

Was bei Fernschach und ausführlichen Partieanalysen längst an der Tagesordnung ist, führt im praktischen Spiel der Schachprogramme untereinander bislang ein Mauerblümchendasein: das gemeinsame Rechnen verschiedener Engines an einer Stellung. Dabei ist es doch nahe liegend, dass zwei Schach"hirne" schlauer sein müssten als eines allein. Um dies herauszufinden, nahm *CSS Online* das *TripleBrain*-Feature der *Shredder Classic*-Oberfläche mal genauer unter die Lupe. Ein großer Praxistest sollte zeigen, was dieses Konzept zu leisten vermag.

Verderben zwei Köche den Brei?

Die Geschichte der Dreihirn-Idee sowie die Arbeitsweise des TripleBrains stellte der Artikel "Dreihirn maschinell" in CSS 4/04 ausführlich vor: Zwei Schachengines rechnen an einer Stellung, während ein Koordinator sich für einen Zugvorschlag entscheiden muss, sofern zwei unterschiedliche Züge zur Auswahl stehen. Die Engines laufen dabei im Daueranalysemodus, rechnen also theoretisch unendlich lange an der vorgegebenen Position. Die Entscheidung, wann welcher Zug ausgespielt wird, trifft allein der Koordinator. Ursprünglich war von Prof. Ingo Althöfer, dem Erfinder



TripleBrain-Autor Stefan Meyer-Kahlen

dieses Konzeptes, ein Mensch als Lenker des Rechenknecht-Gespanns vorgesehen. Das Dreihirn sollte zwei Siliziumhirne und eines aus Kohlenstoff beinhalten. Doch was, wenn der menschliche Drahtzieher in kniffligen Stellungen die Übersicht verliert, wird er dann noch eine adäquate Entscheidung zwischen zwei möglicherweise verlockenden Zugvorschlägen treffen können? Außerdem spielen heutige Programme bereits auf einem Niveau, das weit über dem Schachverständnis eines normalen Wald- und Wiesenspielers liegt. Also lag der Schritt nahe, das menschliche Hirn ebenfalls durch ein künstliches zu ersetzen – eben durch das TripleBrain, welches in Stefan Meyer-Kahlens Schach-GUI *Shredder Classic* den Ingo Althöfer zu simulieren versucht. Wo die jeweiligen Stärken und Schwächen dieses Ansatzes liegen und wie sich das TripleBrain in praktischen Partien schlägt, soll dieser Artikel genauer beleuchten.

Erste Schritte

Für die Dreihirn-Idee bzw. deren maschinelle Umsetzung im TripleBrain interessierte ich mich schon länger. Schließlich besitzt Stefan Meyer-Kahlens Oberfläche das Feature bereits seit den Zeiten von Shredder 5 (damals noch im Millennium WM-Paket 2001). Auch wenn Ossi Weiner das automatische Dreihirn damals als einen "entscheidenden Durchbruch in der Schachprogrammierung" anpries – auf viel Begeisterung bei den Computerschachfans ist das TripleBrain (zu Unrecht) nie gestoßen. Vielleicht lag das einfach daran, dass zur damaligen Zeit kaum jemand einen Dual-Rechner besaß, denn zwei CPUs sind für ein optimales Arbeiten des TripleBrains nun einmal erforderlich. Die beiden Schachengines rechnen schließlich gleichzeitig an der Stellung und müssen sich bei einem einzigen Prozessor dann auch dessen Rechenzeit teilen. Dies führt dazu, dass alles ungefähr doppelt so lange dauert wie bei einem normalen Schachprogramm: das Erreichen konkurrenzfähiger Rechartiefen im praktischen Spiel, Lösezeiten von Teststellungen, akzeptable Analyseergebnisse. Und wer hat in der heutigen Zeit schon Lust, zweimal so lange zu warten wie sonst?

Seit Mitte 2003 gibt es aber Netchess, ein sehr nützliches Programm, mit welchem man via Netzwerk eine WB- oder UCI-Engine fernsteuern kann. Diese läuft dann auf einem anderen Computer und hat somit dessen CPU und RAM ganz für sich allein. "Ein tolle Sache für das TripleBrain", dachte ich mir, und führte sofort einige Experimente damit durch. In der Tat verhält sich die über *NetChess* eingebundene Engine genauso wie die lokal rechnende – das TripleBrain bemerkt keinen Unterschied. Viele Computerschach-Tester sind inzwischen mit zwei Rechnern ausgestattet, einer uneingeschränkten Nutzung des Dreihirns mit dem oben beschriebenen Verfahren steht also nichts mehr im Wege. Seit ein paar Monaten gibt es endlich auch CPUs mit zwei Kernen auf einem Chip. Ein Dual-Rechner ist also von nun an relativ preisgünstig zu haben (im Vergleich zu früher).



Wie Lars Bremer in seinem o.g. Artikel beschrieb, steht man bei der Auswahl der beiden Schachhirne vor dem Dilemma, zwei möglichst gleich starke Engines für das TripleBrain zu finden. Dies ist logischerweise eine Grundvoraussetzung für ein vernünftiges Funktionieren des Dreihirn-Prinzips: Ist eine Komponente wesentlich schwächer, dürfte der Entscheider bei seiner Zugauswahl diese nicht so hoch gewichten. So etwas ist aber beim jetzigen Entwicklungsstand des TripleBrains nicht vorgesehen, also sollte man tunlichst darauf achten, dass die beiden Rechenknechte ungefähr auf gleichem Level spielen. Bis vor wenigen Monaten schied Shredder damit als Komponente praktisch aus, denn dieser spielte mehr oder weniger in einer eigenen Liga. ChessBase-Natives wie Fritz, Junior und Hiarcs kann man leider nicht in *Shredder Classic* einsetzen, weil sie das UCI-Protokoll nicht unterstützen. Erfreulich, dass es inzwischen auch auf absolutem Top-Niveau genug UCI-Engines gibt: Fruit, Toga, Spike, Ktulu, Gandalf, um nur einige zu nennen. Die Erstgenannten machen sogar Shredder starke Konkurrenz, hier wäre sicher eine gute Kombination für das Dreihirn zu finden. Doch Ende 2003 sah die Sache noch anders aus, da stand

Shredder ziemlich allein auf weiter Flur. Die stärksten anderen UCI-Engines waren damals Ruffian, List, Aristarch, Deep Sjeng und SOS. Da ich damals meinen ersten Turniereinsatz in Bernburg mit dem TripleBrain plante, wollte ich trotzdem einen Shredder mit im Boot haben und führte ein paar unbedeutende Tests durch, welcher Partner denn gut zu Stefan Meyer-Kahlens Engine passen würde. Meine Wahl fiel auf Ruffian, dieser schien zusammen mit Shredder ein schlagkräftiges Duo zu sein, welches auch interessante Partien ablieferte.

Beim Bernburger-Turnier im März 2004 spielte das TripleBrain (Shredder 7.04/Ruffian 1.0.5) sehr solide und interessant, verlor keine Partie und landete am Ende mit 5 Punkten aus 9 Runden auf dem dritten Platz. Angesichts der starken Konkurrenz hatte ich mir vorher kaum Chancen ausgerechnet; mit diesem tollen Ergebnis war ich natürlich sehr glücklich.

Auch beim *Frühlings-Online-Turnier 2004* schnitt das TripleBrain in der bewährten Formation recht gut ab (+3 =6 –2, Platz 7 von 24), doch bereits zu diesem Zeitpunkt war mir klar, dass es möglicherweise bessere Kombinationen von Engines geben musste. Was läge da näher als der Versuch, bei einem größeren Test herauszufinden, welche Kombinationen von Schachprogrammen sich besonders für das TripleBrain eignen?



Der Autor dieses Artikels bei der Siegerehrung des Computerschachturniers in Bernburg 2004

Planung

Nachdem die Idee und der Wille für einen derartigen Praxistest vorhanden waren, stellte sich natürlich die Frage, unter welchen Bedingungen dieser Test möglichst aussagekräftig durchzuführen sei. Zunächst einmal sollte die richtige Hardwarebasis geschaffen werden. Diese war glücklicherweise schon vorhanden, sodass ich lediglich noch einige Umbaumaßnahmen vornehmen musste. Für das TripleBrain stand ein Dual-Rechner mit zwei AMD Athlon XP 2400+ und 1,5 GByte RAM zur Verfügung, die Gegner bekamen einen Single-PC mit der gleichen CPU und 786 MByte Arbeitsspeicher. So war gewährleistet, dass die Rechenplattformen für die jeweiligen Programme gleich schnell waren und auch 512 MByte Hash pro Engine vergeben werden konnten. Die beiden Rechner wurden per Nullmodemkabel verbunden und in den Flur verfrachtet, damit sich die Lärmbelastung beim 24-stündigen Dauereinsatz noch in Grenzen hielt. Per Remote-Administration kann man die Rechner ja schließlich bequem via Netzwerk fernsteuern. Da stand sie nun, die Testplattform – bereit, per Autoplayer einen der umfangreichsten Tests vorzunehmen, die zu diesem Thema jemals durchgeführt wurden. Doch halt, was und wie genau sollte eigentlich getestet werden?

Eines habe ich in meiner (vergleichsweise kurzen) Zeit als Computerschächer mit Sicherheit gelernt: Kein Testverfahren ist völlig objektiv. Jeder Test trägt die Handschrift und damit die persönlichen Vorlieben seines Testers. Es gibt immer Argumente, die für den einen oder den anderen Aspekt sprechen, ebenso lassen sich stets Gegenargumente finden. Keine Einzelperson ist in der Lage, alle Voraussetzungen für einen nahezu perfekten Test zu schaffen; dafür fehlen meist die Mittel und auch die Zeit. Deshalb ist aus meiner Sicht jedes Experiment ein Kompromiss, den der Versuchsleiter mit sich selbst und gewissen Vorbedingungen schließt, welche sich bewährt haben und deshalb allgemein anerkannt werden. Mein Credo bei diesen Kompromissen ist: Der Test muss mir Spaß machen, darf also ruhig meine eigene Handschrift tragen (sonst kann ich den Zeit- und Geldaufwand nicht vor mir selbst rechtfertigen). Trotzdem sollte er hinreichend objektiv sein und annähernd den Status einer wissenschaftlichen Untersuchung tragen. Die Testbedingungen werden also vor Beginn des Experiments festgelegt und dürfen im Nachhinein nicht mehr geändert werden. Vor dem Test wird eine Hypothese aufgestellt, während der Durchführung werden nur Messdaten gesammelt. Danach erfolgt eine Auswertung der Daten und der Versuch einer Deutung. Abschließend sollte man noch den Test selbst und seine Ergebnisse kritisch hinterfragen. So habe ich es zumindest im Chemieunterricht gelernt ...



Die Hardware: ein Dual-Prozessor-System mit reichlich Arbeitsspeicher

Bei der Wahl der Bedenkzeit stößt man bereits das erste Mal an einen kritischen Punkt: Höhere Bedenkzeiten bedeuten qualitativ hochwertige Partien, bei denen man auch beim Zusehen noch Spaß hat. Viele Partien und damit aussagekräftigere Statistiken bekommt man hingegen nur mit weniger Bedenkzeit. Als Liebhaber längerer Zeiten schloss ich hier bereits einen Kompromiss, meine Wahl fiel auf die Spielstufe 45 Minuten plus 5 Sekunden Aufschlag pro Zug. Die restlichen Bedingungen ergaben sich aus meiner Sicht beinahe wie von selbst: Eine Auswahl von UCI-fähigen Engines (auf annähernd gleichem Spielstärkeniveau) sollte in verschiedenen Kombinationen im TripleBrain auftreten. Diese Kombinationen spielen dann einen festgelegten Testparcours gegen ein vorher bestimmtes Gegnerkontingent. Als Testparcours kam von vornherein nur eine Auswahl von Vorgabestellungen in Frage, da sonst keine gleichen Bedingungen herrschen können. Das Gegnerkontingent sollte aus den stärksten verfügbaren Engines bestehen. Auch hier mussten aus zeitlichen Gründen wieder Kompromisse geschlossen werden: Sind es zu viele Vorgabestellungen, kann man nicht mehr gegen mehrere Gegner spielen. Bei zu vielen Gegnern ist es jedoch nicht mehr möglich, eine möglichst große Vielfalt an Eröffnungsstellungen zu bieten.

Nach bestem Wissen und Gewissen suchte ich also 20 Stellungen aus dem Noomen-Test aus und entschied mich für die zu dieser Zeit vier elostärksten Gegner: Shredder, Fritz, Hiarcs und Junior. Gern hätte ich mehr Gegner gehabt und noch weitere Teststellungen, doch zeitlich erschien mir einfach nicht mehr möglich. Schließlich muss jede TripleBrain-Formation 160 Partien spielen (20 Teststellungen pro Gegner jeweils mit Weiß und mit Schwarz), was bei einer Bedenkzeit von 45m+5s im Durchschnitt 7–9 Tage dauern kann.

Als Anhaltspunkt für die Spielstärke der Testkandidaten und deren Gegner nahm ich die CSS-Rangliste von Klaus Wlotzka, welcher ebenfalls mit Vorgabestellungen arbeitet. Shredder, Fritz, Hiarcs und Junior waren damals die Top 4, während im oberen Mittelfeld eine ganze Reihe von Engines zur Verfügung standen, die im TripleBrain eingesetzt werden konnten: TheKing (mit Wb2Uci-Adapter), Ruffian, List, Aristarch, Deep Sjeng und SOS fasste ich ins Auge. Doch hier wollte ich mich nicht schon vorher genau festlegen, welche Partner genau miteinander im Test antreten sollten. Der Grund ist folgender: Durch einfache Kombinatorik findet man schnell heraus, dass der zeitliche Aufwand rapide ansteigt, wenn jedes von n Elementen aus einer Menge mit jedem anderen ein Paar bilden soll. Mithilfe der Formel

$$n \cdot (n-1) / 2$$

kann man ausrechnen, wie viele Paare sich bei einer n -elementigen Menge ergeben. Nehme ich 4 Engines, sind bereits 6 Kombinationen möglich, bei 6 Programmen sind es schon 15. Deshalb ließ ich die eigentlichen Testkandidaten vor Beginn des Experiments noch offen; dies sollte sich im Laufe der Durchführung ergeben.

Vor Testbeginn erwartete ich, dass sich bestimmte Kombinationen von Engines im TripleBrain besser eignen als andere und dass dies nicht unbedingt von der Einzelstärke der jeweiligen Engine abhängt. Um Letzteres noch zu belegen, würde wohl nur ein Vergleichstest nach Ende des eigentlichen Experimentes in Frage kommen. Dort müssten dann die Engines

allein im Testparcours gegen das oben bestimmte Gegnerquadrupel antreten. Aber nun sollte es erst einmal losgehen, es war inzwischen Juni 2004.

Die ersten drei Kandidaten, die ins Rennen geschickt wurden, waren Aristarch, SOS und List. Etwas ernüchtert schaute ich nach dem Test aller drei möglichen Kombinationen auf die Ergebnisse: Mit nicht einmal 40% der Punkte schnitten SOS/List und List/Aristarch ab, das Duo Aristarch/SOS holte immerhin fast 43%. Danach wurde Deep Sjeng an Bord geholt, worauf bereits *die* Überraschung des Tests entdeckt wurde: Die Kombination Sjeng/Aristarch erwies sich nämlich im Nachhinein als die schlagkräftigste überhaupt. Mit 84 Punkten aus 160 Partien (52,5%) führen die beiden ein aus meiner Sicht überragendes Ergebnis ein, schließlich muss man bedenken, dass der Elo-Schnitt dieses Duos ca. 90 Punkte unterhalb des Durchschnitts der Gegnerschaft liegt.

In der Folgezeit kamen noch einige Engines mehr dazu; dafür schieden andere aus dem Testlauf aus. Um die noch während des Experimentes gewonnen Erkenntnisse gleich in der Praxis auszuprobieren, sollte das ermittelte Spitzen-Duo einige Male hart auf die Probe gestellt werden.

Pleiten, Pech und Pannen – Teil 1

Während das Jahr 2004 sich mit Riesenschritten seinem Ende zuneigte, näherte sich automatisch ein wichtiger Termin: Mitte November sollte wie immer das traditionelle Computerschachturnier in Thüringen stattfinden. Als begeisterter TripleBrain-Fan war für mich natürlich von vornherein klar, dass ich genau wie in Bernburg mit dem automatischen 3-Hirn antreten würde. Aufgrund der sehr guten Ergebnisse des Duos Sjeng/Aristarch entschied ich mich für diese Kombination. Mit zwei Rechnern im Gepäck machte ich mich auf den Weg ins thüringische Langenwetzendorf, wohl wissend, dass mein „Schützling“ eher experimentellen Charakter hatte. Ernsthafte Siegchancen rechnete ich mir nicht aus; vom Podium bis zur Roten Laterne hielt ich dennoch alles für möglich.

Der Start gelang grandios: Zwei Siege (gegen Crafty und List) in den ersten beiden Runden, gefolgt von zwei Remisen gegen die beiden späteren Plätze eins und zwei (Deep Sjeng, Hiarcs). Voller Zuversicht erwartete ich die fünfte Runde. Junior war der Gegner – das versprach eine spannende Partie. Es kam wie es kommen musste, in dieser Partie erlebte ich die erste große Enttäuschung:

(5) Triple Brain – Junior

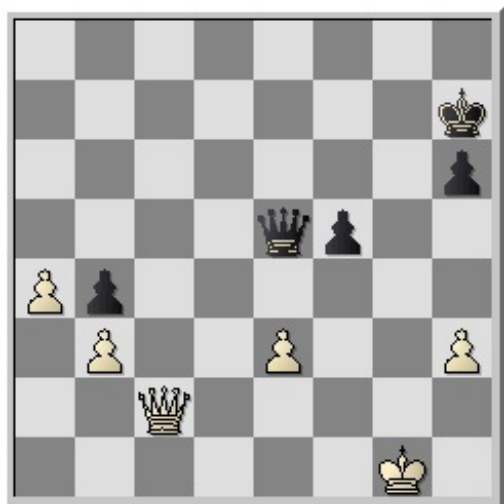
Langenwetzendorf Thüringen (5), 18.11.2004

1.d4 Sf6 2.Sf3 d5 3.c4 c6 4.Sc3 e6 5.e3 Sbd7 6.Ld3 dxc4 7.Lxc4 b5 8.Ld3 a6 9.a4 Lb7 10.De2 b4 11.Se4 Sxe4 12.Lxe4 Le7 13.0–0 0–0 14.b3 a5 15.Td1 La6 16.Ld3 Lxd3 17.Dxd3 Dc7 18.e4 c5 19.Le3 Tad8 20.Tac1 Sf6 21.Lg5 h6 22.Lxf6 gxf6 23.De2 Kh7 24.Tc2 Td7 25.Tdc1 Tfd8 26.h3 Dc6 27.Kh1 Lf8 28.dxc5 Td3 29.Td2 Txd2 30.Sxd2 Ld6 31.Db5 Tc8 32.Dxa5 Lxc5 33.Sf1 Dd6



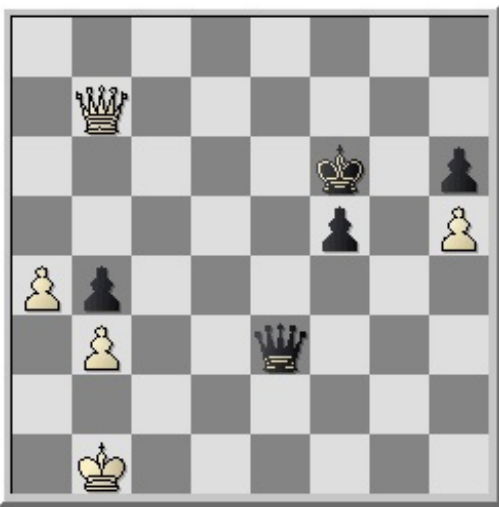
Bisher hat Weiß alles richtig gemacht, ein gesunder und freier Mehrbauer ist vorhanden. Doch das TripleBrain muss am Ball bleiben und darf sich keine Ungenauigkeiten leisten. Das Beste wäre, in nächster Zeit die Damen (und möglichst noch anderes Material) zu tauschen und dann in ein gewonnenes Endspiel abzuwickeln. Deep Sjeng schlug hier mit Gewinnbewertung den Zug 34. Db5 vor und hatte den Damentausch bereits in der Hauptvariante. Auch Junior erwartete diesen Zug und sah ebenfalls den aus seiner Sicht unvorteilhaften Damentausch auf sich zukommen. Leider wurde der Zug von Aristarch ausgewählt, welcher dem Gegner wieder Kontermöglichkeiten bot.

34.Tc4? f5! 35.Db5 fxe4 36.Db7 Tc7 37.Dxe4+ f5 38.Dc2 Lb6 39.Txc7+ Lxc7 40.g4 Dd5+ 41.Kg1 Lf4 42.gxf5 exf5 43.Se3 Lxe3 44.fxe3 De5



Bis hierher kann Weiß noch hoffen: Sein König steht zwar sehr offen da, doch sollte es ihm gelingen, eventuellen Dauerschachgeboten zu entkommen und vielleicht sogar die Damen zu tauschen, wäre die Stellung immer noch gewonnen. Doch prompt folgen die nächsten schwachen Züge, Weiß bleibt passiv und zieht nur unnütz mit dem König umher:

45.Kf1? Dd5 46.Kf2 Kg6 47.Kg3 Kg5 48.h4+ Kf6 49.Kf2 Kg6 50.Kg1 Kf6 51.h5 Dg8+ 52.Kf2 Dd5 53.Ke1 Dh1+ 54.Kd2 Dg2+ 55.Kc1 Dh1+ 56.Dd1 Dc6+ 57.Kb2 Dc3+ 58.Kb1 Dxe3 59.Dd8+ Kg7 60.Dc7+ Kf6 61.Dc6+ Ke7 62.Db7+ Kf6



Nun ist die Stellung Remis, Weiß sollte tunlichst darum bemüht sein, weiter Schachs zu geben. Ansonsten wird der schwarze f-Bauer zu stark. Doch hier wird dem TripleBrain seine Gefräßigkeit zum Verhängnis: Es folgt der finale Fehler. Zwar ist die Stellung nach dem Nehmen auf b4 möglicherweise noch nicht zwingend verloren, doch Weiß muss danach höllisch aufpassen und jede kleine Ungenauigkeit würde zum Verlust führen. Mehr als ein Remis für Weiß wäre sowieso nicht drin, also warum nicht sofort die Chance wahrnehmen, wenn sie sich bietet?

63.Dxb4? f4! 64.Kc2 f3 65.Dd2 Dc5+ 66.Kd3 Kf5 67.b4 Dd6+ 68.Kc2 Dc6+ 69.Kb3 De6+ 70.Kc2 Dc4+ 71.Kd1 Df1+ 72.Kc2 f2 73.Dd5+ Kf4 74.Dd6+ Kf3 75.Dd5+ Kg3 76.De5+ Kg2 77.De4+ Kg1 0-1

Nach dieser Niederlage lief beim TripleBrain nichts mehr so recht zusammen. Es kam mir fast vor, als hätte sich das 3-Hirn „psychologisch“ davon nicht mehr erholt. Zwar folgten noch einige respektable Remisen (gegen ProDeo, Shredder und Aristarch), doch auch eine sang- und klanglose Niederlage gegen Gandalf. Mit 4,5 Punkten aus 9 Partien wurde das TripleBrain letztlich Neunter (von 16 Teilnehmern), damit konnte ich eigentlich ganz zufrieden sein. Doch trotzdem ist man enttäuscht, wenn man selbst in der Lage war zu erkennen, wie und wo es hätte besser laufen können. Das Turnier selbst und die gut gelaunten Teilnehmer waren aber mehr als eine gute Entschädigung für die etwas ermüthenden Ergebnisse meines Schützlings.



Ein MDR-Fernsehteam berichtet über das Thüringer Turnier



Der Stromanschluss des Netzteils ist mit dem Motherboard verschmolzen, die Kabel sind durch die Hitze zum Teil angeschmort

Wieder zu Hause in Paderborn angekommen, sollte natürlich alsbald der große TripleBrain-Praxistest weitergeführt werden. Doch anstatt in stoischer Ruhe durchzulaufen, produzierte mein Dual-Rechner auf einmal tägliche Abstürze. Früher musste man ihn tage-, wochen-, teilweise monatelang nicht neu starten und jetzt das? Ich holte den Rechner vom Flur, um ihn mal richtig durchzuchecken. Nach ein paar Funktionstests rührte sich plötzlich gar nichts mehr. Was nun? Ein prüfender Blick ins Gehäuse ließ mich zunächst ungläubig stutzen, dann machte sich Entsetzen breit: Anscheinend war dem Rechner die Dauerbelastung zu viel geworden, es hatte kräftig geschmort an der Stromversorgung des Motherboards.

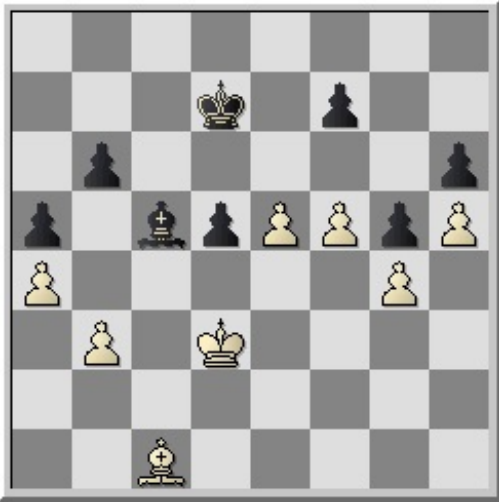
Das Durchstöbern diverser Hardware-Foren offenbarte sich mir schließlich, dass dieser Fehler für mein Mainboard-Modell durchaus typisch ist. Der Hersteller hatte offenbar genau an diesem Punkt gespart, es fehlte nämlich der bei neueren Motherboards obligatorische „ATX 12V-Zusatzstecker“ für die Stromversorgung. Zum ersten Mal war mir eine Computerkomponente während des laufenden Betriebs kaputt gegangen.

Leider befand sich das gute Stück schon außerhalb der Garantiezeit, so dass ein kostenloser Ersatz nicht in Sicht war. Zu meinem Glück waren Dual-Boards für den Athlon MP inzwischen kräftig im Preis gefallen, so dass ich nach ein paar Wochen bei eBay günstig ein fast neuwertiges gebrauchtes Board ersteigern konnte; diesmal natürlich ein anderes Modell. Danach konnte das große Testen weiter gehen.

Pleiten, Pech und Pannen – Teil 2

Beim CSS-Online-Masters 2004/05 trat ich ein weiteres Mal mit dem TripleBrain an; wieder mit der Kombination Sjeng/Aristarch, denn mein Test hatte noch kein besseres Duo zu Tage gefördert. Mein Schützling scheiterte knapp in der Vorrunde, war aber punktgleich mit dem Drittplatzierten dieser Gruppe. Eine besonders schöne Partie aus diesem Turnier ist mir in Erinnerung geblieben, sie eignet sich auch in einer bestimmten Stellung sehr gut für einen Stellungstest:

Aus der Partie TripleBrain (Aristarch / Deep Sjeng) vs. Ktulu:



Weiß am Zug sollte hier das schöne Läuferopfer Lxg5!! spielen und die Partie damit für sich entscheiden. Deep Sjeng fand diesen Zug in der Partie nicht, dafür aber Aristarch mit einer Gewinn bringenden Bewertung. Der Zug wurde dann auch vom Entscheidungsmodul gewählt, so dass das TripleBrain diese Partie gewinnen konnte. Dieses Beispiel zeigt sehr schön einen der Vorteile des TripleBrains auf: Bei Situationen dieser Art ist es eben besser, zwei verschiedene Engines im Boot zu haben, da einfach die Wahrscheinlichkeit größer ist, dass wenigstens eine von beiden den Gewinnzug findet.

Nach dem Turnier fuhr ich mit dem Testen fort. Inzwischen waren mir jedoch zwei Dinge klar geworden: Erstens würde es mir nie und nimmer gelingen, alle von mir ins Auge gefassten Engines in den TripleBrain-Test mit einzubeziehen. Dafür war der Test bereits zu umfangreich und langwierig geworden; die Stromrechnung am Ende des Jahres machte dies nur allzu deutlich. Zweitens waren mir inzwischen einige negative Aspekte am TripleBrain aufgefallen, die mich zu der Überlegung führten, es beim nächsten großen Turnier einmal selbst als Entscheider zu probieren. Schließlich bildete ich mir ein, ab und zu bessere Entscheidungen treffen zu können, wann welcher Zug ausgespielt werden soll. Bernburg 2005 sollte zeigen, ob ich als menschlicher 3-Hirn-Bediener besser zurechtkam als die automatisierte Variante.

Das TripleBrain-Experiment sollte bis zum Turnierstart im März abgeschlossen sein, damit ich dort mit der stärksten gefundenen Kombination antreten konnte. Im Laufe des Tests waren einige neue Kandidaten aufgenommen worden: The King, ProDeo und Gandalf machten sich auf die Jagd nach der TripleBrain-Krone. Ausgeschieden aus dem Test waren hingegen List und SOS, denn Formationen mit deren Beteiligung schnitten eher schlecht ab. Leider konnte ich nicht mehr alle zehn möglichen Kombinationen aus den verbleibenden fünf Kandidaten testen, am Ende wurde die Zeit aufgrund des näher rückenden Turnierbeginns und meines danach geplanten Umzugs nach Hamburg einfach zu knapp. Was ich auch testete, Aristarch und Deep Sjeng blieben das spielstärkste Duo. Somit stand Anfang März meine Meldung für das Bernburger Turnier fest: Ich wollte als klassisches Dreihirn an den Start gehen, mit Aristarch 4.51 und Deep Sjeng 1.6 als Rechenknechte und meiner Wenigkeit als menschlicher Entscheidungsinstanz. Wiederum mit zwei schnellen Rechnern bewaffnet, wollte ich den anderen gemeldeten Blechkästen möglichst viel Schneid abkaufen.



Die Kraft der zwei Herzen in Bernburg: Unter dem Tisch stehen die zwei Recheneinheiten des 3-Hirns.

Nach einem fulminanten Start ging es mir dann ähnlich wie in Thüringen, die Leistung des Dreihirns brach ein. Am Ende gelang es mir nicht, als 3-Hirn überzeugender aufzutreten als mit dem TripleBrain. Für einen ungeübten Dreihirn-Bediener ist so ein 9-rundiges Turnier mit langer Bedenkzeit unglaublich Kräfte raubend. Während die anderen Turnierteilnehmer fröhlich ihren Schützlingen beim mehr oder weniger automatischen Spielen zuschauen können, muss sich der menschliche Part eines 3-Hirns fast immer voll konzentrieren und kann sich kaum eine Pause gönnen. Bei zwei, teilweise drei Runden am Tag ist selbst ein gesunder und ausgeruhter Mensch mit seinen geistigen Kräften am Ende. Doch dies war sicher nicht der einzige Aspekt, der die Leistung des 3-Hirns schmälerte: Schlechte Eröffnungsvarianten, meine geringe Spielstärke und noch einiges mehr verhinderten ein besseres Abschneiden des Trios Sjeng/Aristarch/Klaustermeyer. Ein ausführlicherer Turnierbericht aus der Sicht des 3-Hirns ist hier zu finden.

Ergebnisse, Erkenntnisse und Deutungsversuche

Nach einem dreiviertel Jahr und über 2000 gespielten Partien war das Experiment beendet. Natürlich hätte man noch viel mehr testen können: Etliche Engine-Kombinationen wurden ausgelassen und viele starke Programme gar nicht berücksichtigt. Doch damit sich dieser Marathon nicht noch weiter verlängerte, zog ich einen Schlussstrich. Die Endergebnisse sehen folgendermaßen aus:

TripleBrain-Kandidaten

(Aristarch 4.50, SOS 4 f. Arena, List 5.12, Deep Sjeng 1.6, The King 3.23, ProDeo 1.0, Gandalf 6.0)

Sjeng / Aristarch	84.0 / 160
The King / Aristarch	81.0 / 160
ProDeo / Aristarch	78.0 / 160
Gandalf / Aristarch	78.0 / 160
Sjeng / Gandalf	74.0 / 160
Sjeng / ProDeo	73.5 / 160
SOS / Sjeng	68.5 / 160
Aristarch / SOS	68.5 / 160
List / Sjeng	68.0 / 160
Sjeng / The King	68.0 / 160
ProDeo / The King	63.5 / 160
Aristarch / List	63.0 / 160
SOS / List	60.5 / 160

Gegner

X3D Fritz	300.0 / 520
Shredder 7.04	295.5 / 520
Hiarcs 9	279.0 / 520
Junior 8	277.0 / 520

(2080 Partien)

Die Partien können hier heruntergeladen werden: Chessbase-Format, PGN-Format

So weit, so gut. Welche Erkenntnisse kann man nun aus diesen Ergebnissen gewinnen? Zunächst einmal scheint sich meine vor Testbeginn aufgestellte Hypothese bestätigt zu haben: Das Feld ist recht breit gestreut, obwohl die einzelnen Engines sich in der Spielstärke nicht übermäßig stark unterscheiden. Sehr gut schnitten die ersten vier Ränge mit Aristarch-Beteiligung ab, auch die Kombinationen Sjeng mit Gandalf und ProDeo erzielten ein recht gutes Ergebnis. Die zweite Hälfte des Feldes punktete nicht so toll, die hinteren Plätze sind weit abgeschlagen. Auch das Feld der Gegner scheint zweigeteilt: Fritz und Shredder vor Hiarcs und Junior; dies entspricht ungefähr den Spielstärkeverhältnissen dieser Engines untereinander (mit Ausnahme Shredders, der sonst vorne läge). Es hat sich außerdem gezeigt, dass sich sehr wohl schlagkräftige Kombinationen finden lassen, die den hier als Gegner verwendeten Top-Engines Paroli bieten können. Zum Teil wurden um die 50% Punktausbeute erzielt, was ein beachtliches Ergebnis darstellt. Alles deutet darauf hin, dass das TripleBrain mehr ist, als die Summe seiner Teile – wenn man die richtigen Komponenten findet.

Eine bestimmte Frage drängt sich allerdings auf: Wie würden die Kandidaten einzeln in dem verwendeten Testparcours abschneiden? Schließlich kann es ja sein, dass bestimmten Engines diese Teststellungen besonders liegen, anderen wiederum gar nicht. Um dies herauszufinden, wurde ein Vergleichstest durchgeführt, in welchem die im TripleBrain verwendeten Engines unter sonst gleichen Bedingungen allein auf die Gegnerschaft losgelassen wurden. Daraus sollte hingegen auch kein Bandwurm werden – bei sieben zu testenden Kandidaten wären das wiederum über 1100 Partien. Deshalb beschränkte ich mich bei der Auswahl auf die beiden „Top-Scorer“ Aristarch und Sjeng, den in diesem Test erstaunlich schwachen List und schließlich auf ProDeo, der viel Licht und viel Schatten zeigte. Dieser Test förderte aufschlussreiche Ergebnisse zu Tage:

TripleBrain-Vergleichstest

Shredder 7.04	103.0 / 160
---------------	-------------

X3D Fritz	96.5 / 160
Hiarcs 9	92.0 / 160
Junior 8	91.5 / 160
Deep Sjeng 1.6	70.0 / 160
Aristarch 4.50	67.0 / 160
ProDeo 1.0	61.5 / 160
List 5.12	58.5 / 160

(640 Partien)

Die Partien können hier heruntergeladen werden: ChessBase-Format, PGN-Format

Zunächst einmal zeigt sich, dass ein Klassenunterschied zwischen der Gegnerschaft (diesmal in der gewohnten Reihenfolge) und den Testkandidaten besteht. Weiterhin ist festzustellen, dass List offenbar gerade die ausgewählten Teststellungen nicht besonders gut beherrscht. Ansonsten hätte hier ein besserer Score erzielt werden müssen, denn List 5.12 ist in diversen Ranglisten vor den anderen Kandidaten platziert. Möglicherweise liegt List auch die gewählte Bedenkzeit nicht (45m+5s), doch das halte ich für sehr unwahrscheinlich. Das schlechte Abschneiden der Engine-Kombinationen mit List-Beteiligung ist sicher auf den Testparcours zu schieben.

Sjeng und Aristarch schneiden im Einzeltest deutlich schwächer ab als zusammen im TripleBrain; hier hat sich offenbar ein „Traumpaar“ gefunden. Das Durchschnittsergebnis aller Kombinationen mit Aristarch-Beteiligung liegt bei ca. 75,5 Punkten (ohne die schwachen Ergebnisse mit SOS und List wären es 80 Punkte = 50%!); Aristarch fühlt sich im TripleBrain offensichtlich wohler als im Einzelkampf. Sjeng erzielte im TripleBrain durchschnittlich 72,5 Punkte, hier ist der Unterschied zum Einzeltest nicht ganz so deutlich. Trotzdem scheint sich auch dieses Programm gut für das automatische 3-Hirn zu eignen. Bei ProDeo klafft hingegen eine größere Lücke: 71,5 Punkte Durchschnitt im TripleBrain stehen nur 61,5 Punkte im Einzeltest gegenüber. Allerdings war diese Engine auch nur dreimal an Kombinationen beteiligt, so dass sich hier eventuell eine Messungenauigkeit ergeben hat.

Wenn man diesen Test nun kritisch hinterfragt, ergeben sich eine ganze Reihe von ausbesserungsfähigen Punkten: Die Teststellungen waren möglicherweise nicht repräsentativ und zahlreich genug (siehe List), es hätten mehr Gegner sein können, die Bedenkzeit war nicht lang genug, es wurden nicht alle möglichen Kombinationen getestet und nicht alle Engines tauchen im Vergleichstest auf. Beinahe allen Argumenten schließe ich mich vorbehaltlos an – hätte ich mehr Zeit (und Geduld) gehabt, wäre der Test noch besser geworden und die Ergebnisse aussagekräftiger. Doch wie ich bereits oben erwähnte, muss ein Computerschächer bei fast jedem Experiment Kompromisse schließen. Mit diesem ausführlichen TripleBrain-Test und dessen Ergebnissen kann ich trotz aller selbst auferlegter Kritik dennoch gut leben. Auch wenn es über einige Phasen recht langwierig zuzuging, hat mir der Test viel Spaß gemacht. Meiner Meinung nach hat sich gezeigt, dass ein automatisches 3-Hirn eine beeindruckende Spielstärke an den Tag legen kann, wenn man es mit den richtigen Komponenten bestückt. Falls das TripleBrain durch diesen Test den einen oder anderen zusätzlichen Fan gefunden hat, wäre das Ziel des Autors bereits erreicht ...

Tipps und Verbesserungsvorschläge zum TripleBrain

Für alle, die einmal mit dem TripleBrain experimentieren möchten, habe ich folgende Tipps:

- Setzen Sie zwei in etwa gleichstarke Engines ein, da sonst die stärkere Komponente möglicherweise durch die andere geschwächt wird.
- Die verwendeten Engines sollten einen unterschiedlichen Spielstil aufweisen, damit es für das TripleBrain auch etwas zu entscheiden gibt.
- Bei zu ähnlicher Spielweise wird sehr häufig derselbe Zug favorisiert.
- Achten Sie darauf, dass die Ausgabeinformationen der Engines (Zeit, Bewertung, Rechentiefe, Hauptvariante) korrekt angezeigt werden. Das Entscheider-Modul bezieht seine Informationen lediglich aus diesen Anzeigen, deshalb sollten hier keine störenden Einflüsse beobachtet werden (wie sie manchmal bei bestimmten Adaptern auftreten). Außerdem ist es von Vorteil, wenn sich die Engines beim Anzeigen neuer Informationen ähneln (z.B. gleiches Verhalten beim Ausgeben neuer Hauptvarianten).
- Das TripleBrain braucht Zeit. Nach meiner Beobachtung macht es nicht viel Sinn, das automatische 3-Hirn bei Blitzpartien einzusetzen. Ein sinnvolles Urteil des Entscheider-Moduls über den besten Zug ist nur dann möglich, wenn diesem viele Informationen (in Form von Engine-Ausgaben) zur Verfügung stehen. Diese Informationen werden aber erst dann umfangreicher und präziser, wenn den Rechenkomponenten genügend Zeit zur Verfügung steht.
- Verteilen Sie die Rechenlast der zwei Engines möglichst auf verschiedene CPUs. Es eignet sich entweder ein Multiprozessor-System oder der Einsatz von zwei separaten Rechnern. Für letztere Möglichkeit empfiehlt sich das o.g. Netzwerkprogramm *NetChess*. Beim Einsatz des TripleBrains auf einem Single-Prozessor kommt es zwangsläufig zu einer Schwächung der beiden Engines.

Zum Schluss möchte ich noch einige Verbesserungsvorschläge einbringen, die bei zukünftigen Versionen des TripleBrains oder neuen Multi-Hirn-Ansätzen von Vorteil sein könnten:

- Es wäre schön, wenn dem User Eingriffe in die Parameter des TripleBrains zur Verfügung stünden. Wenn man beispielsweise von außen bestimmen könnte, wie hoch die einzelnen Schwellen für ein Umschwenken der Zugentscheidung sind, wäre man möglicherweise dazu in der Lage, bessere Einstellungen zu finden bzw. seine eigenen Vorstellungen umzusetzen.
- Vielleicht ist es sinnvoll, eine konfigurierbare Gewichtung der einzelnen Rechenkomponenten zu integrieren. Dadurch könnte man die Schwierigkeiten mit zwei unterschiedlich starken Engines zum Teil ausgleichen. Wenn z.B. Engine A 50 Elo stärker ist als B, könnte man A einen bestimmten Bonus zuteilen, welcher sich dann auf die Zugauswahl auswirkt.
- Interessant wäre ein "Multi-Hirn", bei welchem beliebig viele Recheneinheiten eingesetzt werden können. Bei Auslagerung der Komponenten auf verschiedene Computer macht dies Sinn, da die Rechenlast aufgeteilt wird. So wäre es möglich, die Stärken von mehr als zwei Engines zu vereinen. Selbstverständlich würde dies eine Verkomplizierung des Programmcodes mit sich bringen, welcher bei zwei Einheiten sicherlich einfacher gehalten werden kann als bei n Einheiten.
- Um die ursprünglich als Mensch gedachte Komponente des Entscheiders noch weiter zu verbessern, könnte es zweckdienlich sein, dem Entscheidungsmodul mehr Schachwissen an die Hand zu geben. Soweit ich weiß, werden bereits jetzt sinnlose "Computerzüge" (wie z.B. die Zugfolge Kg8-h8, Kh8-g8 und wieder zurück) erkannt und darauf der Alternativvorschlag höher bewertet. Wenn dieses Wissen irgendwie weiter ausgebaut werden könnte, wäre das sicher von Vorteil für die Gesamtspielstärke des 3-Hirns.
- Eine weitere Variante des 3- oder Multihirns könnte so aussehen: Die Engines rechnen im 2- oder 3-best-Modus, während das Entscheidungsmodul die Anzeigen vergleicht und auf Gemeinsamkeiten untersucht. Taucht z.B. ein bestimmter Zug bei der einen Komponente als erstbesten Zug und bei der zweiten als zweitbesten Zug (nur minimal schwächer bewertet als der beste), spricht einiges dafür, dass dieser Zug gut sein könnte. Hier sind sicherlich viele Variationen möglich; wünschenswert wäre es, dem User verschiedene konfigurierbare Parameter anzubieten, mit denen er selbst die optimale Einstellung vornehmen kann.

Die Zeiteinteilung des TripleBrains ist leider nicht immer optimal. Schon öfters habe ich in Partien mit längerer Bedenkzeit (sogar Turnierbedenkzeit) beobachtet, dass ein Zug ausgespielt wurde, obwohl eine der Rechenkomponenten noch in einem Fail-high oder -low steckte. Hier sollte meiner Meinung nach erst die Auflösung dieses Ereignisses abgewartet werden, bevor eine Zug-Entscheidung getroffen wird. Natürlich kann die Zeit bis zum Überwinden eines Fail-highs oder -lows mitunter sehr lange dauern, so dass man ab einer bestimmten Gesamtrestzeit vielleicht eine andere Regelung treffen sollte.
