Scalability and Parallelization of Monte-Carlo Tree Search

Amine Bourki, Guillaume Chaslot, Matthieu Coulm, Vincent Danjean, Hassen Doghmen, Jean-Baptiste Hoock, Thomas Hérault, Arpad Rimmel, Fabien Teytaud, Olivier Teytaud, Paul Vayssière, and Ziqin Yu

TAO (Inria), LRI, UMR 8623(CNRS - Univ. Paris-Sud),
bat 490 Univ. Paris-Sud 91405 Orsay, France
teytaud@lri.fr

Abstract. Monte-Carlo Tree Search is now a well established algorithm, in games and beyond. We analyze its scalability, and in particular its limitations and the implications in terms of parallelization. We focus on our Go program MoGo and our Havannah program Shakti. We use multicore machines and message-passing machines. For both games and on both type of machines we achieve adequate efficiency for the parallel version. However, in spite of promising results in self-play there are situations for which increasing the time per move does not solve anything. Therefore parallelization is not a solution to all our problems. Nonetheless, for problems where the Monte-Carlo part is less biased than in the game of Go, parallelization should be quite efficient, even without shared memory.

1 Introduction

Since 2006, Monte-Carlo Tree Search (MCTS[5,8,14]) is a revolution in games and planning, with applications in many fields. It is widely said that MCTS has some scalability advantages.

It is quite natural, then, to parallelize MCTS, both on multicore machines [18] and on clusters [10,4]. In this paper, after an introduction to MCTS (section 2), we (i) discuss the scalability of MCTS, showing big limitations to this scalability, and not only due to RAVE (section 3); (ii) compare existing algorithms on clusters (section 4). Finally, conclusions are given (section 5).

2 Monte-Carlo Tree Search

We below introduce Monte-Carlo Tree Search, i.e., MCTS. We here present the MCTS variant termed UCT [14], which is shorter to present and quite general; the formulas involved in our programs are more tricky and can be found in [11,15,10,16]; these details do not affect the parallelization, and UCT is a trustable algorithm in the general case of games and planning.

UCT is presented in Algorithm 1. The reader is referred to [14] for a more detailed presentation, and to [11,18,8,6] for a more comprehensive introduction in particular for the specific case of binary rewards and two-player games.

J. van den Herik, H. Iida, and A. Plaat (Eds.): CG 2010, LNCS 6515, pp. 48–58, 2011.

Algorithm 1. Overview of the UCT algorithm for two-player deterministic games. The adaptation to stochastic cases or one-player games is straightforward. UCT takes as input a situation $s \in \mathcal{S}$, and outputs a decision. For any situation s and any decision d, $s' = s.d$ denotes the situation s' subsequent to decision d in situation s. T is made of two mappings (initially identically 0), N_T and S_T: N_T is a mapping from $\mathcal{S}$ to $\mathbb{N}$ (i.e., maps situations to integers) and S_T is a mapping from $\mathcal{S}$ to $\mathbb{R}$. $\mathcal{S}$ is the set of states, S_T stands for the sum of rewards at a given state and N_T stands for the number of visits at a given state. Inspired by [8,17], we propose $PW(n) = Kn^{1/4}$.

Function $UCT(s)$
$T \leftarrow 0$
while TimeLeft> 0 **do**
  $PerformSimulation(T, s)$
end while
Return r maximizing $N_T(s.r)$
Function $reward = PerformSimulation(T, s)$
if s is a final state **then**
  return the reward of s
else
  if $N_T(s) > 0$ **then**
    Choose the set of admissible decisions thanks to progressive widening PW and the heuristic H as follows:
    $R = PW(N_T(s))$ // $R \in \mathbb{N}$ is the size of the considered pool of moves
    $W = \{H(s, i); i \in [[1, R]]\}$ // W is the size of the considered pool of moves
    Choose the move to be simulated as follows:
    if Color(s)=myColor **then**
      $\epsilon = 1$
    else
      $\epsilon = -1$
    end if
    $d = \arg\max_{d \in W} Score(\epsilon.S_T(s.d), N_T(s.d), N_T(s))$
  else
    $d = MC(d)$ /* $MC(d)$ is a heuristic choice of move */
  end if
end if
$reward = PerformSimulation(T, s.d)$ // reward$\in \{0, 1\}$
Update the statistics in the tree as follows:
$N_T(s) \leftarrow N_T(s) + 1$
$S_T(s) \leftarrow S_T(s) + reward$
Return $reward$
Function $Score(a, b, c)$
Return $a/b + \sqrt{2\log(c)/b}$ /* plenty of improvements are published in the literature for specific problems*/
Function $H(s, i)$
Return the i^{th} best move according to the heuristic in situation s.

3 Scalability of MCTS

The scalability of MCTS, i.e., its ability to play better when additional computational power or time is provided, is often given as an argument in favor of it. Also, it is said that the parallelization is quite efficient; the conclusion of these two statements is that with big clusters, programs should now be much stronger than humans in games in which single computers are already at the level of beginners. We will here give more information (limitations) on this scalability.

The number of simulations per move is usually much larger in real games than in experimental results published in papers, because of limited computational power - it is difficult, even with a cluster, to have significant results corresponding to the computational power associated to realistic time settings on a big machine. In this section, we investigate the behavior of MCTS when the time per move is increased (section 3.1), followed by counter-examples to scalability (section 3.2).

3.1 The Limited Scalability by Numbers

It is usually said that MCTS is highly scalable, and provides improvements of constant order against the baseline when the computational power is doubled. We here show that things are not so constant; results are presented in Table 1 for the game of Go. The numbers show a clear decrease of scalability as the

Table 1. Scalability of MCTS for the game of Go. These results show a decrease of scalability as the computational power increases.

N =Number of simulations	Succes rate of 2N simulations against N simulations in 9x9 Go	Success rate of 2N simulations against N simulations in 19x19 Go
1,000	71.1 ± 0.1 %	90.5 ± 0.3 %
4,000	68.7 ± 0.2 %	84.5 ± 0.3 %
16,000	66.5 ± 0.9 %	80.2 ± 0.4 %
256,000	61.0 ± 0.2 %	58.5 ± 1.7 %

computational power increases. It is not specific to Go; Table 2 shows that the situation is similar in Havannah. This holds even so, when the opponent is an MCTS too; please note, this is not equivalent to the case of the scalability study `http://cgos.boardspace.net/study/index.html` which considers non-MCTS opponents as well; we here see that just against the same MCTS program, we have a limit in scalability; this even happens in 19x19 Go. In Havannah with slow simulations (the operational case, with the best performance in practice), 10,000 simulations per move give only a 52% winning rate against 5,000 simulations per move (Table 2). It suggests that the scalability is smaller than expected from small-scale experiments. Usually people do not publish experiments with so many simulations because it is quite expensive; nonetheless, real games are played with more than this kind of numbers of simulations. Note that the numbers in the tables above are probably larger than the scalability in realistic scenarios.

Table 2. Scaling for the game of Havannah, for fast (left) and slow (right) simulations. As we can see, the success rate is not constant, but decreases when the number of simulations increases.

Number of fast simulations	Success rate
100 vs 50	68.60 ± 0.68%
1000 vs 500	63.57 ± 0.76%
2,000 vs 1000	59.00 ± 1.0%
4,000 vs 2,000	53.90 ± 1.6%
10,000 vs 5,000	55.20 ± 1.6%
20,000 vs 10,000	54.89 ± 1.25%

Number of slow simulations	Success rate
100 vs 50	63.28 ± 0.4%
1000 vs 500	57.37 ± 0.9%
2,000 vs 1000	56.42 ± 1.1%
4,000 vs 2,000	53.24 ± 1.42%
10,000 vs 5,000	52.00 ± 1.6%

A particularity of these numbers is that they are from self-play; this provides a limitation even in the ideal case in which we only consider an opponent of the same type; it is widely known that the improvement is much smaller when considering humans or programs of a different type. Interestingly, Kato [13] has shown that his MCTS implementation reaches a plateau against GnuGo when the number of simulations goes to infinity. This shows limited scalability, to be confirmed by situations (practically) unsolved by Monte-Carlo Tree Search, presented in section below.

3.2 Counter-Examples to Scalability

The RAVE heuristic [3,11] is known as quite efficient in several games: it introduces a bias in H. It is nonetheless suspected that RAVE is responsible for the bad asymptotic behavior of some MCTS programs. Below we recall a well known counter-example when RAVE is included, and then give a detailed presentation of other counter-examples which do not depend on RAVE.

Counter-examples based on RAVE values. Martin Müller posted in the computer-Go mailing list the situation shown in Fig. 1(left, `http://fuego.svn.sourceforge.net/viewvc/fuego/trunk/regression/sgf/rave-problems/`) in which their MCTS implementation Fuego does not find the good move due to RAVE (discussed below), because the only good move is good only if played first (the RAVE value[11] does not work in this case) - such cases are clearly moderately sensitive to computational time or computational power, and this has impacts in terms of scalability.

Other counter-examples. Importantly, Fig. 1(right) from [2] shows that there are some bad behaviors even without RAVE values. Below, we propose new clear examples of limited speed-up, that have the following suitable properties:

- these situations are extremely easy for human players. Even a beginner can solve them;
- these counter-examples are independent of RAVE, as shown in our experiments.

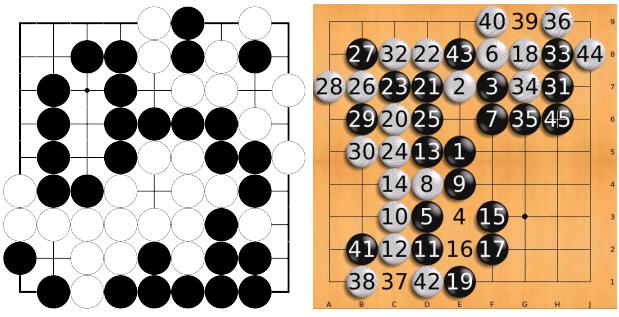

Fig. 1. Left: White to play, an example by M. Müller of bad scalability due to RAVE. RAVE gives a very bad value to the move B2 (second row, second column), because it only makes sense if it is the first move, whereas this is the only move avoiding the seki (otherwise, black A5 and the two black stones A2 and B1 are alive). Right, white to play: an example of bad behavior shown in [2], independently of RAVE values: in many cases (yet not always, this depends on the first simulations), MoGo is almost sure that he is going to win as white by playing C1, whereas it is a loss for white.

Such situations are given in Fig. 2. These situations are semeais; it is known since [9,15] that MCTS algorithms are weak in such cases. We show that this weakness remains without RAVE and even so with the inclusion of specific tactical solvers.

It is often said that classical solvers are able to solve semeais. Therefore including expert modules should improve MCTS algorithms by including a semeai solver. We thus tested two ways of including expertise in MCTS.

- (1) *Expertise*: we introduce a bias in the score, as usually performed in MCTS algorithms [5,8,15]. Some virtual wins are added to UCT statistics so that moves which are good according to our tactical semeai solver are more simulated; the idea, detailed in [5,8,15] consists in increasing the score of moves evaluated as necessary by the semeai solver, so that the heuristic H is more favorable to them. Only moves necessary for solving the semeai are given a bonus; no move at all is given a bonus if the semeai is won even if the player to play passes.
- (2) *Conditionning*: then, all simulations which are not consistent with the solver are discarded and replayed. This means that when the solver predicts that the semeai is won for Black (the solver is called at the end of the tree part, before the MC part), before the Monte-Carlo part, then the Monte-Carlo simulation is replayed until it gives a result consistent with this prediction. Human experts could validate the results (i.e., only simulations consistent with the semeai solver were included in the Monte-Carlo) and the quality of the solver is not the cause for results in Table 3; the coefficients have been tuned in order to be a minimum perturbation for having a correct solving for Fig. 2, left: the coefficients are with respect to (i) the size of semeais considered, and (ii) the weight of the expertise in the function H (for versions with expertise).

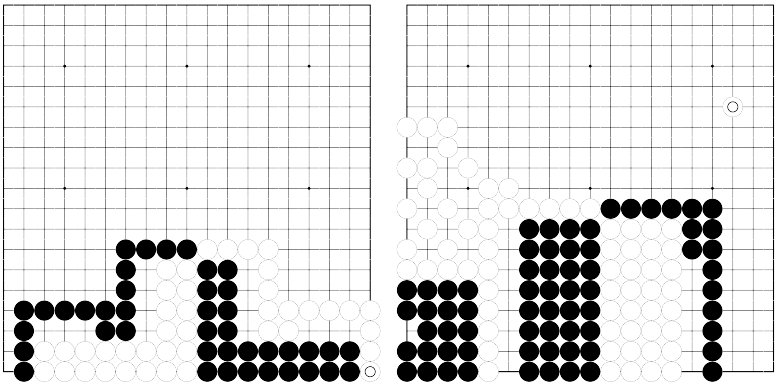

Fig. 2. Left: Black to play. It is here necessary to play in the semeai. Right: Black to play: playing in the semeai is useless as the semeai is won anyway (Black has two more liberties than White) - good moves are outside the semeai. MoGo often makes the mistake of playing in the semeai.

The results are presented in Table 3. In order to be implementation-independent, we consider the performance for fixed numbers of simulations; the slowness of the tactical solver cannot be an explanation for poor results. From these negative results, and also for many trials with various tunings, all of them leading to success rates lower than 50 % against the baseline, we may conclude that including expert knowledge is quite difficult for semeais; it is true that tactical solvers can solve semeais, but they do not solve the impact of semeais on the rest of the board: in conditionning, if simulations are accepted as soon as they are consistent with the semeai solver, then the result of the semeai will be understood by the program, but the program might consider cases in which Black played two more stones than necessary - this is certainly not a good solving of the semeai.

These examples of bad behavior are not restricted to MoGo. Fig. 3 is a game played by Fuego and Aya in the 56^{th} KGS tournament (february 2010); Fuego (a strong program by Univ. of Alberta) played (1) and lost the game.

4 Message-Passing Parallelization

Multi-core machines are increasingly efficient, but the bandwidth is nonetheless limited, and the number of cores is much bigger when we consider clusters than when we consider a single machine. This is why message-passing parallelization (in which communications are explicit and limited) must be considered. We will see here that, in particular in 19x19, the technique is quite efficient from a parallelization point of view: the main issue for MCTS is not the computational power, but the limits to scalability emphasized in section 3.

Table 3. These results are for Fig. 2; Black should or should not play in the semeai (left or right situation in Fig. 2). All results are averaged over 1000+ runs. Bold is for results with more than 75 % on correct moves. We point out that the Go situations under consideration are very easy, understandable by very beginners. We see that (i) with 30K sims/move, many versions play the semeai whenever it is useless, and all versions play it with significant probability, what is a disaster as in real situations there are many time steps at which the MCTS program can have the opportunity of such a bad move and even only one such move is a disaster (losing one stone!), (ii) removing RAVE does not solve the problem, and (iii) adding a tactical solver can work better (moderately better) with the traditional solution of adding expertise as virtual wins, but results remain quite moderate, and far from what can do even a beginner. We also tested many parameterizations in self-play and none of these tests provided more than 50 % of success rate in self-play.

Version of the algorithm	Percentage of "good" moves
Situation in which the semeai should be played 1K sims per move	
MoGo	32 %
MoGo with expertise	**79 %**
MoGo with conditioning	24 %
MoGo with exp.+condit.	**84 %**
Situation in which the semeai should not be played 1K sims per move / 30K sims per move	
MoGo	**100%** / 58 %
MoGo with expertise	**95 %** / 51 %
MoGo with conditioning	**93 %** / 0 %
MoGo with exp.+condit.	**93 %** / 54 %

The various published techniques for the parallelization of MCTS are as follows.

- *Fast tree parallelization* consists in simulating the multi-core process on a cluster; there is still only one tree in memory, on the master; slaves (i) compute the Monte-Carlo part, and (ii) send the results to the master for updates. This is sensitive to Amdahl's law, and is quite expensive in terms of communication when RAVE values are used[12,10].
- *Slow tree parallelization* consists in having one tree on each computation node, and to synchronize these trees slowly, i.e., not at each simulation but with frequency, e.g., three times per second [10]. The synchronization is not on the whole tree; it is typically performed as follows:
  - select all the nodes with
    - at least 5% of the total number of simulations of the root;
    - depth at most d (e.g. $d = 3$);
  - average the number of wins and the number of simulations for each of these nodes.

  This can be computed recursively (from the root), using commands like $MPI\_AllReduce$ which have a cost logarithmic in the number of nodes. A

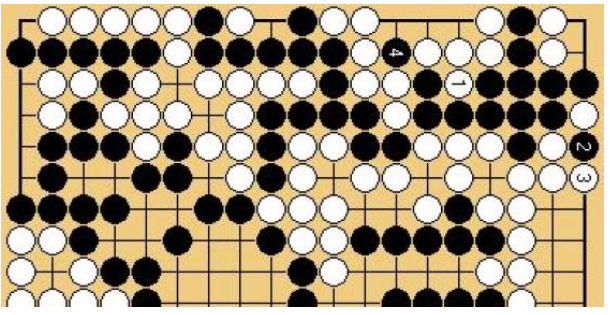

Fig. 3. FUEGO as White played the very bad move (1) during the 56^{th} KGS tournament and lost the game. This is an example of a situation very poorly handled by computers.

special case is **slow root parallelization:** this is slow tree parallelization, but with depth at most $d = 0$; this means that only the root is considered.
- *Voting schemes.* This is a special case of tree parallelization advocated in [7], that we will term here for the sake of comparison with other techniques above **very slow root parallelization:** this is slow root parallelization, but with frequency $f = 1/t$ with t the time per move: the averaging is only performed at the end of the thinking time. There is no communication during the thinking time, and the drawback is that consequently there's no load balancing.

It is usually considered that fast tree parallelization does not perform well; we will consider only other parallelizations. We present in Table 4 the very good results we have in 19x19 and the moderately good results we have in 9x9 for slow tree parallelization.

Table 4. Experiments showing the speed-up of "slow-tree parallelization" in 9x9 and 19x19 Go. We see that a plateau is reached somewhere between 8 and 16 machines in 9x9, whereas the improvement is regular in 19x19 and consistent with a linear speed-up - a 63% success rate is equivalent to a speed-up 2, therefore the results still show a speed-up 2 between 16 and 32 machines in 19x19. Experiments were reproduced with different parameters with strong difference; in this table, the delay between two calls to the "share" functions is 0.05s, and x is set to 5%. The results with high numbers of machines will be confirmed in Table 5.

Configuration of game	Winning rate in 9x9	Winning rate in 19x19
32 against 1	75.85 ± 2.49 %	95.10±1.37 %
32 against 2	66.30 ± 2.82 %	82.38±2.74 %
32 against 4	62.63 ± 2.88 %	73.49±3.42 %
32 against 8	59.64 ± 2.93 %	63.07±4.23 %
32 against 16	52.00 ± 3.01 %	63.15±5.53 %
32 against 32	48.91 ± 3.00 %	48.00±9.99 %

Below we compare **slow root parallelization** to **the "voting scheme" very slow root parallelization.** With 40 machines and 2 seconds per move in 9x9 and 19x19, the slow root parallelization wins clearly against the version with very slow root parallelization, as shown in Table 5, using a frequency 1/0.35

Table 5. The very good success rate of slow tree parallelization versus very slow tree parallelization. The weakness of voting schemes appears clearly, in particular for the case in which huge speed-ups are possible, namely 19x19.

Framework	Success rate against voting schemes
9x9 Go	63.6 % ± 4.6 %
19x19 Go	94 % ± 3.2 %

against the very slow root parallelization. As a rule of thumb, it is seemingly good to have a frequency such that at least 6 "averagings" are performed; 3 per second is a stable solution as games have usually more than 2 seconds per move; with a reasonable cluster 3 times per second is a negligible cost.

We now compare **slow tree parallelization** with depth $d = 1$, to the case $d = 0$ (slow root parallelization) advocated in [4]. Results are as follows and show that $d = 0$ is a not such a bad approximation.

Time per move	Winning rate of slow-tree parallelization (depth=1) against slow-root parallelization
2	50.1 ± 1.1 %
4	51.4 ± 1.5 %
8	52.3 ± 1 %
16	51.5 ± 1 %

These experiments are performed with 40 machines. The results are significant but very moderate.

5 Conclusion

We revisited scalability and parallelism in MCTS.

The scalability of MCTS has often been emphasized as a strength of these methods; we show that when the computation time is already huge, then doubling it has a smaller effect than when it is small. This completes results proposed by Hideki Kato[13] or the scalability study `http://cgos.boardspace.net/study/index.html`; the scalability study was stopped at $524,288$ simulations, and shows a concave curve for the ELO rating in a framework including different opponents; Kato's results show a limited efficiency, when computational power goes to infinity, against a non-MCTS algorithm. Seemingly, there are clear limitations to the scalability of MCTS; even with huge computational power, some particular cases cannot be solved. We also show that the limited speed-up

exists in 19x19 Go as well, and not with much more computational time than in 9x9 Go. In particular, cases involving visual elements (like big yose) and cases involving human sophisticated techniques around liberties (like semeais) are not properly solved by MCTS, as well as situations involving multiple unfinished fights. Our experiments also show that the situation is similar in Havannah with good simulations. The main limitation of MCTS is clearly the bias, and for some situations (as those proposed in Fig. 2) introducing a bias in the score formula is not sufficient; even discarding simulations which are not consistent with a tactical solver is not efficient for semeai situations or situations in which liberty counting is crucial.

Several parallelizations of MCTS on clusters have been proposed. We clearly show that communications during the thinking time are necessary for optimal performance; voting schemes ("very" slow root parallelization) do not perform so well. In particular, slow tree parallelization wins with probability 94 % against very slow root parallelization in 19x19, showing that the slow tree parallelization from [10] or the slow root parallelization from [4] are probably the state of the art. Slow tree parallelization performs only moderately better than slow root parallelization when MCTS is used for choosing a single move, suggesting that slow root parallelization (which is equal to slow tree parallelization simplified to depth= 0) is sufficient in some cases for good speed-up - when MCTS is applied for proposing a strategy (as in, e.g., [1] for opening books), tree parallelization naturally becomes much better.

Acknowledgements. We thank the various people who inspired the development of MoGo: Bruno Bouzy, Tristan Cazenave, Albert Cohen, Sylvain Gelly, Rémi Coulom, Rémi Munos, the computer-go mailing list, the KGS community, the Cgos server, the IAGO challenge, the Recitsproque company, the Spanish Federation of Go, and in particular the organizers of the Spanish Open 2009. We thank Grid5000 for providing computational resources for experiments presented in this paper.

References

1. Audouard, P., Chaslot, G., Hoock, J.-B., Perez, J., Rimmel, A., Teytaud, O.: Grid coevolution for adaptive simulations: Application to the building of opening books in the game of go. In: Giacobini, M., Brabazon, A., Cagnoni, S., Di Caro, G.A., Ekárt, A., Esparcia-Alcázar, A.I., Farooq, M., Fink, A., Machado, P. (eds.) EvoCOMNET. LNCS, vol. 5484, pp. 323–332. Springer, Heidelberg (2009)
2. Berthier, V., Doghmen, H., Teytaud, O.: Consistency modifications for automatically tuned monte-carlo tree search. In: Blum, C., Battiti, R. (eds.) LION 4. LNCS, vol. 6073, pp. 111–124. Springer, Heidelberg (2010)
3. Bruegmann, B.: Monte carlo go (1993) (unpublished draft), `http://www.althofer.de/bruegmann-montecarlogo.pdf`
4. Cazenave, T., Jouandeau, N.: On the parallelization of UCT. In: Proceedings of CGW 2007, pp. 93–101 (2007)

5. Chaslot, G., Saito, J.-T., Bouzy, B., Uiterwijk, J.W.H.M., van den Herik, H.J.: Monte-Carlo Strategies for Computer Go. In: Schobbens, P.-Y., Vanhoof, W., Schwanen, G. (eds.) Proceedings of the 18th BeNeLux Conference on Artificial Intelligence, Namur, Belgium, pp. 83–91 (2006)
6. Chaslot, G., Winands, M., Uiterwijk, J., van den Herik, H., Bouzy, B.: Progressive strategies for monte-carlo tree search. In: Wang, P., et al. (eds.) Proceedings of the 10th Joint Conference on Information Sciences (JCIS 2007), pp. 655–661. World Scientific Publishing Co. Pte. Ltd., Singapore (2007)
7. Chaslot, G., Winands, M., van den Herik, H.: Parallel Monte-Carlo Tree Search. In: van den Herik, H.J., Xu, X., Ma, Z., Winands, M.H.M. (eds.) CG 2008. LNCS, vol. 5131, pp. 60–71. Springer, Heidelberg (2008)
8. Coulom, R.: Efficient selectivity and backup operators in monte-carlo tree search. In: Ciancarini, P., van den Herik, H.J. (eds.) Proceedings of the 5th International Conference on Computers and Games, Turin, Italy (2006)
9. Coulom, R.: Criticality: a monte-carlo heuristic for go programs. Invited talk at the University of Electro-Communications, Tokyo, Japan (2009)
10. Gelly, S., Hoock, J.B., Rimmel, A., Teytaud, O., Kalemkarian, Y.: The parallelization of monte-carlo planning. In: Proceedings of the International Conference on Informatics in Control, Automation and Robotics (ICINCO 2008), pp. 198–203 (2008)
11. Gelly, S., Silver, D.: Combining online and offline knowledge in UCT. In: ICML 2007: Proceedings of the 24th International Conference on Machine Learning, pp. 273–280. ACM Press, New York (2007)
12. Hill, M.D., Marty, M.R.: Amdahl's law in the multicore era. Computer 41(7), 33–38 (2008)
13. Kato, H.: Post on the computer-go mailing list (October 2009)
14. Kocsis, L., Szepesvári, C.: Bandit based monte-carlo planning. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) ECML 2006. LNCS (LNAI), vol. 4212, pp. 282–293. Springer, Heidelberg (2006)
15. Lee, C.-S., Wang, M.-H., Chaslot, G., Hoock, J.-B., Rimmel, A., Teytaud, O., Tsai, S.-R., Hsu, S.-C., Hong, T.-P.: The Computational Intelligence of MoGo Revealed in Taiwan's Computer Go Tournaments. IEEE Transactions on Computational Intelligence and AI in games (2009)
16. Teytaud, F., Teytaud, O.: Creating an upper-confidence-tree program for havannah. In: van den Herik, H.J., Spronck, P. (eds.) ACG 2009. LNCS, vol. 6048, pp. 65–74. Springer, Heidelberg (2010)
17. Wang, Y., Audibert, J.-Y., Munos, R.: Algorithms for infinitely many-armed bandits. In: Advances in Neural Information Processing Systems, vol. 21 (2008)
18. Wang, Y., Gelly, S.: Modifications of UCT and sequence-like simulations for Monte-Carlo Go. In: IEEE Symposium on Computational Intelligence and Games, Honolulu, Hawaii, pp. 175–182 (2007)