

## [Home](#) \* Engine Testing



The ever-optimistic [Wile E. Coyote](#) <sup>1</sup>

**Engine Testing**, the process either to eliminate [bugs](#) and to measure [performance](#) of a chess engine. New implementations of [move generation](#) are tested with [Perft](#), while new features and [tuning](#) of [search](#) and [evaluation](#) are verified by [test-positions](#) and by playing [matches](#) against other engines.

## Table of Contents

[Bug Hunting](#)

[Analyzing](#)

[Tuning](#)

[Test-Positions](#)

[Matches](#)

[Time Controls](#)

[Opening](#)

[Interfaces](#)

[Frameworks](#)

[Chess Server](#)

[Statistics](#)

[Ratings](#)

[Test Results](#)

[Notable Bugs](#)

[Publications](#)

## [Forum Posts](#)

[1995 ...](#)

[2000 ...](#)

[2005 ...](#)

[2010 ...](#)

[2015 ...](#)

## [External Links](#)

## [References](#)

[What links here?](#)

# Bug Hunting

- [Perft \(Perft Results\)](#)
- [Debugging](#)

# Analyzing

- [Logging](#)
- [Profiling](#)
- [Search Statistics](#)

# Tuning

- [Automated Tuning](#)

# Test-Positions

Running sets of test-positions with number of solutions per fixed time-frame is useful to prove whether things are broken after program changes or to get hints about missing knowledge. But one should be careful to tune engines based on test-position results, since solving (possible tactical) test-positions does not necessarily correlate with practical [playing strength](#) in matches against other opponents.

- [Test-Positions](#)

## Matches

Most testing involves running different versions of a program in matches, and comparing results.

## Time Controls

Generally speaking, for testing changes that don't alter the search tree itself, but only affect performance (eg. [move generation](#)) can be tested with given fixed nodes, fixed time or fixed depth. In all other cases the [time management](#) should be left to the engine to simulate real tournament conditions. On the other hand, [debugging](#) is much easier under fixed conditions as the games become deterministic.

A side from the type of [time control](#) one also has to decide on how much time should be spent per game, ie. what the average quality of the games should be like. While one can test more changes in the a certain time at short time controls, it is also relevant how a certain change scales to different strengths. So for example should one increase the [R](#) in [Null move pruning](#) to 3 in depths > 7, this change may only be effectively tested on time controls where this new condition is triggered frequently enough, ie. where the average search depth is far greater than seven. It is hard to generalize, but on average changes of the search functions ([LMR](#), [nullmove](#), [futility](#) or similar [pruning](#), [reductions](#) and [extensions](#) ) tend to be more sensitive to the time control than the tuning of [evaluation](#) parameters.

## Opening

During testing the engines should ideally play the same style of openings they would play in a normal tournament, so not to optimize them for different types of positions. One option is to use the engines own [opening book](#) or one can use [opening suites](#), a set of quiet test positions. In the latter case the same opening suit would be used for each tournament conducted and furthermore each position is played a second time with colors reversed. With these measures one can try to minimize the disparity between tests caused by different openings.

## Interfaces

Free [graphical user interfaces](#) or [command line tools](#) for [UCI](#) and [Chess Engine Communication Protocol](#) compatible engines in engine-engine matches are:

- [Amoeba Tournament Manager](#)
- [Arena](#) by [Martin Blume](#) <sup>[2]</sup>
- [Cute Chess](#) by [Arto Jonsson](#) and [Ilari Pihlajisto](#)  
[Cutechess-cli](#)

- [LittleBlitzer](#) by [Nathan Thom](#) <sup>[3]</sup>

## Frameworks

- [Fishtest](#)

## Chess Server

One can also test an engine's performance by comparing it to other programs on the various internet platforms <sup>[4]</sup>. In this case the different hardware and features like different [Endgame Tablebases](#) or [Opening Books](#) have to be considered.

- [Chess Server](#)
- [Tournaments](#)

## Statistics

The question whether certain results actually indicates a [strength](#) increase or not, can be answered with

- [Match Statistics](#)
- [Pawn Advantage, Win Percentage, and Elo](#)
- [LOS Table](#)

## Ratings

- [Rating System](#)
- [Engine Rating Lists](#)

## Test Results

- [Null Move Pruning Test Results](#)
- [Late Move Reduction Test Results](#)

## Notable Bugs

- [Brute Force \(Program\)](#), [En passant bug](#), [ACM 1977](#) and [ACM 1978](#)
- [Coko - Mate in One?](#), [ACM 1971](#)
- [Chess 2175X vs. Genesis](#), [Promotion](#) bug, [4th Computer Olympiad 1992](#)
- [Nimzo's winning white-black bug](#), [WMCCC 1993](#)
- [Novag Micro Chess - Castling bug](#), [CPWTIPC 1981](#)
- [Proscha](#) capturing its own king versus [Daja](#), [First GI Computer Chess Tournament 1975](#)
- [System Tal vs. XXXX](#), [Promotion](#) bug, [WMCCC 1995](#)

- [Xinix - Mate in One](#), [DOCCC 2000](#) <sup>[5]</sup>

## Publications

- [Tony Marsland](#), [Paul Rushton](#) (1973). [Mechanisms for Comparing Chess Programs](#). [ACM Annual Conference](#), pdf
- [Tim Breitreutz](#), [Jonathan Schaeffer](#) (1984). *Computer vs Computer via Computer*. [ICCA Journal](#), Vol. 7, No. 4
- [John Stanback](#) (1990). *Supercomputing '90: Computer-Chess Testing and Programming Session*. [ICCA Journal](#), Vol. 13, No. 4 » [ACM 1990](#)
- [Larry Kaufman](#) (1993). *How Our PC Chess Programs Are Developed*. [Computer Chess Reports](#) 1992-93, Vol. 3, No. 2, pp. 12
- [Thomas Mally](#) (1993). *Matt in Wieviel?* [PC Schach](#) 3/93 (German)
- [Jeff Rollason](#) (2007). [Statistical Minefields with Version Testing](#). [AI Factory](#), Winter 2007 » [Match Statistics](#)
- [Jónheiður Ísleifsdóttir](#) (2007). *GTQL: A Query Language for Game Trees*. M.Sc. thesis, [Reykjavík University](#), pdf
- [Jónheiður Ísleifsdóttir](#), [Yngvi Björnsson](#). (2008). [GTQ: A Language and Tool for Game-Tree Analysis](#). [CG 2008](#), pdf

## Forum Posts

### 1995 ...

- [Testing Chess Programs](#) by [Jan Eric Larsson](#), [rgcc](#), February 09, 1996
- [Self-test and others rating stuffs...](#) by [Christophe Théron](#), [CCC](#), January 01, 1998
- [Proposal: New testing methods for SSDF \(1\)](#) by [Jeroen Noomen](#), [CCC](#), April 13, 1998

### 2000 ...

- [Using 2 machines for matches \(Linux\)](#) by [Jon Dart](#), [CCC](#), June 24, 2001 » [XBoard](#), [Linux](#)
- [A proposed WAC replacement for testing](#) by [Gian-Carlo Pascutto](#), [CCC](#), September 18, 2001 » [Win at Chess](#)
- [Value of playing different versions of a program against each other](#) by [Tom King](#), [CCC](#), January 06, 2003
- [testing of evaluation function](#) by [Steven Chu](#), [CCC](#), April 17, 2003 » [Evaluation](#)
- [Testing the reliability of forward pruning](#) by [Russell Reagan](#), [CCC](#), May 15, 2003 » [Pruning](#)
- [To programmers: Hints for testing after a partial rewrite](#) by [Federico Corigliano](#), [CCC](#), December 08, 2003
- [Is there a way?](#) by [Ed Schröder](#), [CCC](#), December 13, 2004

### 2005 ...

- [table for detecting significant difference between two engines](#) by [Joseph Ciarrochi](#), [CCC](#), February

03, 2006

- [test methodology](#) by [Giuseppe Cannella](#), [Winboard Forum](#), November 13, 2006
- [Testing and debugging chess engines](#) by [Patrice Duhamel](#), [Winboard Forum](#), December 03, 2006
- [Programmer bug hunt challenge](#) by [Ed Schröder](#), [CCC](#), May 04, 2007 » [Portable Game Notation, En passant](#)
- [a beat b.b beat c.c beat a question](#) by [Uri Blass](#), [CCC](#), May 16, 2007 » [Playing Strength](#)
- [An objective test process for the rest of us?](#) by [Nicolai Czempin](#), [CCC](#), September 12, 2007
- [My new testing scheme](#) by [Zach Wegner](#), [CCC](#), November 20, 2007
- [New testing thread](#) by [Robert Hyatt](#), [CCC](#), August 07, 2008
- [Cutechess-cli: A command line tool for engine-engine matches](#), by [Ilari Pihlajisto](#), [CCC](#), March 16, 2009
- [Cutechess-cli version 0.1.8 released](#) by [Ilari Pihlajisto](#), [CCC](#), September 29, 2009
- [A reason for testing at fixed number of nodes](#) by [J. Wesley Cleveland](#), [CCC](#), November 06, 2009
- [different kinds of testing](#) by [Don Dailey](#), [CCC](#), November 09, 2009
- [more on fixed nodes](#) by [Robert Hyatt](#), [CCC](#), November 10, 2009

## 2010 ...

- [XBoard and epd tournament](#) by [Vlad Stamate](#), [CCC](#), January 31, 2010 » [Chess Engine Communication Protocol](#)
- [Long game vs short game testing](#) by [Vlad Stamate](#), [CCC](#), April 08, 2010
- [Pairings generation based on a big PGN file](#) by [Harun Taner](#), [CCC](#), July 22, 2010
- [hiatus good for bug-finding](#) by [Stuart Cracraft](#), [CCC](#), June 27, 2010

## 2011

- [testing question](#) by [Larry Kaufman](#), [CCC](#), June 01, 2011
- [Debugging regression tests](#) by [Onno Garms](#), [CCC](#), June 16, 2011 <sup>[6]</sup>

## 2012

- [fast game testing](#) by [Jon Dart](#), [CCC](#), January 08, 2012
- [Your best bug ?](#) by [Ed Schröder](#), [CCC](#), August 06, 2012
- [Yet Another Testing Question](#) by [Brian Richardson](#), [CCC](#), September 15, 2012
- [Another testing question](#) by [Larry Kaufman](#), [CCC](#), September 23, 2012
- [A word for casual testers](#) by [Don Dailey](#), [CCC](#), December 25, 2012

## 2013

- [A poor man's testing environment](#) by [Ed Schröder](#), [CCC](#), January 04, 2013 <sup>[7]</sup> » [Match Statistics](#)
- [engine-engine testing issues](#) by [Jens Bæk Nielsen](#), [CCC](#), January 20, 2013
- [Beta for Stockfish distributed testing](#) by [Gary](#), [CCC](#), March 05, 2013 » [Fishtest](#)

- [Fishtest Distributed Testing Framework](#) by [Marco Costalba](#), [CCC](#), May 01, 2013 » [Fishtest](#)
- [cutechess-cli 0.6.0 released](#) by [Ilari Pihlajisto](#), [CCC](#), July 12, 2013
- [fast testing NIT algorithm](#) by [Don Dailey](#), [CCC](#), August 22, 2013
- [OICS: Computers Only ICS based Chess server for anyone](#) by [Joshua Shriver](#), [CCC](#), August 26, 2013 » [OICS](#)

## 2014

- [testing procedure](#) by [Daniel José Queralto](#), [CCC](#), February 23, 2014

## 2015 ...

- [Bullet vs regular time control, say 40/4m CCRL/CEGT](#) by [Ed Schröder](#), [CCC](#), August 29, 2015
- [Static evaluation test positions](#) by [Shawn Chidester](#), [CCC](#), November 25, 2015  
[Re: Static evaluation test positions](#) by [Ferdinand Mosca](#), [CCC](#), November 26, 2015 » [Python](#)

## 2016

- [Ordo 1.0.9 \(new features for testers\)](#) by [Miguel A. Ballicora](#), [CCC](#), January 25, 2016
- [cluster versus single server](#) by [Folkert van Heusden](#), [CCC](#), April 28, 2016
- [Testing using many computers and architectures](#) by [Andrew Grant](#), [CCC](#), September 14, 2016
- [command line engine match?](#) by [Erin Dame](#), [CCC](#), November 06, 2016 » [CLI](#)
- [Testing with different EPD suits for search vs eval changes](#) by [Michael Sherwin](#), [CCC](#), December 23, 2016

## 2017

- [sptr tourney manager](#) by [Richard Delorme](#), [CCC](#), January 24, 2017 » [Amoeba Tournament Manager](#), [SPRT](#)
- [how to properly test the changes to the engine ?](#) by [Mahmoud Uthman](#), [CCC](#), February 01, 2017
- [How to go about chasing a bug like this?](#) by [Colin Jenkins](#), [CCC](#), February 09, 2017 » [Debugging](#)
- [How to find SMP bugs ?](#) by [Lucas Braesch](#), [CCC](#), March 15, 2017 » [Debugging](#), [Lazy SMP](#)
- [Testing for Move Ordering Improvements](#) by [Cheney Nattress](#), [CCC](#), March 25, 2017 » [Move Ordering](#), [Search Statistics](#)
- [Testing endgame strength](#) by [Álvaro Begué](#), [CCC](#), June 21, 2017 » [Endgame](#), [RuyDos](#)
- [Opening testing suites efficiency](#) by [Kai Laskos](#), [CCC](#), June 21, 2017 » [Opening](#), [Match Statistics](#)
- [Testing A against B by playing a pool of others](#) by [Andrew Grant](#), [CCC](#), June 24, 2017 » [Match Statistics](#)
- [Core behaviour](#) by [Ed Schroder](#), [CCC](#), June 28, 2017 » [Process](#), [Thread](#)
- [Engine testing & error margin ?](#) by [Mahmoud Uthman](#), [CCC](#), July 05, 2017
- [Engines for testing \(Linux, fast time control\)](#) by [Jon Dart](#), [CCC](#), November 18, 2017 » [Linux](#)

## External Links

- [Cute Chess](#)
- [cutechess · GitHub](#)
- [Testing a chess engine from the ground up](#) from [Home of the Dutch Rebel](#) by [Ed Schröder](#) <sup>[8]</sup> »  
[Match Statistics](#)
- [Regression testing from Wikipedia](#)
- [SPCC](#) by [Stefan Pohl](#)
- [CHESS - Microsoft Research](#) a tool for finding and reproducing [Heisenbugs](#) in concurrent programs.
- [Engine test stand from Wikipedia](#)
- [Terje Rypdal Group](#) feat. [Palle Mikkeltorg](#), [Håkon Graf](#), [Sveinung Hovensjø](#) and [Jon Christensen](#)  
- [Per Ulv](#), 1978, [YouTube](#) Video

## References

1. [^ Hope Springs Eternal](#)
2. [^ Free chess graphical user interface \(GUI\) Arena for chess engines](#)
3. [^ Fast Games](#) by [Manuel Diaz](#), [CCC](#), July 02, 2010
4. [^ Internet chess servers from Wikipedia](#)
5. [^ Defending Humanity's Honor](#) by [Tim Krabbé](#)
6. [^ Regression testing from Wikipedia](#)
7. [^ Testing a chess engine from the ground up](#) from [Home of the Dutch Rebel](#) by [Ed Schröder](#)
8. [^ A poor man's testing environment](#) by [Ed Schröder](#), [CCC](#), January 04, 2013

## What links here?

Page	Date Edited
<a href="#">4th Computer Olympiad</a>	Jul 15, 2017
<a href="#">AI Factory</a>	Dec 26, 2015
<a href="#">Albert Silver</a>	Feb 25, 2018
<a href="#">Albrecht Heeffer</a>	Sep 6, 2016
<a href="#">Aldebaran</a>	Feb 10, 2013
<a href="#">Ananse</a>	Jan 8, 2016
<a href="#">Aquarium</a>	Dec 4, 2016
<a href="#">Arena</a>	Mar 16, 2018
<a href="#">Behting Study</a>	Feb 1, 2015
<a href="#">Bratko-Kopec Test</a>	Jun 1, 2015
<a href="#">CCR One Hour Test</a>	Jun 14, 2016
<a href="#">Chess Engine Communication Protocol</a>	Mar 4, 2018
<a href="#">Chess Player 2150</a>	Jan 1, 2017
<a href="#">Chess Server</a>	Jun 10, 2016
<a href="#">ChessGUI</a>	Jan 17, 2018



Page	Date Edited
<a href="#">ChessPartner</a>	Nov 13, 2017
<a href="#">Christophe Théron</a>	Jan 7, 2016
<a href="#">CLI</a>	Jan 20, 2018
<a href="#">Coko</a>	Jul 13, 2015
<a href="#">Cute Chess</a>	Jul 20, 2017
<a href="#">Cutechess-cli</a>	Jan 6, 2018
<a href="#">Daniel José Queralto</a>	Dec 31, 2017
<a href="#">Debugging</a>	Jan 20, 2018
<a href="#">DOCCC 2000</a>	Aug 15, 2015
<a href="#">Don Dailey</a>	Dec 7, 2016
<a href="#">Ed Schroder</a>	Feb 26, 2018
<a href="#">Eigenmann Endgame Test</a>	Jun 1, 2017
<a href="#">En passant</a>	Jan 16, 2018
<a href="#">Endgame</a>	Sep 18, 2017
<a href="#">Engine Testing</a>	Feb 1, 2018
<a href="#">Evaluation</a>	Feb 1, 2018
<a href="#">Extended Position Description</a>	Dec 2, 2017
<a href="#">Federico Andrés Corigliano</a>	Feb 13, 2014
<a href="#">Ferdinand Mosca</a>	Jul 28, 2017
<a href="#">Fire</a>	Oct 12, 2017
<a href="#">Folkert van Heusden</a>	Nov 24, 2017
<a href="#">Getting Started</a>	Jan 20, 2014
<a href="#">Hannibal</a>	Oct 4, 2017
<a href="#">Home</a>	Jul 2, 2018
<a href="#">Ilari Pihlajisto</a>	Sep 17, 2016
<a href="#">Jeff Rollason</a>	Dec 23, 2016
<a href="#">Jens Bæk Nielsen</a>	Mar 3, 2015
<a href="#">John Nunn</a>	Dec 9, 2017
<a href="#">Jónheiður Ísleifsdóttir</a>	Jun 22, 2015
<a href="#">Kai Laskos</a>	Jan 21, 2018
<a href="#">Kaufman Test</a>	Jan 25, 2013
<a href="#">Kirill Kryukov</a>	Apr 16, 2014
<a href="#">Larry Kaufman</a>	Jan 28, 2018
<a href="#">Lasker-Reichhelm Position</a>	Oct 24, 2017
<a href="#">Late Move Reduction Test Results</a>	Sep 25, 2017
<a href="#">LCT II</a>	Feb 25, 2018
<a href="#">Leila</a>	May 8, 2017
<a href="#">Linux</a>	Jan 21, 2018
<a href="#">LittleBlitzer</a>	Jan 24, 2017
<a href="#">Logging</a>	Jan 20, 2018
<a href="#">LOS Table</a>	Jan 10, 2015
<a href="#">Match Statistics</a>	Mar 31, 2018
<a href="#">Move Ordering</a>	Feb 27, 2018
<a href="#">Nimzo</a>	Dec 30, 2017
<a href="#">Novag Micro Chess</a>	Jan 16, 2013

Page	Date Edited
<a href="#">Null Move Pruning Test Results</a>	May 16, 2015
<a href="#">Null Move Test-Positions</a>	Nov 22, 2015
<a href="#">Opening</a>	Jun 25, 2017
<a href="#">Opening Book</a>	Feb 26, 2018
<a href="#">Perft</a>	Sep 26, 2017
<a href="#">Playing Strength</a>	Mar 31, 2018
<a href="#">Process</a>	Jul 31, 2017
<a href="#">Profiling</a>	Sep 15, 2016
<a href="#">Promotions</a>	Jun 6, 2017
<a href="#">Proscha</a>	May 20, 2015
<a href="#">Pruning</a>	Jan 1, 2018
<a href="#">Python</a>	Jan 31, 2018
<a href="#">Réti Endgame Study</a>	Feb 25, 2014
<a href="#">Russell Reagan</a>	Aug 31, 2013
<a href="#">RuyDos</a>	Feb 17, 2018
<a href="#">Search</a>	Feb 1, 2018
<a href="#">Search Statistics</a>	Jan 16, 2018
<a href="#">Silent but deadly</a>	Jun 17, 2015
<a href="#">space.menu</a>	Oct 9, 2015
<a href="#">Strategic Test Suite</a>	Dec 19, 2016
<a href="#">Stuart Cracraft</a>	May 19, 2017
<a href="#">Test-Positions</a>	Feb 25, 2018
<a href="#">The Nolot Suite</a>	Sep 17, 2012
<a href="#">Thread</a>	Apr 22, 2018
<a href="#">Tim Breitzkreutz</a>	Sep 23, 2017
<a href="#">Tom King</a>	Aug 19, 2017
<a href="#">Tournaments</a>	Dec 21, 2017
<a href="#">UCI</a>	Mar 4, 2018
<a href="#">Uri Blass</a>	Feb 26, 2018
<a href="#">Vlad Stamate</a>	Dec 31, 2014
<a href="#">Win at Chess</a>	Sep 25, 2017
<a href="#">WinBoard</a>	Mar 4, 2018
<a href="#">WMCCC 1993</a>	Dec 30, 2017
<a href="#">XBoard</a>	Jan 21, 2018
<a href="#">XiniX</a>	Jan 7, 2016
<a href="#">Zugzwang</a>	Dec 8, 2017

[Up one Level](#)