Ensemble Monte-Carlo Planning: An Empirical Study

Alan Fern and Paul Lewis

Computer Science

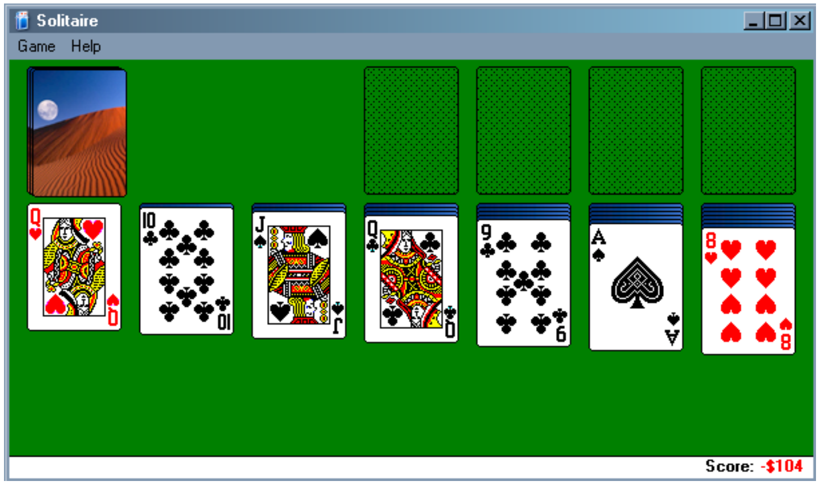
Oregon State University

Talk Outline

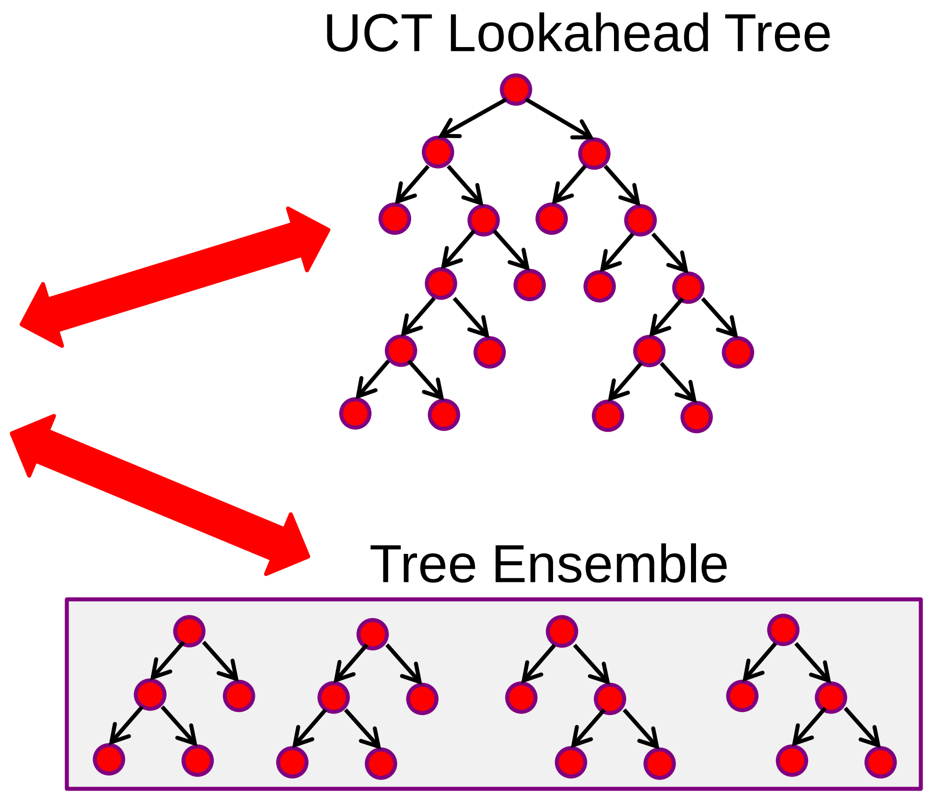
- Motivation
- Monte-Carlo Planning with UCT
- Ensembles
- Domains & Results

Klondike Solitaire

Bjarnason et. al., ICAPS-2009



HOP Planner (27% win rate)



- UCT planner worked surprisingly well (34.4%)
- Voting across small UCT trees worked better (37%)
  - Using less total time!

Parallel UCT in Go

- Several proposals to parallelize UCT with experiments in Go
  (Cazenave, et. al., 2007) (Gelly, et. al., 2008)
- Simply voting of independent UCTs worked best!

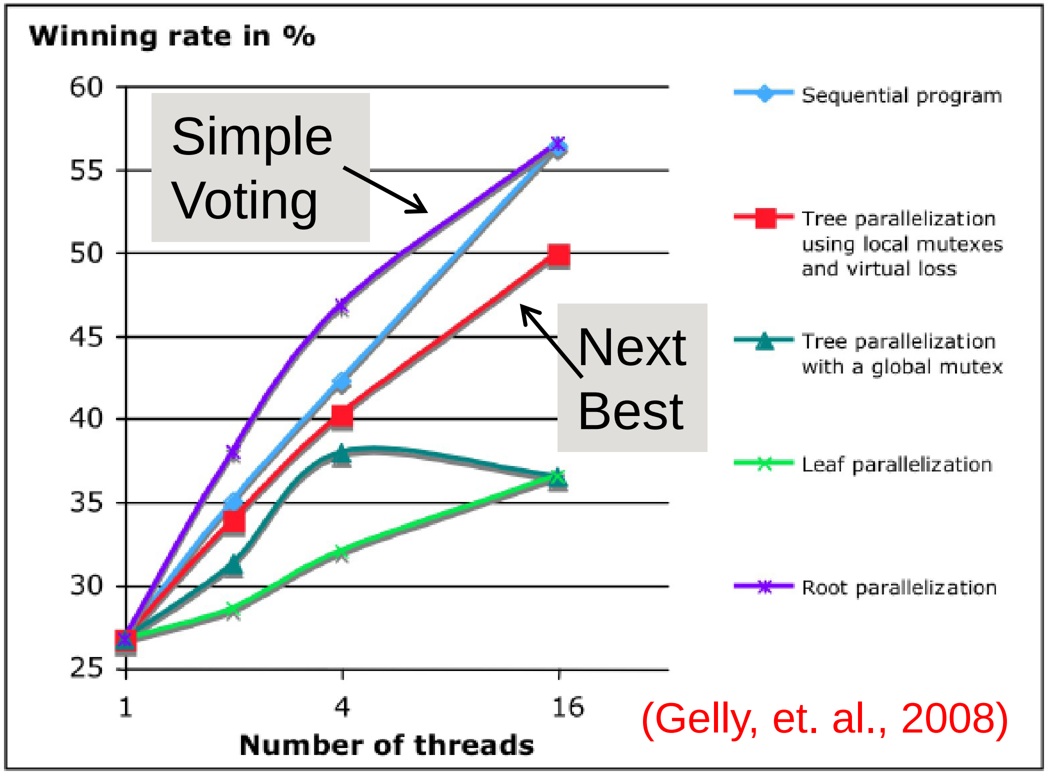
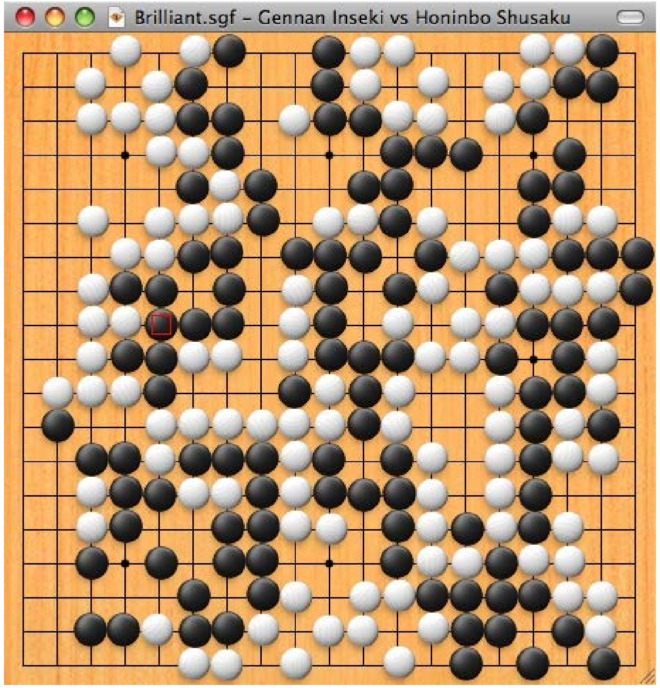
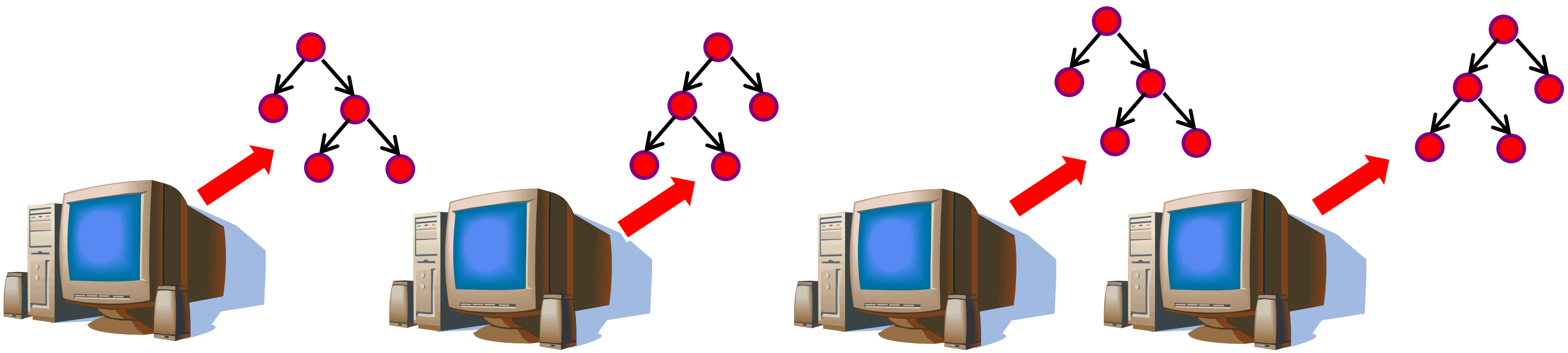


Fig. 5. Performance of the different parallelization methods

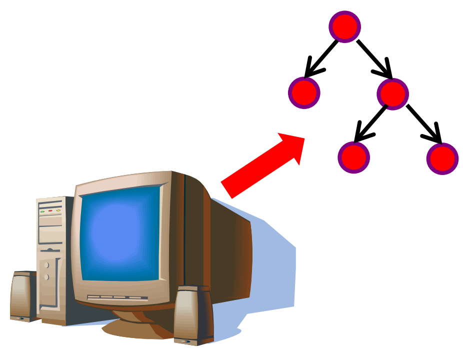


Prior Observations: Multi-Core

- Parallel Time Advantage
  - More CPUs showed significant improvement

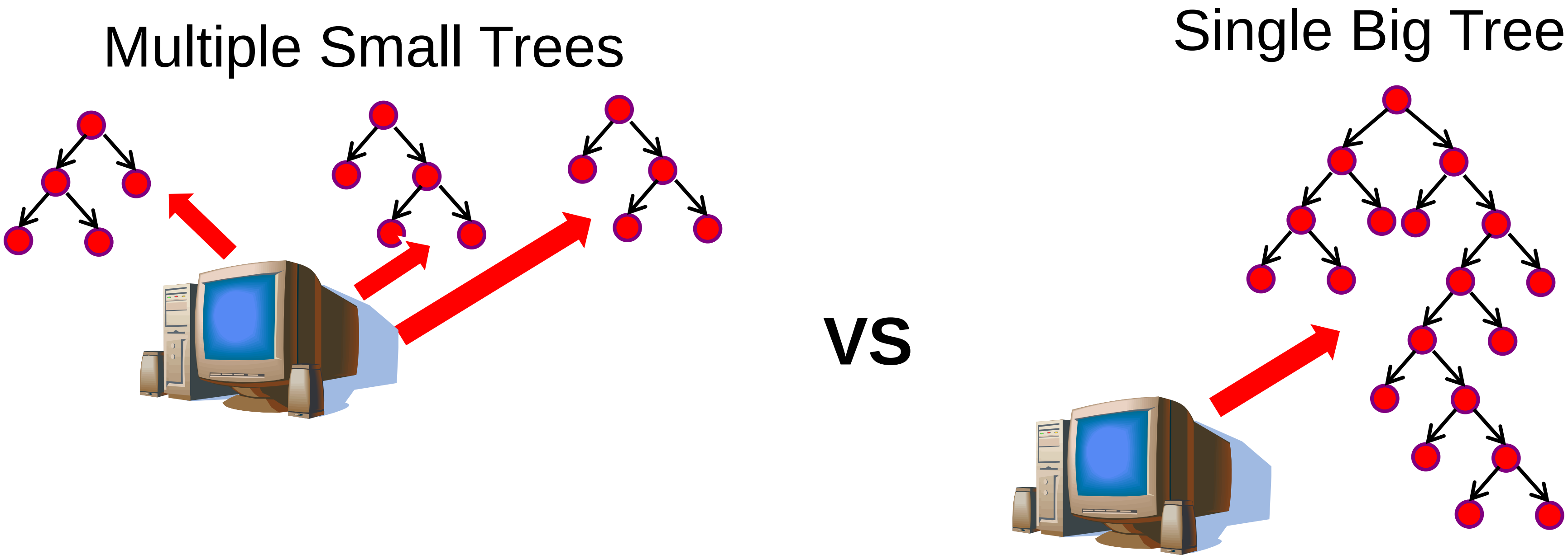


VS



Prior Observations: Single-Core

- Single-Core Space Advantage
  - Single core only needs memory for single (smaller) tree
- Single-Core Time Advantage
  - Ensembles show advantages w.r.t. total CPU time



Objective

- Prior observations about UCT ensembles are limited in scope
  - Domains limited to Go and Solitaire
  - Limited ensemble configurations
- **Our Goal:** provide evidence for or against prior observations
  - Consider 6 domains (other than Go and Solitaire)
  - Test a regular grid of ensemble configurations

Talk Outline

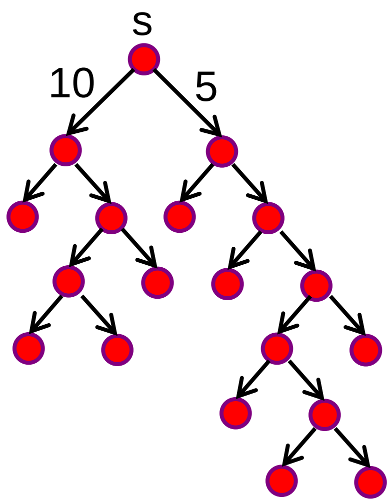
- Motivation
- Monte-Carlo Planning with UCT
- Ensembles
- Domains & Results

UCT Algorithm [Kocsis & Szepesvari, 2006]

- UCT is an instance of Monte-Carlo Tree Search
  - Single-agent problems or games
  - Stochastic and deterministic problems
- Major advance in computer Go
- A growing number of success stories
- Practical successes not well understood

Monte-Carlo Tree Search

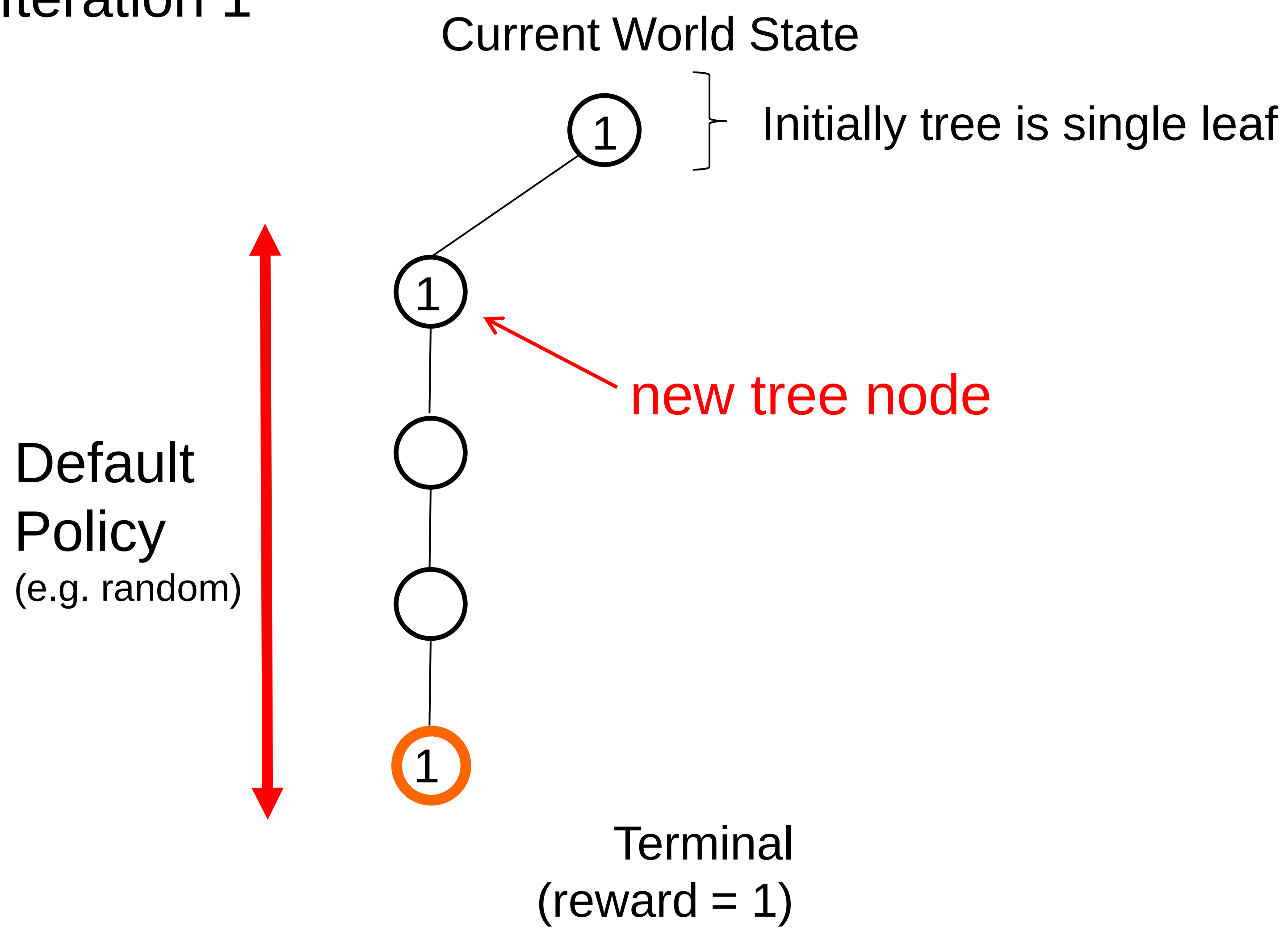
- **<u>Online Action Selection:</u>**
  - Build a sparse lookahead tree rooted at current state s
  - Select root action that looks best



- **<u>Tree Building:</u>** repeatedly executes a **rollout policy** from root
  - Add one fringe node to tree each time
  - Updates statistics of tree nodes based on return
- **<u>Key Idea:</u>** rollout policy biased by previous runs to expand tree in more promising directions

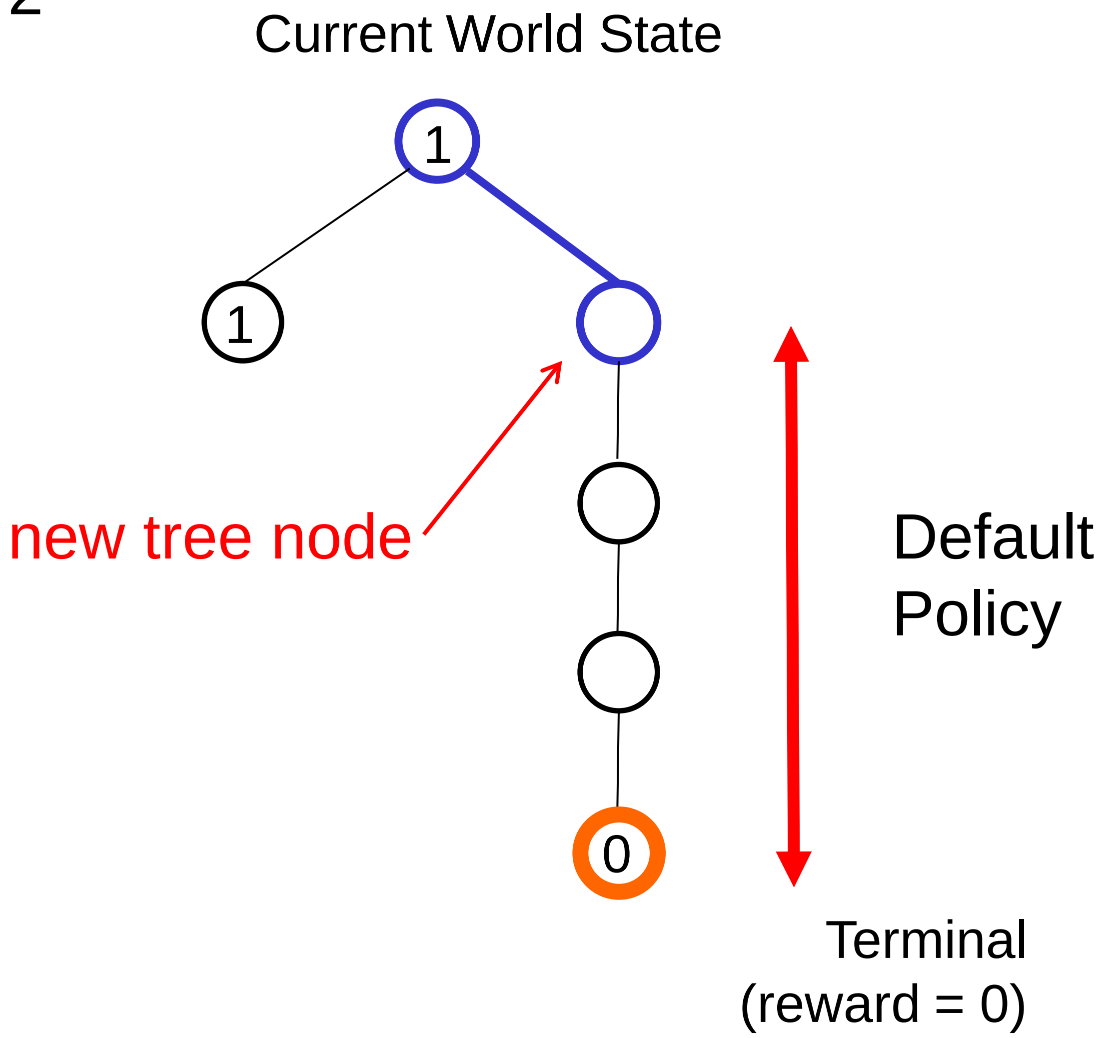
UCT Example

Iteration 1



UCT Example

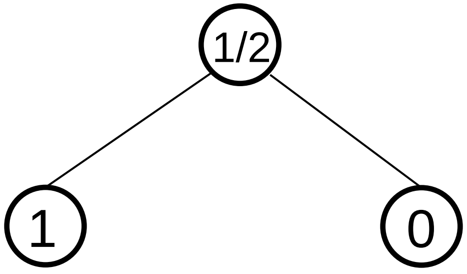
Iteration 2



UCT Example

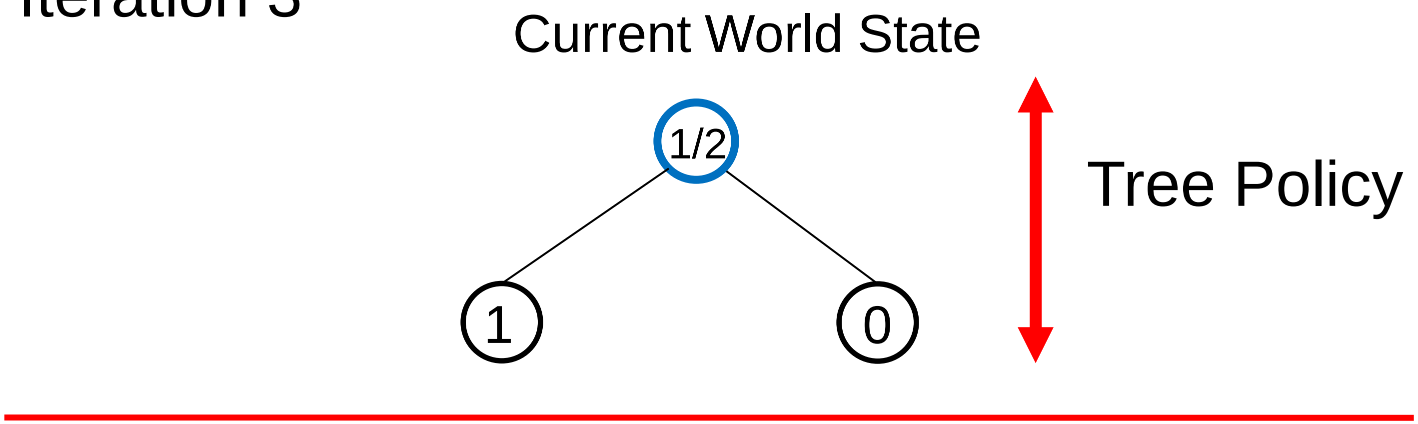
Iteration 3

Current World State



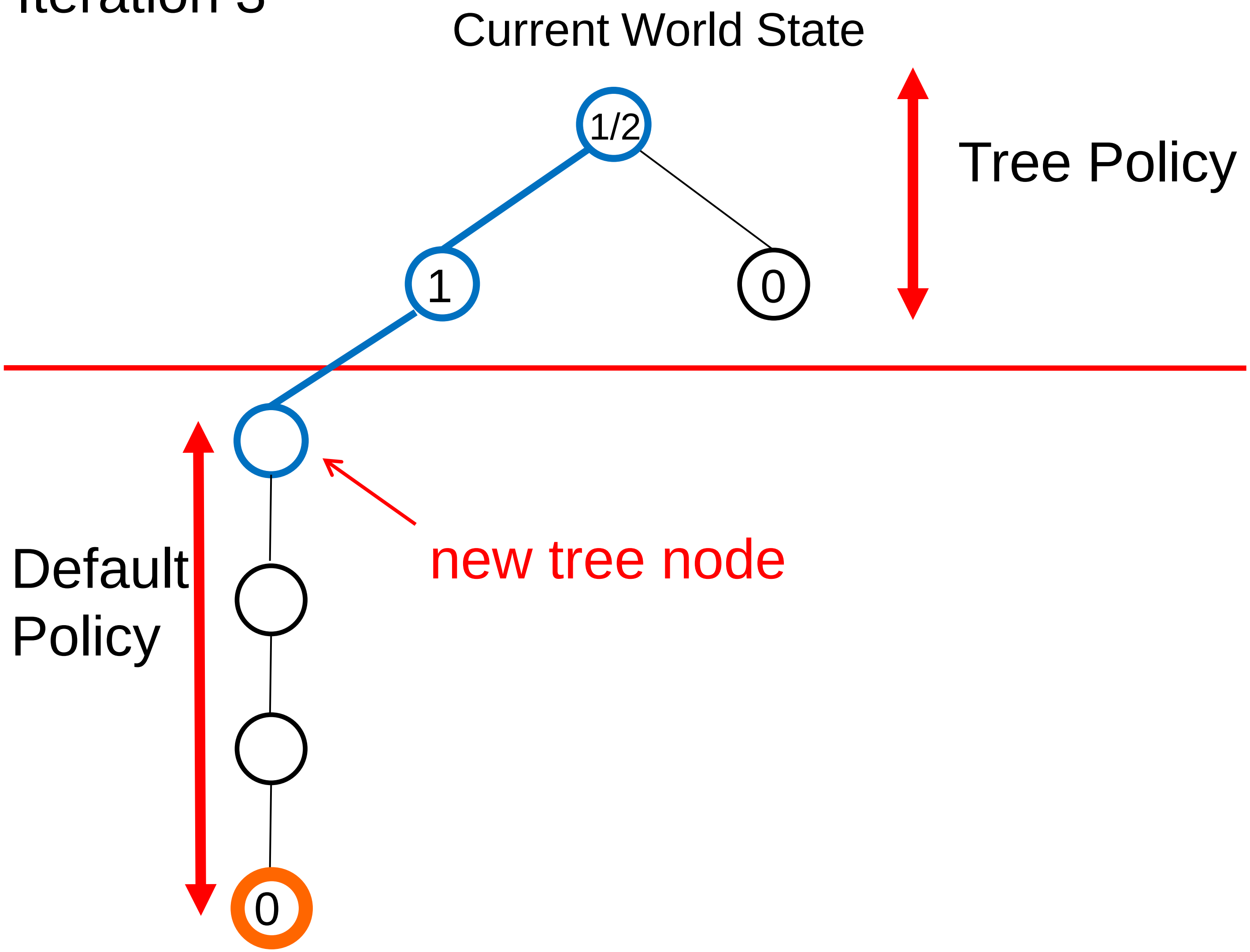
UCT Example

Iteration 3



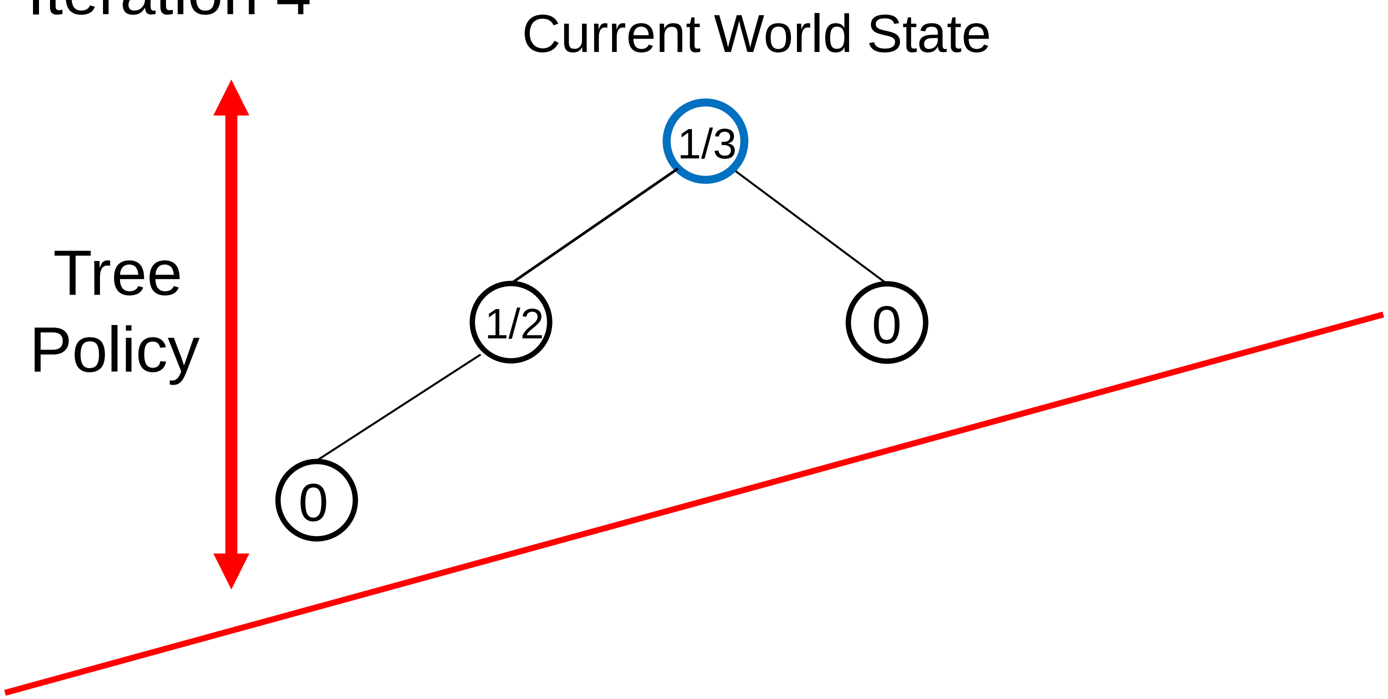
UCT Example

Iteration 3



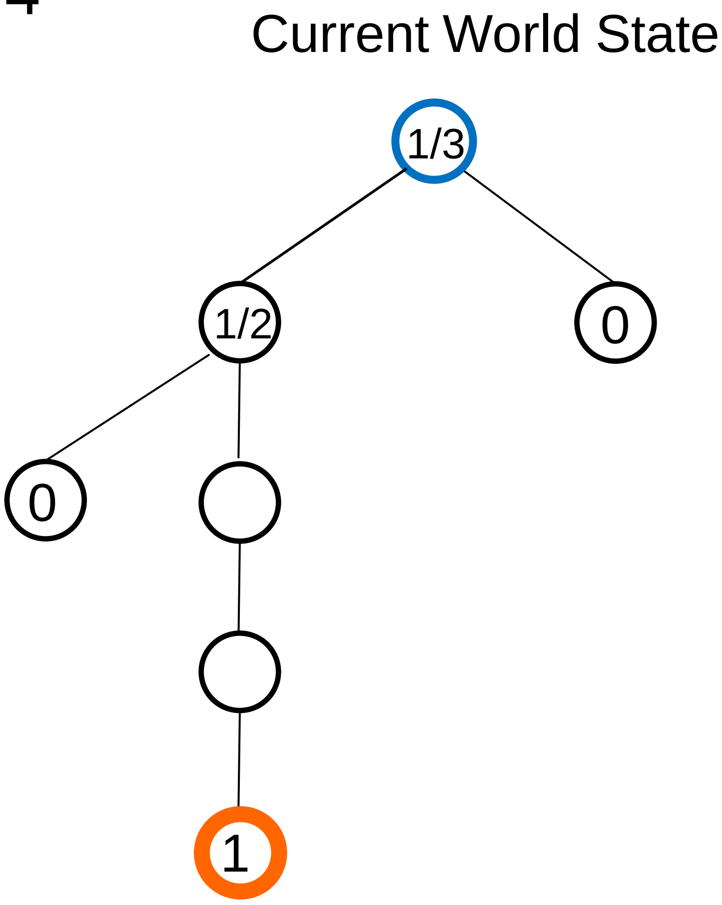
UCT Example

Iteration 4

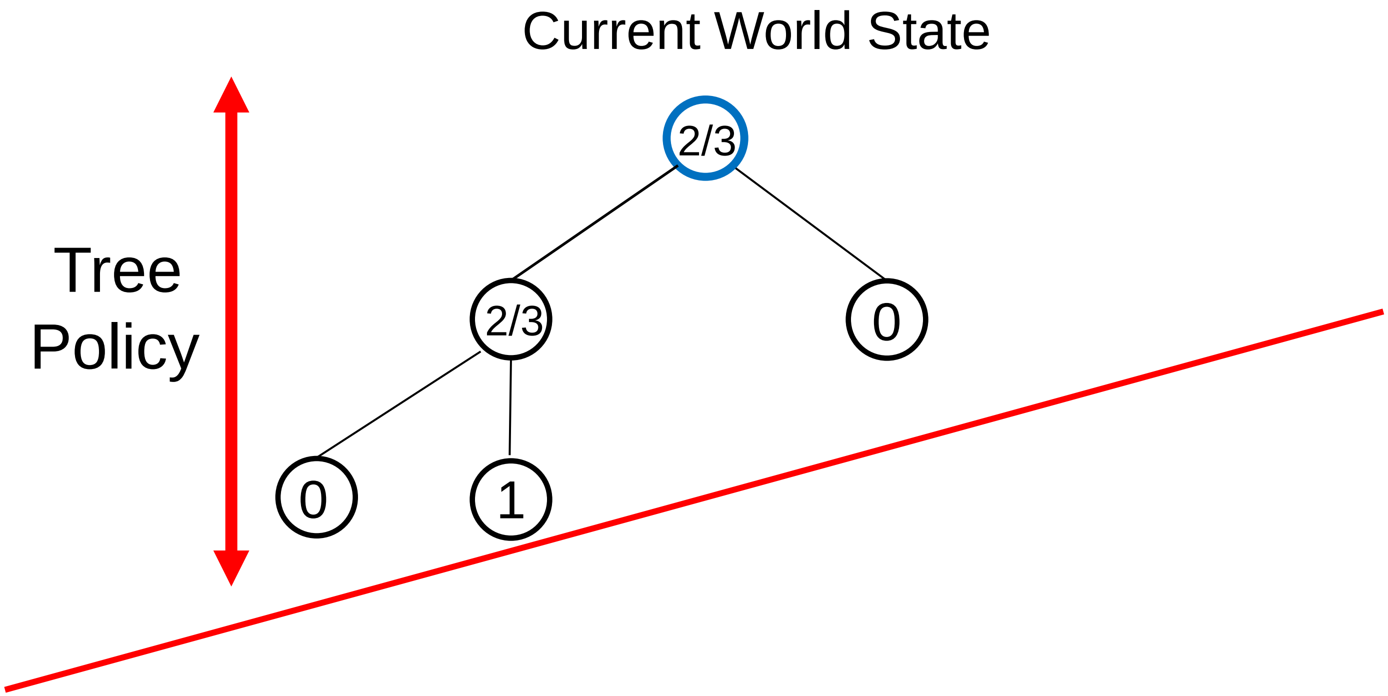


UCT Example

Iteration 4



UCT Example

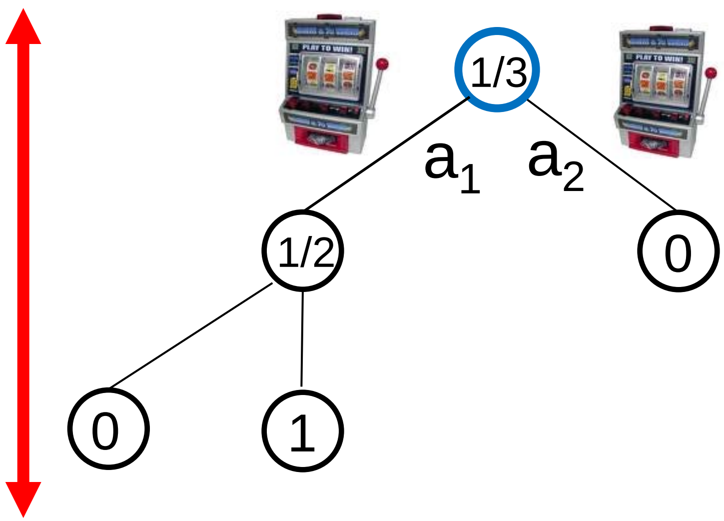


What is an appropriate tree policy?

UCT Example

Current World State

Tree Policy



Treat action selection as **bandit problem**

Upper Confidence Bound (UCB) Rule:

$$\pi_{UCT}(s) = \arg\max_a Q(s,a) + c\sqrt{\frac{\ln n(s)}{n(s,a)}}$$

Exploitation Term

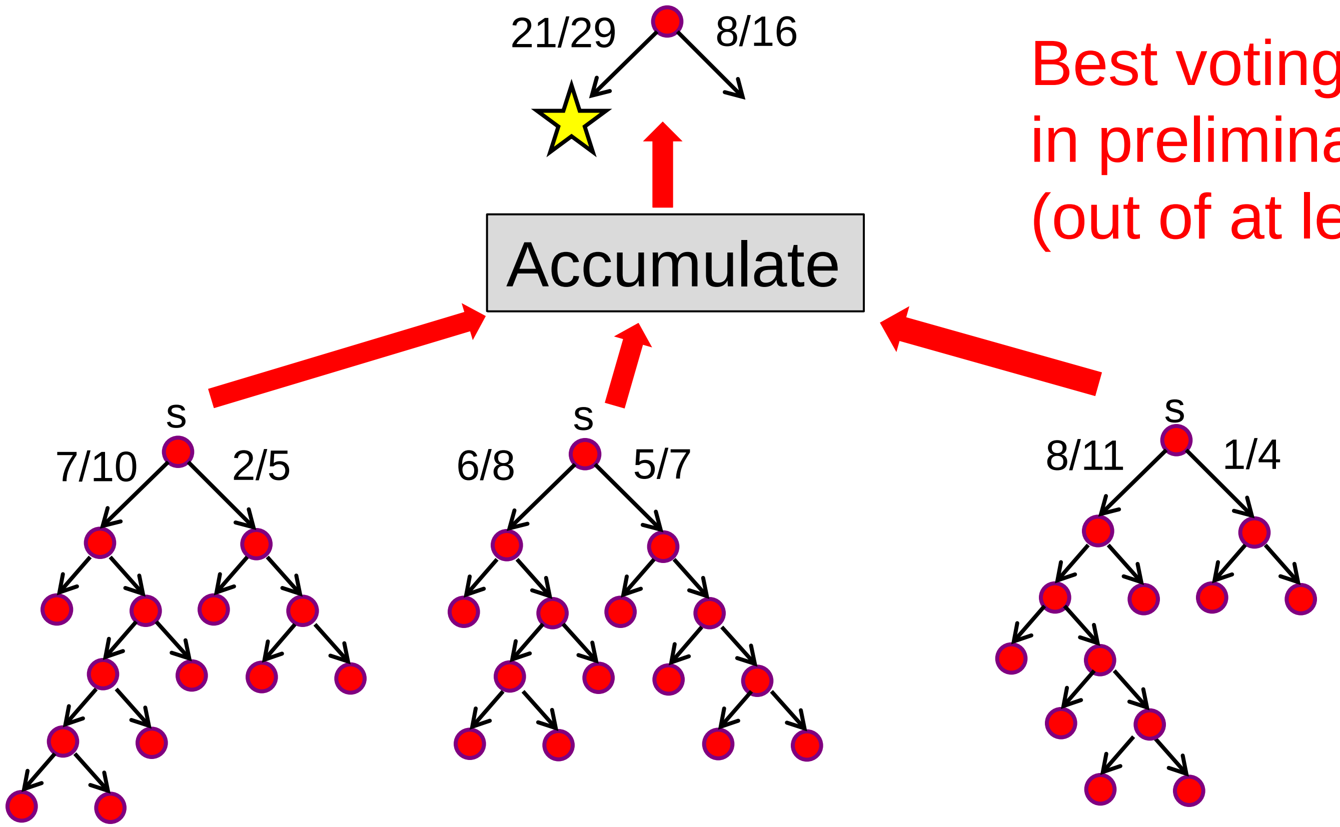
Exploitation Term

Talk Outline

- Motivation
- Monte-Carlo Planning with UCT
- Ensembles
- Domains & Results

Ensemble UCT

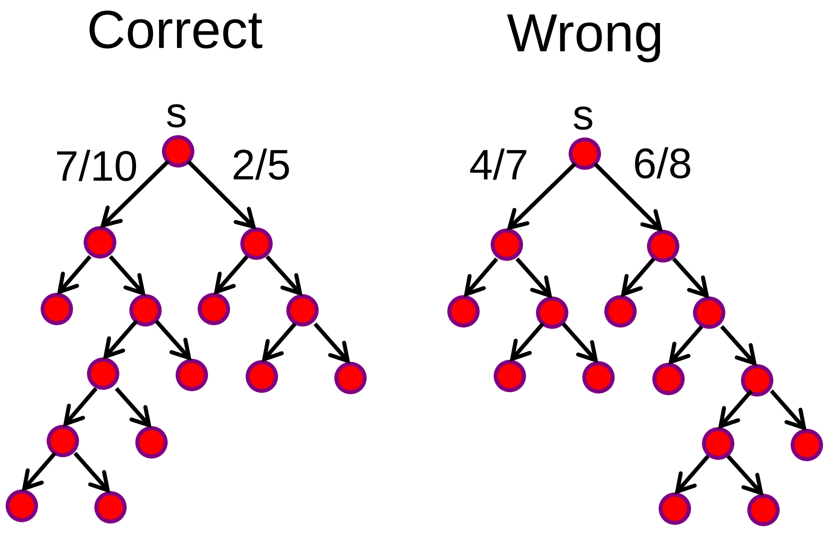
1. Build *T* independent UCT trees rooted at current state
2. Accumulate action statistics at root nodes of trees
3. Execute action with best average return



Best voting method in preliminary study (out of at least 5)

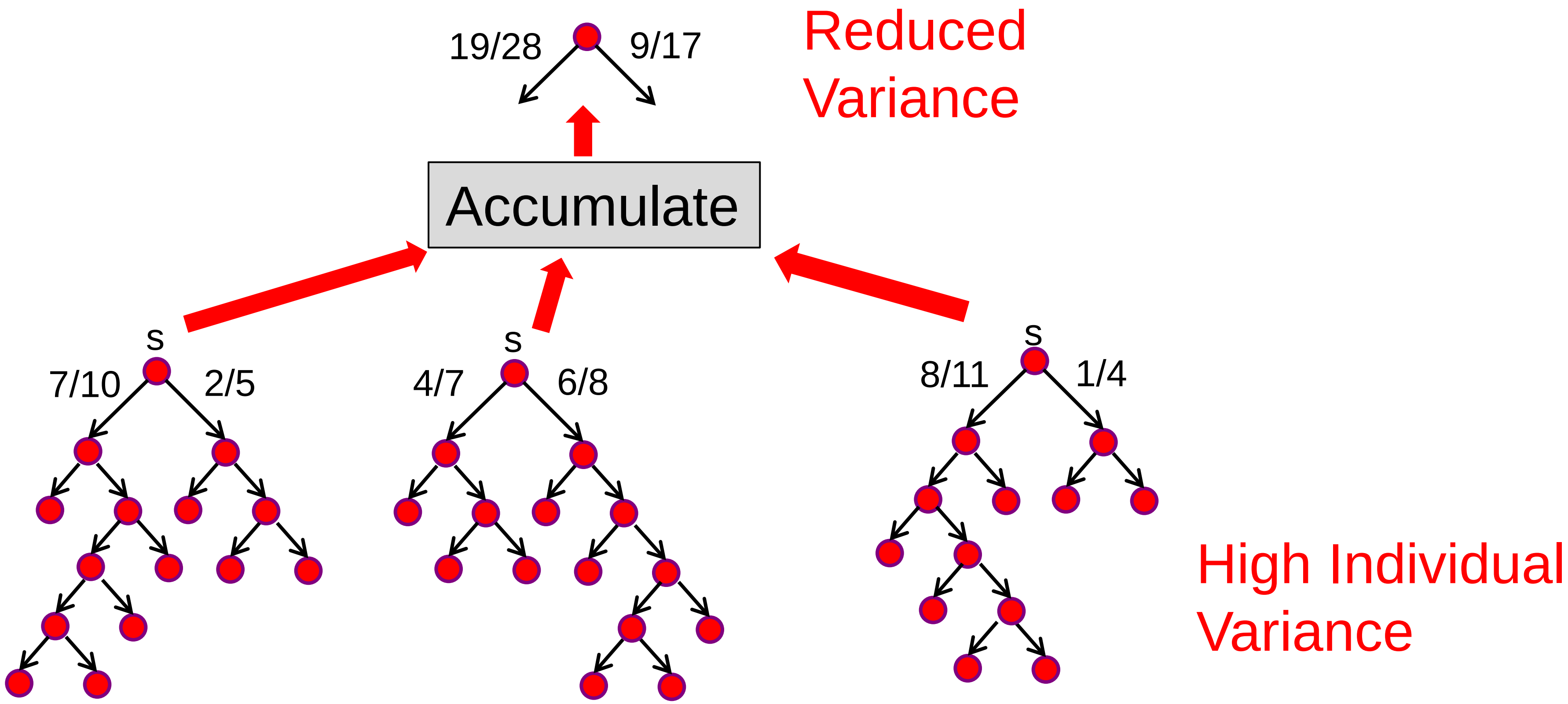
Why might ensembles work?

- UCT is stochastic – unlucky runs can choose bad actions



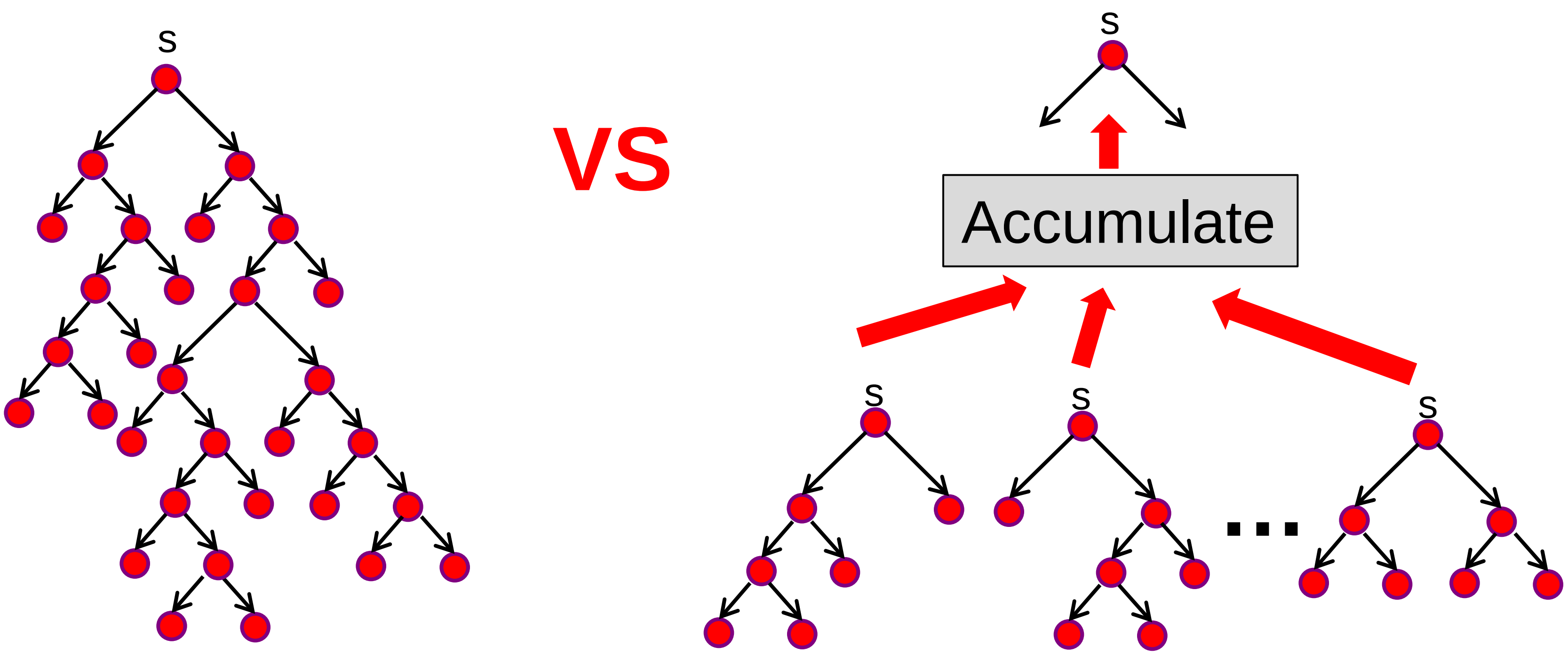
Why might ensembles work?

- UCT is stochastic – unlucky runs can choose bad actions
- **Variance Reduction:** average reduces variance (and bad luck)
- Likely explanation for observed parallel-time advantage



Why might ensembles work?

- Sequential-time advantage
  - Single large tree must be worse than multiple smaller trees (w/ equal # nodes)
  - Assumes time is reflected by total number of tree nodes
- **Smaller trees:** higher variance and bias at root
  - Can averaging make up for the lower quality individuals?



Talk Outline

- Motivation
- Monte-Carlo Planning with UCT
- Ensembles
- Domains & Results

Backgammon

- 2 player
- Stochastic

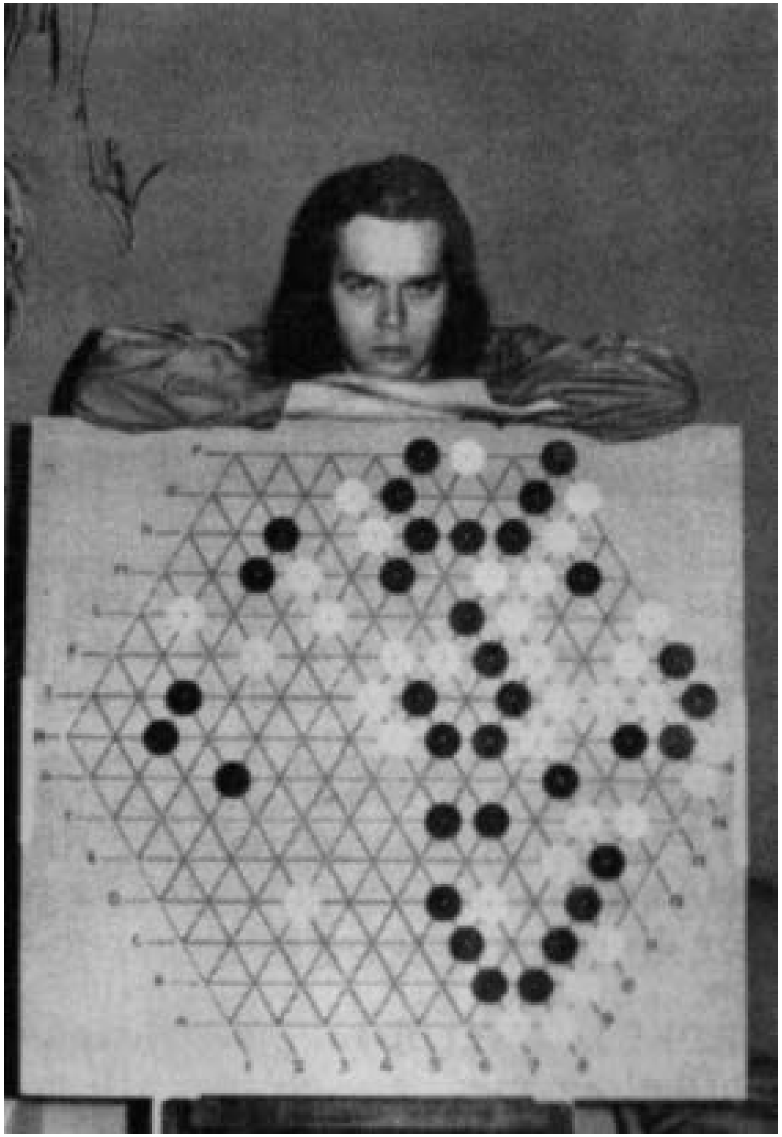
Biniax

- 1 player
- stochastic

Connect 4

- 2 player
- Deterministic

Havannah



- 2 player
- deterministic

- 1 player
- stochastic

Two variants with different scoring schemes

UPPER SECTION	HOW TO SCORE	GAME #1	GAME #2	GAME #3	GAME #4	GAME #5	GAME #6
Aces = 1	Count and Add Only Aces						
Twos = 2	Count and Add Only Twos						
Threes = 3	Count and Add Only Threes						
Fours = 4	Count and Add Only Fours						
Fives = 5	Count and Add Only Fives						
Sixes = 6	Count and Add Only Sixes						
TOTAL SCORE	→						
BONUS If total score is 63 or over	SCORE 35						
TOTAL Of Upper Section	→						
LOWER SECTION							
3 of a kind	Add Total Of All Dice						
4 of a kind	Add Total Of All Dice						
Full House	SCORE 25						
Sm. Straight Sequence of 4	SCORE 30						
Lg. Straight Sequence of 5	SCORE 40						
YAHTZEE 5 of a kind	SCORE 50						
Chance	Score Total Of All 5 Dice						
YAHTZEE BONUS	✓ FOR EACH BONUS						
	SCORE 100 PER ✓						
TOTAL Of Lower Section	→						
TOTAL Of Upper Section	→						
GRAND TOTAL	→						

Experiment Setup

- All ensembles run as a single thread
- UCT constant set per domain (same for all ensembles)
  - 24G Ram
- Varied ensemble configurations
  - Ensemble size = # of trees
  - Trajectories per tree = Size of individual trees
- Averaged results over 1000-4000 runs (usually 4000)
  - Show 99% confidence intervals

Results

Yahtzee

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^7					
2^8					
2^9					
2^{10}					
2^{11}					
2^{12}					
2^{13}					
2^{14}					
2^{15}					
2^{16}					

Results

Yahtzee

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^7	160.3 ± 2.5	167.9 ± 1.5	175.3 ± 2.8	186.3 ± 2.8	193.5 ± 3.3
2^8					
2^9					
2^{10}					
2^{11}					
2^{12}					
2^{13}					
2^{14}					
2^{15}					
2^{16}					

Results

Yahtzee

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^7	160.3 ± 2.5	167.9 ± 1.5	175.3 ± 2.8	186.3 ± 2.8	193.5 ± 3.3
2^8	172.3 ± 2.8	179.2 ± 1.6	185.9 ± 2.8	193.7 ± 3.0	202.2 ± 3.7
2^9					
2^{10}					
2^{11}					
2^{12}					
2^{13}					
2^{14}					
2^{15}					
2^{16}					

Results

Yahtzee

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^7	160.3 ± 2.5	167.9 ± 1.5	175.3 ± 2.8	186.3 ± 2.8	193.5 ± 3.3
2^8	172.3 ± 2.8	179.2 ± 1.6	185.9 ± 2.8	193.7 ± 3.0	202.2 ± 3.7
2^9	183.1 ± 2.7	190.2 ± 1.8	197.0 ± 3.4	205.0 ± 3.9	208.3 ± 3.2
2^{10}	191.8 ± 2.8	199.9 ± 1.9	204.0 ± 3.3	207.9 ± 3.2	214.2 ± 3.7
2^{11}	197.9 ± 2.5	206.2 ± 2.0	211.0 ± 3.6	214.7 ± 3.8	217.4 ± 3.7
2^{12}	208.1 ± 3.7	211.1 ± 2.1	214.9 ± 3.9	215.6 ± 3.5	220.6 ± 2.7
2^{13}	209.0 ± 3.3	214.9 ± 1.8	216.4 ± 3.4	218.9 ± 4.0	221.4 ± 2.9
2^{14}	215.2 ± 4.0	217.1 ± 2.2	219.8 ± 2.8	223.4 ± 3.1	221.3 ± 4.0
2^{15}	215.0 ± 3.5	220.7 ± 2.1	220.9 ± 3.7		
2^{16}	216.6 ± 3.7	221.0 ± 3.2			

- Consistent improvement as ensemble size grows
- Parallel-time and single-core space advantage

Results

Yahtzee

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^7	160.3 ± 2.5	167.9 ± 1.5	175.3 ± 2.8	186.3 ± 2.8	193.5 ± 3.3
2^8	172.3 ± 2.8	179.2 ± 1.6	185.9 ± 2.8	193.7 ± 3.0	202.2 ± 3.7
2^9	183.1 ± 2.7	190.2 ± 1.8	197.0 ± 3.4	205.0 ± 3.9	208.3 ± 3.2
2^{10}	191.8 ± 2.8	199.9 ± 1.9	204.0 ± 3.3	207.9 ± 3.2	214.2 ± 3.7
2^{11}	197.9 ± 2.5	206.2 ± 2.0	211.0 ± 3.6	214.7 ± 3.8	217.4 ± 3.7
2^{12}	208.1 ± 3.7	211.1 ± 2.1	214.9 ± 3.9	215.6 ± 3.5	220.6 ± 2.7
2^{13}	209.0 ± 3.3	214.9 ± 1.8	216.4 ± 3.4	218.9 ± 4.0	221.4 ± 2.9
2^{14}	215.2 ± 4.0	217.1 ± 2.2	219.8 ± 2.8	223.4 ± 3.1	221.3 ± 4.0
2^{15}	215.0 ± 3.5	220.7 ± 2.1	220.9 ± 3.7		
2^{16}	216.6 ± 3.7	221.0 ± 3.2			

- Consistent improvement for larger ensemble sizes
- Parallel-time and single-core space advantage
- 16 x 2^{11} on par w/ 1 x 2^{16} = 32x improvement

Results

Connect 4

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^{10}	$-.522 \pm .048$	$-.370 \pm .052$	$-.299 \pm .053$	$-.233 \pm .055$	$-.189 \pm .055$
2^{11}	$-.256 \pm .054$	$-.139 \pm .055$	$-.102 \pm .056$	$-.011 \pm .057$	$-.056 \pm .056$
2^{12}	$.011 \pm .056$	$.121 \pm .056$	$.227 \pm .055$	$.253 \pm .054$	$.284 \pm .076$
2^{13}	$.234 \pm .054$	$.413 \pm .051$	$.507 \pm .048$	$.543 \pm .067$	$.608 \pm .064$
2^{14}	$.470 \pm .049$	$.646 \pm .043$	$.765 \pm .051$	$.842 \pm .042$	$.841 \pm .042$
2^{15}	$.648 \pm .042$	$.793 \pm .048$	$.859 \pm .040$	$.899 \pm .034$	$.918 \pm .031$
2^{16}	$.727 \pm .054$	$.884 \pm .037$	$.886 \pm .036$	$.926 \pm .029$	
2^{17}	$.811 \pm .045$	$.898 \pm .035$	$.917 \pm .024$		
2^{18}	$.871 \pm .038$	$.910 \pm 0.31$			
2^{19}	$.903 \pm .032$				

- Similar observations across other domains
- Except Binax

Results

Binax

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^8	102.1 ± 1.2	102.0 ± 1.2	100.9 ± 1.2	101.2 ± 1.4	101.8 ± 2.4
2^9	103.9 ± 1.2	104.0 ± 1.2	104.4 ± 1.2	103.0 ± 1.4	103.9 ± 2.4
2^{10}	105.9 ± 1.2	105.3 ± 1.2	105.0 ± 1.2	106.6 ± 2.4	107.7 ± 2.4
2^{11}	108.0 ± 1.2	107.9 ± 1.2	107.4 ± 1.2	108.3 ± 2.4	108.7 ± 2.4
2^{12}	109.0 ± 1.2	109.5 ± 1.2	110.6 ± 2.4	110.5 ± 2.4	
2^{13}	110.6 ± 1.2	112.1 ± 1.2	113.8 ± 2.4	114.0 ± 2.4	
2^{14}	111.9 ± 1.2	113.9 ± 1.2			
2^{15}	113.2 ± 1.2				

- **Small trees:** no improvement

Results

Binax

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^8	102.1 ± 1.2	102.0 ± 1.2	100.9 ± 1.2	101.2 ± 1.4	101.8 ± 2.4
2^9	103.9 ± 1.2	104.0 ± 1.2	104.4 ± 1.2	103.0 ± 1.4	103.9 ± 2.4
2^{10}	105.9 ± 1.2	105.3 ± 1.2	105.0 ± 1.2	106.6 ± 2.4	107.7 ± 2.4
2^{11}	108.0 ± 1.2	107.9 ± 1.2	107.4 ± 1.2	108.3 ± 2.4	108.7 ± 2.4
2^{12}	109.0 ± 1.2	109.5 ± 1.2	110.6 ± 2.4	110.5 ± 2.4	
2^{13}	110.6 ± 1.2	112.1 ± 1.2	113.8 ± 2.4	114.0 ± 2.4	
2^{14}	111.9 ± 1.2	113.9 ± 1.2			
2^{15}	113.2 ± 1.2				

- **Small trees:** no improvement
- **Larger trees:** very small improvement
- **Binax Properties:** UCT has very low variance
  Small trees are quite biased

Results: Single Core

Connect 4

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^{10}	$-.522 \pm .048$	$-.370 \pm .052$	$-.299 \pm .053$	$-.233 \pm .055$	$-.189 \pm .055$
2^{11}	$-.256 \pm .054$	$-.139 \pm .055$	$-.102 \pm .056$	$-.011 \pm .057$	$-.056 \pm .056$
2^{12}	$.011 \pm .056$	$.121 \pm .056$	$.227 \pm .055$	$.253 \pm .054$	$.284 \pm .076$
2^{13}	$.234 \pm .054$	$.413 \pm .051$	$.507 \pm .048$	$.543 \pm .067$	$.608 \pm .064$
2^{14}	$.470 \pm .049$	$.646 \pm .043$	$.765 \pm .051$	$.842 \pm .042$	$.841 \pm .042$
2^{15}	$.648 \pm .042$	$.793 \pm .048$	$.859 \pm .040$	$.899 \pm .034$	$.918 \pm .031$
2^{16}	$.727 \pm .054$	$.884 \pm .037$	$.886 \pm .036$	$.926 \pm .029$	
2^{17}	$.811 \pm .045$	$.898 \pm .035$	$.917 \pm .024$		
2^{18}	$.871 \pm .038$	$.910 \pm 0.31$			
2^{19}	$.903 \pm .032$				

- Ensembles along diagonals have same total nodes

Results

Connect 4

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^{10}	$-.522 \pm .048$	$-.370 \pm .052$	$-.299 \pm .053$	$-.233 \pm .055$	$-.189 \pm .055$
2^{11}	$-.256 \pm .054$	$-.139 \pm .055$	$-.102 \pm .056$	$-.011 \pm .057$	$-.056 \pm .056$
2^{12}	$.011 \pm .056$	$.121 \pm .056$	$.227 \pm .055$	$.253 \pm .054$	$.284 \pm .076$
2^{13}	$.234 \pm .054$	$.413 \pm .051$	$.507 \pm .048$	$.543 \pm .067$	$.608 \pm .064$
2^{14}	$.470 \pm .049$	$.646 \pm .043$	$.765 \pm .051$	$.842 \pm .042$	$.841 \pm .042$
2^{15}	$.648 \pm .042$	$.793 \pm .048$	$.859 \pm .040$	$.899 \pm .034$	$.918 \pm .031$
2^{16}	$.727 \pm .054$	$.884 \pm .037$	$.886 \pm .036$	$.926 \pm .029$	
2^{17}	$.811 \pm .045$	$.898 \pm .035$	$.917 \pm .024$		
2^{18}	$.871 \pm .038$	$.910 \pm 0.31$			
2^{19}	$.903 \pm .032$				

- Ensembles along diagonals have same total nodes
- **Small Trees:** ensembles of very small trees hurt performance

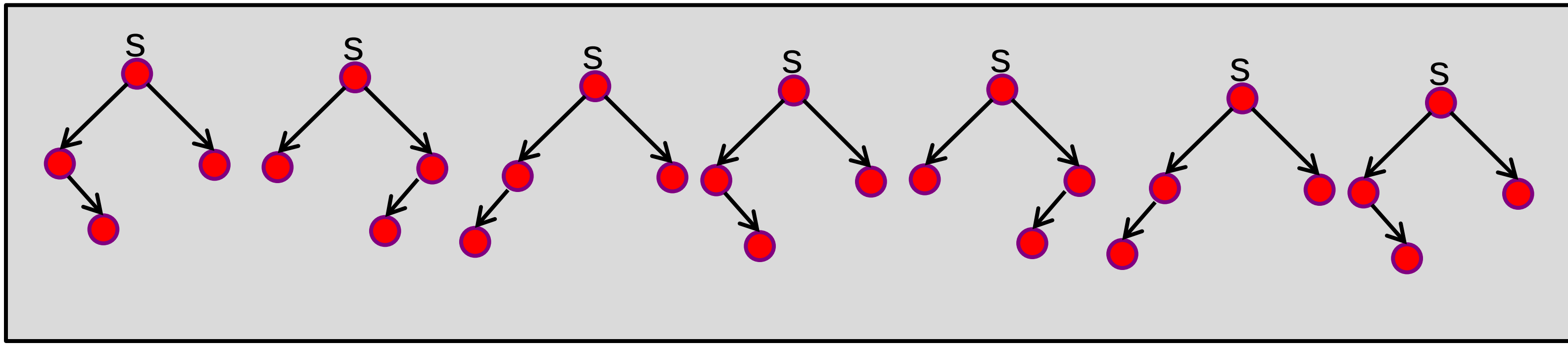
Results

Connect 4

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^{10}	$-.522 \pm .048$	$-.370 \pm .052$	$-.299 \pm .053$	$-.233 \pm .055$	$-.189 \pm .055$
2^{11}	$-.256 \pm .054$	$-.139 \pm .055$	$-.102 \pm .056$	$-.011 \pm .057$	$-.056 \pm .056$
2^{12}	$.011 \pm .056$	$.121 \pm .056$	$.227 \pm .055$	$.253 \pm .054$	$.284 \pm .076$
2^{13}	$.234 \pm .054$	$.413 \pm .051$	$.507 \pm .048$	$.543 \pm .067$	$.608 \pm .064$
2^{14}	$.470 \pm .049$	$.646 \pm .043$	$.765 \pm .051$	$.842 \pm .042$	$.841 \pm .042$
2^{15}	$.648 \pm .042$	$.793 \pm .048$	$.859 \pm .040$	$.899 \pm .034$	$.918 \pm .031$
2^{16}	$.727 \pm .054$	$.884 \pm .037$	$.886 \pm .036$	$.926 \pm .029$	
2^{17}	$.811 \pm .045$	$.898 \pm .035$	$.917 \pm .024$		
2^{18}	$.871 \pm .038$	$.910 \pm 0.31$			
2^{19}	$.903 \pm .032$				

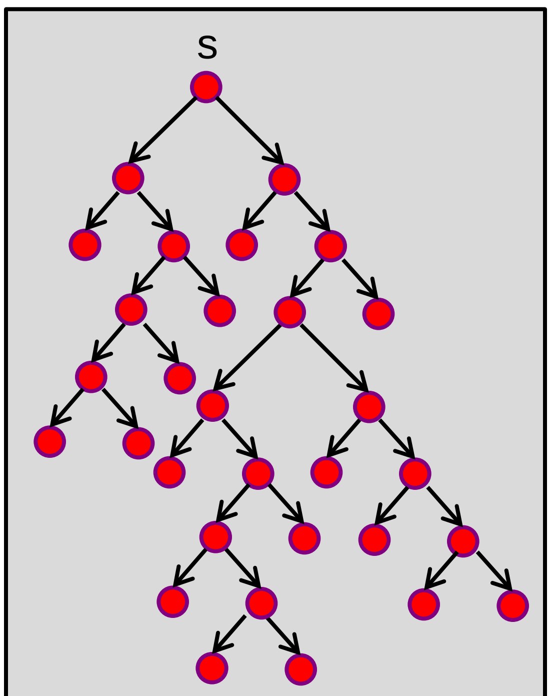
- Ensembles along diagonals have same total nodes
- **Small Trees:** ensembles of very small trees hurt performance
- **<u>Medium Trees:</u>** performance not hurt for larger trees

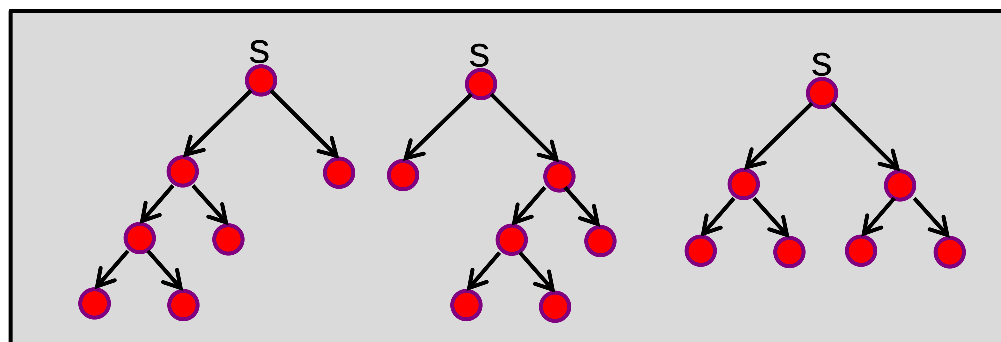
Small Trees



High bias trees
Variance is not main problem

VS





Results

Connect 4

Trajectories per Tree	Ensemble Size				
	1	2	4	8	16
2^{10}	$-.522 \pm .048$	$-.370 \pm .052$	$-.299 \pm .053$	$-.233 \pm .055$	$-.189 \pm .055$
2^{11}	$-.256 \pm .054$	$-.139 \pm .055$	$-.102 \pm .056$	$-.011 \pm .057$	$-.056 \pm .056$
2^{12}	$.011 \pm .056$	$.121 \pm .056$	$.227 \pm .055$	$.253 \pm .054$	$.284 \pm .076$
2^{13}	$.234 \pm .054$	$.413 \pm .051$	$.507 \pm .048$	$.543 \pm .067$	$.608 \pm .064$
2^{14}	$.470 \pm .049$	$.646 \pm .043$	$.765 \pm .051$	$.842 \pm .042$	$.841 \pm .042$
2^{15}	$.648 \pm .042$	$.793 \pm .048$	$.859 \pm .040$	$.899 \pm .034$	$.918 \pm .031$
2^{16}	$.727 \pm .054$	$.884 \pm .037$	$.886 \pm .036$	$.926 \pm .029$	
2^{17}	$.811 \pm .045$	$.898 \pm .035$	$.917 \pm .024$		
2^{18}	$.871 \pm .038$	$.910 \pm 0.31$			
2^{19}	$.903 \pm .032$				

- **<u>Large Trees:</u>** improvement trend – not statistically significant
- Only two domains show such a trend
- Most show no significant single-core time advantage

Summary

- Parallel time and single-core space advantage?
  - Yes – consistently significant
  - Except when individual tree variance is very small
- Sequential-time advantage?
  - Nothing significant
  - Suggestive trends in 2 domains for large trees
- Sequential-time disadvantage
  - Happens when trees are too small (high bias)

Future Work

- Better understand bias-variance trade-off
- Experiments for larger ensembles
  - When does improvement stop?
- Are trends the same for “enhanced” versions of UCT?
- Optimize configuration for a cluster’s space and time constraints

Thanks

Time Overhead for Large Trees

Connect 4 Ensemble Timing Table (ms)

Total Trajectories	Ensembles				
	1	2	4	8	16
4096					694 ± 6
8192				714 ± 6	
16384			740 ± 6		
32768		773 ± 6			
65536	792 ± 6				

Ensemble Parameter Sensitivity

Connect 4 Ensemble Parameter Sensitivity

UCT Constant	Ensembles				
	1	2	4	8	16
0.5	0.205 ± 0.075	0.318 ± 0.075	0.378 ± 0.073	0.413 ± 0.072	0.434 ± 0.072
0.6	0.412 ± 0.070	0.565 ± 0.065	0.618 ± 0.062	0.619 ± 0.062	0.525 ± 0.068
1	0.727 ± 0.054	0.793 ± 0.048	0.765 ± 0.051	0.543 ± 0.067	0.284 ± 0.076

Yahtzee Ensemble Parameter Sensitivity

UCT Constant	Ensembles				
	1	2	4	8	16
2	173 ± 2.6	195 ± 2.9	207 ± 3.4	208 ± 2.9	207 ± 3.0
4	187 ± 2.6	201 ± 2.9	209 ± 3.2	211 ± 3.2	208 ± 3.2
64	215 ± 3.5	217 ± 2.2	216 ± 3.4	216 ± 3.5	217 ± 3.7