Two-player Games – Part 2

ZUI 2016/2017

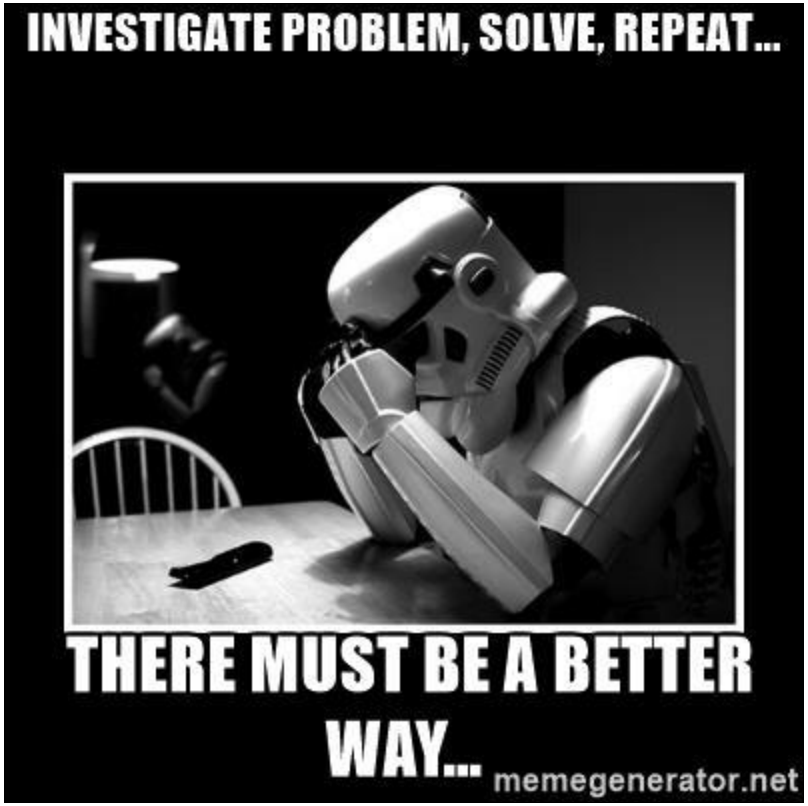
Branislav Bošanský

bosansky@fel.cvut.cz

Previously ... on Two-Player Games

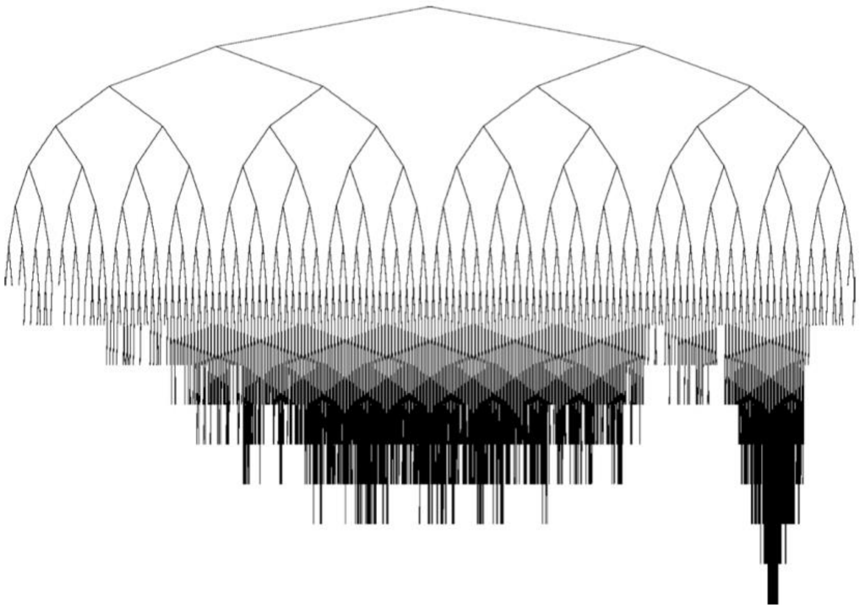
- minimax search
- alpha-beta pruning
- Negascout

- problems with long horizon
  - evaluation function
  - iterative deepening



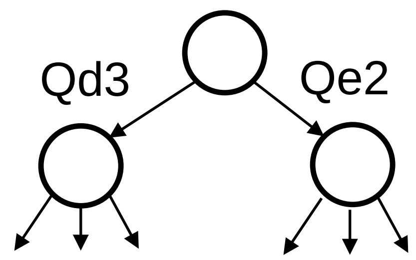
Towards better algorithms …

- we do not want to evaluate all paths equally
- we want to search more deeply (thoroughly) more prospect variants
- we do not want to spend time with bad variants



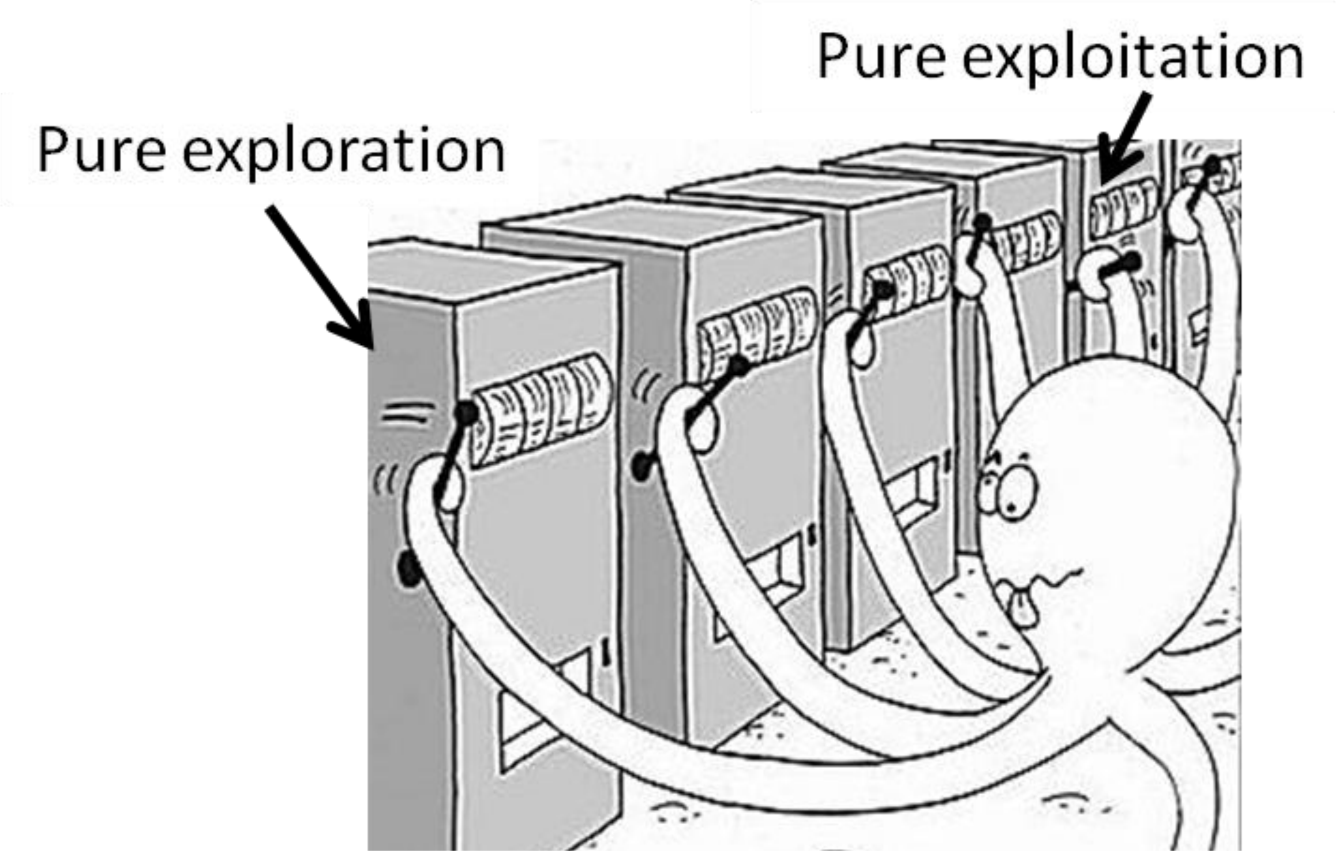
Towards better algorithms …

- let’s start from the beginning



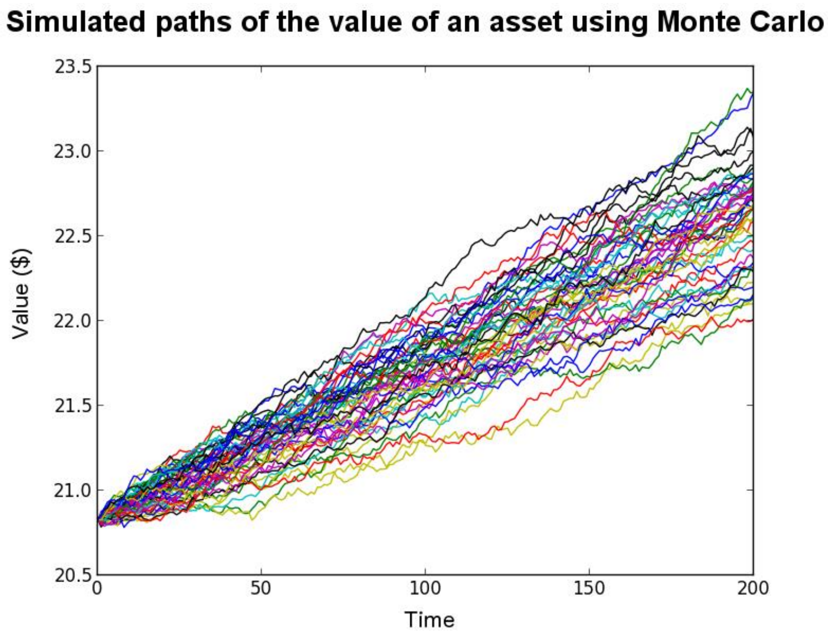
- what if we estimate that Qd3 is (right now) a better move than Qe2
- there is a dilemma
  - either we want to get a better further plan (and thus also an estimate) of the better move (Qd3)
  - or we want to find a better continuation for the worse move (Qe2) – maybe there is one and we’ve just missed it before

Exploration vs. Exploitation



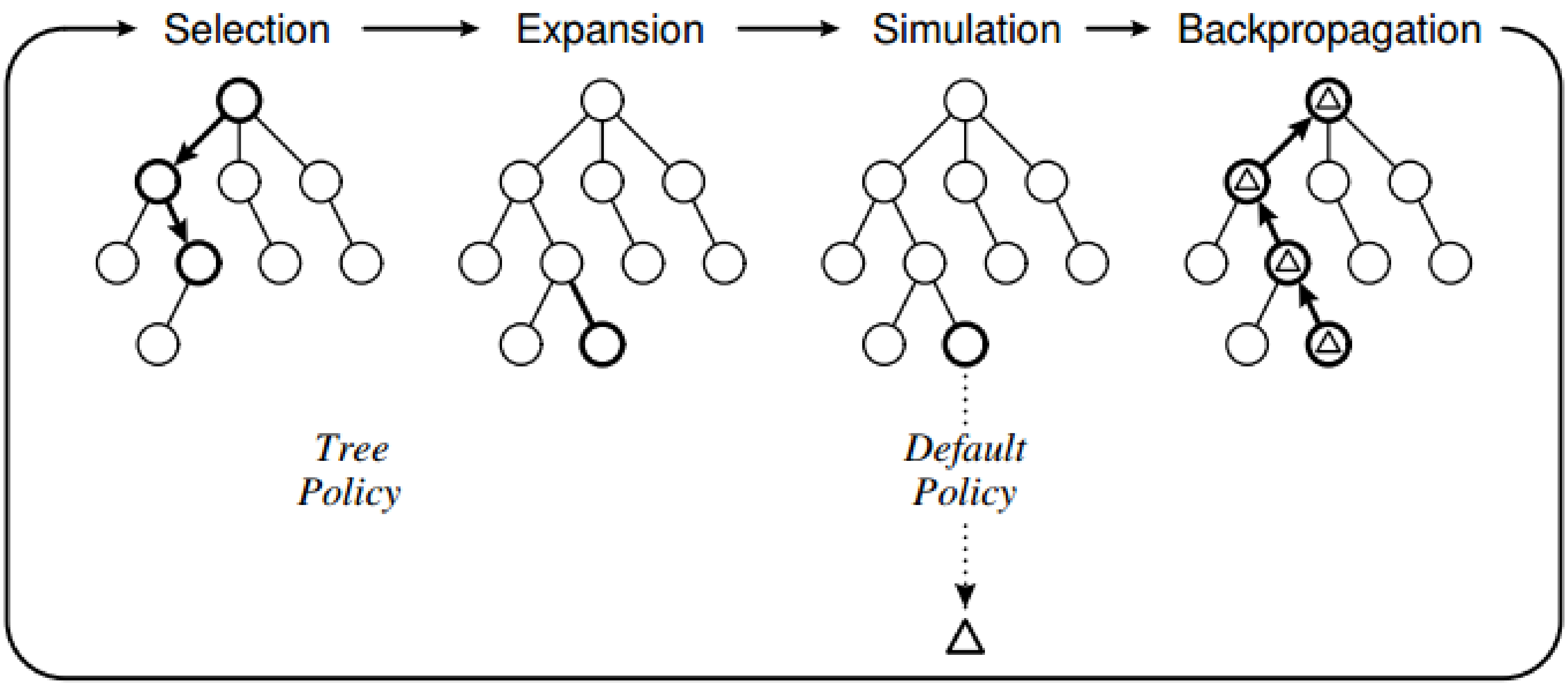
Monte Carlo Methods

- what if we do not have evaluation function?
  - we can estimate the value of the position with a Monte Carlo method
  - from a given position we perform random samples until the terminal position of the game
  - the more samples we perform, the better estimate of the true value we get

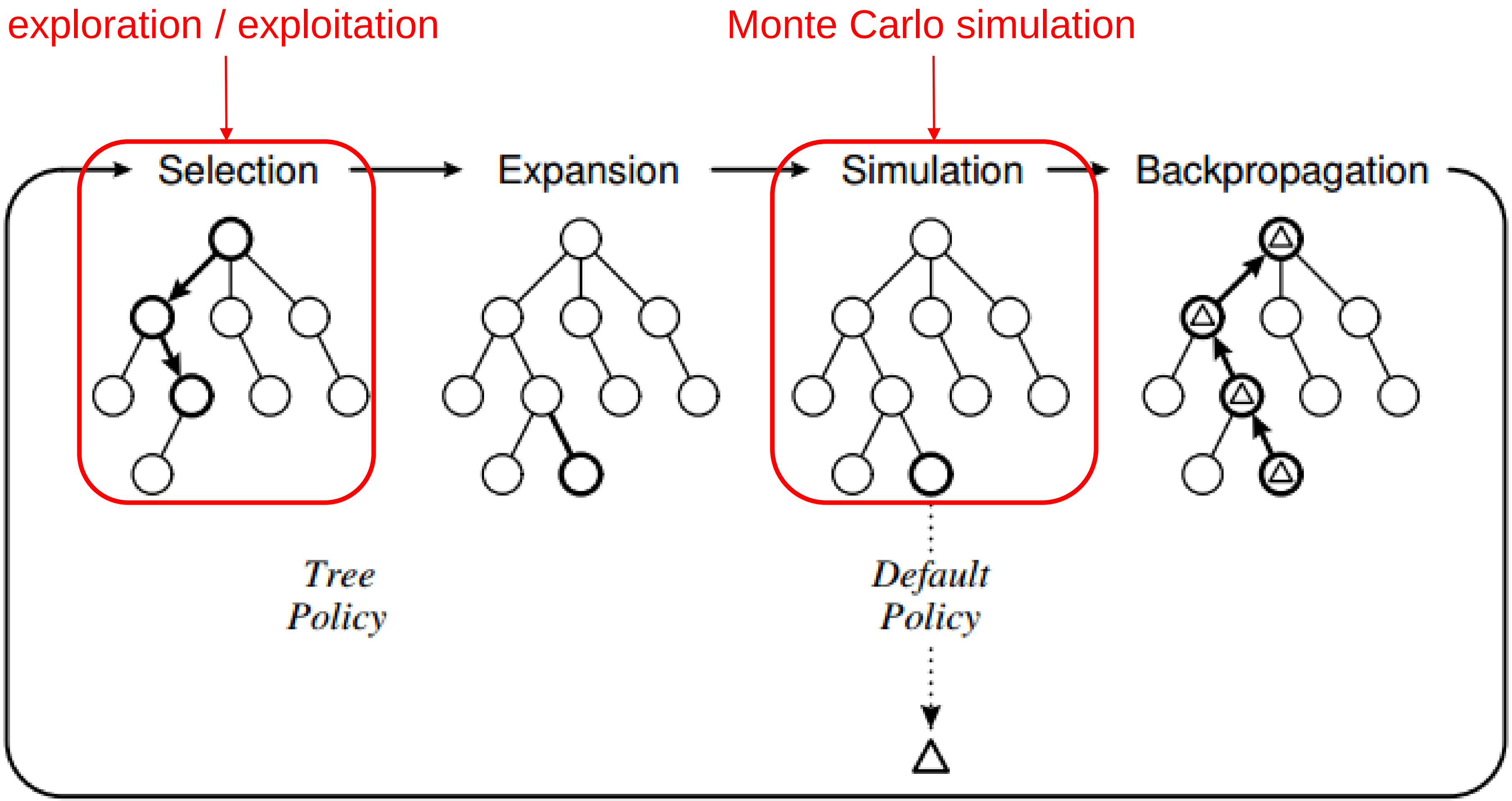


Monte Carlo Tree Search

putting it all together

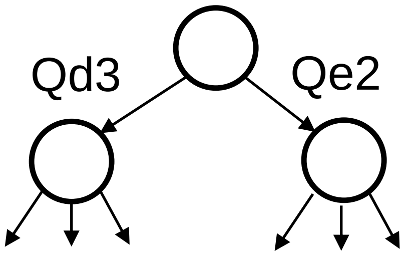


Monte Carlo Tree Search



Exploration vs. Exploitation

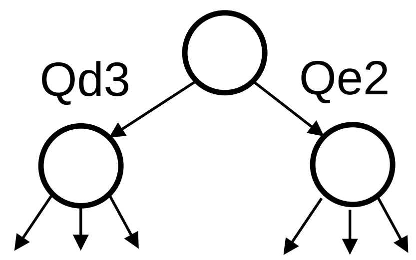
- bandit theory



- UCB – upper confidence bounds
  - $\arg\max_{v' \in children(v)} \frac{Q(v')}{N(v')} + c\sqrt{\frac{2\ln N(v)}{N(v')}}$

Exploration vs. Exploitation

- bandit theory



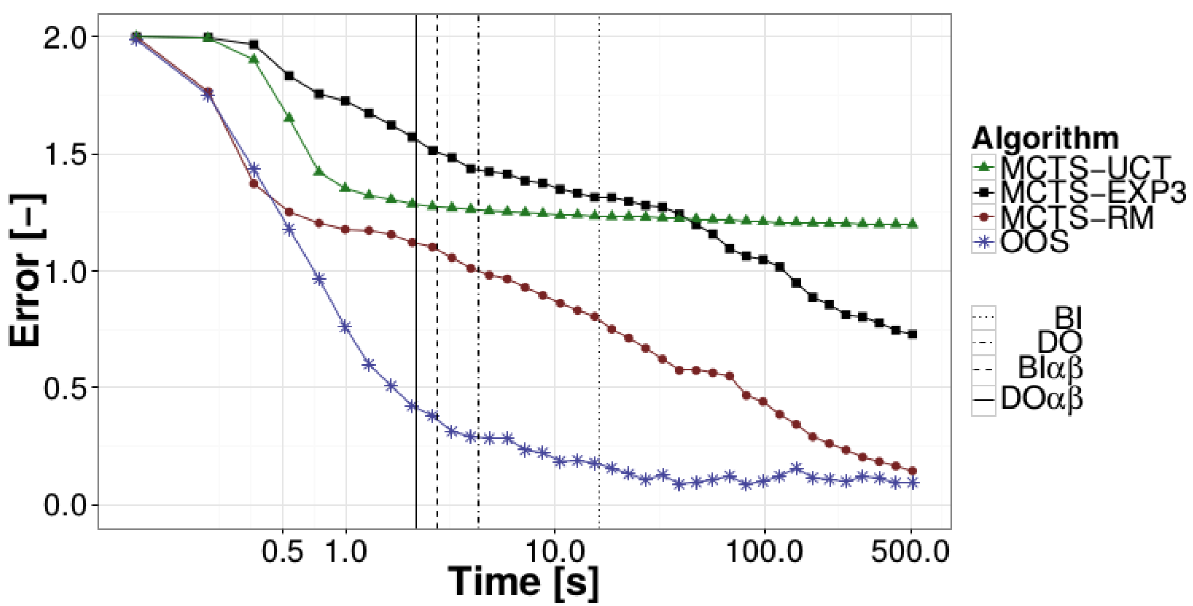
- UCB – upper confidence bounds
  - $\arg\max_{v' \in children(v)} \frac{Q(v')}{N(v')} + c\sqrt{\frac{2 \ln N(v)}{N(v')}}$

average utility

exploration factor

Exploration vs. Exploitation

- many existing variants for the bandit problem
  - UCB1
  - EXP3
  - UCB-V
  - …
- can have a very different performance in practice



MCTS and Parameter Tuning

- Different bandit methods can have different parameters
- Practical performance depends on the correct choice
- The choice is domain dependent
- The choice is opponent dependent (!)

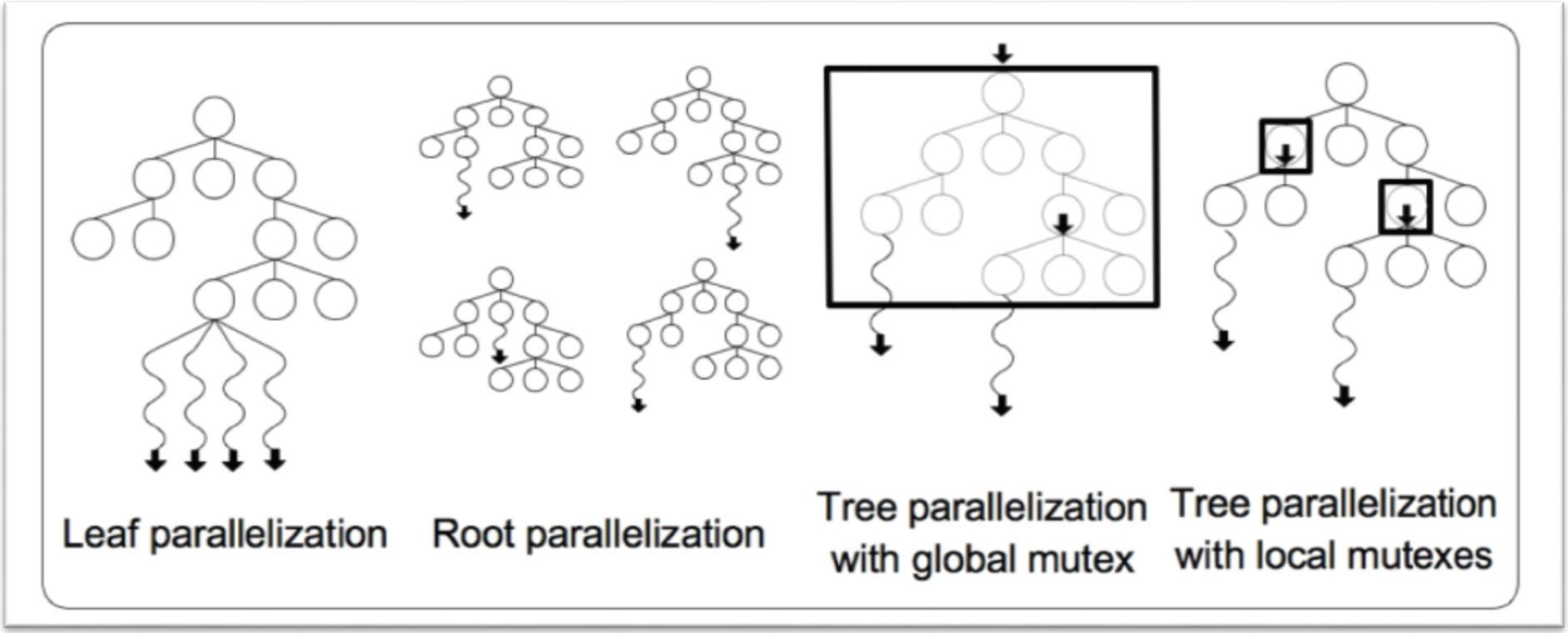
MCTS and Parameter Tuning

		DOαβ	OOS(0.6)	UCT(2)	EXP3(0.2)	RM(0.1)	Mean
OOS	0.5	35.3(2.9)	50.9(3.6)	28.5(3.3)	54.9(3.6)	43.7(3.5)	42.66
OOS	0.4	35.0(2.9)	56.0(3.6)	26.6(3.2)	56.1(3.6)	42.6(3.6)	43.26
OOS	**0.3**	36.5(3.0)	57.8(3.5)	27.7(3.2)	55.7(3.6)	44.8(3.6)	**44.5**
OOS	0.2	35.0(2.9)	53.1(3.6)	26.8(3.2)	54.1(3.6)	41.4(3.5)	42.08
OOS	0.1	34.6(2.9)	55.6(3.6)	24.1(3.1)	56.2(3.6)	43.0(3.6)	42.7
UCT	1.5	83.2(2.2)	74.0(3.8)	79.1(2.9)	87.4(2.9)	70.6(3.9)	78.86
UCT	1	83.8(2.1)	74.8(3.7)	81.4(2.7)	89.8(2.6)	68.8(4.0)	79.72
UCT	**0.8**	86.5(2.0)	77.9(3.6)	77.1(3.0)	89.2(2.7)	74.1(3.8)	**80.96**
UCT	0.6	89.4(1.8)	75.7(3.7)	54.9(3.9)	90.0(2.6)	74.1(3.7)	76.82
UCT	0.4	75.8(2.6)	75.0(3.7)	31.4(3.7)	89.8(2.6)	70.6(3.9)	68.52
EXP3	0.9	47.8(3.1)	68.2(2.8)	23.1(2.4)	67.2(2.8)	55.2(2.8)	52.3
EXP3	**0.8**	46.9(3.1)	68.4(3.6)	23.0(3.1)	74.2(3.4)	61.5(3.7)	**54.8**
EXP3	0.6	42.5(3.1)	67.6(3.7)	20.4(3.1)	65.4(3.7)	59.4(3.8)	51.06
EXP3	0.5	38.7(3.0)	60.9(3.8)	15.1(2.7)	64.7(3.7)	52.9(3.9)	46.46
EXP3	0.4	35.9(3.0)	57.5(3.9)	17.5(3.0)	64.1(3.8)	54.9(3.9)	45.98
RM	0.5	44.5(3.0)	41.1(3.5)	31.7(3.3)	49.4(3.6)	34.3(3.3)	40.2
RM	0.3	42.8(3.0)	52.1(3.5)	33.8(3.4)	61.2(3.5)	43.7(3.5)	46.72
RM	0.2	41.8(3.0)	55.7(3.6)	30.7(3.3)	59.2(3.5)	46.4(3.6)	46.76
RM	**0.1**	37.0(2.9)	58.1(3.5)	34.9(3.4)	57.6(3.6)	54.1(3.6)	**48.34**
RM	0.05	36.4(3.0)	59.6(3.5)	29.7(3.3)	59.3(3.5)	51.1(3.6)	47.22

Heuristics and MCTS

- there are several points where MCTS can benefit from domain-specific heuristic
  - progressive unpruning/widening
    - standard MCTS adds all children
  - heavy rollout simulations
    - simulations do not have to be completely random
    - tradeoff between bias and complexity vs. speed
  - using evaluation function instead of simulation
    - often combined with previous

Parallelization of MCTS



Variants of MCTS

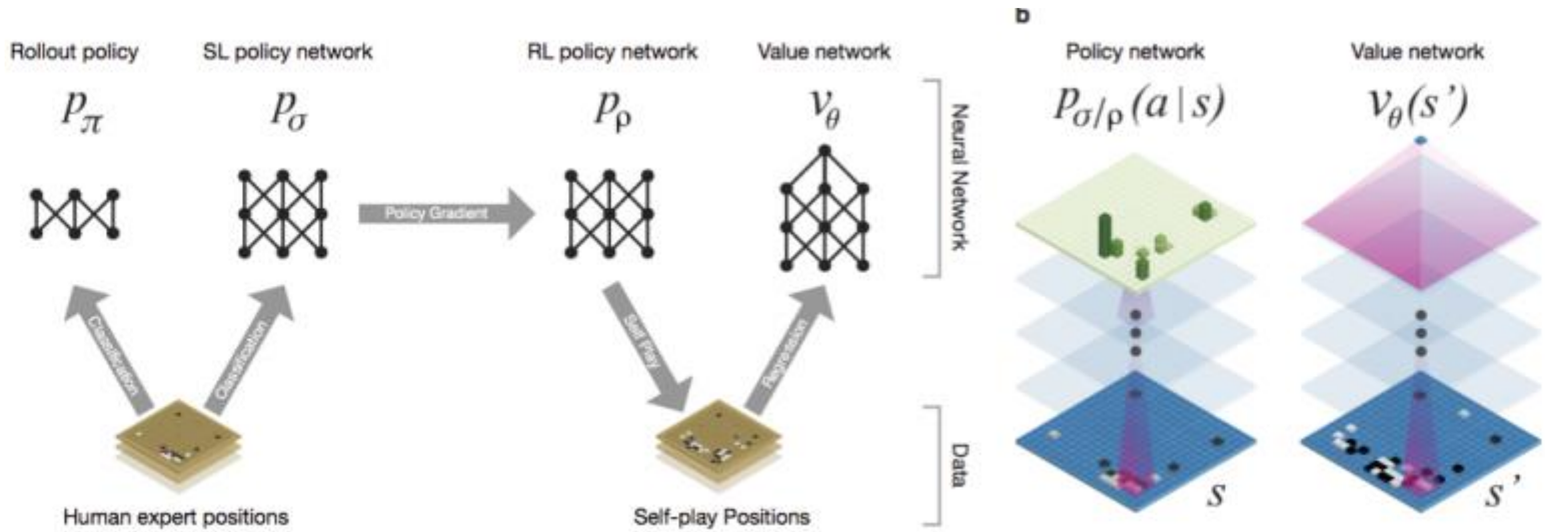
- there are **many** improvements and variants of MCTS
  - (see ``A Survey of Monte Carlo Tree Search Methods'' by Browne et al. 2012)

	Go	Phantom Go	Blind Go	NoGo	Multi-player Go	Hex	Y, Star, Renkula!	Havannah	Lines of Action	P-Game	Clobber	Othello	Amazons	Arimaa	Khet	Shogi	Mancala	Blokus Duo	Focus	Chinese Checkers	Yavalath	Connect Four	Tic Tac Toe	Sum of Switches	Chess	LeftRight Games	Morpion Solitaire	Crossword	SameGame	Sudoku, Kakuro	Wumpus World	Mazes,Tigers,Grids	CadiaPlayer	Ary
Flat MC/UCB	+										+		+																					
BAST																																		
TDMC(λ)												+																						
BB Active Learner																																		
UCT	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			+	?								+	+
SP-MCTS																													+					
FUSE																																		
MP-MCTS																			+	+														
Coalition Reduction																																		
Multi-agent MCTS	+																																	
Ensemble MCTS	+																																	
HOP																																		
Sparse UCT																																		
Info Set UCT																																		
Multiple MCTS																							+											
UCT+																																		
$MC_{\alpha\beta}$									+																?									
MCCFR																																		
Reflexive MC																											+							
Nested MC																										+	+		+	+				+
NRPA																											+	+						
HGSTS																											+		+					
FSSS, BFS3																															+			
TAG																																		
UNLEO																																		
UCTSAT																																		
ρUCT																							+									+		
MRW																																		
MHSP																																		
UCB1-Tuned																																		
Bayesian UCT																																		
EXP3																																		
HOOT																																		
First Play Urgency	+																																	
(Anti)Decisive Moves	+			+				+													+													
Move Groups	+									+																								
Move Ordering	+											+				+																		
Transpositions	+									+				+	+											+								
Progressive Bias	+																																	
Opening Books	+												+																					
MCPG																																		
Search Seeding	+																																	
Parameter Tuning	+																																	
History Heuristic	+													+		+			+	+													+	
AMAF	+	+						+																			+							
RAVE	+	+	+	+	+	+		+						+										+									+	
Killer RAVE								+																										
RAVE-max																								+										
PoolRAVE	+							+																										
MCTS-Solver									+																									
MC-PNS	+																																	
Score Bounded MCTS	+																					+												
Progressive Widening	+												+																					
Pruning	+					+							+									+												
Contextual MC	+							+																										
Fill the Board	+					+	+																											
MAST, PAST, FAST																																	+	
Simulation Balancing	+																																	
Last Good Reply	+							+																										
Patterns	+					+	+	+																										
Score Bonus																								+										
Decaying Reward										+																								
Leaf Parallelisation	+													+			+																+	
Root Parallelisation	+													+																				
Tree Parallelisation	+																																	+
UCT-Treesplit	+																																	+

	Tron	Ms. Pac-Man	Pocman, Battleship	Dead-End	Wargus	ORTS	Skat	Bridge	Poker	Dou Di Zhu	Klondike Solitaire	Magic: The Gathering	Phantom Chess	Urban Rivals	Backgammon	Settlers of Catan	Scotland Yard	Roshambo	Thurn and Taxis	OnTop	Security	Mixed Integer Prog.	TSP, CTP	Sailing Domain	Physics Simulations	Function Approx.	Constraint Satisfaction	Schedul. Benchmarks	Printer Scheduling	Rock-Sample Problem	PMPs	Bus Regulation	Large State Spaces	Feature Selection	PCG
Flat MC/UCB		+	+	+				+	+			+							+	+						+									
BAST																										+									
TDMC(λ)																																			
BB Active Learner																																			
UCT	+	+		+	+	+			+		+		+	+	+				+	+	+	+	+	+	+		+	+		+	+		+		+
SP-MCTS																													+						
FUSE																																		+	
MP-MCTS		+														+																			
Coalition Reduction																	+																		
Multi-agent MCTS																																			
Ensemble MCTS																																			
HOP											+																								
Sparse UCT													+																						
Info Set UCT							+			+																									
Multiple MCTS																																			
UCT+									+																										
$MC_{\alpha\beta}$																																			
MCCFR									+																										
Reflexive MC																																			
Nested MC																							+			+						+			
NRPA																																			
HGSTS																																			
FSSS, BFS3																																	+		
TAG																										+									
UNLEO																																			
UCTSAT																											+								
ρUCT			+						+									+																	
MRW																										+									
MHSP																												+							
UCB1-Tuned	+																																		
Bayesian UCT																																			
EXP3														+																					
HOOT																									+										
First Play Urgency																																			
(Anti)Decisive Moves																																			
Move Groups																																			
Move Ordering																																			
Transpositions																																			
Progressive Bias	+																																		
Opening Books																																			
MCPG																																			+
Search Seeding																																			
Parameter Tuning																																			
History Heuristic																																			
AMAF													+																						
RAVE													+																						
Killer RAVE																																			
RAVE-max																																			
PoolRAVE																																			
MCTS-Solver	+																																		
MC-PNS																																			
Score Bounded MCTS																																			
Progressive Widening																															+				
Pruning				+																															
Contextual MC																																			
Fill the Board																																			
MAST, PAST, FAST																																			
Simulation Balancing																																			
Last Good Reply																																			
Patterns																																			
Score Bonus																																			
Decaying Reward																								+											
Leaf Parallelisation																																			
Root Parallelisation																																			
Tree Parallelisation																																			
UCT-Treesplit																																			

MCTS and Alpha Go

- a combination of statistical and symbolic AI
- integration of learned heuristic functions in a MCTS framework



Games and Game Theory

- one shot simultaneous-move games
  - Rock-Paper-Scissors
- sequential games with simultaneous moves
  - Tron, many card games, …
  - alpha-beta algorithm can be generalized
- games with imperfect information

Two Player Games

- Important test environment for AI algorithms
- Benchmark of AI
  - Chinook (1994/96) – world champion in checkers
  - Deep Blue (1997) – beats G. Kasparov in chess (3.5 – 2.5)
  - …
  - Alpha Go (2016) – beats Lee Sedol in Go (4 – 1)
  - **DeepStack (2016/2017) – beats Poker Pros (https://www.deepstack.ai/)**
  - …

Invitation

Artificial Intelligence Goes All-In: Computers Playing Poker

Photos by John Ulan from the University of Alberta

Lecture by prof. Michael Bowling, head of Computer Poker Research Group at University of Alberta.

Prof. Michael Bowling

- World-famous expert on AI and reinforcement learning
- Led many outstanding computer poker results:
  - Polaris, beating pros in heads-up limit poker
  - Cepheus, playing optimally heads-up limit poker
  - **DeepStack, beating pros in heads-up no-limit**
  - Two publications on poker in prestigious Science
- Proposed Atari games as a benchmark for AI
- Won one of the first RoboCup challenges

March 30, 2017 at 16:00

Auditorium KN:E-107, FEL CTU,

Karlovo nám. 13, Prague 2

UNIVERZITA KARLOVA
Matematicko-fyzikální fakulta

COMPUTER SCIENCE

FACULTY OF ELECTRICAL ENGINEERING CTU IN PRAGUE

Games and Game Theory in AIC

- more fundamental research
  - general algorithms for solving sequential games with imperfect information
  - implementation of domain independent algorithms