Using Modern Chess Software for Opening Preparation

Dmitri A Gusev
Purdue University College of Technology
4444 Kelly St
Columbus, IN 47203
812-348-2029
dgusev@purdue.edu
Fax (812) 348-2016

Abstract

A wide variety of modern chess software products is available to the modern professional and amateur chess players alike, helping them improve their chess skills and prepare for online and traditional tournaments. These products include chess user interfaces (UIs), traditional Alpha-Beta (AB) and emergent Neural Network (NN) chess engines, game databases, opening databases and electronic books, chess-specific cloud services, tournament broadcast tools, online tutorials, tactical problem collections, and endgame tablebases (EGTBs). All of these tools except the last two categories can be used to work on opening preparation, an important component of chess training. In this paper, the author presents his computer-based approach to opening preparation tested in chess classes at the Russian School of Indiana for advanced beginner players. The materials used to develop the approach included game openings from the games played in the Free Open-Source Chess Engine Contest (FOSCEC) broadcast online by the author's CIT students at Purdue Polytechnic Columbus. We will discuss the choices of tools and equipment, how the more popular and/or promising opening variations were identified and analyzed, the lessons learned, and the future work.

INTRODUCTION

Chess is a classical turn-based strategy game played on an 8x8 physical or virtual board with white and black pieces (pawns, knights, bishops, rooks, queens, and kings). The game enjoys broad popularity worldwide, as its rules are pretty easy to learn, the design is well-balanced, and the gameplay can be a lot of fun. However, the task of mastering the deeper intricacies of chess presents a formidable challenge. Many software applications have been developed to implement chess as a video game on numerous electronic platforms and help millions of novices, seasoned amateurs, and chess professionals improve and maintain their skills. These applications include the following.

1. Chess engines capable of analyzing the game's positions and playing the game. They utilize variations of the traditional *Alpha-Beta (AB)* minimax algorithm, emergent techniques based on *Neural Networks (NN)*, and, most recently (since 2020), extremely successful hybrids thereof, such as Stockfish 12+ with *NNUE (Efficiently Updatable Neural Network)* for position evaluation (Stockfish, 2021).

2. Game databases, such as Mega Database (2020) of human over-the-board (OTB) games, Tim Harding's UltraCorr 2021 database of correspondence games (UtraCorr, 2021), and CCRL's database of games played by computer chess engines (CCRL, 2021).

3. Chess user interfaces (UIs) that facilitate communication between the user and the engine(s) and/or databases. They include both commercial products, such as ChessBase 16 (2021), and free solutions, such as Arena (2021) or Shane's Chess Information Database (SCID, 2021).

4. Chess tournament broadcast tools, such as Norman Schmidt's CCCC (2018).

5. Opening databases and electronic books for chess engines to use early in the game and for human players to explore and learn chess openings. They include the Cerebellum opening book for Brainfish by Thomas Zipproth (Cerebellum, 2021) and Fauzi Dabat's opening book (Fauzi, 2021).

6. Chess-specific cloud services, such as the ChessBase Engine Cloud (2021).

7. Tactical problem/puzzle collections, such as those available at chess.com (2021) and lichess.org (2021).

8. Endgame tablebases (EGTBs), such as Ronald de Man's Syzygy (2021) and the pioneering 7-piece Lomonosov tablebases (2012) calculated at the Computer Science Department of Moscow State University.

A chess game is traditionally divided into three stages: *opening*, where a lot of attention is paid to piece development and control over the center of the board; the *middlegame*, with its tactics and strategies of attack and defense; and the *endgame*, where few pieces are left on the board, so the kings become active and pawn promotion gains utmost importance. All of the software tools listed above except the last two categories can be used to work on opening preparation, an important component of chess training.

In this paper, the author, an International Chess Federation (FIDE) National Instructor, presents his computer-based approach to opening preparation tested in chess classes taught online at the Russian School of Indiana (2021). The materials used to develop the approach include openings from the computer chess games played in the Free Open-Source Chess Engine Contest (FOSCEC, 2014) broadcast online by the author's CIT undergraduate students at Purdue Polytechnic Columbus as part of the chess-related projects described in detail in the author's previous work (Gusev, 2018).

In the next sections of the paper, we will discuss the choices of tools and equipment, how the more popular and/or promising opening variations were identified, ordered, illustrated, and analyzed. We will then present our conclusions and discuss plans for the future work.

TOOLS AND EQUIPMENT

Even though many modern chess engines have been ported to other platforms, such as Linux and Android (Abshire and Gusev, 2015), the author used Windows laptops, desktop workstations, and servers for this project to take advantage of the convenient chess GUI tools — ChessBase 13, Deep Fritz 14 (2014), and Arena 3.5.1. The author enhanced the truly massive Computer Chess Rating Lists (CCRL) database of 3,172,504 games played in 2005-2019 by adding 215,485 engine games from numerous other sources, including 4 seasons of FOSCEC, 44 themed opening tournaments ran by the author in 2012-2019, the first 15 seasons of the Top Chess Engine Championship (TCEC) (2021), the Chess Engines Grand Tournament (CEGT) archive, 8 Computer Chess Championship (CCC) events (CCC, 2019), the FastGM (2021) archive, World Computer Chess Championship (WCCC) and World Chess Solving Championship (WCSC) events held by the International Computer Games Association (ICGA) (2021), the AlphaZero vs. Stockfish 8 match, Frank Quisinsky's FEOBOS project (2018), etc. This tool dubbed CCRL+ was then used, along with Mega Database 2019 (7,519,541 Over the Board (OTB) games)) and the engine evaluations from the ChessBase Cloud, combined with the local Cfish evaluations (Cfish, 2017) to identify the more popular and/or promising variations to bring to the attention of the author's advanced beginner chess students. This part of the work was completed using ChessBase 13.

For the purpose of illustration, the popular variations were extended using Brainfish version from February 8, 2019 configured to use its Cerebellum opening book under Arena 3.5.1. The details of the process will be explained in the subsequent sections.

The rationale for tool and equipment selection, besides the obvious availability considerations, involved the realization that the current situation is opposite to what the author experienced back in the 1980s, when it was hard to find comprehensive information on chess openings. We have too much data! No person can view, much less analyze, millions of games. Some of the modern Big Data is of much better quality than what was available in the old days, while some other information is just as unreliable as it used to be. We have selected tools and equipment suitable for quick massive statistical processing of the chess game data.

IDENTIFICATION OF POPULAR VARIATIONS

The two big categories of what the students of chess should learn are *what to do* and *what not to do* in the openings. We will concentrate on the former aspect, leaving the latter to the authors of books on opening traps and catastrophes, such as (Wall, 2010). That other kind of books has great entertainment value, but should not be substituted for serious opening research of more practical value.

Initially, we aimed at selecting 1,023 popular and/or promising variations, this "magic number" being the limit of how many games Arena 3.5.1 can play in one match. Our rule of thumb was, therefore, to stop splitting Mega Database 2019 variations once they were down to approximately 7,519,541:1,023 = 7,350 games per variation. We took into consideration the percentages of points won by the white and provided by ChessBase, along with the cloud and local engine evals. The *Principal Variation* (PV) is a sequence of moves that an engine considers best and therefore expects to be played. The number of PVs to be analyzed locally in a given position remained set at 5 most of the time, so as to explore at most 5 possibilities at a time.

As expected, the variations would not split evenly on popularity. Furthermore, even as we kept track of how many variations we expected to pick starting from each position, we kept encountering "promising" continuations, which, while not being popular, showed good percentages and good computer evals. Once we were down to four variations to pick starting from a given position, we were able to complete selection directly, with a little bit of effort.

One major nuisance that made our recursive process more complex was that, from time to time, we would stumble upon a popular *transposition* of moves leading to the same position that we had encountered or were about to encounter elsewhere in the search tree. (Not all move transpositions are legal, and not every legal transposition of moves is safe.) We kept track of such popular transpositions and added their contributions to the corresponding positions to allow more branching to happen afterwards, according to the total combined popularity of the position.

With those practical considerations in place, the manual selection process produced a set of 1,442 popular and/or promising variations and 167 popular transpositions. Given that the first move 1. e4 occurs in 50.9% (~1/2) of the OTB games, and the move second to that in popularity, 1. d4, happens in 31.9% (~1/3) of the games, the variations were divided nearly uniformly into six volumes:

> Volume 1. Sicilian Defense (1. e4 c5) — 304 variations (21.1% of all selected variations) to cover 20.4 % of all OTB games in Mega Database 2019.

Volume 2. Open Game (1. e4 e5) — 219 variations (15.2% of variations) to cover 12.4% of the games.

Volume 3. Non-Sicilian Semi-Open Games: 1. e4 e6 (French Defense), 1. e4 c6 (Caro-Kann Defense), 1. e4 d6 (Pirc Defense), 1. e4 d5 (Scandinavian Defense), 1. e4 g6 (Modern Defense), 1. e4 Nf6 (Alekhine Defense), 1. e4 Nc6 (Nimzowitsch Defense), 1. e4 b6 (Owen Defense), and 1. e4 a6 (St. George Defense) — 254 variations (17.6% of variations) to cover 18.0% of the games.

Volume 4. Indian Defense (1. d4 Nf6) — 249 variations (17.3% of variations) to cover 17.0% of the games.

Volume 5. Non-Indian responses to Queen's Pawn Game (1. d4 followed by moves other than 1… Nf6) — 238 variations (16.5% of variations) to cover 14.8% of the games.

Volume 6. Openings that do not begin with 1. e4 or 1. d4: 1. Nf3 (Reti Opening), 1. c4 (English Opening), 1. f4 (Bird Opening), 1. g3 (Benko Opening), etc. — 178 variations (12.3% of variations) to cover 17.2% of the games. (~0.2% of the games "fell through the cracks", due to their very unusual openings. This number will grow considerably, once we take into consideration unusual continuations ignored later in the tree search.)

ORDERING THE POPULAR VARIATIONS

Many traditional books on chess openings, including the famous Encyclopedia of Chess Openings (ECO, 2000-2008), begin their discourse with rare variations to proceed gradually to the more common ones. Even as we stored the ECO codes assigned automatically by ChessBase and Arena, we have opted for the lexicographic ordering based on our hex *line codes* that capture priorities of the variations according to their Mega Database popularity, with some exceptions made for promising variations. Figure 1 shows a fragment of an Excel worksheet illustrating our greedy Bottom Line Up Front (BLUF) approach, where the "bottom line" is the *first line*, in the conventional chess terminology.

For example, Line 300 (not seen in the figure) has the hex line code of 11C, where the hexadecimal digit C corresponds to the decimal number 12. The corresponding variation is 1. e4 c5 2. b4 (B20 Sicilian: Wing Gambit). Indeed, 1… c5 is the most popular response to 1. e4, and 2. b4 is the 12th most popular reply to that, after 2. Nf3 (1), 2. Nc3 (2), 2. c3 (3), 2. d4 (4), 2. f4 (5), 2. d3 (6), 2. c4 (7), 2. b3 (8), 2. Bc4 (9), 2. Ne2 (A), 2. g3 (B), and before 2. a3 (D). Even though we do not envision the need to consider more than 15 continuations in a practical opening position, we could use subsequent letters of the English alphabet after F (the largest hexadecimal digit corresponding to the decimal 15) if we needed to — G for the hypothetical 16th choice, H for the 17th choice, and so on, thus trivially extending our approach.

Notice that Line 1 in Figure 1 corresponds to the ECO code of B98 (Sicilian: Najdorf, 7.f4 Be7), which means that the information on this line would be found near the end of the second volume (Volume B) of the Encyclopedia of Chess Openings. This line was played in 12,360 Mega Database 2020 OTB games (0.154%), 8,580 UltraCorr 2021 correspondence games (0.379%), and 3,196 CCRL+ engine games (0.094%).

Figure 1. Lexicographic ordering of the selected popular variations by hex line codes.

For comparison, Line 300 (Wing Gambit) has occurred in 5,306 Mega Database 2020 games (0.082%), 1,528 UltraCorr games (0.068%), and 1,268 CCRL+ games (0.037%). Notice that the end positions of Line 1 and Line 300 are not at the same tree depth relative to the classical starting position. The term *ply* in chess denotes half of a move. Line 1 goes 14 plies deep, while Line 300 is only 3 plies deep. Arena screenshots of the end positions of these two selected popular variations are shown in Figure 2.

Figure 2. End positions of 2 selected popular variations — Line 1 (top) and Line 300 (bottom).

Our job is not done here, because we haven't given our chess students an idea how the events may unfold in each of the selected popular and/or promising variations. Clearly, no advanced beginner is going to study thousands (or hundreds) of games per variation, and neither should they attempt it. Our approach to extending the selected popular variations using Brainfish with Cerebellum will be explained in the next section of the paper.

ILLUSTRATING THE POPULAR VARIATIONS

We ran Arena 3.5.1 matches of Brainfish version from February 8, 2019, configured with its Cerebellum opening book playing blitz games of chess against itself to extract the first lines for the previously selected 1,442 popular and/or promising variations and 167 popular transpositions. Brainfish was configured to always play the best move from its opening book until it ran out of the book moves. The way Arena records the games has allowed us to distinguish Cerebellum's first line continuation of the selected popular variation from how the middlegame and endgame stages subsequently played out under the short time control of 3 minutes per game plus 2 seconds of time increment per move (3' +2" in the conventional chess notation).

Figure 3 shows a position after Move 18 and chess notation for Game 111 of Volume 1 of the resulting First Lines collection. The game illustrates Line 111 of the selected popular variations set, our hex line code 1112111111111111111111, ECO code B33, Sicilian: Pelikan, Chelyabinsk, 9. Nd5 Be7, 11. c3 O-O.

Figure 3. Screenshot of Game 111 of the First Lines collection viewed in ChessBase 13.

You can observe that Brainfish has extended the 11-move (22-ply) popular variation with its 13-ply Cerebellum first line. After that, you see a series of blitz game moves, starting with the 18th move of the black, 18… Qd7, complete with its computer evaluation of -0.14 at a modest depth of 26 plies and the expected PV. The game result was a draw, owing both to the nearly even evaluation of the end position of the popular variation (0.26 at Depth 49 by Stockfish 13) and to the exactly even strength of the computer opponents (Brainfish playing against itself). Interestingly enough, not a single game from Mega Database 2020, UltraCorr 2021, or CCRL+ has reached the position after Move 18. 17. a3 appears to be a novelty.

We decided to look at the win/draw/loss statistics of the 1,442 main games of the First Lines collection grouped into six Volumes. The stats by Volume are illustrated in Figure 4.
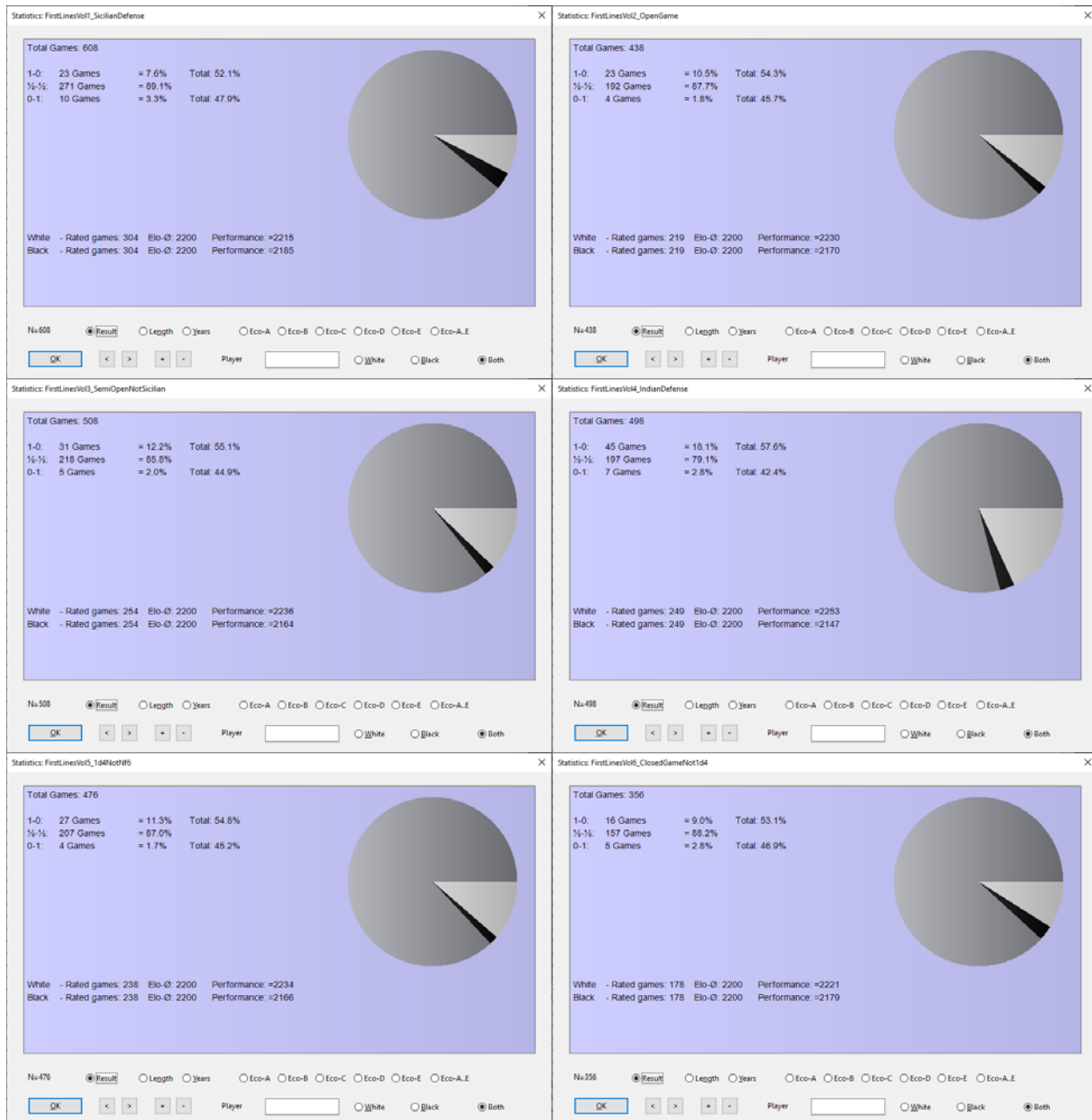
Figure 4. Win/draw/loss statistics for the six Volumes of the First Lines collection of games.

The numbers of "Total Games" displayed above include the selected lines of popular variations that have no game results recorded for them. The stats in Figure 4 confirm that our approach to selecting variations was sound, overall. Indeed, the white got 53.6% of points in Mega Database 2020, 53.9% in UltraCorr 2021, and 53.9% in CCRL+ (with our Brainfish games added). In the next section, we will present and discuss more detailed statistical analysis of the material of Volume 1, Sicilian Defense.

ANALYSIS OF THE POPULAR VARIATIONS

For the popular variations in Volume 1 of our First Lines collection (Sicilian Defense), we have retrieved numerous computer evaluations, primarily by Stockfish 12 and Stockfish 13, from ChessBase Engine Cloud, and filled the remaining few gaps with our local Stockfish 12 evals. We have added evaluations for 30 transpositions and 48 extra variations aimed at the students interested in in-depth research of the Sicilian. The depths of the engine analysis ranged from 32 to 82 plies, with the median value of 47 plies. We then used Excel to produce a linear regression fit of the computer evals to the corresponding percentages of points scored by the white in the CCRL+ database. The resulting graph is shown in Figure 5.
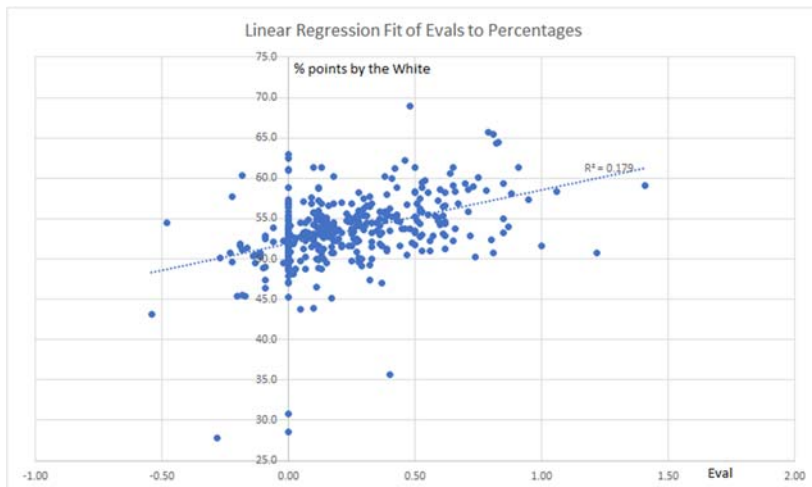
Figure 5. The linear regression fit of evals to CCRL+ percentages for Vol. 1 of First Lines.

The most curious observation here is that the engines that played with the white pieces in CCRL+ games have managed to score so well in the variations given the zero eval ("even game") by Stockfish 12 or Stockfish 13. Our follow-up analysis that seems to indicate that the computer evals for nearly even positions seemingly trend toward zero as the depth of computer search increases is illustrated in Figure 6.

Figure 6. Evals vs. Depth.

In other words, the chess engine sees fewer and fewer chances of beating itself starting with a nearly even position as its analysis gets deeper and deeper. This does not worsen its chances of beating a weaker engine starting

from the same position. It's also important to realize that the computer eval of zero fails to distinguish a game with even chances from a "dead draw" that's practically unavoidable.
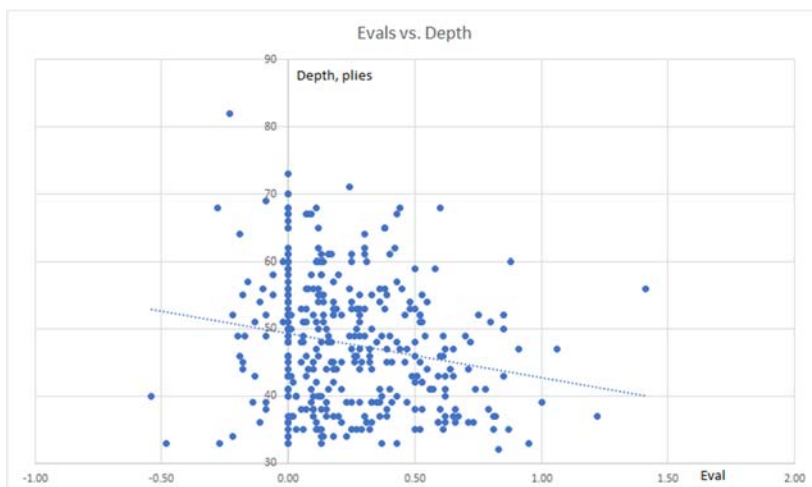
From a practical player's standpoint, it appears that some of these high-depth zero evals are misleading, as they make their variations look less promising in comparison to some others that have not been analyzed to the same depth. Meanwhile, the assessments derived from the statistics of Monte Carlo tree search (CPW, 2021) may sometimes be proven unreliable by encounters with hard-to-find refutations that lurk like the proverbial "skeletons in the closet", ready to jump out and wreak havoc on the board. In other cases, evaluations may optimistically reflect existence of a narrow path to an acceptable position that has to be navigated extremely carefully to avoid the many "landmines". A player naturally gifted with an exceptionally good memory may still choose to learn and memorize the intricate details to be able to play the corresponding variation successfully against those less knowledgeable.

We have also estimated that the main lines of Volume 1 were played in 86.1% of the Mega Database 2020 OTB Sicilians and 87.7% of the UltraCorr 2021 correspondence Sicilians.

CONCLUSIONS AND FUTURE PLANS

The First Lines collection is a novel and useful tool for showing chess students a map of the modern openings' complicated landscape to help them pick openings and variations that suit their emerging individual styles for future in-depth study. The collection can then serve them as a good starting point for building their personal opening repertoires by concentrating on some of the openings and specific variations when playing the white pieces and preparing to play other openings and variations with the black pieces. Many unwanted openings and variations can be avoided, sometimes by cleverly selecting the right transposition of moves. We believe that it is to the students' advantage to learn what works well first, along with the general principles of development in the opening, and only then study opening traps for fun at their leisure. At the same time, the First Line collection helps us avoid the situation in which a trainer would naturally tend to push students toward the openings that the trainer knows best and prefers to play. Those openings may or may not fit the students' styles — attacking, defensive, or balanced.

The author plans to continue to refine and update the First Lines collection. Among possible future projects, the author considers the possibility of converting the First Lines collection into a modern book on openings for beginners and chess instructors, in an electronic and/or traditional paper format.

REFERENCES

Abshire, C., & Gusev, D. A. (2015). Firenzina: Porting a Chess Engine to Android. In: *Mobile and Wireless Technology 2015: Proceedings of the 2nd International Conference on Mobile and Wireless Technology (ICMWT2015)* (pp. 163-171). Berlin, Germany: Springer.

Arena website (2021). Retrieved May 30, 2021 from http://www.playwitharena.de/.

CCC website (2019). Retrieved May 30, 2021 from https://www.chess.com/news/view/ccc-pgn-archive.

CCCC website (2018). Retrieved May 30, 2021 from https://github.com/FireFather/cccc.

CCRL games website (2021). Retrieved May 30, 2021 from https://www.computerchess.org.uk/ccrl/4040/games.html.

Cerebellum and Brainfish website (2021). Retrieved May 30, 2021 from https://zipproth.de/Brainfish/download/.

Cfish website (2017). Retrieved May 30, 2021 from https://github.com/syzygy1/Cfish.

Chess.com puzzles website (2021). Retrieved May 30, 2021 from https://www.chess.com/puzzles

ChessBase 16 website (2021). Retrieved May 30, 2021 from https://shop.chessbase.com/en/products/chessbase_16_mega_package?ref=RF191-8I2RXB2L67.

ChessBase Engine Cloud website (2021). Retrieved May 30, 2021, from https://en.chessbase.com/post/tutorial-how-does-the-engine-cloud-work.

CPW website (2021). *Monte-Carlo Tree Search*. Retrieved May 31, 2021 from https://www.chessprogramming.org/Monte-Carlo_Tree_Search.

Deep Fritz 14 website (2014). Retrieved May 30, 2021 from https://shop.chessbase.com/en/products/deep_fritz_14_english.

ECO (2000-2008). *Encyclopedia of Chess Openings*, 4th Edition, in 5 Volumes. Beograd, Serbia: Šahovski Informator.

Fauzi opening book website (2021). Retrieved May 30, 2021 from https://gumroad.com/l/fzctg.

FastGM website (2021). Retrieved May 30, 2021 from http://www.fastgm.de/.

FEOBOS website (2018). Retrieved May 30, 2021 from https://www.amateurschach.de/.

FOSCEC website (2018). Retrieved May 30, 2021 from http://www.foscec.org/.

Gusev, D. A. (2018). Using Chess Programming in Computer Education. *The 51th ASCUE Annual Conference*, June 10-14, 2018 (pp. 92-102). Myrtle Beach, SC: The Association Supporting Computer Users in Education (ASCUE).

ICGA website (2021). https://icga.org/.

Lichess.org puzzles website (2021). Retrieved May 30, 2021 from https://lichess.org/training.

Lomonosov Tablebases website (2012). Retrieved May 30, 2021 from https://tb7.chessok.com/.

Mega Database website (2020). Retrieved May 30, 2021 from https://en.chessbase.com/post/new-mega-database-2021.

Russian School of Indiana website (2021). Retrieved May 30, 2021 from https://www.indyrussianschool.com/.

SCID website (2021). Retrieved May 30, 2021 from http://scid.sourceforge.net/.

Stockfish website (2012). Retrieved May 30, 2021 from https://stockfishchess.org/.

Syzygy website (2021). Retrieved May 30, 2021 from https://github.com/syzygy1/tb.

TCEC website (2021). Retrieved May 30, 2021 from https://tcec-chess.com/.

UltraCorr website (2021). Retrieved May 30, 2021 from http://www.chessmail.com/UC-2021/UC2021-intro.html.

Wall, B. (2010). *700 Opening Traps*. Seraphim Press.

Presenter's Bio:

Dmitri A. Gusev is an Associate Professor in Computer and Information Technology (CIT) at the Purdue Polytechnic Institute. His primary research interests are imaging, graphics, game development, visualization, and computational linguistics.