Google

Inside AlphaGo

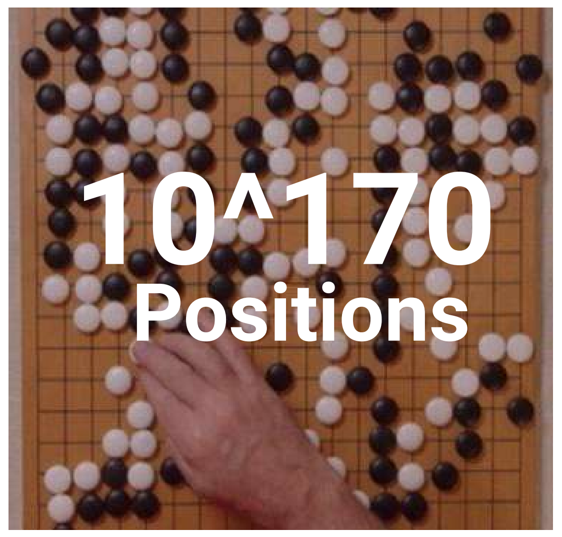
David Silver, Research Scientist and AlphaGo Team Lead

DeepMind

Go in numbers

3,000
Years Old

40M
Players

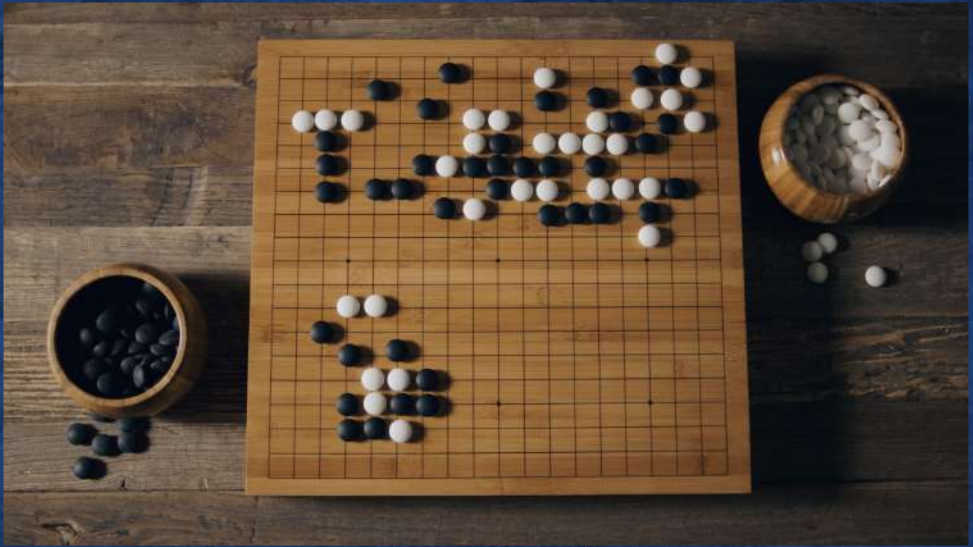


10^170
Positions

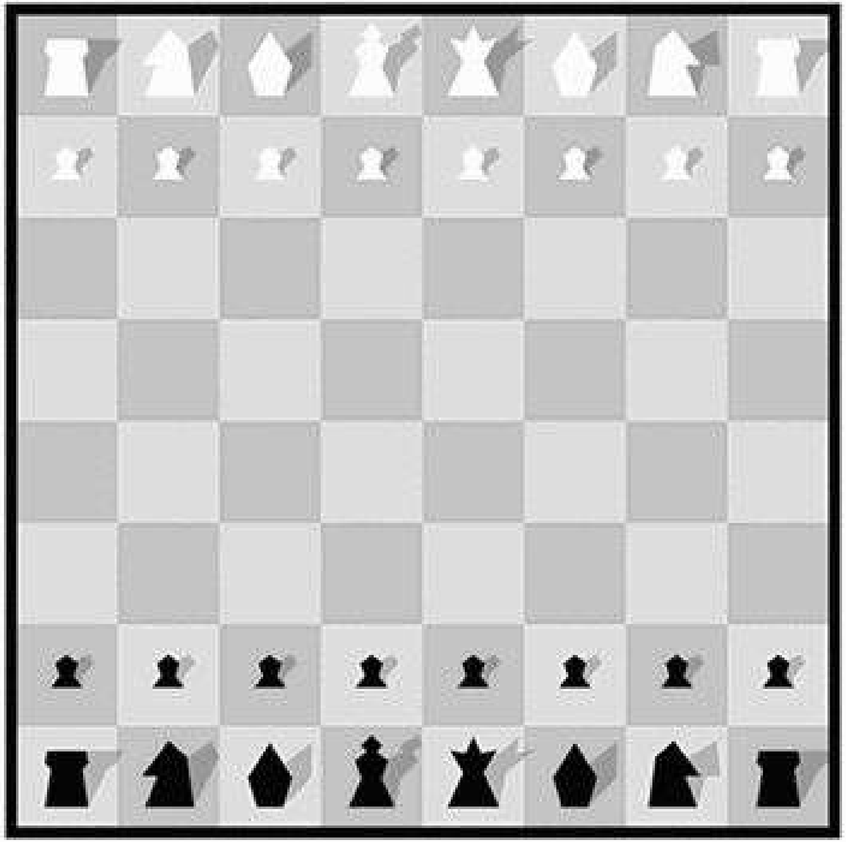
Why is Go hard for computers to play?

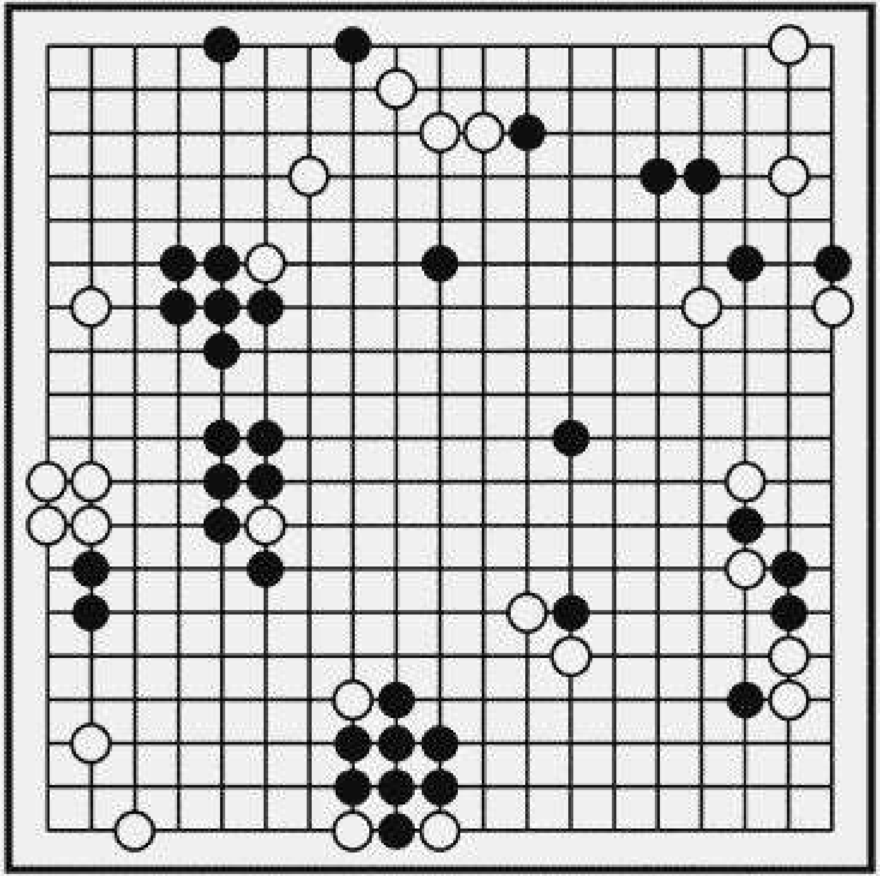
Brute force search intractable:

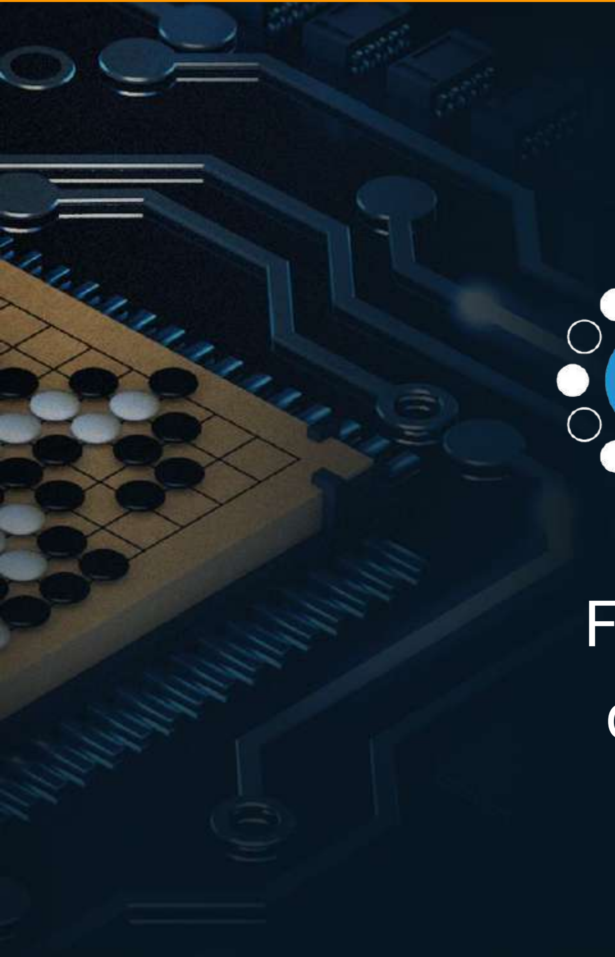
1. Search space is huge
2. "Impossible" for computers to evaluate who is winning



Game tree complexity = b^d



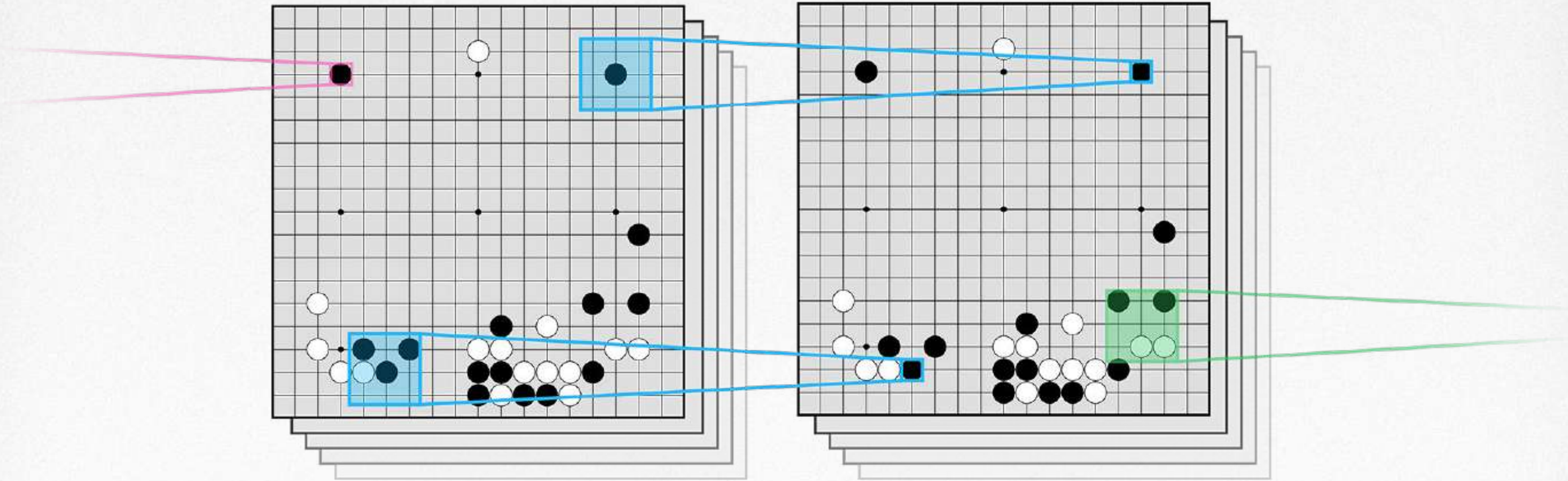




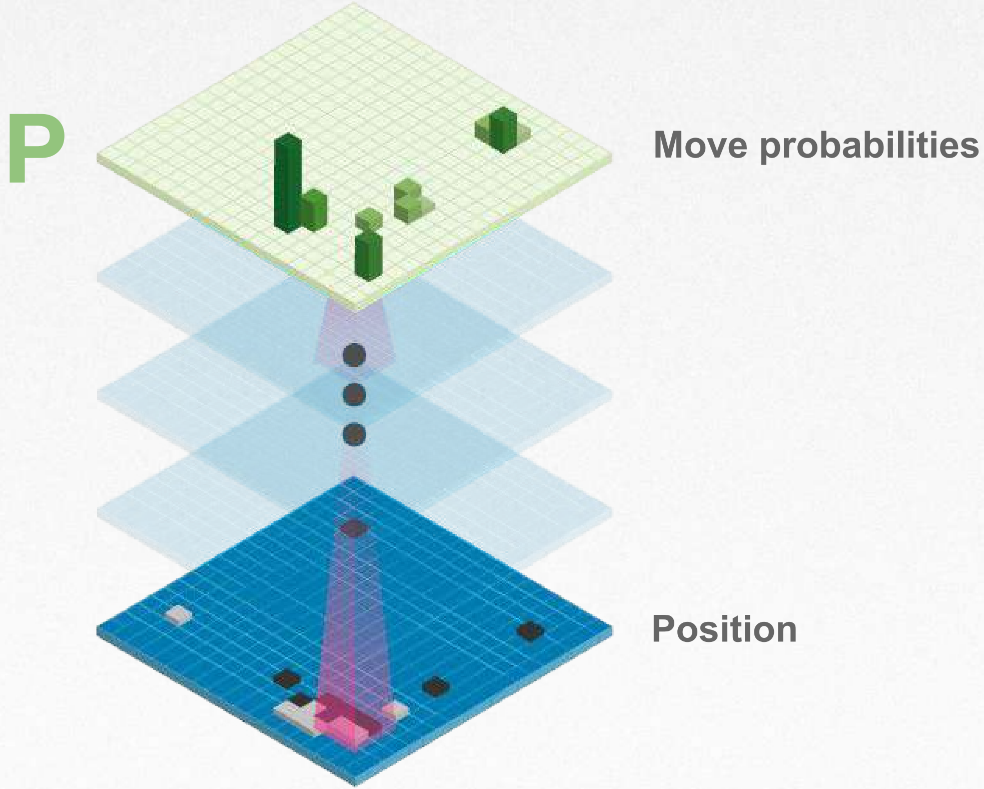
AlphaGo

First computer program to defeat a world champion

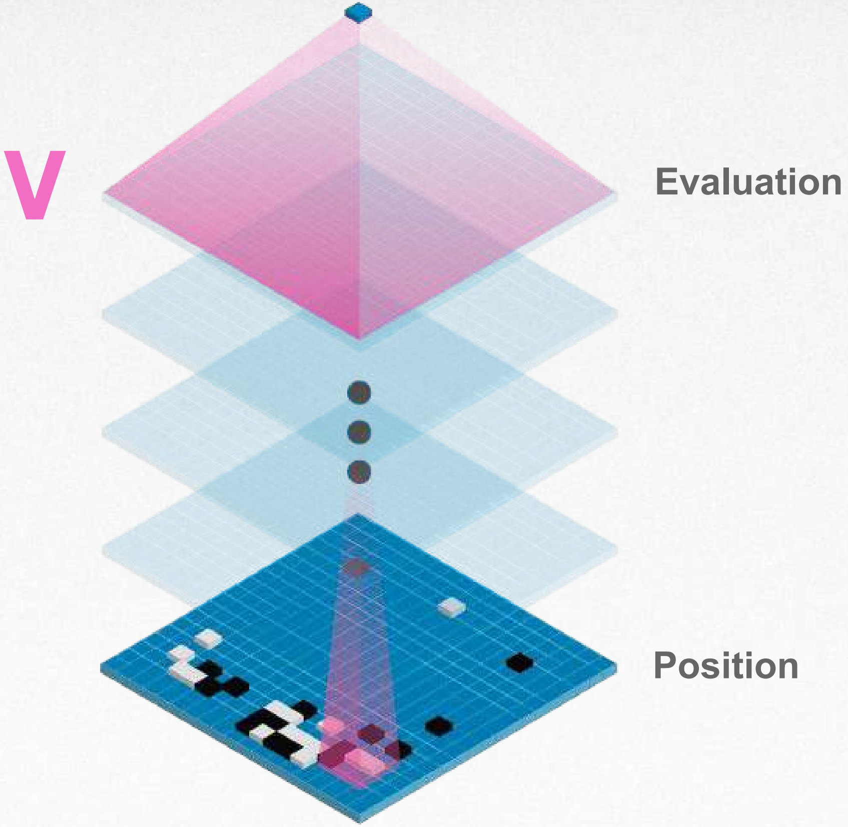
Convolutional neural network



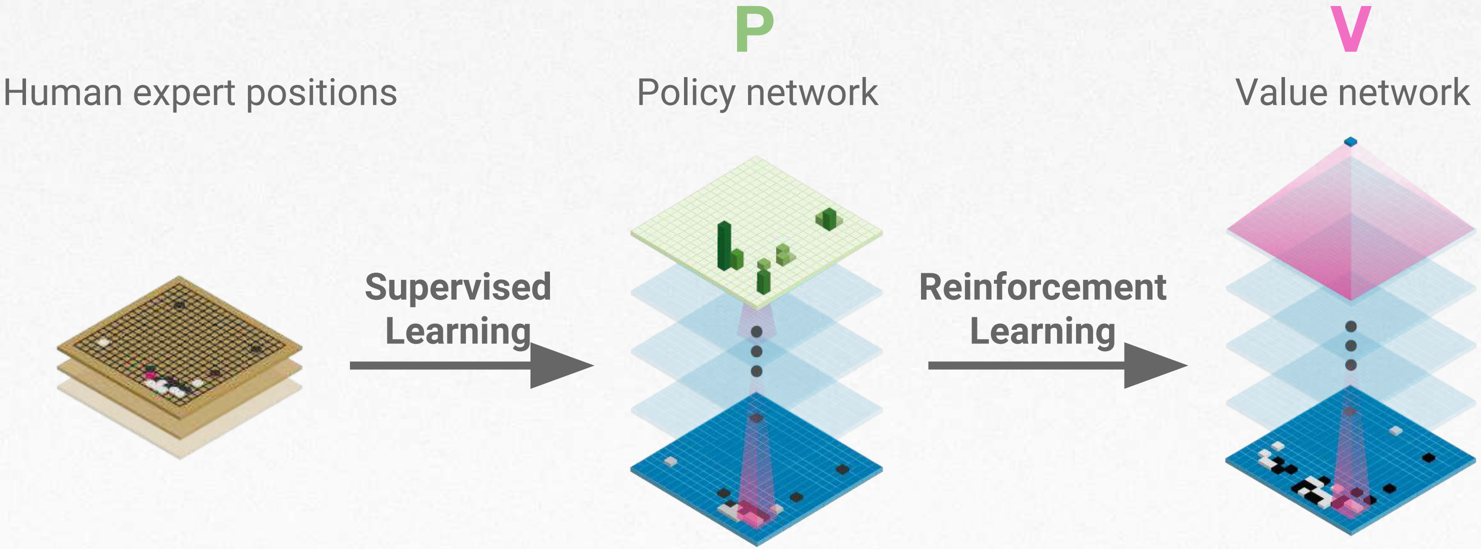
Policy network



Value network



Training AlphaGo



Human expert positions

Supervised Learning

P

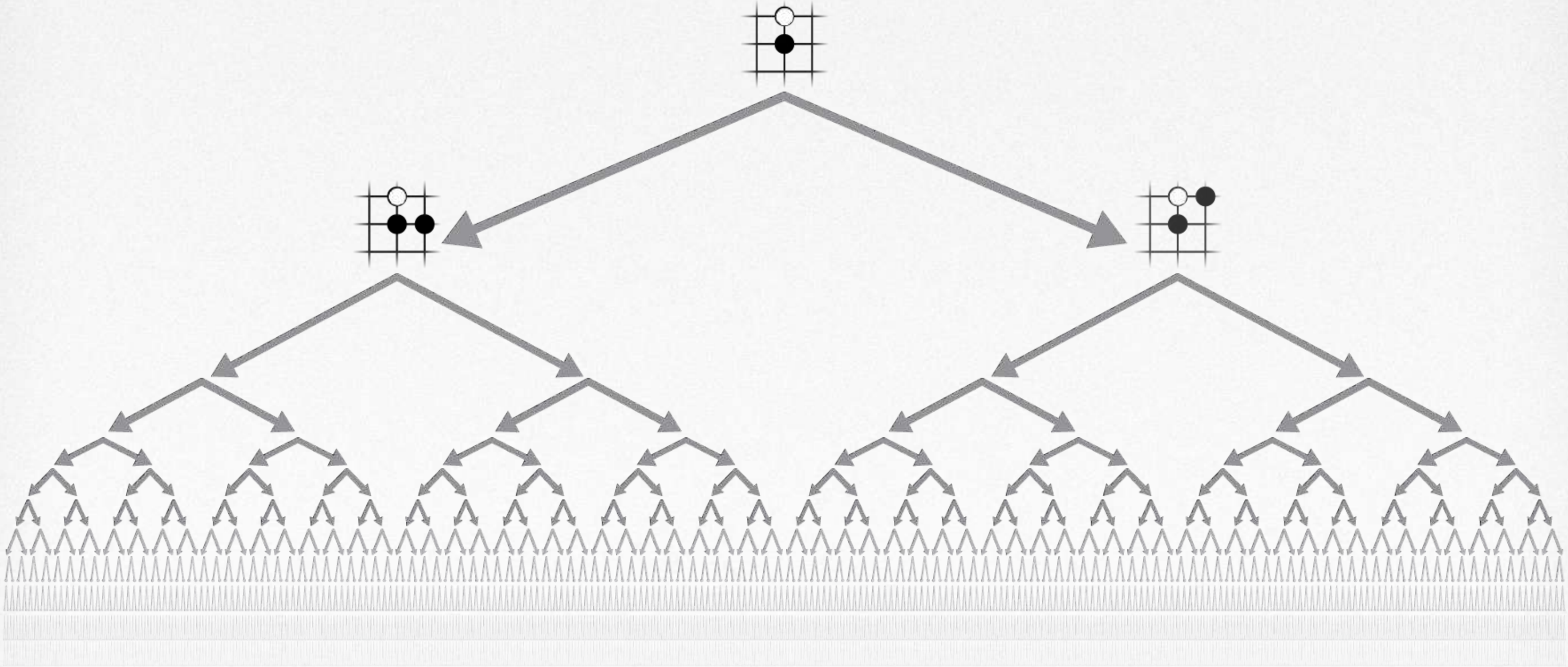
Policy network

Reinforcement Learning

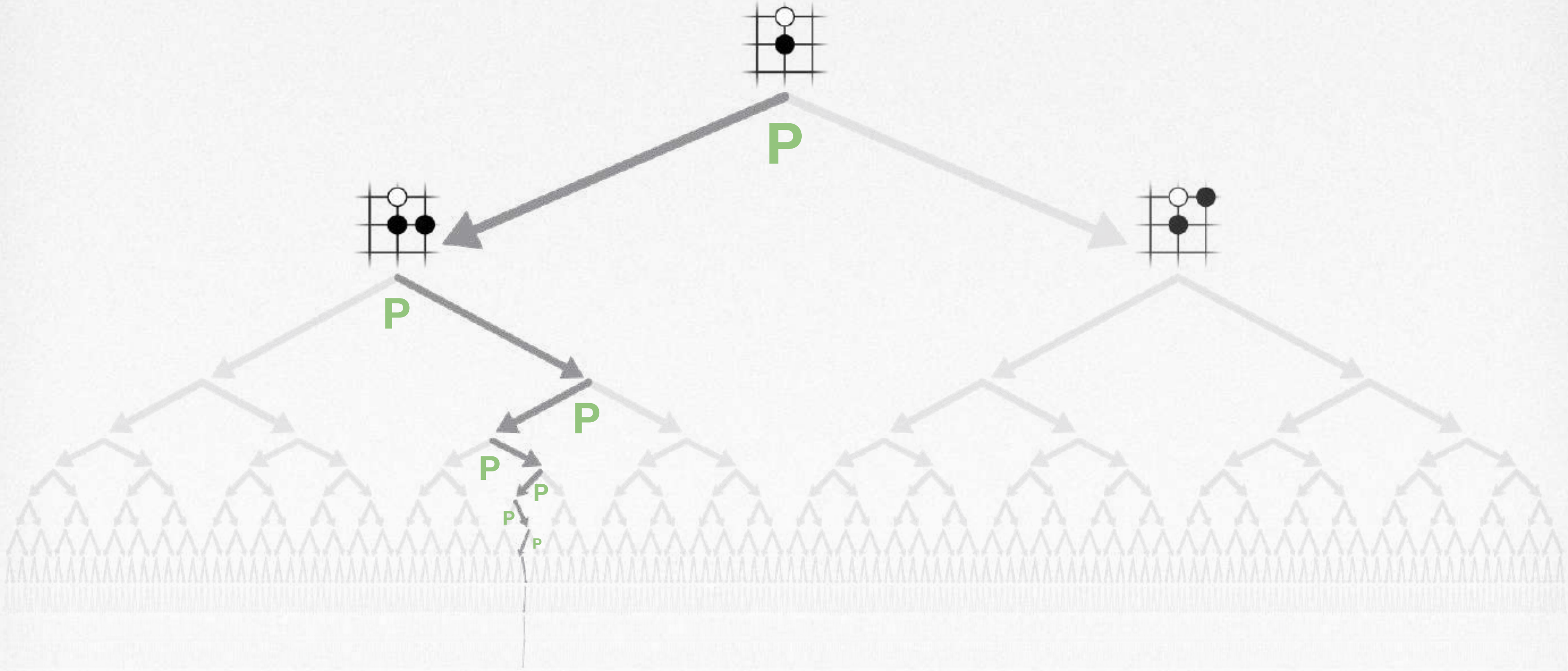
V

Value network

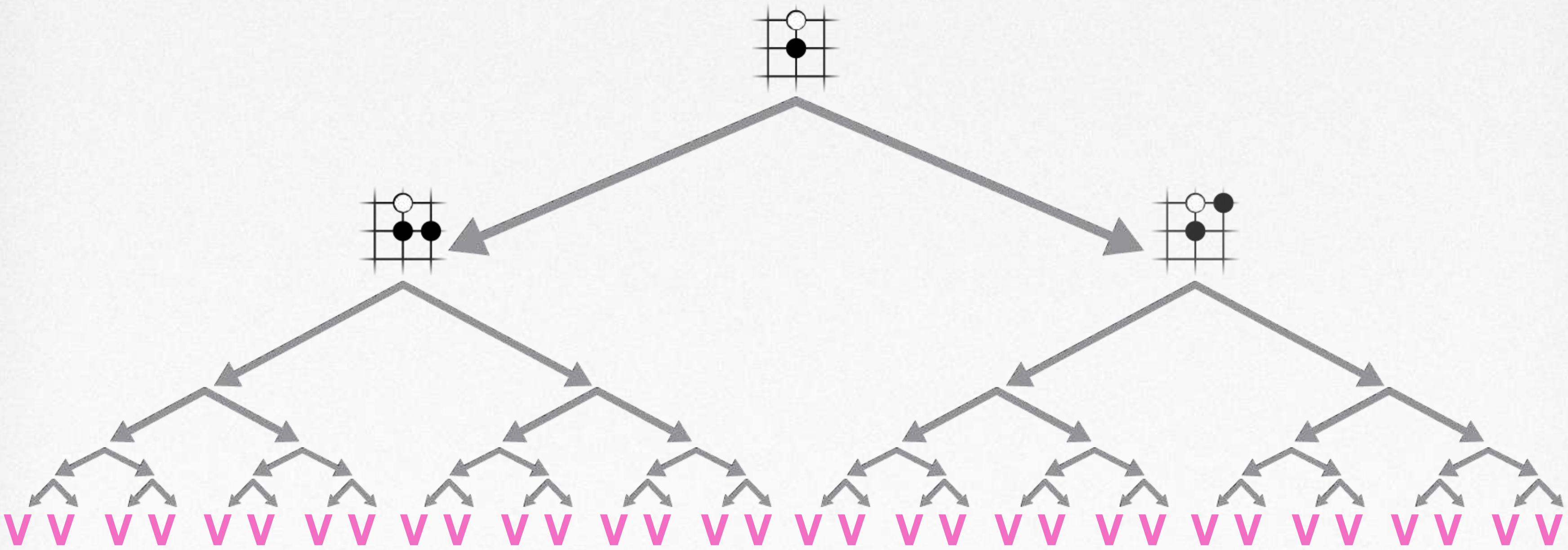
Exhaustive search



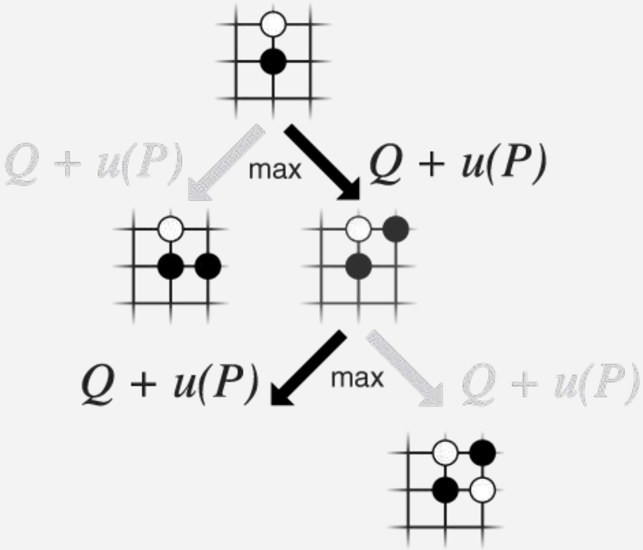
Reducing breadth with policy network



Reducing depth with value network


V V V V V V V V V V V V V V V V V V V V V V V V V V V V V V V V

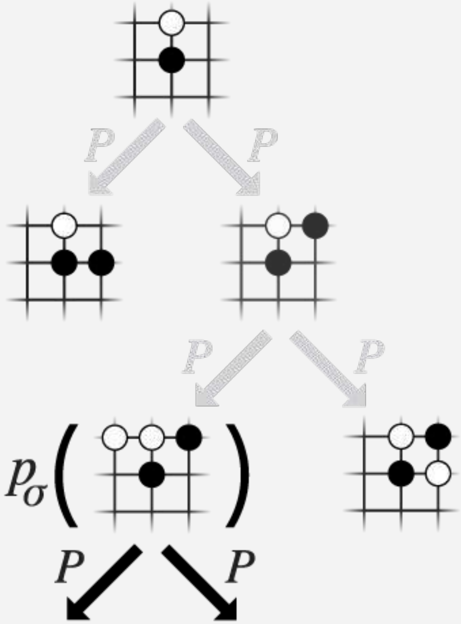
Monte-Carlo tree search in AlphaGo: **selection**



P prior probability
Q action value

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

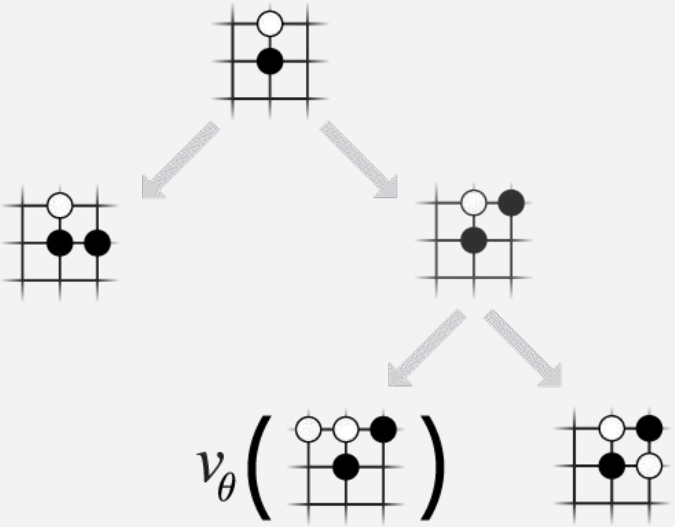
Monte-Carlo tree search in AlphaGo: **expansion**



p_σ Policy network

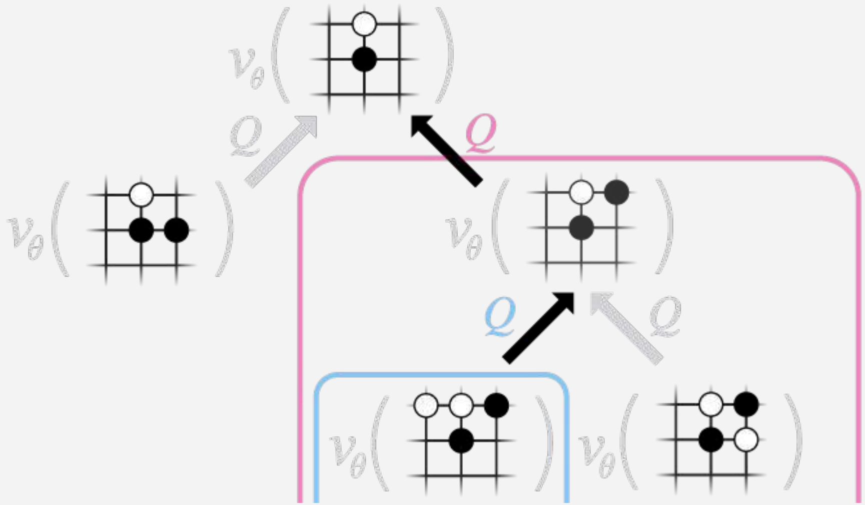
P prior probability

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



v_θ Value network

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



Q Action value

v_θ Value network

AlphaGo

- Plays on 50 TPUs on Google Cloud
- Searches ~50 moves deep
- ~100,000 positions per second

AlphaGo vs Lee Sedol

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

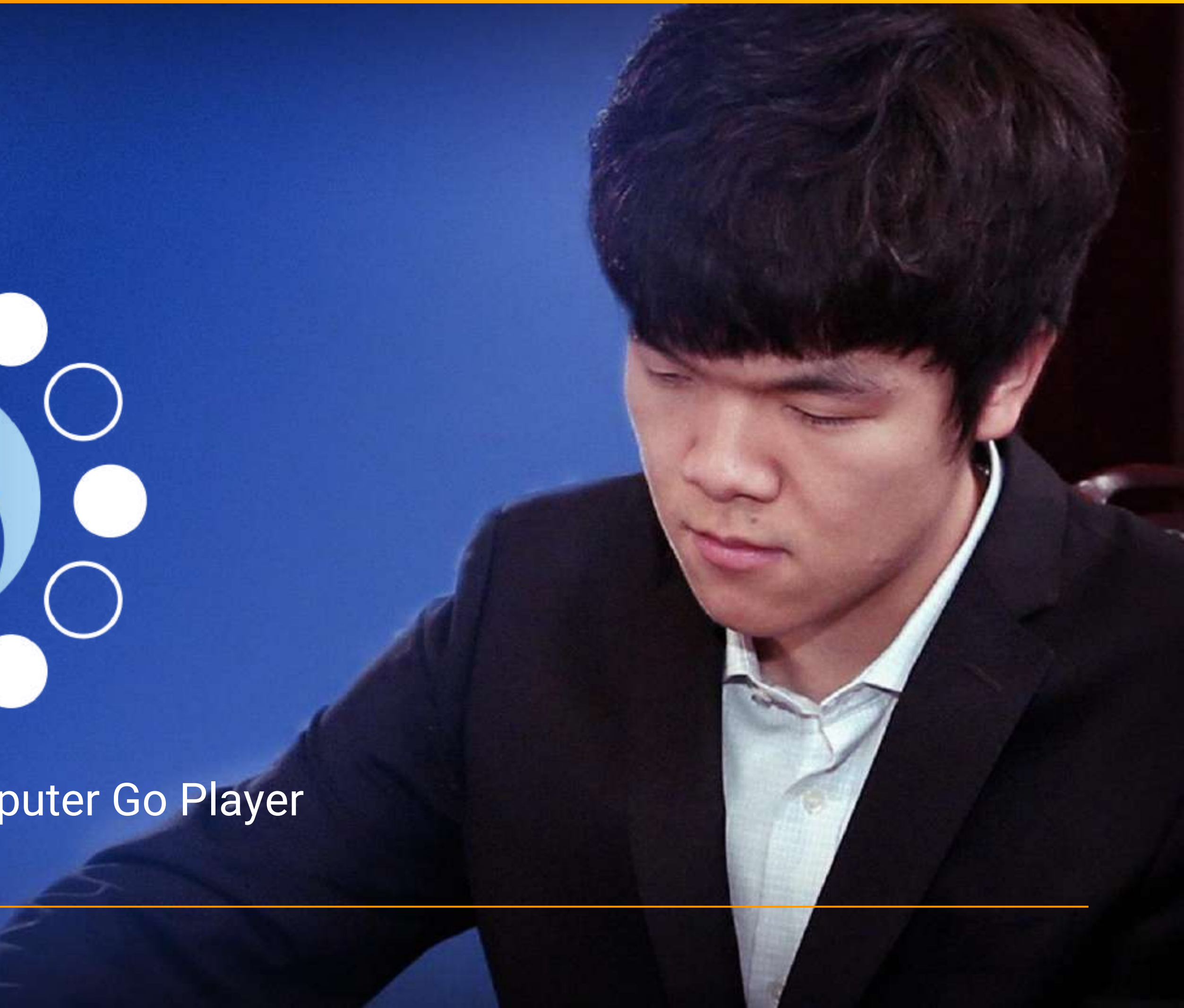
Match was played in Seoul, March 2016

AlphaGo won the match 4-1

KBA
Google DeepMind
Challenge Match
8 - 15 March 2016
AlphaGo

phaGo
KBA
Google DeepMind
Challenge Match
8 - 15 March 2016

AlphaGo Master



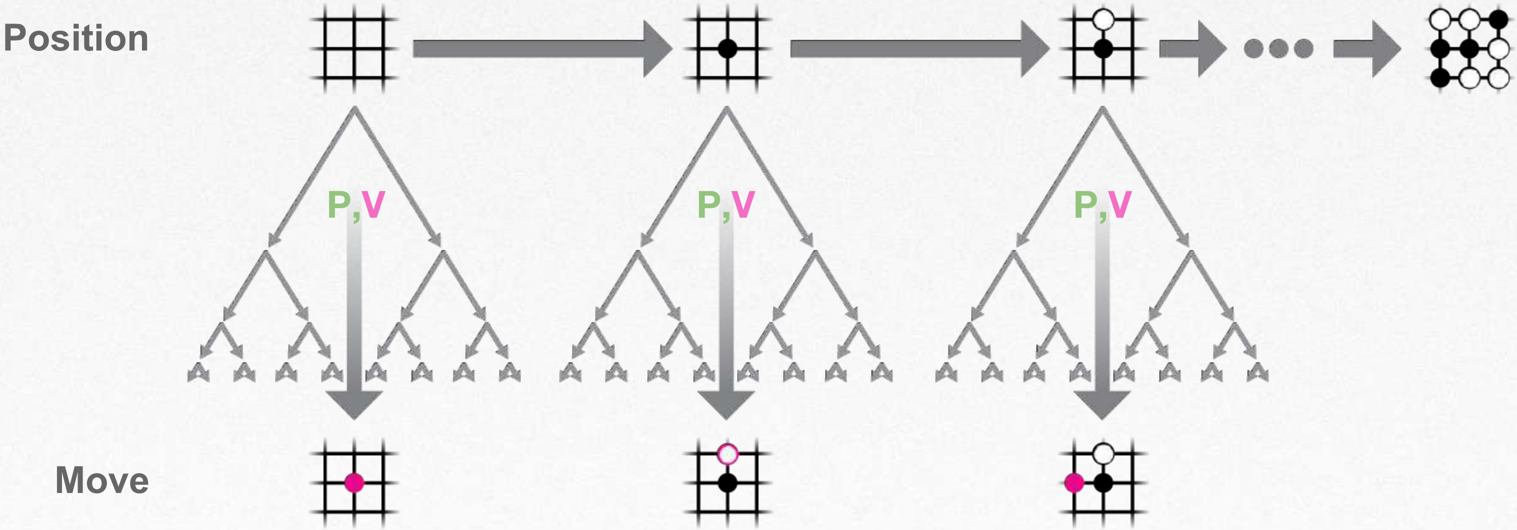
The world’s strongest computer Go Player

Search-Based Policy Iteration

- **AlphaGo becomes its own teacher**
  It learns from its own searches

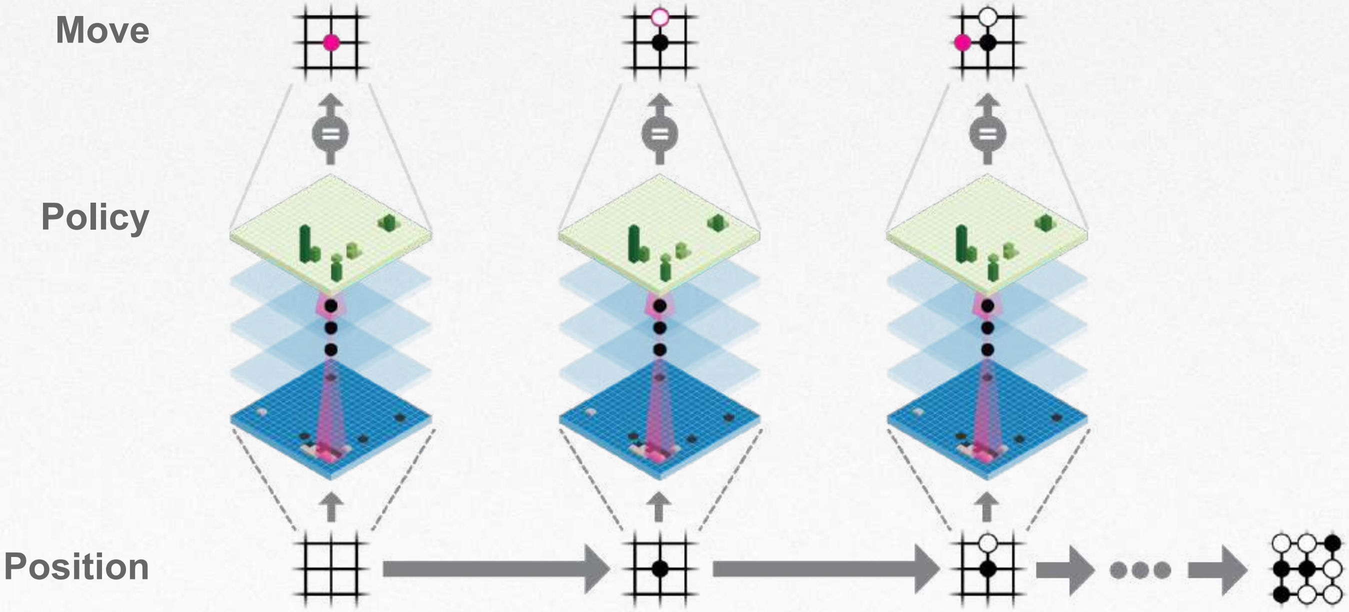
- Policy is **improved** by AlphaGo search
- Policy is **evaluated** according to outcome of AlphaGo vs AlphaGo games

Reinforcement Learning in AlphaGo Master



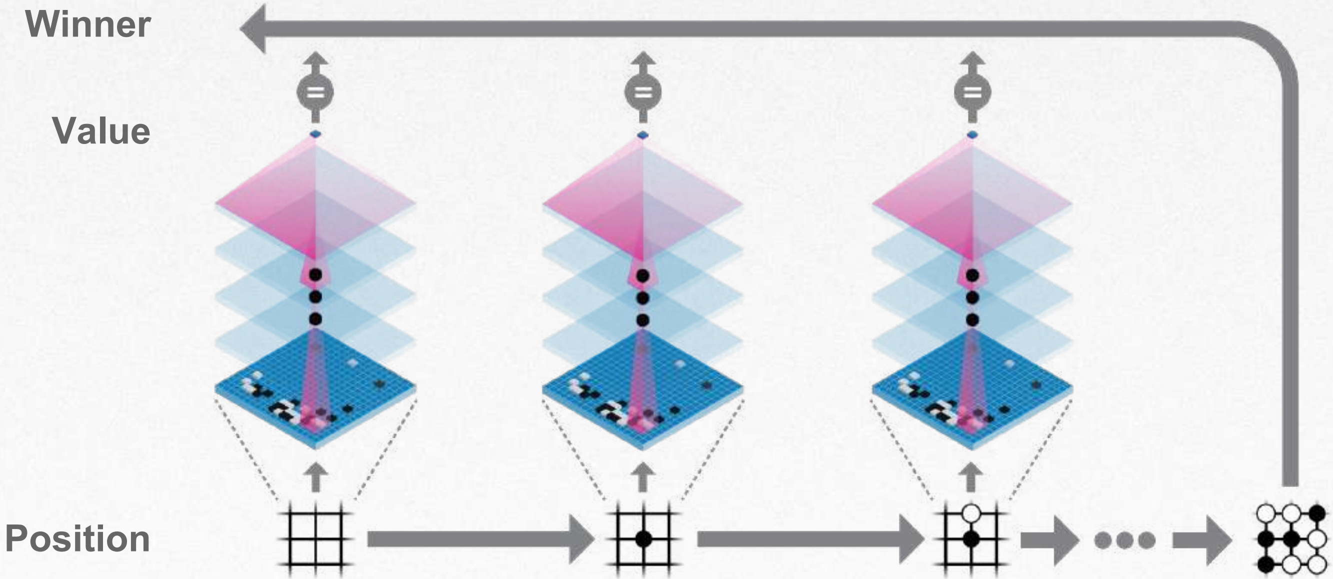
AlphaGo plays games against itself

Reinforcement Learning in AlphaGo Master



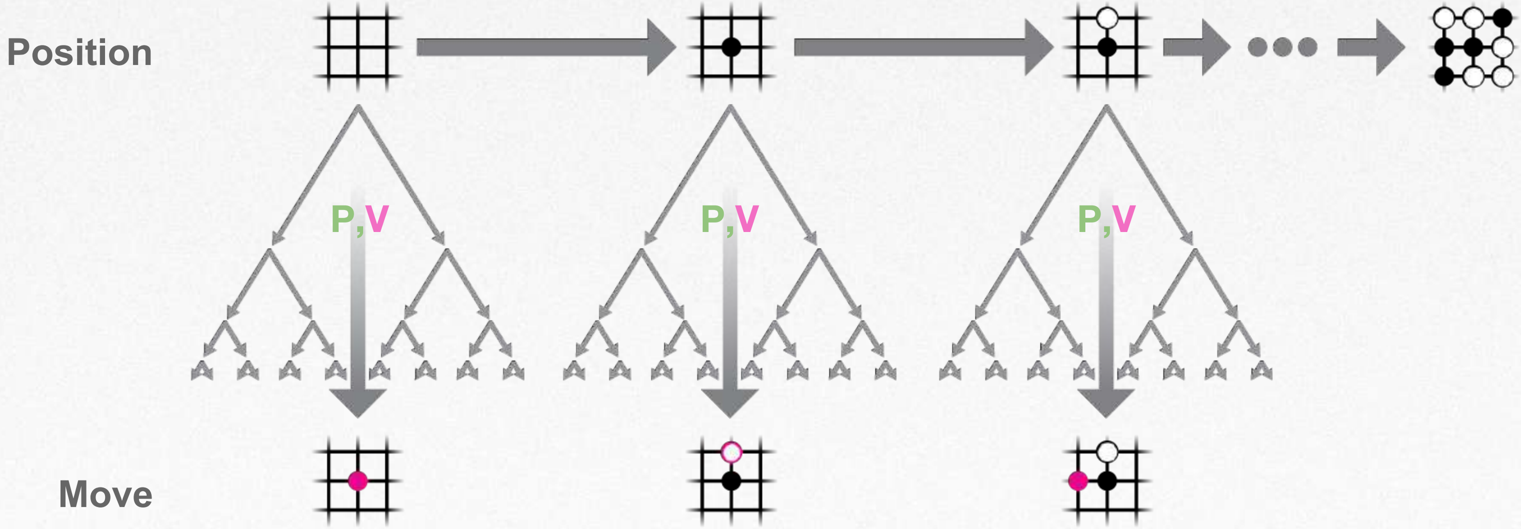
Policy network P is trained to predict AlphaGo’s moves

Reinforcement Learning in AlphaGo Master



Value network **V** is trained to predict winner

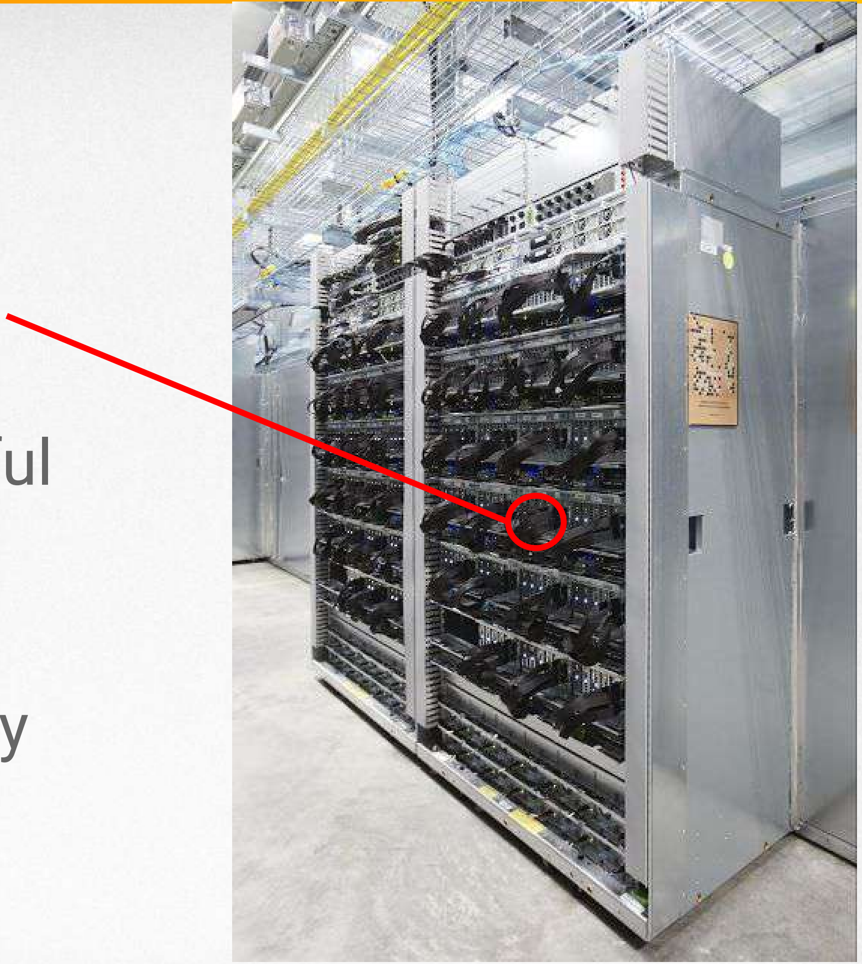
Reinforcement Learning in AlphaGo Master



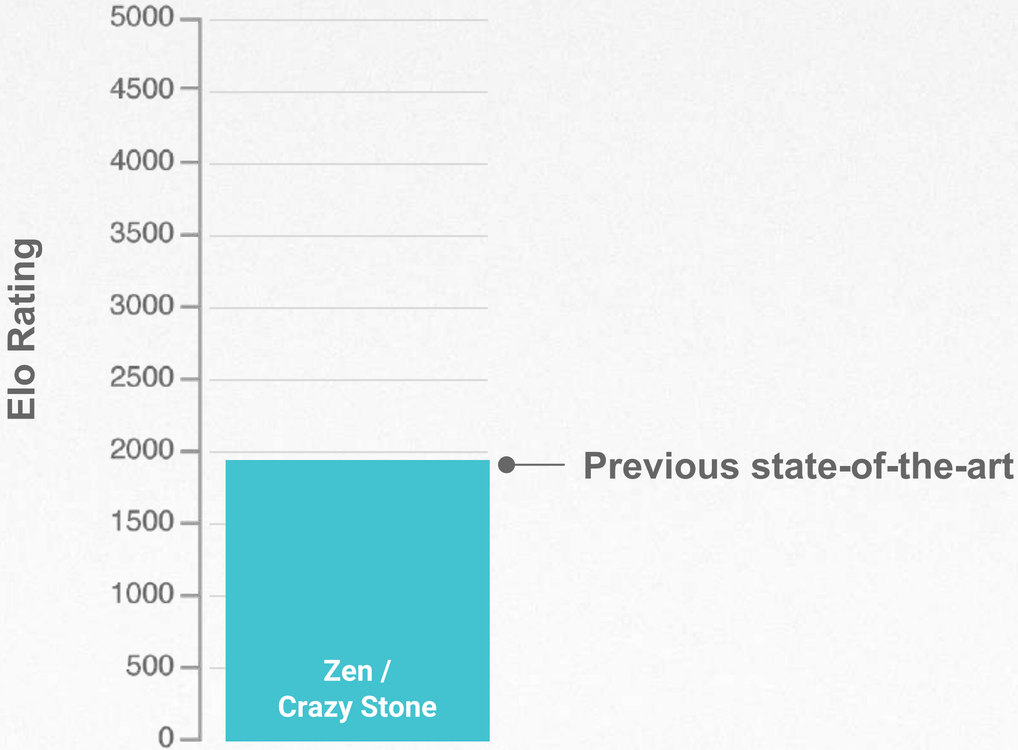
New policy and value network are used in next iteration of AlphaGo

AlphaGo Master

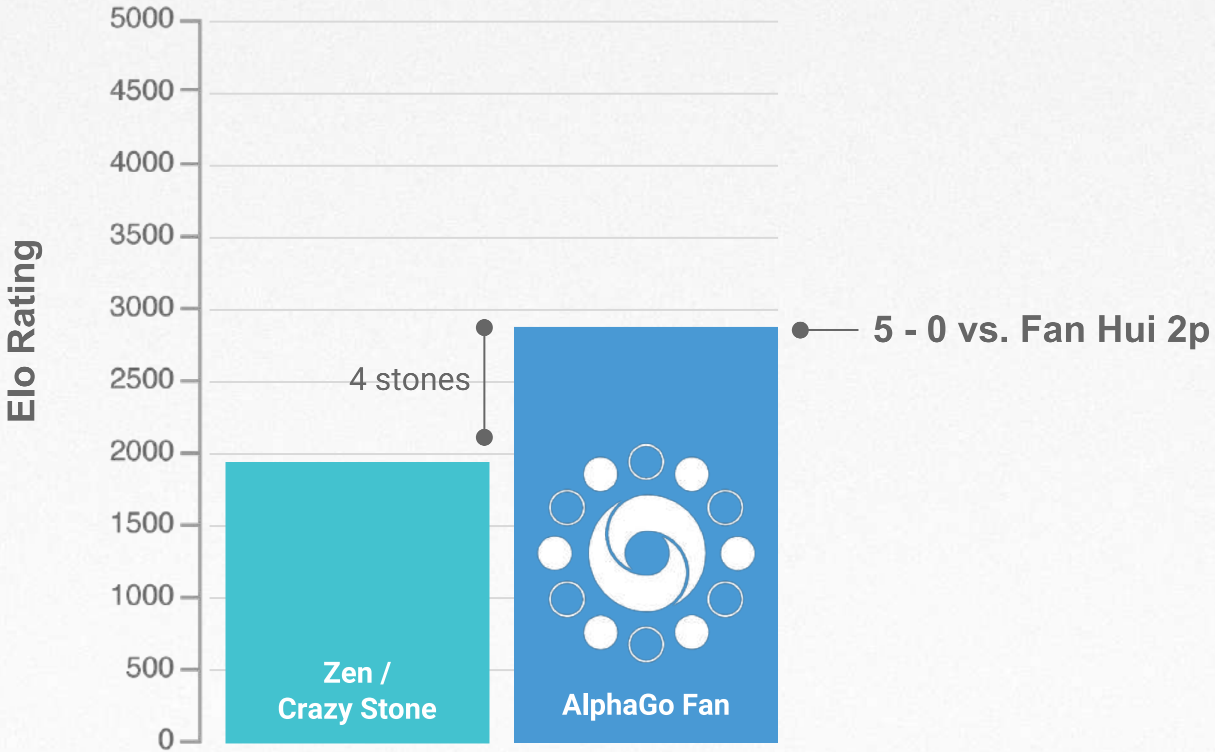
- Plays on single TPU machine
- Uses deeper and more powerful policy/value networks
- Trained by search-based policy iteration



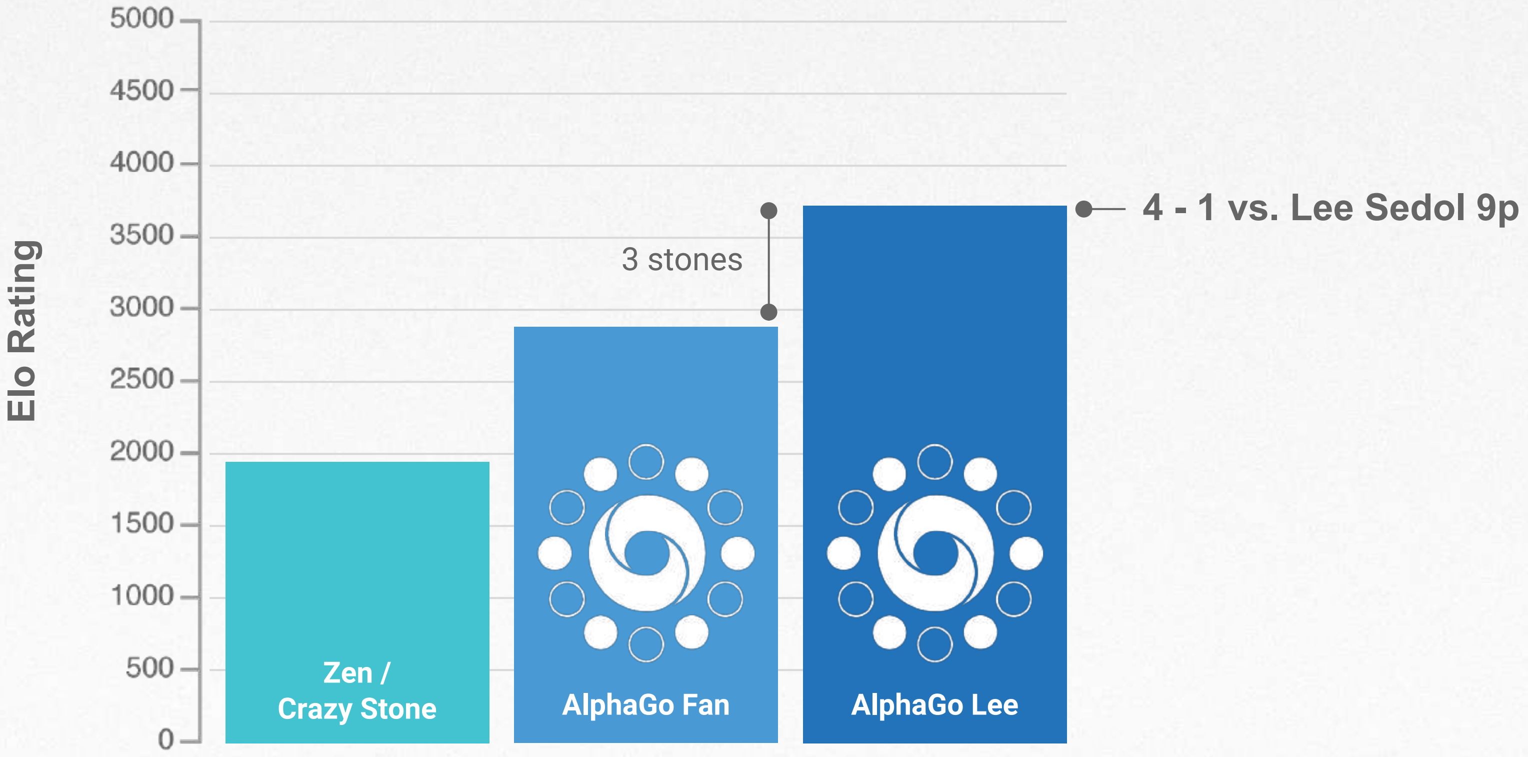
Evaluation of AlphaGo



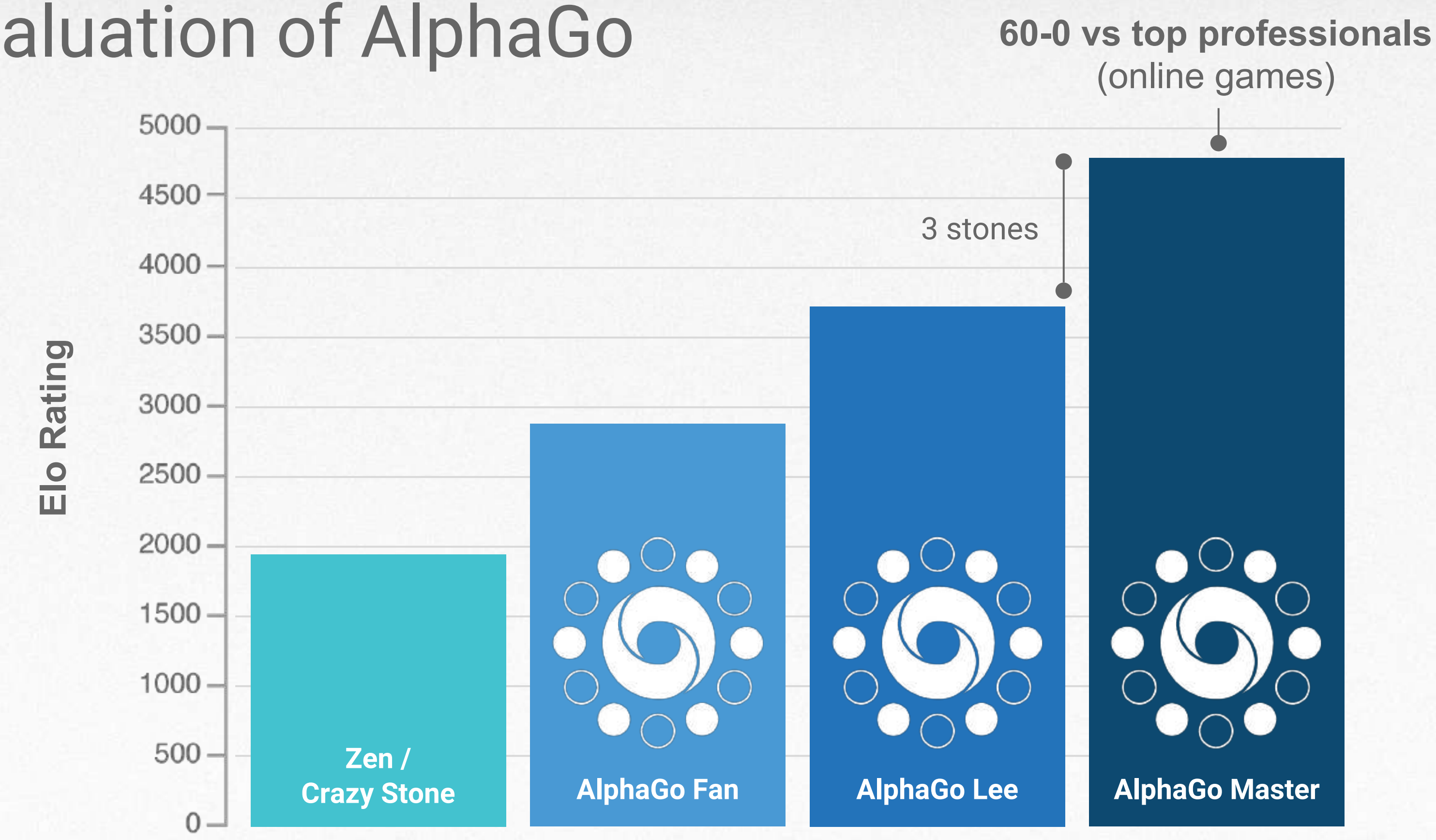
Evaluation of AlphaGo



Evaluation of AlphaGo



Evaluation of AlphaGo



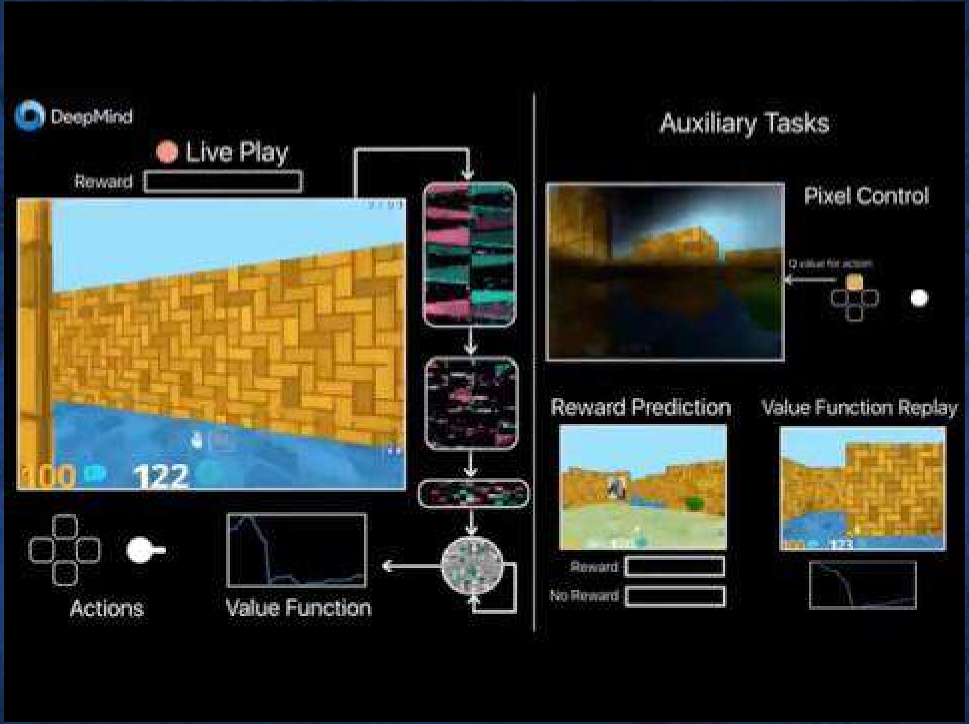
AlphaGo vs Ke Jie

Ke Jie (9p): player ranked #1 in world

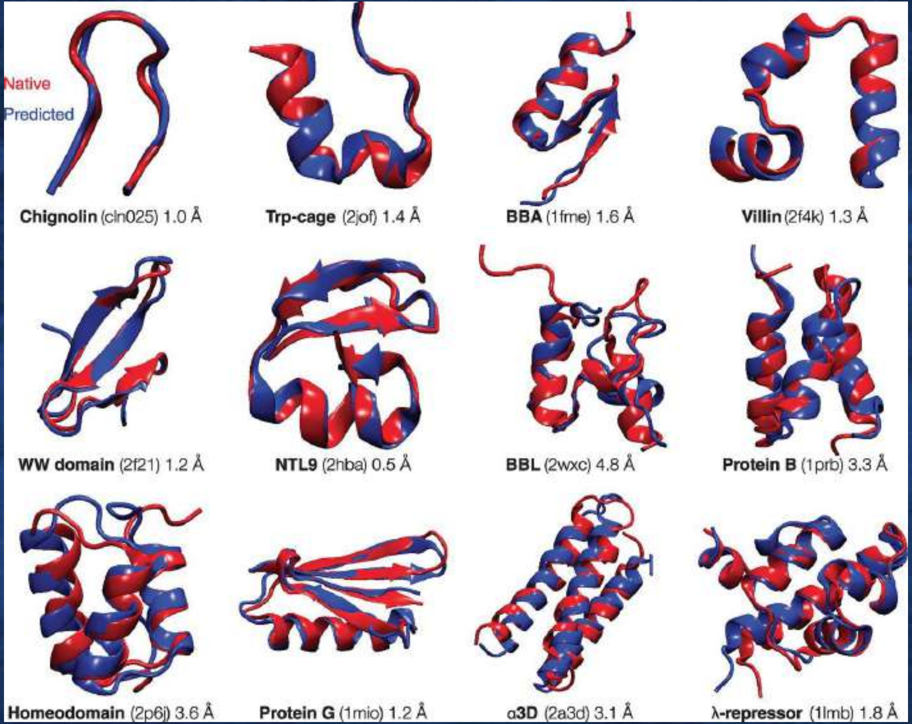
Match was played in China, May 2017

AlphaGo won the match 3-0

Deep Reinforcement Learning: Beyond AlphaGo



Deep Reinforcement Learning: Beyond AlphaGo



Chignolin (cln025) 1.0 Å

Trp-cage (2jof) 1.4 Å

BBA (1fme) 1.6 Å

Villin (2f4k) 1.3 Å

WW domain (2f21) 1.2 Å

NTL9 (2hba) 0.5 Å

BBL (2wxc) 4.8 Å

Protein B (1prb) 3.3 Å

Homeodomain (2p6j) 3.6 Å

Protein G (1mio) 1.2 Å

α3D (2a3d) 3.1 Å

λ-repressor (1lmb) 1.8 Å

Thank you

DeepMind

Google