

Computer ohne Buch

Ein neuer Test für Schachprogramme von John Nunn

Wie gut spielen Schachprogramme, wenn sie auf ihre zum Teil mit ausgeklügelten Killervarianten bewaffneten Eröffnungsbibliotheken verzichten müssen? Gibt es Programme, die ohne die Hilfeleistung der menschlichen Eröffnungsspezialisten einen starken Leistungsabfall zeigen? Mit einem neuen Test, der zehn von GM Dr. John Nunn ausgewählte thematische Eröffnungen verwendet, wollen wir versuchen, diese Frage möglichst objektiv zu beantworten. Der folgende Beitrag stammt von John Nunn, Frederic Friedel, Dieter Steinwender und Christian Liebert.

Will man die Spielstärke von schachspielenden Programmen möglichst genau ermitteln, dann läßt man sie am besten Partien spielen. Ideal wäre es, wenn sie an vielen menschlichen Turnieren teilnehmen und dort ganz regulär eine Elozahl erwerben könnten. Aber das ist kaum praktikabel, gibt es doch viele Menschen, die die elektronische Konkurrenz bei solchen Turnieren schlichtweg ablehnen. Außerdem wären die Bedingungen sehr unterschiedlich und die Ergebnisse nur bedingt zuverlässig. Zum Beispiel müßte man berücksichtigen, ob die Gegnerschaft Erfahrung mit Computern besitzt oder gar die Programme alle bei sich zu Hause hat. Nicht von ungefähr lassen viele Hersteller ihre Schöpfungen viel lieber gegen einen osteuropäischen GM als gegen einen 2300 Elo starken und »Karateschach-erproben« Amateur aus dem Ruhrgebiet antreten.

Anders ist es, wenn Rechner gegeneinander spielen. Bei solchen Partien kann man – zumindest theoretisch – für optimale Bedingungen sorgen und ein absolut faires Ergebnis erwarten. Das versuchen auch Tester in der ganzen Welt. Die größte und bekannteste Testreihe wird von der schwedischen Schachcomputer-Vereinigung (SSDF) durchgeführt. Seit zwölf Jahren spielen nach Angaben des Initiators der Gruppe, Göran Grotting, weit über 100 SSDF-Mitglieder mehr als 40.000 Testpartien. Einige führten 700 Partien im Verlaufe eines Jahres durch – ein gigantisches Arbeitspensum für einen Hobby-Anwender, auch wenn Donningers Auto-player zuletzt zu einer erheblichen Erleichterung führte.

Die »schwedische Liste« findet weltweit große Beachtung (auch über das Internet), wird überall zitiert und abgedruckt. CSS veröffentlicht den aktuellen Stand regelmäßig im Nachrichtenteil. Aber die Er-

gebnisse sind nicht ganz unproblematisch. Bis vor kurzem wurden keine Partien veröffentlicht, so daß eine Überprüfung der Resultate von außen nicht möglich war. Man mußte schon darauf vertrauen, daß alles mit rechten Dingen zugeht und keine falschen Turnierergebnisse zur Auswertung kamen. Bei den inzwischen veröffentlichten Partien zeigte sich dann auch gleich, daß z.B. Doppelzählungen identischer Partien grundsätzlich in Kauf genommen werden. Dahinter steht der Gedanke, daß statistisch alle Programme davon gleichermaßen betroffen sind und man sich manueller Eingriffe grundsätzlich enthalten will.

Aber auch ohne den Schweden unsaubere Arbeit zu unterstellen, gibt es eine Reihe weiterer Kritikpunkte, die an der Durchführung und Auswertung der Testarbeit ansetzen. Der wichtigste betrifft die Eröffnungsbibliotheken der Programme. Diese spielen bei den reinen Partien Computer gegen Computer eine tragende Rolle und lassen sich zudem vorzüglich von außen – zum Teil an den Testern vorbei – manipulieren.

Killerbibliotheken

Über Killerbibliotheken wird in der Szene so lange diskutiert, wie es Schachcomputer gibt. Es geht um Zugvarianten, die von menschlichen Eröffnungsexperten ausgetüfelt wurden, um einzelne Konkurrenzprogramme gezielt in ungeliebte Stellungen zu bringen und dadurch die Siegwahrscheinlichkeit zu erhöhen. Bibliotheken heutiger Programme enthalten mitunter Varianten mit vierzig und mehr Zügen! Wann, fragt man sich, sollen derart spezifische Zugfolgen je zum Tragen kommen – bis man eines Tages eine Partie gegen ein anderes Programm erlebt, bei der der Konkurrent brav die Gegenseite in voller Länge herunterspielt (nicht aus der Bibliothek, natürlich) und die Partie sang- und klanglos verliert. Oder wie bei der WM in Paris, wo in der Partie Chess System Tal gegen MChess Pro 7.0 eine 50 Halbzüge umfassende Theorievariante mit einem Remis durch Zugwiederholung aufs Brett gekommen sein soll.

Nun könnte man meinen, daß Killervarianten, die genau für einen Gegner bereitgehalten werden, zum Spiel gehören. Auch Menschen legen sich Fallen für den Gegner zurecht – sogar für spezifische Gegner, von denen man weiß, daß sie eine bestimmte Eröff-

Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 – Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)

GM Dr. John Nunn



John Nunn, geboren 1955, lernte Schach mit vier Jahren. Als Fünfzehnjähriger begann er ein Mathematikstudium in Oxford, als jüngster Student in Großbritannien seit vierhundert Jahren. Er spielte weiterhin Schach und gewann mit Neunzehn die Europameisterschaft der Junioren.

1978 promovierte er mit einer Dissertation über algebraische Topologie und wurde im gleichen Jahr Großmeister. Die zweite Leistung hielt er für ungleich schwerer als die erste.

In seiner schachlichen Karriere war er zeitweilig unter den ersten zehn der Weltrangliste zu finden. Vor allem machte er sich aber einen Namen als Autor von Eröffnungs- und Endspielbüchern, und als Komponist bzw. einer der weltbesten Löser von Schachproblemen. Inzwischen besitzt er einen eigenen Verlag (Gambit) und produziert einige der besten Schachbücher der Welt. Er ist auch ein hervorragender Computer-Spezialist, der selber programmiert (vorwiegend in Delphi) und die Arbeitsweise von Schachprogrammen sehr gut kennt.

nung spielen, in der man eine Widerlegung gefunden hat. Aber bei Computern gibt es einen entscheidenden Unterschied. Während der menschliche Gegner sich nur einmal in dieser Weise hereinlegen läßt, wird ein Computerprogramm sich beliebig oft hintereinander von der gleichen Variante überlisten lassen. Und da ein Programm von vielen Testern gleichzeitig eingesetzt wird, gilt das im Prinzip sogar für den Fall, daß das Programm eine Lernbibliothek besitzt. Bei den Schweden sind über 100 Tester zugange, und jede installierte Version durchläuft ihren eigenen Lernprozeß und kann irgendwann die gleiche Killerpartie aufs Brett bekommen.

Elopunkte aus dem Buch

Hinter dieser Möglichkeit birgt sich eine recht unangenehme Gefahr. Nehmen wir an, der Autor oder Hersteller möchte, daß sein Programm etliche Ränge in der Liste aufsteigen soll, wofür allerdings die Spielstärke alleine nicht reicht. Also heuert er einen Schachmeister oder gar einen Hobbyspieler an, der eine Reihe von Killerpartien gegen die Hauptkonkurrenz austüfelt. Was heißt austüfelt, der Autoplayer und zwei zeitgemäße Computer reichen. Er

braucht nur Programm X gegen Y 50 Partien spielen zu lassen und anschließend alle zum Gewinn führenden Varianten mit hohen Ausspielpräferenzen zu versehen. Genau genommen wird damit der Lernprozeß, den neuere Programme mit der Zeit durchlaufen (aber dann zu Lasten der Elozahl), im Zeitraffer vorweggenommen. Abschließend baut er die Züge und Ausspielpräferenzen in eine neue Update-Bibliothek ein und stellt sie den Testern zur Verfügung.

Nehmen wir weiter an, daß es in der Bibliothek für jeden Konkurrenten, der deutlich höher in der Ratingliste steht (das bringt am meisten Elo-Punkte!) jeweils eine Killerpartie gibt. In der neuen Testreihe bekommen nun alle Tester mit einer gewissen Wahrscheinlichkeit die K.o.-Partie aufs Brett, und aus einem 10:10 wird ein 11:9, oder bei erneuter Wiederholung gar ein 12:8. Daraus ergeben sich, bezogen auf das jeweilige Duell, ganz locker »Elo-Zuschläge« von 50 Punkten! Das Programm schneidet wesentlich besser ab und steigt in der Liste rasch nach oben, ohne daß sich an der Spielstärke etwas geändert hätte. Das ist gar nicht so abwegig, wie wir bei einer Untersuchung der schwedischen Partien in CSS bereits nachgewiesen haben. Nicht einmal Mehrfach-Dubletten einer Serie eliminieren die Tester, sei doch das Programm selbst schuld, so die schwedischen Betreuer auf der SSDF-Homepage, wenn es sich immer wieder mit der gleichen Variante über den Tisch ziehen läßt.

Natürlich wird der Effekt, so auch die Argumentation in Schweden, durch die große Anzahl von Partien wieder mehr oder weniger nivelliert. Wenn alle Hersteller die Killer-Taktik verwenden, spielt sie letztendlich eine recht geringe Rolle. Aber das gilt nur für den Fall, daß alle Hersteller und Autoren den gleichen Aufwand betreiben. Wir wissen nur zu gut, daß diese Annahme nicht zutrifft. Es gibt Autoren, die sich über die Zeit schon fast den Ruf eines wahren Killer-Spezialisten erarbeitet haben, und es gibt andere, die überhaupt keine Anstrengungen unternehmen, die Konkurrenz zu studieren und ihre Programme für den Computer-Computer-Vergleich vorzubereiten.

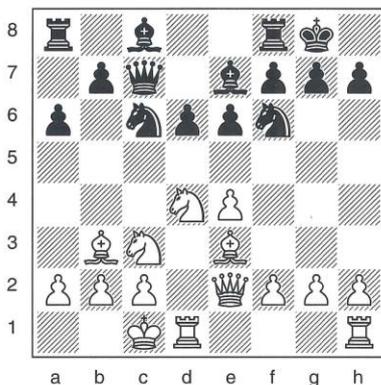
Eine andere Asymmetrie stellt man fest, wenn man Amateurprogramme und die professionellen Produkte gegenüberstellt. Die letztgenannten können sich kaum auf ihre nicht allgemein verfügbaren Konkurrenten vorbereiten, während die Amateure ihre Gegner aus dem Profilage, die sie in der vorletzten Version im Laden kaufen können, mit akribischer Sorgfalt untersuchen können. Der Rebel-Autor Ed Schröder hat erklärt, daß er genau aus diesem Grund nur sehr ungern zu den Computer-Turnieren fährt. Rebel, Fritz und Genius kennen alle, aber wer hatte schon die Gelegenheit, vor Paris gegen den

Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 – Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)

Eröffnung 2

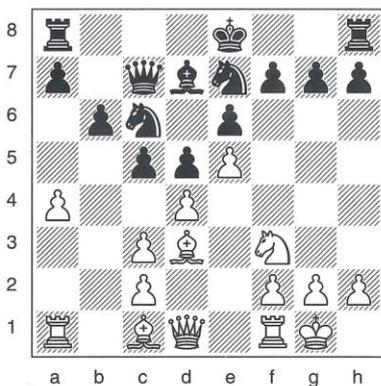
B89: Sizilianische Verteidigung, Velimirovic-Angriff: 1.e4 c5 2.Sf3 Sc6 3.d4 cxd4 4.Sxd4 Sf6 5.Sc3 d6 6.Lc4 e6 7.Le3 a6 8.De2 Dc7 9.0-0-0 Le7 10.Lb3 0-0.



Eine zweischneidige Stellung aus dem Velimirovic-Angriff. Das Hauptmerkmal ist, daß beide Seiten entgegengesetzt rochiert haben, so daß ein Wettlauf um den Angriff auf den beiden Flügeln entsteht. Die Theorie bevorzugt **11.Thg1**, was 12.g4 vorbereitet, oder gar das sofortige **11.g4**.

Eröffnung 3

C19: Französische Verteidigung, Winawer (Nimzowitsch) Variante: 1.e4 e6 2.d4 d5 3.Sc3 Lb4 4.e5 c5 5.a3 Lxc3+ 6.bxc3 Dc7 7.Sf3 Se7 8.a4 b6 9.Lb5+ Ld7 10.Ld3 Sbc6 11.0-0.



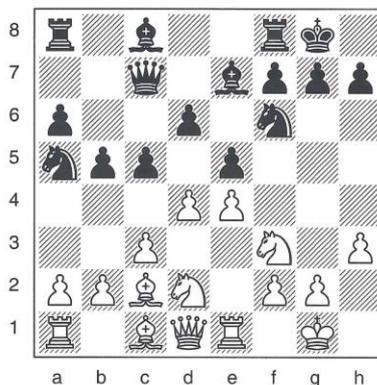
Eine theoretische Schlüsselstellung aus der Französischen Verteidigung. Weiß hat Raumvorteil und das Läuferpaar, aber er hat langfristige Schwächen am Damenflügel. Er muß seine Vorteile rasch nutzen und einen Angriff starten, während Schwarz bestrebt sein sollte, Figuren abzutauschen, um die weißen Bauernschwächen später auszunutzen. Die Theoretiker kennen in erster Linie **11...h6**. Die Fortsetzung 11...0-0? 12.Lxh7+! ist eine Falle, die man vermeiden muß.

Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 – Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)

Eröffnung 4

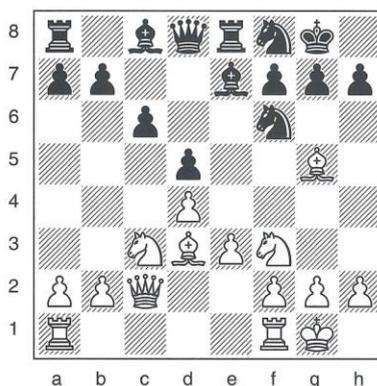
C97: Spanische Abweichungen von der Tschigorin-Verteidigung: 1.e4 e5 2.Sf3 Sc6 3.Lb5 a6 4.La4 Sf6 5.0-0 b5 6.Lb3 Le7 7.Te1 0-0 8.c3 d6 9.h3 Sa5 10.Lc2 c5 11.d4 Dc7 12.Sbd2.



Diese Stellung aus dem Spanischen ist einer der Klassiker der Eröffnungstheorie. Sie wird seit 75 Jahren ohne abschließendes Ergebnis untersucht. Die Situation im Zentrum ist sehr flexibel: Durch einen Abtausch auf d4 kann eine offene Stellung entstehen, während Weiß mit d4-d5 das Zentrum blockieren kann. Die Schachtheorie kennt eine Reihe von verschiedenen Alternativen, darunter **11...cxd4**, **11...Ld7** und **11...Sc6**.

Eröffnung 5

D36: Damengambit, Abtauschvariante: 1.d4 d5 2.c4 c6 3.Sf3 e6 4.cxd5 exd5 5.Sc3 Sf6 6.Lg5 Le7 7.Dc2 Sbd7 8.e3 0-0 9.Ld3 Te8 10.0-0 Sf8.

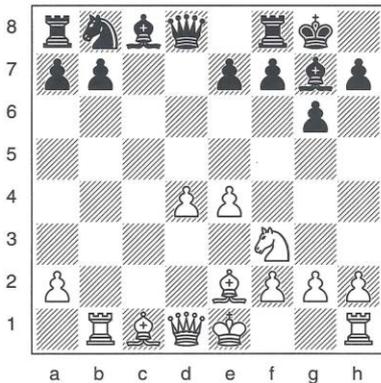


Eine weitere beliebte Stellung aus der orthodoxen Verteidigung des Abgelehnten Damengambits. Diese Variante ist besonders interessant, weil Weiß eine große Auswahl an Plänen hat. Er kann mit **11.Tae1** im Zentrum spielen und e3-e4 anstreben, oder er kann mit **11.Tab1** gefolgt von b4 einen Minoritätsangriff am Damenflügel starten. Spitzenspieler haben beide Züge in extenso analysiert. Interessant ist

außerdem der Zug **11.h3**, der Schwarz die Entwicklung des c8-Läufers erschwert, während gleichzeitig die beiden oben erwähnten Pläne noch zur Verfügung stehen.

Eröffnung 6

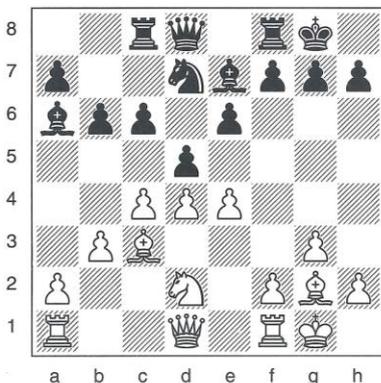
D85: Grünfeld-Verteidigung, Abtausch-Variante:
1.d4 Sf6 2.c4 g6 3.Sc3 d5 4.cxd5 Sxd5 5.e4 Sxc3 6.bxc3 Lg7 7.Sf3 c5 8.Tb1 0-0 9.Le2 cxd4 10.cxd4.



Eine offene Stellung aus dem Grünfeld-System, die eine Reihe von typischen Themen der heutigen Eröffnungstheorie verkörpert: Schwarz gestattet, daß Weiß das Zentrum besetzt, in der Erwartung, dieses später unterminieren zu können. Außerdem setzt er seine Bauernmehrheit am Damenflügel gegen die weißen Bauern im Zentrum. Wir Menschen halten **10...Da5+** für das beste, worauf 11.Dd2 gleich ins Endspiel abwickelt.

Eröffnung 7

E15: Damenindische Verteidigung, Fianchetto-System:
1.d4 Sf6 2.c4 e6 3.Sf3 b6 4.g3 La6 5.b3 Lb4+ 6.Ld2 Le7 7.Lg2 c6 8.Lc3 d5 9.Se5 Sfd7 10.Sxd7 Sxd7 11.Sd2 0-0 12.0-0 Tc8 13.e4.

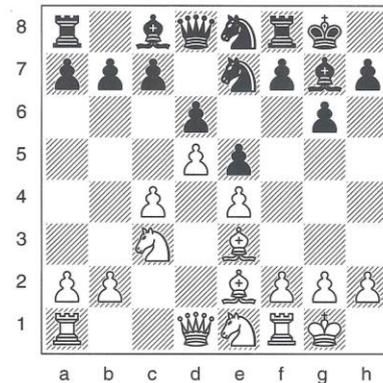


Eine komplexe, heute sehr populäre Stellung aus dem Damenindisch. Die Spannung im Zentrum ist

gewaltig, und man kann sie auf verschiedene Weise lösen, je nachdem, wie viele Bauern abgetauscht werden. Die Eröffnungstheorie kennt hauptsächlich **13...b5**, **13...dxe4** und **13...c5**.

Eröffnung 8

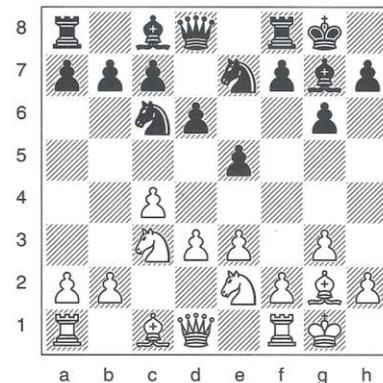
E98: Königsindische Verteidigung, Klassisches System:
1.d4 Sf6 2.c4 g6 3.Sc3 Lg7 4.e4 d6 5.Le2 0-0 6.Sf3 e5 7.0-0 Sc6 8.d5 Se7 9.Se1 Se8 10.Le3.



Eine blockierte Stellung aus dem Königsindischen. Menschliche Spieler erkennen sofort, daß Schwarz am Königsflügel und Weiß am Damenflügel angreifen muß, aber werden die Computer auch die richtigen Ideen entdecken? Obwohl die Theorie nur **10...f5** untersucht hat, kann das weitere Abspiel in verschiedene Richtungen gehen.

Eröffnung 9

A25: Englische Eröffnung:
1.c4 e5 2.Sc3 Sc6 3.g3 g6 4.Lg2 Lg7 5.e3 Sge7 6.Sge2 0-0 7.0-0 d6 8.d3.



Eine typische Flankeneröffnung. Die Zentrumsbauern haben noch keinen Gefechtskontakt, und jede Aktivität hier wird erst später in der Partie stattfinden. Schwarz kann auf Königsangriff spielen, oder er wird einfach den ehrgeizigen Aufbau am gegnerischen Damenflügel einzudämmen versuchen. Fast jeder vernünftige Zug wurde von menschlichen

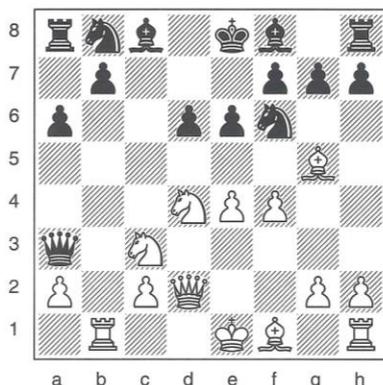
Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 – Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)

Spielern ausprobiert, aber 11...Le6 und 11...Lg4 sind die beliebtesten.

Eröffnung 10

B97: Sizilianische Verteidigung, Najdorf-Bauernraub-Variante: 1.e4 c5 2.Sf3 d6 3.d4 cxd4 4.Sxd4 Sf6 5.Sc3 a6 6.Lg5 e6 7.f4 Db6 8.Dd2 Dxb2 9.Tb1 Da3.



Zum Schluß die einzige Gambitstellung in dieser Auswahl. Es ist ein schwerer Test für die heutigen Rechner, die häufig das langfristige Angriffspotential nicht zu nutzen verstehen. Computer sind schließlich wie Termininvestoren, die eine rasche Rendite für das eingesetzte Kapital suchen. Trotz extensiver Analyse kennt keiner den richtigen Ausgang dieser Stellung. Heute glaubt man, daß bei perfektem Spiel die Partie remis enden sollte. Wir Menschen haben in erster Linie 10.f5, 10.e5 und 10.Le2 analysiert, wobei die erstgenannte Möglichkeit für die beste gehalten wird.

Wie man den Test durchführt

Wenn Sie den Nunn-Test durchführen wollen, müssen Sie die folgenden Punkte beachten.

- Die Programme spielen jede Eröffnung ab der angegebenen Diagrammstellung (also nach Eingabe der Eröffnungszüge). Dabei muß die Bibliothek (und ggf. Lerndateien) dauerhaft ausgeschaltet bleiben. Kein Programm darf z.B. über Zugumstellung später darauf wieder zurückgreifen.
- Jedes Programm spielt jede Eröffnung einmal mit Schwarz und einmal mit Weiß. Damit ein Test gültig ist, müssen alle zehn Eröffnungen bei gleicher Konfiguration durchgenommen werden.
- Das Spieltempo beträgt entweder fünf Minuten für alle Züge (Blitz), 30 Minuten für alle Züge (Aktivschach) oder zwei Stunden für 40, danach eine Stunde für 20 Züge (Turnier).
- Sie sollten zwei möglichst gleichwertige Rechner verwenden. Wenn beide Rechner unterschiedlich schnell sind, so müssen Sie mit einer Zeitvorgabe für die Seite mit dem langsameren Rechner spielen. Dazu verwenden Sie die untenstehende Tabelle.
- Wenn Sie mit unterschiedlichen Rechnern spielen, sollten Sie bei jeder neuen Eröffnung die Hardware tauschen, so daß jedes Programm fünf Eröffnungen mit dem schnelleren und fünf mit dem langsameren Rechner spielt. Wer in Eröffnung Nr. 1 die schnellere Hardware bekommt, entscheidet das Los. Danach wird abwechselnd gespielt.
- Sie können die Partien auch auf einem einzigen Rechner im Multitasking-Verfahren spielen. Dann müssen Sie das Permanent Brain (Rechnen in der

Umrechnungstabelle für Rechner unterschiedlicher Leistungsklassen

	Drhy	PII-300	P-233	P-200	P-166	P-150	P-133	P-120	P-90	P-60	486-66	486-50	486-33
486-33	16200	18.21	13.96	10.2	8.46	7.78	7.26	6.22	4.35	2.90	2.07	1.54	1.00
486-50	24900	11.85	9.08	6.61	5.51	5.06	4.72	4.05	2.83	1.89	1.35	1.00	0.65
486-66	33600	8.78	6.73	4.90	4.08	3.75	3.50	3.00	2.10	1.40	1.00	0.74	0.48
P-60	47000	6.28	4.81	3.50	2.92	2.68	2.50	2.14	1.50	1.00	0.71	0.53	0.34
P-90	70550	4.18	3.21	2.33	1.94	1.79	1.67	1.43	1.00	0.67	0.48	0.35	0.23
P-120	100800	2.93	2.24	1.63	1.36	1.25	1.17	1.00	0.70	0.47	0.33	0.25	0.16
P-133	117629	2.51	1.92	1.40	1.17	1.07	1.00	0.86	0.60	0.40	0.29	0.21	0.14
P-150	126000	2.34	1.79	1.31	1.09	1.00	0.93	0.80	0.56	0.37	0.27	0.20	0.13
P-166	137122	2.15	1.65	1.20	1.00	0.92	0.86	0.74	0.51	0.34	0.25	0.18	0.12
P-200	164547	1.79	1.37	1.00	0.83	0.77	0.71	0.61	0.43	0.29	0.20	0.15	0.10
P-233	226143	1.30	1.00	0.73	0.61	0.56	0.52	0.45	0.31	0.21	0.15	0.11	0.07
PII-300	295000	1.00	0.77	0.56	0.46	0.43	0.40	0.34	0.24	0.16	0.11	0.08	0.05

Benutzen Sie diese Tabelle, um den hardwarebedingten Leistungsunterschied zwischen zwei Rechnern auszugleichen. Wenn Sie z.B. einen Pentium 90 MHz gegen einen Pentium 166 MHz spielen lassen, bekommt das Programm auf dem langsameren Rechner $5 \times 1,94 = 9,7$ min für eine Blitzpartie, $30 \times 1,94 = 58,2$ min für Aktivschach und $2 \times 1,94 = 3,88$ Stunden für 40 Züge in Turnierpartien. Sie können auch dem schnelleren Rechner Zeit wegnehmen. Dazu verwenden Sie die Faktoren aus dem dunklen Bereich, also $5 \times 0,51 = 2,55$ min für Blitzpartien, $30 \times 0,51 = 15,3$ min für Aktivschach und $2 \times 0,51 = 1,02$ Stunden für Turnierpartien. Grundsätzlich ist, wenn vertretbar, das erste Verfahren vorzuziehen.

Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 - Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)

Bedenkzeit des Gegners) für beide Seiten ausschalten und die Timesharing-Parameter von Windows so einstellen, daß beide Programme gleiche Prozessorpower zur Verfügung gestellt bekommen. Bitte halten Sie alle Einstellungen und Parameter sorgfältig fest, damit man später die genauen Spielbedingungen rekonstruieren kann.

- Die Hashtables sollen bei den kürzeren Bedenkzeiten mindestens 4 MB betragen, auf der Turnierstufe mindestens 16 MB.
- Bitte notieren Sie zu jeder Partie alle Parameter: Prozessor, Taktfrequenz, Hashtabellen, Zeitvorgaben, genaue Programmversion etc.
- Um sicherzugehen, daß die Bibliothek ausgeschaltet ist, spielen Sie den Zug 1.e4 in der Turnierstufe aus. Achten Sie darauf, daß der Gegenzug voll durchgerechnet wird. Notieren Sie zur Kontrolle auch die Zeit, die das Programm braucht, um eine bestimmte Suchtiefe zu erreichen (z.B. n Sek., um acht Halbzüge brute force abzuschließen). Das gilt als Plausibilitäts-Check für das einwandfreie Funktionieren des Programms.
- Überprüfen Sie auch vor der Partierserie das einwandfreie Funktionieren der Hashtables. Lassen Sie dazu das Programm in der Analysestufe lange an einer Stellung rechnen. Dabei dürfen keine Festplattenzugriffe erfolgen. Auch während der Partien sind Festplattenzugriffe ein Warnzeichen: treten sie bei längeren Rechenvorgängen auf, deutet das auf überdimensionierte Hashtables hin (die auf die Festplatte ausgelagert werden).
- Bitte speichern Sie alle Partien ab und schicken Sie sie auf Diskette an die Redaktion. Ergebnisberichte ohne die dazugehörigen Partien können nicht berücksichtigt werden.

Reproduzierbarkeit

Ein wesentlicher Vorteil des Nunn-Tests ist seine eindeutige Reproduzierbarkeit. Immer mehr Programme erlauben das Abspeichern von Hauptvarianten, Rechentiefen und Zeiten mit den Zügen, so daß letztendlich jede einzelne Partie überprüft werden kann. Bei beliebig zusammengestellten Testpartien, wie sie auch diverse Hersteller auf ihren Produkt-CDs oder Internetseiten veröffentlichen, bleibt immer ein Restzweifel, ob da nicht womöglich doch ein bißchen am Ergebnis gedreht wurde. Wer möchte, bekommt jedes Ergebnis zusammengemixt, auch ein 10:0 von Fritz3 gegen Rebel9 wäre kein Problem. Mit dem Nunn-Test hat jeder Hersteller und jeder Tester die Möglichkeit, derartige Zweifel schon im Keim zu ersticken. Nunn-Ergebnisse mit Partiennachweis haben ein wesentlich stärkeres Gewicht, gleich, wer sie veröffentlicht.

Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 – Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)

Unsere Leser sind herzlich eingeladen, den Nunn-Test auch mit älteren Programmen und vor allem mit Brettcomputern durchzuführen. Hierfür gelten folgende Sonderregeln:

- Spielen zwei Brettcomputer gegeneinander, findet kein Ausgleich der Rechengeschwindigkeit statt, d.h. beide Rechner spielen auf der nominellen Spielstufe (Turnier, Aktiv, Blitz).
- Spielt ein Brettcomputer gegen einen PC, dann wird so getan, als wenn der Brettcomputer ein 486-66 wäre. Ist der Gegner z.B. ein Pentium 133, dann erhält der Brettcomputer eine um den Faktor 3,5 erhöhte Bedenkzeit bzw. der PC eine mit dem Faktor 0,29 verminderte.

Als Anhaltspunkte für das Kräfteverhältnis anderer Prozessoren zueinander mögen die in CSS 4/97 zu den Prozessoren veröffentlichten Werte Verwendung finden. Alternativ dazu könnten Sie vergleichen, um welchen Faktor das mit dem langsameren PC »gehandicapt« Programm auf dem schnelleren PC z.B. bis zum Erreichen einer bestimmten Rechentiefe nach 1.a4 zulegen würde. Diese Differenz gilt es im gleichen Verhältnis auszugleichen. Im Falle von 60 und 30s gibt es dann auf dem schnelleren Rechner die halbe Bedenkzeit.

Ein Test im Test

Fritz5 war Dank der Option, bestimmte Eröffnungszüge für das Autoplay vorgeben zu können, das erste Programm, mit dem wir den Nunn-Test in größerem Umfange durchführen konnten. Die Ergebnisse von Nunn-Engine-Engine-Vergleichen sind in CSS 5/97 auf Seite 16 nachzulesen. Inzwischen liegen weitere Partien vor, die über den Donninger-Autoplayer zwischen Fritz5 und einer ganzen Reihe externer Konkurrenten auf jeweils zwei 200 oder 233 MHz-Intel-Rechnern durchgeführt wurden. Die Ergebnisse aus Sicht von Fritz5:

Fritz5 gegen	Stufe	Resultat
Fritz 4.1	15 min/Partie	14,5:5,5
Fritz 4.1	10 min/Partie	12,5:7,5
Fritz 3.1	5 min/Partie	14,5:5,5
Genius 5	30 min/Partie	13,5:6,5
MChess Pro 6	30 min/Partie	14,5:5,5
Rebel 8	60 min/Partie	12,5:7,5
Rebel 9	60 min/Partie	13,5:6,5
Genius 5	Turnier *	13,5:6,5
Rebel 9	Turnier *	13,5:6,5

* 2 h für 40 Züge, dann 1 h für 20 Züge und 2 h für den Rest

Im Nunn-Spiel ohne Bücher beherrscht Fritz5 die besten Konkurrenten deutlich, eigentlich schon zu deutlich. Irgendwo ist hier der Wurm drin, oder?!

Wie jeder Test, ob BT-, BS- oder SSDF, testet auch der Nunn-Test zunächst einmal sich selbst. Möglicherweise liegen Fritz die Nunn-Stellungen zufällig besonders gut? Eine Frage, die wir schon jetzt zumindest teilweise beantworten können.

Eine erste, für manche naheliegende Vermutung kann sofort ausgeschlossen werden: Die Eröffnungen sind nicht in Hinblick auf irgendwelche Fritz-Stärken ausgewählt, bzw. das Hamburger Programm ist nicht auf die Stellungen getunt. Der Nunn-Test wurde von CSS Ende 1996 in Auftrag gegeben, Dr. Nunn hat die Eröffnungen gewählt, ohne sie mit einem Computer zu überprüfen. Die Entwickler von Fritz5 bekamen die Stellungen erstmalig kurz vor der Fertigstellung der Master-CD zu sehen. Für ein eventuelles »Tuning« blieb keine Zeit.

Um sicherzugehen, haben wir aber ein Kontrollexperiment durchgeführt. Mit dem SEE-Test von Arno Nickel (s. CSS 5/97) steht nämlich ein weiteres Testverfahren nach dem gleichen Strickmuster zur Verfügung. Der Nunn-Test beinhaltet 10, der SEE-Test 16 Vorgabestellungen, d.h. beim Spiel mit vertauschten Farben ergeben sich 20 bzw. 32 Testpartien. Abgesehen vom Umfang unterscheiden sich diese Tests bei der Tiefe der vorgegebenen Varianten. Bei Nunn sind es 8 bis 13 Züge, bei Nickel 1 bis 5.

Wie wir in CSS 5/97 (S.16) festgestellt haben, beeinflusst auch der Zufall der Stellungsauswahl das Ergebnis. Natürlich um so mehr, je weniger Stellungen es sind. Und sehr wahrscheinlich ist auch die Tiefe der Vorgabe-Varianten nicht ergebnisneutral. Taktisch orientierte Programme, die normalerweise von ihren Büchern in gesunde Mittelspielstellungen entlassen werden, sollten mit einer Stellung nach e4 d5 eigentlich schlechter zurechtkommen als Programme mit vergleichsweise positionellerem Spielbau. Beim SEE-Test werden nun die Programme sehr viel früher zum Rechnen gezwungen als beim Nunn-Test. Also, machen wir doch einfach die Gegenprobe. Ein Kandidat wie Fritz5 sollte demzufolge gegen Programme wie Rebel und Genius schlechter abschneiden als beim Nunn-Test.

Fritz5 gegen	Stufe	SEE-Test
Genius 5	60 min/Partie	18,5:13,5
Rebel 9	60 min/Partie	18,5:13,5

Beide SEE-Duelle bestätigen unsere Annahme, mit jeweils 58% statt zuvor jeweils 67,5% bewegt sich das (zufällig exakt gleiche) Ergebnis schon in normaleren Bahnen. Insgesamt aber immer noch ein eindrucksvoller Sieg gegen die Creme de la Creme aus Schweden (2504 bzw. 2543 Elo). Schon bei früheren theorienneutralen Testreihen (z.B. Wels 1996) fiel auf, daß Fritz besser abschneidet, als seine schwedische Plazierung vermuten ließ. Und mit

welcher Begründung sollten wir nach den obigen Testresultaten nicht von einer deutlichen Spielstärkesteigerung ausgehen? Spielen die Bücher vielleicht eine größere Rolle als bisher angenommen?

Spätestens, wenn Fritz5 in Schweden auf 200MMX/64 MB so abfallen sollte wie seinerzeit Fritz4, dürfte das theoriebereinigte Testen für reichlich Gesprächsstoff sorgen. Vielleicht werden wir auch schon bald die Ergebnisse aus Schweden in einem anderen Licht sehen. Warten wir es ab. Der Nunn-Test, darauf haben wir bereits hingewiesen, ist ein erster Forschungsansatz und nicht der Weisheit letzter Schluß. Das grundsätzliche Prinzip halten wir für richtig, und jetzt geht es darum, diese Idee weiterzuentwickeln und abzusichern. Sollte der Nunn-Test mit seinen längerzügigen Vorgaben erweitert werden? Oder analog der kurzzügigere SEE-Test? Oder halten Sie einen kombinierten SEE-Nunn-Test mit 52 Spielen für einen geeigneten Bewertungsmaßstab? Wir hoffen auf eine lebhaftige Diskussion und eine rege Beteiligung unserer Leser.

Es versteht sich von selbst, daß sämtliche in diesem Artikel angesprochenen Partien einschließlich aller Fritz5-Zeiten auf der Monatsdiskette zu finden sind.

Computer ohne Buch - Ein neuer Test für Schachprogramme von John Nunn

(Quelle: Computer-Schach & Spiele Nr. 1 – Februar-März 1998) (photo copyright © by <https://www.schaakcomputers.nl/>) (600 dpi)