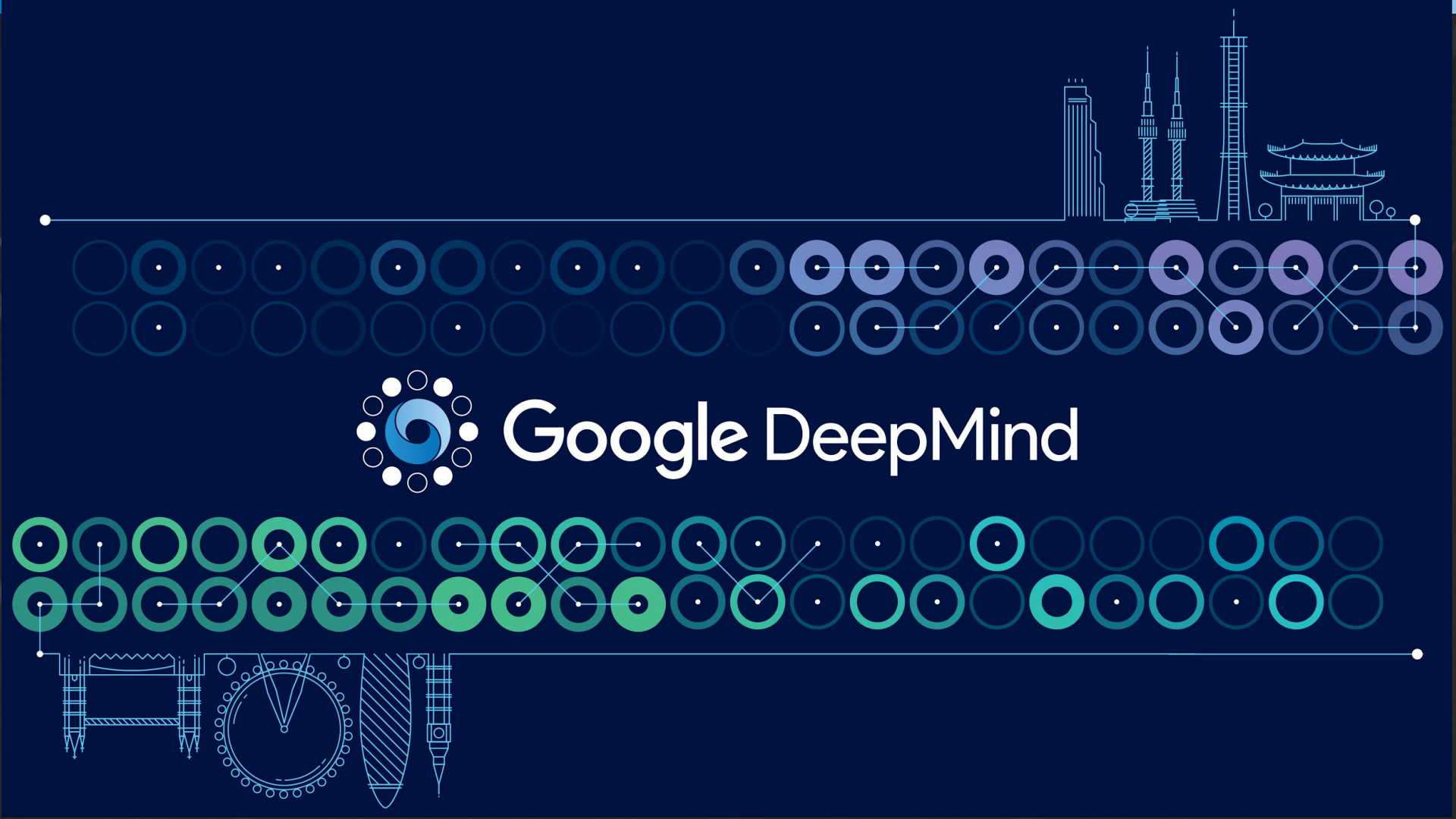
Deep Reinforcement Learning: Beyond AlphaGo

What's Next?



AlphaGo Team

With thanks to: Lucas Baker, David Szepesvari, Malcolm Reynolds, Ziyu Wang, Nando De Freitas, Mike Johnson, Ilya Sutskever, Jeff Dean, Mike Marty, Sanjay Ghemawat.


Google DeepMind