Fuego-GB Prototype at the Human machine competition in Barcelona 2010: a Tournament Report and Analysis*

Martin Müller

Abstract

A Human vs Computer Go competition took place in Barcelona, Spain on July 20, 2010. This report provides a report and some analysis of the games played by FUEGO-GB PROTOTYPE in this event. The program played well in its 9×9 games with White, winning against professional 4 Dan Ping-Chiang Chou and losing after achieving a winning middle game position against Chun-Hsun Chou 9 Dan. However, its games with Black and its handicap games on 13×13 were one-sided wins for the humans. Some reasons for these losses are analyzed.

1 Background

The 2010 Human vs Computer Go competition in Barcelona is the latest in a series of man-machine matches organized by a Taiwanese group around Professor Lee Chang-shing of the National University of Tainan. It was held in conjunction with the WCCI 2010 conference. For an overview of previous man-machine competitions in Go, please see Nick Wedd's comprehensive web page at `http://www.computer-go.info/h-c/index.html`.

1.1 The Humans

The human competitors included two professionals and two amateur players. Chun-Hsun Chou 9 Dan[1] is a world-class player who has won the 2007 LG cup, a world-championship level international tournament. Ping-Chiang Chou 4 Dan is his brother. He is apparently considered a 9×9 specialist in Taiwan. On the day before the match, he told me that he has played many games against FUEGO.

Shi-Jim Yen and Shang-Rong Tsai are both strong amateur 6 Dan players with extensive experience playing against Go programs.

*Technical Report TR 10-08. July 2010. Dept. of Computing Science, University of Alberta. Canada.

[1] Names are given in western format here. Chou is the last name.

1.2 The Programs

The same programs as last year competed: ZEN, MOGO, MANY FACES OF GO and FUEGO (competing as FUEGO-GB PROTOTYPE). These programs were the recent winners of the Computer Olympiad competitions. One idea for the future would be to hold an online qualifier, to give more strong programs a chance to participate in these events.

This report focuses on the performance of one program. FUEGO-GB PROTOTYPE is an experimental program based on FUEGO. In addition to the normal FUEGO libraries and Go engine, it uses machine-learned pattern knowledge in form of an extra additive term in the UCT formula. This knowledge base was created by Chris Rosin and originally used in his program GREENPEEP. The program also uses IBM's BLUEFUEGO parallel library, written by Rich Segal, with experimental changes to the FUEGO shared memory search code which improve its performance on large-scale shared memory architectures.

Finally, the program used a new experimental *autobook*, an automatically constructed opening book. The code to construct the book, written by Broderick Arneson, is game-independent and part of the open source FUEGO code base. At the time of the competition, the Go-specific book, containing the specific opening moves for 9×9 Go with 7.5 komi, was pre-release and still in a somewhat immature condition.

For this competition, FUEGO-GB PROTOTYPE ran on an IBM System p5 595 with 56 Power5+ cores, running at 2.1 GHz and using 2-way SMT (hyperthreading). The machine, located at IBM Research in Austin, had 200 GB of memory and was remotely operated by Rich Segal for the competition.

1.3 Competition Format

Program authors were given a choice of playing either on 9×9 or on 19×19 against the professional players. Against the amateurs, 13×13 was an option as well. For FUEGO-GB PROTOTYPE, we decided to play 9×9 games against professionals and 13×13 against amateurs. This is the first time that 13×13 games have been played in these competitions.

2 The Competition

There were four rounds of play. Rounds 1 and 4 were on large board sizes. ZEN and MANY FACES OF GO played 19×19 games against professionals, taking 6 handicap from the 4 Dan and 7 handicap from the 9 Dan. MOGO and FUEGO-GB PROTOTYPE played 13×13 games with 2 handicap and 0.5 komi against the amateurs.

Rounds 2 and 3 were on 9×9 boards, except for MANY FACES OF GO which played on 13×13. Each round consisted of two 9×9 games, one with each color. Games were even with 7.5 komi. MOGO and FUEGO-GB PROTOTYPE played the professionals while ZEN took on the amateurs.

All game records in sgf format are accessible on KGS under user accounts `http://www.gokgs.com/gameArchives.jsp?user=NAME`, where `NAME` is one of

{FuegoGB, MoGoBot5, Zen9, Zen19, ManyFaces1}. Look for games played on July 20, 2010 and ignore the test games that only contain a few moves.

On 9×9, ZEN won 3 out of 4 games against the amateurs. MOGO and FUEGO-GB PROTOTYPE both scored one win and one loss with white and lost their games on black against the professionals.

On 13×13, MOGO won both its games, MANY FACES OF GO won one of two and FUEGO-GB PROTOTYPE lost both against the amateurs. ZEN did not play on this board size.

On 19×19, ZEN won its handicap 6 game against the 4 Dan professional and lost its handicap 7 game. MANY FACES OF GO lost both its handicap games, but played very well against the 9 Dan.

A brief report by Olivier Teytaud, leader of the MOGO team, can be found on his blog: http://teytaud.over-blog.com/article-mogo-vs-humans-in-barcelona-wcci-2010-morning-54191608.html

The current report first discusses FUEGO-GB PROTOTYPE's 9×9 games in rounds 2 and 3, followed by its 13×13 games in rounds 1 and 4. It analyzes some program weaknesses at key moments in some of these games.

3 The 9×9 Games of FUEGO-GB PROTOTYPE

3.1 Round 2: Games against Ping-Chiang Chou 4 Dan

3.1.1 First game with Black - a quick loss

FUEGO-GB PROTOTYPE's first 9×9 game on Black, shown in Figure 1, was a quick loss and a failure for the program's Monte-Carlo search. The opening with 5, 9 and 11 does not put any real pressure on White, and this game never gets off the ground for the program. White has two safe groups and Black cannot create enough territory. This game illustrates two old but still unsolved weaknesses of FUEGO:

1. Even though White has two safe groups, the program remains optimistic, since one of the groups often dies in the playouts.

2. The program does not clearly see that White is alive in seki after the attack with 27-29.

The program's evaluation, shown in Figure 2, is completely off the mark, hovering around 50%. It even becomes more optimistic after White lives in seki. I needed to manually resign for Fuego.

A long time ago, I developed a fix that greatly improves FUEGO's play in seki situations. However, this fix has a small negative overall impact on playing strength, and is therefore not used.

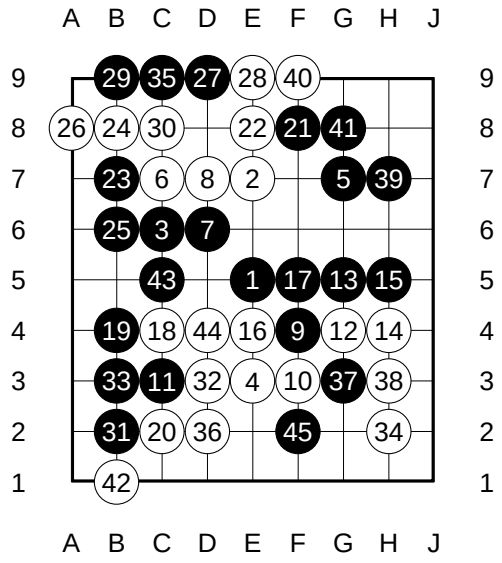

Figure 1: Round 2 game 1, Ping-Chiang Chou 4 Dan (W) - FUEGO-GB PROTOTYPE (B). White wins by resignation.

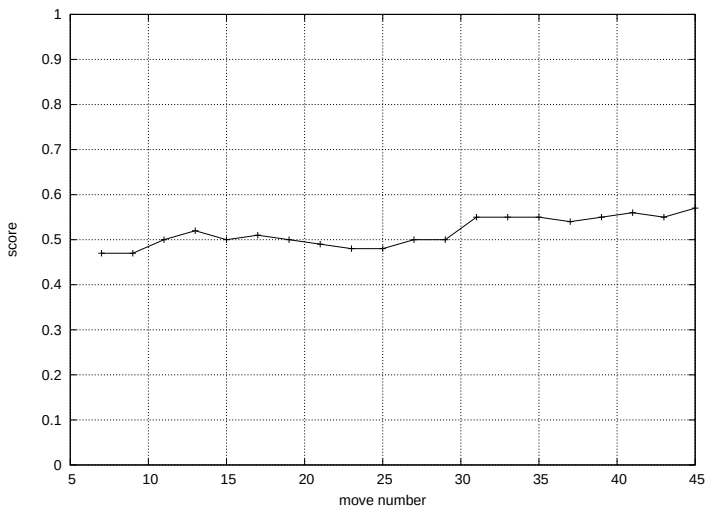

Figure 2: FUEGO-GB PROTOTYPE's evaluation of round 2 game 1.

3.1.2 First game with White - a hard-earned win by half a point

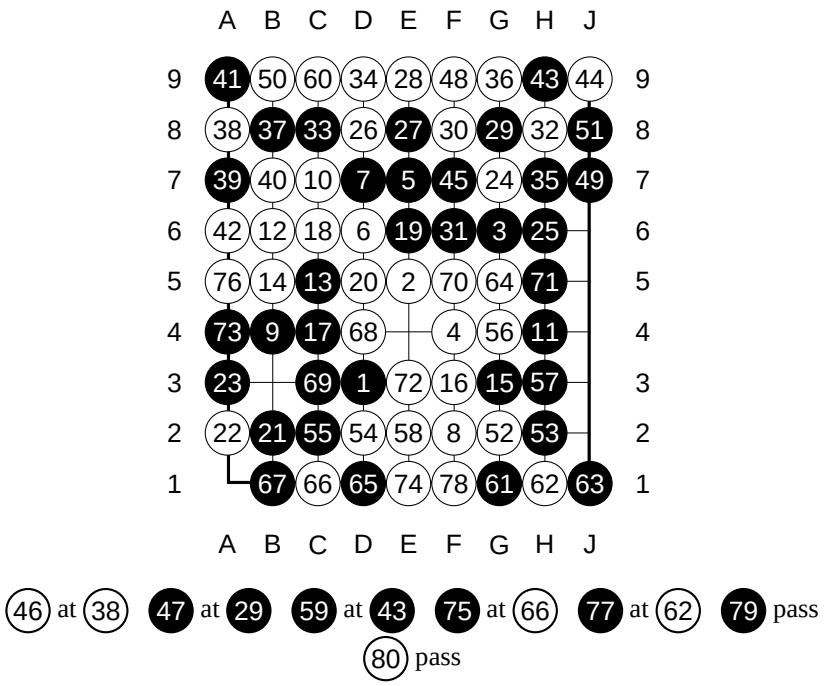

Figure 3: Round 2 game 2, Ping-Chiang Chou 4 Dan (B) - FUEGO-GB PROTOTYPE (W). White wins by 0.5.

Figure 3 shows the only win of FUEGO-GB PROTOTYPE in this competition. I do not have professional comment on this game so all remarks are my own. The opening looks difficult for White, with Black taking solid profit on both sides. The new book did not contain this opening, and the program was out of book on move 3. Still, the program likes these influence-oriented openings, and its evaluation was around 0.58. Black takes profit with moves 7, 11 and 15, then lives with 21. The game looks difficult for White but FUEGO-GB PROTOTYPE found a way to win. Move 14 is an interesting reply against the peep. Move 19 is a forcing move but not without drawbacks because Black becomes short of liberties. I do not understand the program's probe at 22. FUEGO-GB PROTOTYPE's principal variation here is extremely deep, 63 ply: A2 A3 G7 F6 D2 C2 H6 H5 H7 F5 G4 G2 E4 G5 F8 E9 C8 C9 J8 D9 B8 G1 F1 E2 E1 D1 J2 H3 B9 H9 F9 J6 E8 D8 J5 H2 J1 F7 G8 J4 H1 C1 A7 A1 E3 G9 A5 H8 J9 H9 D2 G9 H8 H9 E2 A4 G9 J3 D4 C3 J1 H1 A9. Only the first few moves make sense. Move 24 leads to the decisive fight. The professionals looked at a few variations here after the game, but I do not think they found an improvement for Black. Up to 32, White successfully reduces this corner. Up to 51, Black wins the corner back by double ko, but the reduction ends

in sente for White and is therefore a success. 52 is a very solid move and guarantees victory. 70 looks strange, since the push at 71 is natural, but it does not change the final score. White wins by 0.5 points. A nice victory for FUEGO-GB PROTOTYPE from a difficult opening. As Figure 4 shows, the program's evaluation climbed steadily throughout the game.

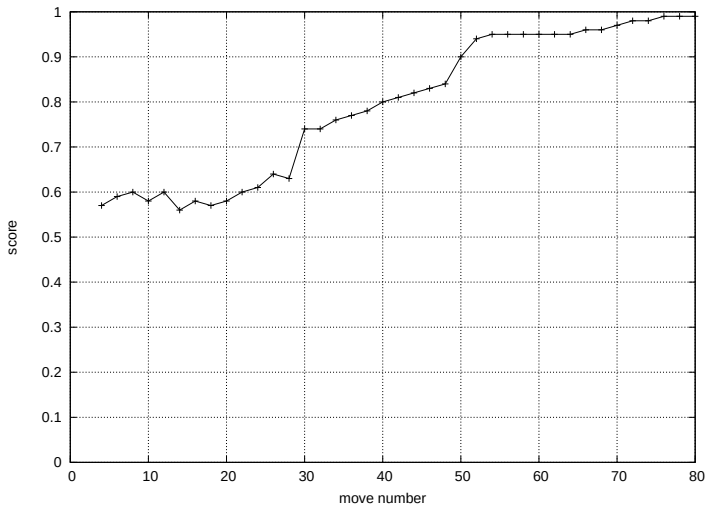

Figure 4: FUEGO-GB PROTOTYPE's evaluation of round 2 game 2.

3.2 Round 3: Games against Chun-Hsun Chou 9 Dan

3.2.1 Second game with Black - another quick loss

Like his brother before, Chun-Hsun Chou 9 Dan had no trouble winning his game with White, shown in Figure 5. After move 10, both white groups are alive. The moves 11 and 13 already look desperate. They actually make White's job easier by giving up a large corner. White concentrates on living with the other group and coasts to an easy victory. In contrast to the human, the capturing race in the top right is not so clear to the program. While always below 50%, its evaluation score is reasonably close to even up to move 35, as shown in Figure 6. Only after White's move 36, which simplifies the tactics in the lower left, the program's evaluation starts to drop rapidly.

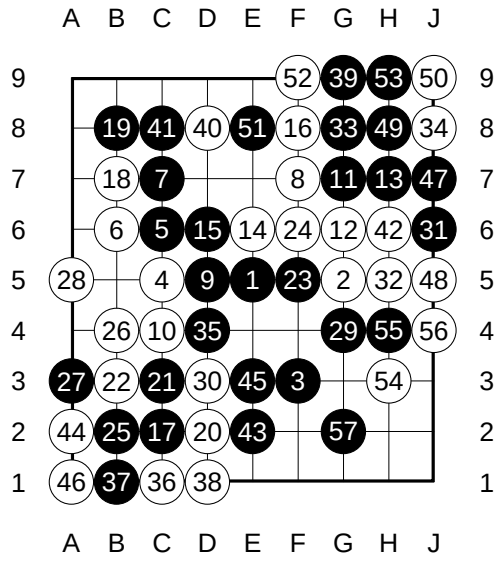

Figure 5: Round 3 game 1, Chun-Hsun Chou 9 Dan (W) - FUEGO-GB PROTOTYPE (B). White wins by resignation.

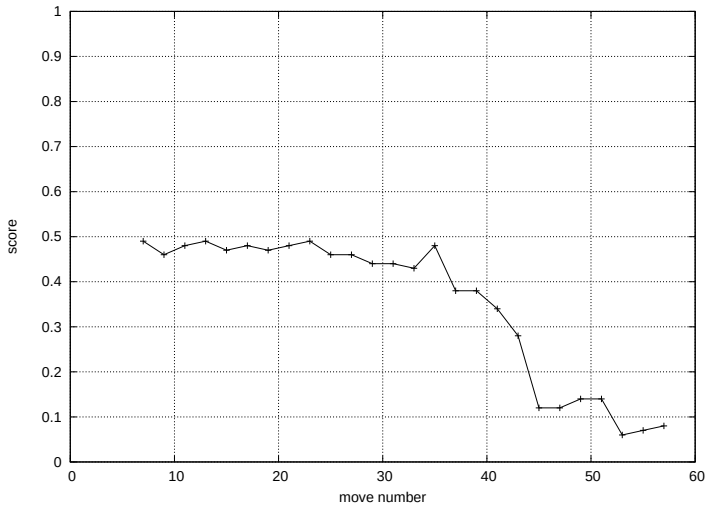

Figure 6: FUEGO-GB PROTOTYPE's evaluation of round 3 game 1.

3.2.2 Second game with White - a strong performance spoiled by one bad mistake

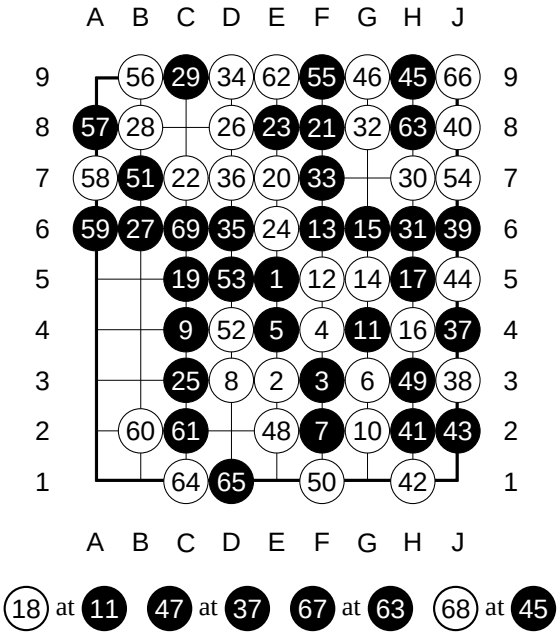

Figure 7: Round 3 game 2, Chun-Hsun Chou 9 Dan (B) - FUEGO-GB PROTOTYPE (W). Black wins by resignation.

FUEGO-GB PROTOTYPE played well in this second game with White, shown in Figure 7, but still lost the game. At move 36, Chun-Hsun Chou 9 Dan knew that he was in a losing position, and thought for a long time to set up a trap ("make a hole" as he called it afterwards). With move 40, the program promptly fell into the hole. This is the critical mistake that lost the game. In Figure 8, if White simply connects the ko at A, there are different possible variations but in any case the position ends up as a ko that White will win because Black has no threats. However, move B as played in the game leads to a different ko that White loses. White has to give way and let Black capture in sente with 49. This gives Black enough liberties to win the capturing race and the game.

Analyzing the loss, FUEGO-GB PROTOTYPE did not foresee Black's good move 45. Still, even at move 46 the program estimated its winning probability at 67% after 10 million simulations. All PV's up to move 50 show white getting the key point of B7. After Black played there with move 51, the winning probability immediately dropped to 45%. At move 56, the probability inexplicably jumped up again to 56%, with a PV starting with B9 C6 A8. FUEGO-GB PROTOTYPE can not resolve this complex capturing race, which involves multiple blocks, under competition time limits.

In contrast, the professional player could use his vast knowledge and feel for the game to focus on reading out the crucial variations.

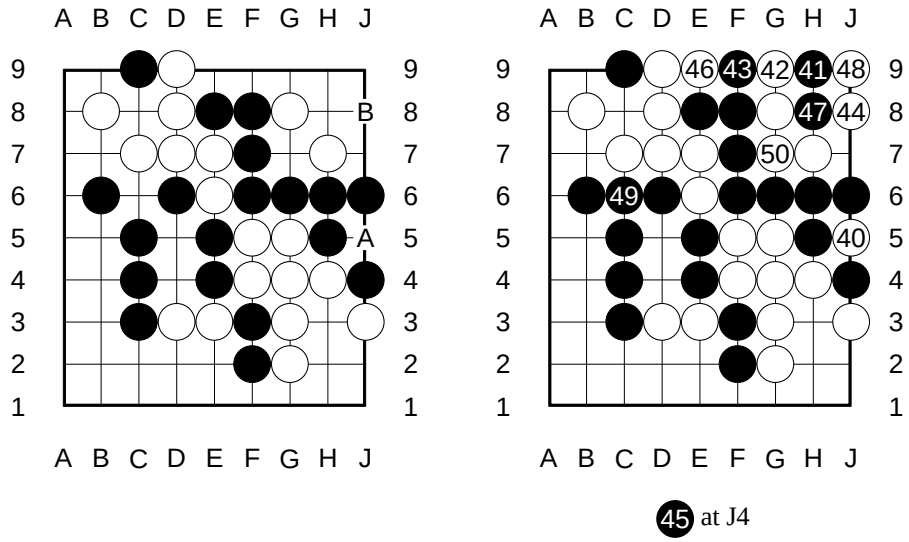

Figure 8: Left: The mistake. White to play. A would win. B, as played in the game, loses. Right: one sample continuation after A. White wins the all-or-nothing ko fight.

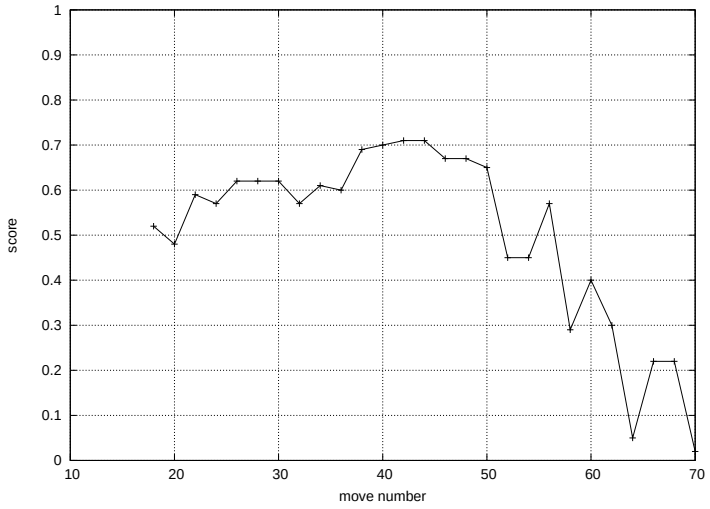

Figure 9: FUEGO-GB PROTOTYPE's evaluation of round 3 game 2.

4 The 13×13 Games of FUEGO-GB PROTOTYPE

4.1 Round 1: Game against Prof. Shang-Rong Tsai, Amateur 6 Dan

Prof. Tsai commented that the program played human-like, without big mistakes. He pointed out that move 58 was slow. In typical Monte-Carlo style, the program gives up three corners quickly and tries to build influence. Black wins the ko fight in the forth corner, but with the successive moves of 57 and 59 White greatly reduces it. However, the program's evaluation, shown in Figure 11, remains above 70% until late in the game. The next section analyzes the reasons for this behavior.

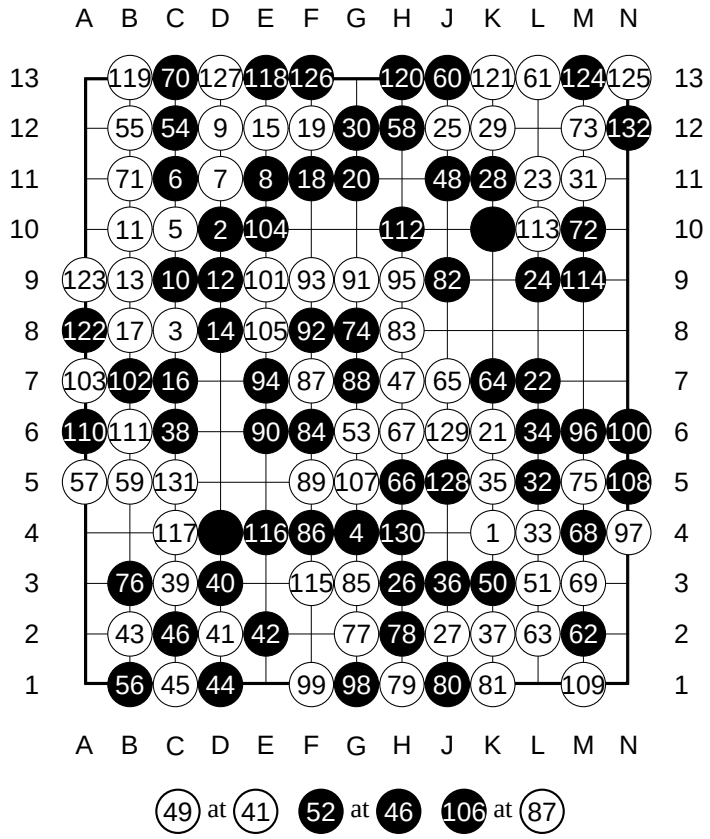

Figure 10: Round 1, Shang-Rong Tsai 6 Dan - FUEGO-GB PROTOTYPE (2 handicap). White wins by resignation.

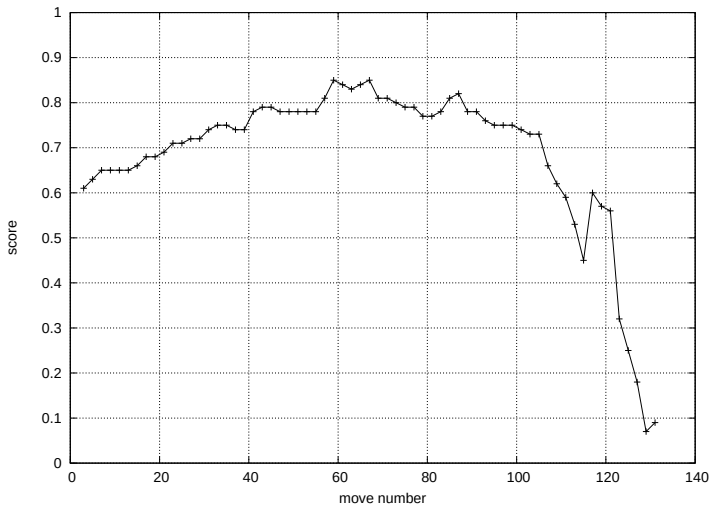

Figure 11: FUEGO-GB PROTOTYPE's evaluation of its round 1 game against Prof. Tsai.

4.1.1 Analysis of Evaluation Problems

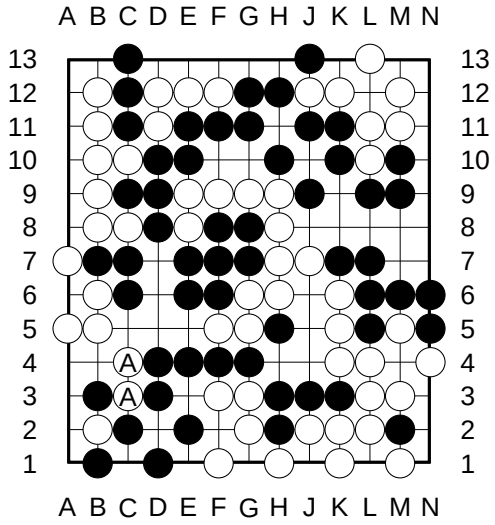

Figure 12: Round 1 game vs Shang-Rong Tsai 6 Dan, position after move 117. Black to play.

Figure 12 shows the position after White's move 117. The game is already completely lost for the program. White has 13 points in the top left, 5 in the top right, 8 in the bottom right, plus 0.5 komi for a score of 26.5. Black has 6 in the bottom left, 2-3 in the center, 9-10 in the top and right for a total of about 18. Yet after searching over 3 million simulations, FUEGO-GB PROTOTYPE has a 60% win rate in its search. This can be explained by looking at the territory statistics.

Figure 13 shows the pointwise evaluation after 10000, 100000, 1 and 10 million simulations. These statistics were computed with FUEGO revision 1148, which is the version used as the basis for FUEGO-GB PROTOTYPE in Barcelona. In this experiment it is running single-threaded and without Greenpeep knowledge, but the same problem occurs in all versions.

The statistics show clear problems evaluating the safety of stones. The semeai in the top left is won by White, but the territory score at each local point is only about -0.2 instead of the true value -1. Worse, the score for the safe white stones on C3 is +0.25, indicating that more often than not those stones are captured by Black in the simulations.

Looking at the changes in evaluation with increasing number of simulations, the numbers seem to improve for the fight at the top left, but not for the stones in the bottom left. Most of the time, these fights are not resolved in the in-tree phase, and the simulations often misplay them.

4.1.2 A Sample Simulation

Figure 14 shows a simulation starting from the position after move 117. Move 1 is a capture, moves 2 is atari-defense and moves 3 and 4 are patterns which also happen to be atari-defense and capture. Moves 4 and 5 are captures. Move 5 accidentally affects the tactical status of the C4 block. It can now be captured in ko. 6-8 are all random moves and 9 is a pattern reply. After the 6-7 exchange the status of C4 becomes unsettled - it can be captured or escape without ko, depending on who plays first locally. After the 10-11 exchange it is still unsettled. White can defend at A8 or A6, and Black can capture at B4. None of these moves are generated by a FUEGO playout policy other than random.

In the sequence from 13 to 24, moves 13, 14, 16, 17, 20 and 22 are random, moves 15 and 19 are atari-defenses, 18, 21 and 24 are patterns and 23 is capture. With move 13, the fate of blocks C11 and D11 becomes unsettled. White needs to reply at B13. Moves 14-15 and 17-19 are reasonable local sequences. 16 erases half an eye for White. Move 20 does not increase White's liberties but triggers the pattern reply 21 which captures white. Move 22 again reduces its own eye space and at the end of this diagram the whole white group is dead. In the rest of the simulation, several similar fights are played out correctly, but the simulated game ends in a win for Black because of the reversals in the two fights before.

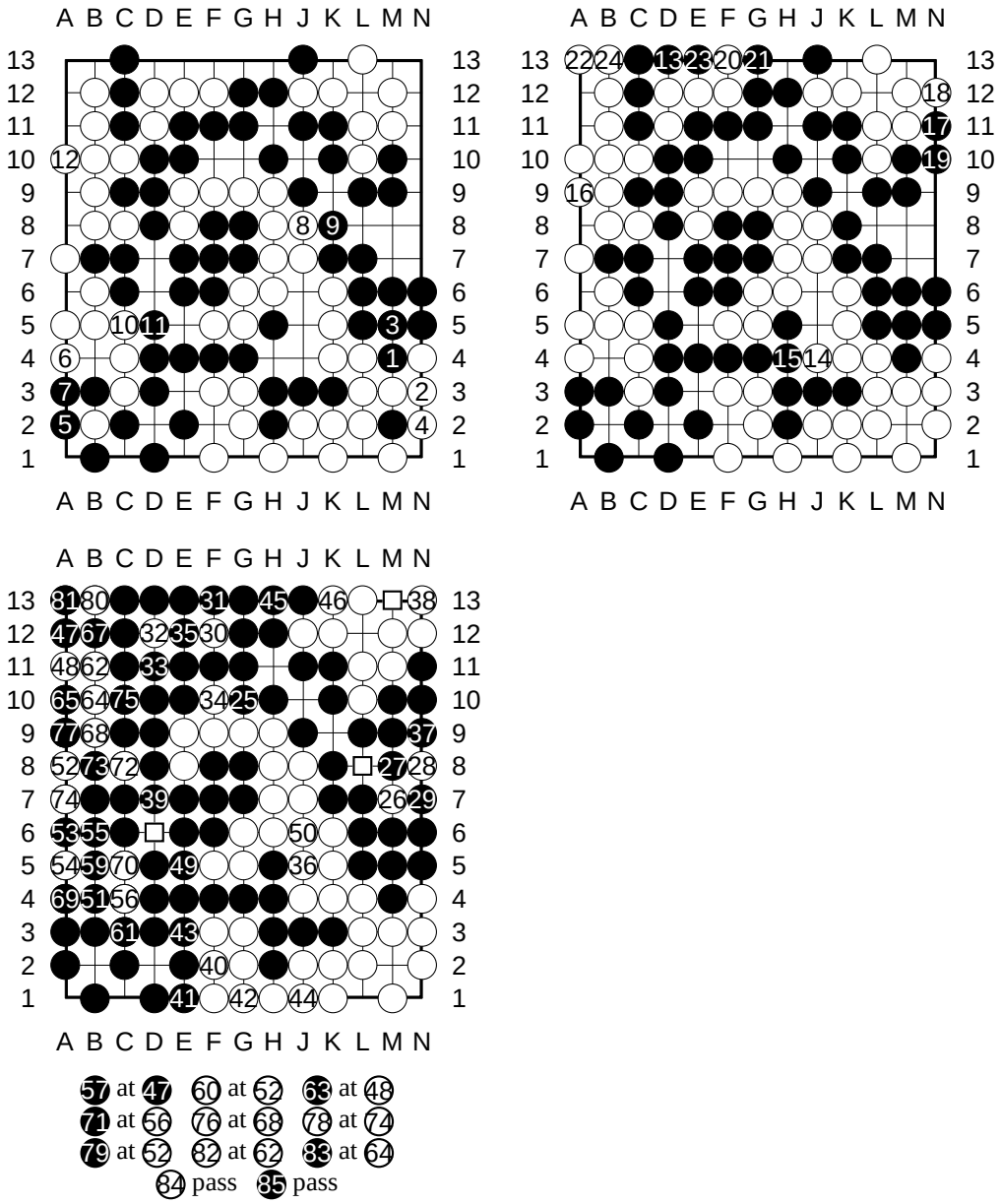

Figure 14: Sample simulation from position after move 117.

4.2 Round 4: Game against Prof. Shi-Jim Yen, Amateur 6 Dan

Even though he won the game, Prof. Yen commented that he was not confident playing with handicap 2. Move 22 should connect at 26. The attack with 36 and 38 is too much - playing at 41 instead would be normal. Move 51 is huge. 56 should be at 58. 70 is bad and loses many points. 86 loses points and must be at 87. Black needs to protect the center with move 100 at 103, but White is already leading by a safe margin.

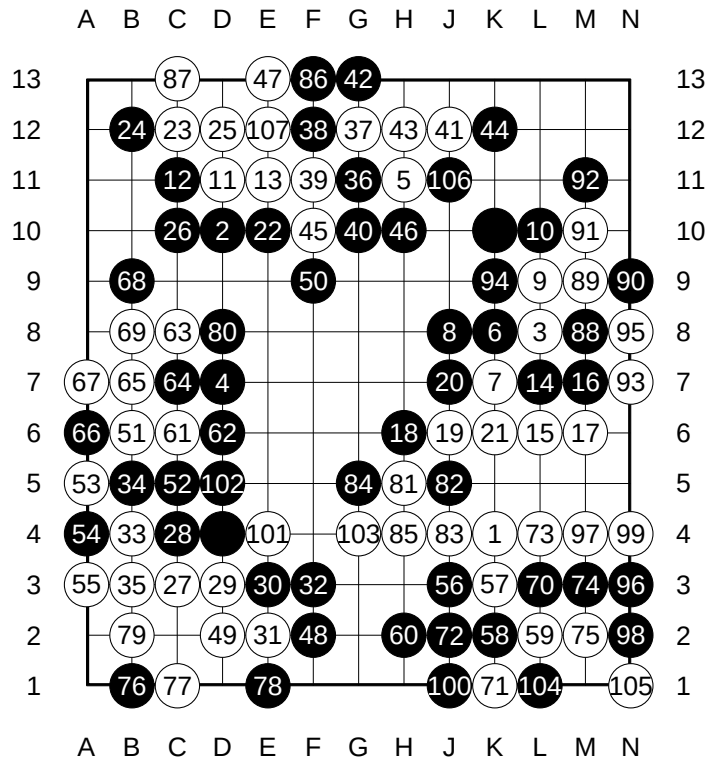

Figure 15: Round 4, Shi-Jim Yen 6 Dan - FUEGO-GB PROTOTYPE (2 handicap). White wins by resignation.

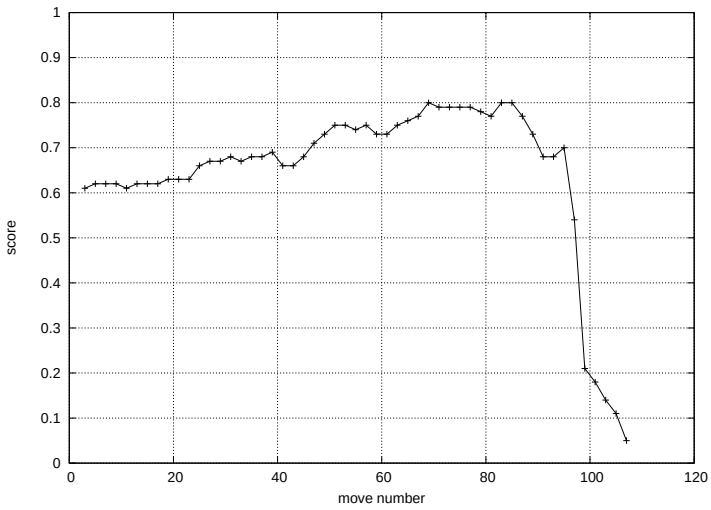

Figure 16: FUEGO-GB PROTOTYPE's evaluation of its round 4 game against Prof. Yen.

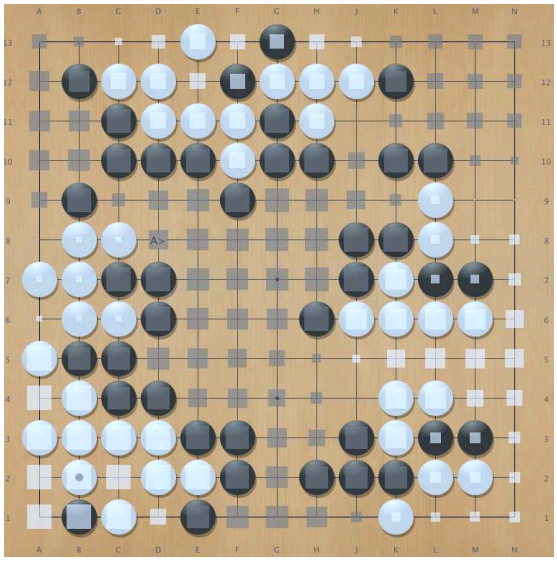

Figure 17: Illustration of Evaluation Problems in the position after move 79 in the game vs. Shi-Jim Yen.

Figure 17 reveals similar evaluation problems as in the game against Prof. Tsai. All of the low-liberty fights were won by White, but none of the territory evaluations is

close to 100% for White. Especially the white stones on B7 are evaluated only slightly in favor of White, even though they are completely safe. Combined, all these "fuzzy" evaluations introduce a large bias which causes the program to be overoptimistic and lose without a fight.

5 Summary and Conclusions

Chou 9 Dan praised the playing strength of MOGO and FUEGO-GB PROTOTYPE on 9×9. If the programs took white in all games, he thinks that they would win 6 or more games out of 10 against him. He did not comment on their performance as Black, but it is clear from this match that professionals have no trouble winning with that color when the komi is 7.5. Mr Chou also praised MANY FACES' play, and said that this was the best 19×19 game played by a program against him since MOGO's win in 2008. He thinks that the best programs are of 1 Dan amateur level on the big board.

I think that FUEGO-GB PROTOTYPE played well on 9×9 with White. It pulled off a half point win in a very tough position against the 4 Dan professional and achieved a winning position against the 9 Dan before going wrong in a complex semeai, which is a well-known weakness of MCTS-based programs.

Including last year's games, in 9×9 against professionals FUEGO's score is now two wins and one loss with White, and three straight losses with Black. All these games used 7.5 komi.

About opening books: the MOGO team is focusing their efforts on building a very strong book for white. In tests against their old book, their program achieves a winning rate of over 90%, according to Olivier Teytaud. In Barcelona, MOGO won a good game on White against the 9 Dan but lost against the 4 Dan with both Black and White despite their huge book. FUEGO's autobook is still under construction. It did not play a major role in this competition. The main questions for computer competitions are how to create a very solid book for White, and how to create a book for Black that poses the most practical problems for fallible opponents. However, any current autobook is still based on the playout results of a particular MCTS engine with all its biases. Positions that are very promising according to the book evaluation may still be objectively lost. To build a really strong book, a *portfolio* approach combining the evaluations of different engines may yield better results. It would also be very interesting to directly compare the books and evaluations of FUEGO and other strong programs. If both programs use their book, positions will result that reflect each program's biases.

On 13×13, the same typical weaknesses of FUEGO that are seen on big boards are apparent: influence-oriented but territorially weak play, overconfidence when facing weak-looking but safe opponent groups, overoptimism in attack, and missing consistency in defending large frameworks. Still, programs can give strong humans a good fight with only 2 handicap stones, which would have been unthinkable only a few years ago.

5.1 Future Work

Playout results should reflect both the relative advantage of different moves from the same starting position and the absolute quality (in terms of winning or losing) of positions. The 13×13 simulations suggest work on local tactics, preserving the escape/capture status of stones. The 9×9 games with seki and semeai problems also suggest that work on improved simulations is critical. It is known informally that programs such as ZEN and VALKYRIA use much more informed simulations than FUEGO. However, technical details are unknown at this time. From a research perspective, large-scale game-specific engineering in order to improve simulation policies is unsatisfying, and more generic approaches are desirable.

Acknowledgements

First of all, thanks to all team members who worked on the program, and especially Rich Segal who also obtained access to and operated the machine. Research contributing to FUEGO-GB PROTOTYPE is supported in part by NSERC and by the DARPA GALE project, Contract No. HR0011-08-C-0110. IBM provided computer time for development, testing and the competition on the IBM System p5 595. Jonathan Schaeffer at the University of Alberta let us use his 80-core cluster for opening book generation and testing.

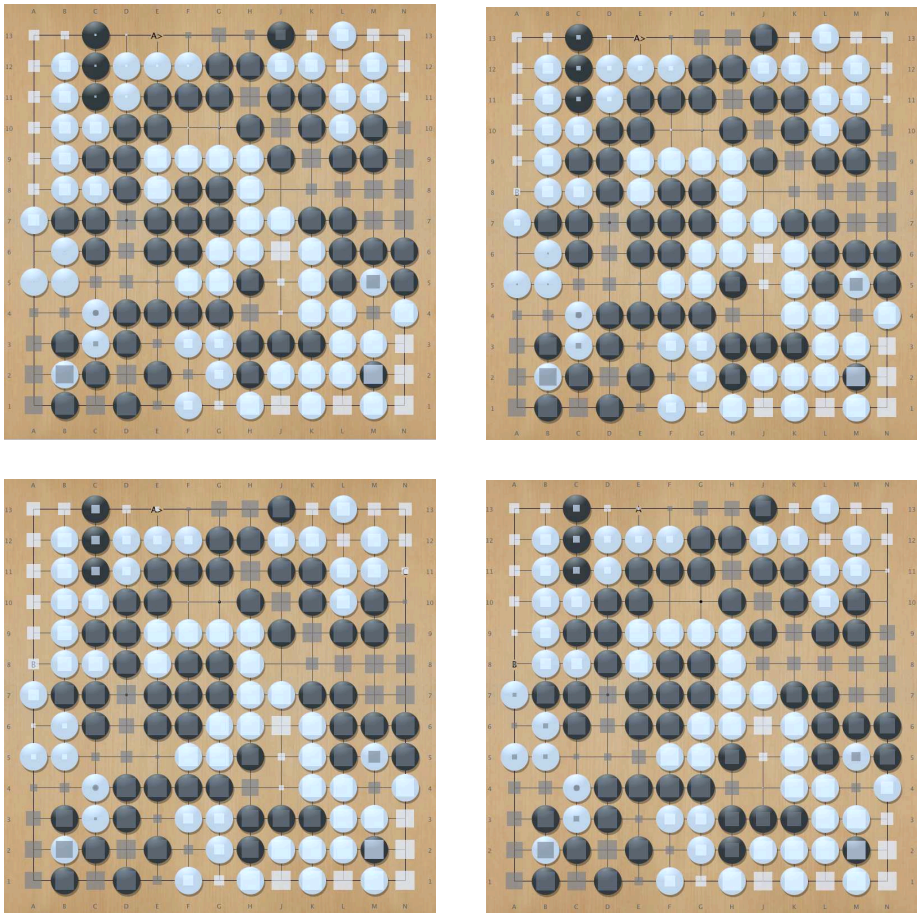

Figure 13: Territorial evaluation of the position in Figure 12 after 10000, 100000, 1 million and 10 million simulations.