

---

## Kapitel 4: Die synchrone Dynamik des Netzwerkes mit pseudoinverser Kopplungsmatrix

---

In diesem Kapitel wird die Dynamik des Netzwerkes mit pseudoinverser Kopplungsmatrix im Falle hoher Speicherdichte, also  $\alpha > 0$ , untersucht. Während die Eigenschaften dieses Netzwerkes im thermischen Gleichgewicht wohlbekannt sind, zumindest in einer replika-symmetrischen Analyse, existieren bislang keine Ansätze zur Beschreibung der Dynamik des Netzwerkes, welche einem Nichtgleichgewichtsvorgang entspricht. Insbesondere die deterministische Dynamik, bei der nicht durch die Einführung einer Temperatur die komplexen Strukturen des Phasenraumes geglättet werden, entzieht sich bisher jeder analytischen Beschreibung.

Für das Hopfield-Modell wurde zwar schon in [GA87] eine exakte dynamische Theorie formuliert, die aber wegen der Komplexität der dynamischen Gleichungen nur für wenige Zeitschritte gelöst werden konnte. Auch die hier, im nächsten Abschnitt, vorgestellte exakte dynamische Theorie für das Netzwerk mit pseudoinverser Kopplungsmatrix trifft auf die gleichen Schwierigkeiten. Wir verwenden sie deshalb als Ausgangspunkt für eine neue, approximative Theorie, die uns erlauben wird, dynamisches Verhalten der Pseudoinversen über lange Zeitskalen vorherzusagen.

### 4.1 Exakte dynamische Theorie

#### 4.1.1 Die Verteilung der inneren Felder — Erzeugende Funktion

Ausgangspunkt unserer Betrachtungen ist wieder die gemittelte Wahrscheinlichkeitsverteilung der inneren Felder

$$P_t(h) = \frac{1}{N} \sum_i \left[ \delta(h - h_i(t)) \right]_{S_i(0)}, \quad (4.1)$$

aus welcher dann der Überlapp zum kondensierten Muster, das wieder auf eins gesetzt ist, also  $\xi_i^1 = 1$ , im nächsten Zeitschritt durch

$$m(t+1) = \int dh P_t(h) \operatorname{sign}(h) \quad (4.2)$$

folgt.

Um diese Wahrscheinlichkeitsverteilung zu berechnen, führen wir die erzeugende Funktion

$$\begin{aligned} \left[ Z(\underline{l}) \right]_{S_i(0)} &= \left[ \operatorname{Tr}_{S_i(t)} \int \prod_{it} \left\{ dh_i(t) \cdot \Theta(S_i(t+1)h_i(t)) \right. \right. \\ &\quad \left. \left. \cdot \delta(h_i(t) - \sum J_{ij} S_j(t)) \right\} \cdot \exp \left( i \sum_{it} l_i(t) h_i(t) \right) \right]_{S_i(0)} \end{aligned}$$

ein. Die Spur läuft über alle Neuronenkonfigurationen  $S_i(t)$  mit  $t > 0$ . Dadurch gilt  $\left[ Z(\underline{0}) \right]_{S_i(0)} = 1$  und Momente und Korrelationsfunktionen der Feldverteilung

können direkt aus

$$\left[ h(t)^n \right]_{S_i(0)} = (-i)^n \frac{\partial^n}{\partial l_k(t)^n} \left[ Z(\underline{l}) \right]_{S_i(0)} \Big|_{\underline{l}=\underline{0}}$$

berechnet werden.

### 4.1.2 Dynamische Einteilchengleichungen

Wir ersetzen nun das Ortsmittel in (4.1) wieder (vergl. Abschnitt 3.2.1) durch ein Mittel über die Kopplungen  $J_{ij}$ . Wie in Anhang A gezeigt wird, kann diese Mittelung im Limes  $N \rightarrow \infty$  wiederum durch ein Mittel über gaußsche Rauschterme  $\underline{w}$  ersetzt werden, und man erhält

$$\begin{aligned} \left[ Z(\underline{l}) \right]_J &\sim \left[ \text{Tr}_{S_i(t)} \int \prod_{it} \{ dh_i(t) \} \cdot \prod_{it} \left\{ \Theta(S_i(t+1)h_i(t)) \right. \right. \\ &\quad \cdot \left. \delta \left( N^{-1} \sum_i h_i(t) - (1-\alpha)S_i(t) \right) \right\} \cdot \exp \left( i \sum_{it} l_i(t)h_i(t) \right) \\ &\quad \left. \cdot \prod_{it} \left\{ \delta(h_i(t) + \alpha S_i(t) - w_i(t) - \sum_{\tau} K_{t\tau} f_i(\tau) - a_1(t)) \right\} \right]_{\underline{w}} \end{aligned}$$

Das kann nun als die erzeugende Funktion eines dynamisches Systems von  $N$  ungekoppelten Neuronen, angetrieben durch gaußsche Rauschterme  $w_i(t)$  interpretiert werden. Die dynamischen Gleichungen eines Neurons lauten

$$S_i(t+1) = \text{sign}(h_i(t)) , \quad (4.3)$$

$$h_i(t) + \alpha S_i(t) = m(t) + \sum_{\tau} K_{i\tau} f_i(\tau) + w_i(t) , \quad (4.4)$$

$$f_i(t) = h_i(t) - (1-\alpha)S_i(t) . \quad (4.5)$$

Hierbei sind die Korrelationen des gaußschen Rauschen  $w_i(t)$  durch

$$\overline{w_i(t)w_i(\tau)} = \alpha \overline{a_{\mu}(t)a_{\mu}(\tau)} \quad (4.6)$$

gegeben. Die hier auftretenden Hilfsvariablen  $a_{\mu}(t)$  stellen die zufälligen Überlappungen zu den unkondensierten Mustern dar und sind durch

$$a_{\mu}(t) = -\alpha^{-1} \sum_{\tau} K_{t\tau} v_{\mu}(\tau) \quad (4.7)$$

festgelegt. Die  $v_{\mu}(t)$  sind wiederum zusätzliche gaußsche Rauschterme mit Mittelwert null; ihre Korrelationen sind durch die Projektionen  $f_i(t)$  in den Orthogonalraum der Muster festgelegt:

$$\overline{v_{\mu}(t)v_{\mu}(\tau)} = \overline{f_i(t)f_i(\tau)} . \quad (4.8)$$

Das Kernel  $K_{i\tau}$  erhält man aus der Gleichung

$$\sum_{t'} K_{tt'} \widehat{K}_{t'\tau} = -\alpha \delta_{i\tau} , \quad (4.9)$$

mit der Antwortfunktion

$$\widehat{K}_{t\tau} = \frac{\overline{\partial f_i(t)}}{\partial w_i(\tau)} . \quad (4.10)$$

Die Gleichungen (4.3) bis (4.10), zusammen mit Gleichung (4.2), können im Prinzip sukzessiv gelöst werden, um die Feldverteilung zu späteren Zeitpunkten zu erhalten.

### 4.1.3 Die Struktur des inneren Feldes

Uns interessiert zunächst nicht die genaue Feldverteilung  $P_t(h)$ , sondern vielmehr die Struktur des inneren Feldes  $h_t$ . Diese Struktur wird als Ausgangspunkt für die zu entwickelnde, approximative dynamische Theorie dienen. Aus Gleichung (4.4) folgt sofort die allgemeine Struktur des inneren Feldes eines Neurons zu einer beliebigen Zeit  $t$  als:\*

$$\begin{aligned}
h_t &= (1 - \alpha) z_t \\
&+ (1 - \alpha)^2 \sum_{t_1 < t} K_{tt_1} (z_{t_1} - S_{t_1}) \\
&+ (1 - \alpha)^3 \sum_{t_2 < t_1 < t} K_{tt_1} K_{t_1 t_2} (z_{t_2} - S_{t_2}) \\
&\vdots \\
&+ (1 - \alpha)^k \sum_{t_k < t_{k-1} < \dots < t_2 < t_1 < t} K_{tt_1} K_{t_1 t_2} \dots K_{t_{k-1} t_k} (z_{t_k} - S_{t_k}) \\
&\vdots \\
&+ (1 - \alpha)^{(t+1)} K_{tt-1} K_{t-1 t-2} \dots K_{10} (z_0 - S_0) .
\end{aligned} \tag{4.11}$$

Die  $z_t$  sind hierbei gaußsche Rauschterme mit Mittelwert

$$\overline{z_t} = m_t$$

und Varianz

$$\overline{z_t z_\tau} = \alpha \overline{a_t a_\tau} + m_t m_\tau .$$

Wir können (4.11) zu der wichtigen Formel

$$h_t = \tilde{z}_t + \sum_{\tau < t} \tilde{K}_{t\tau} S_\tau \tag{4.12}$$

zusammenfassen, mit einem neuen gaußschen Rauschterm  $\tilde{z}_t$  und neuen Kernen  $\tilde{K}_{t\tau}$ , welche Summen von Produkten der ursprünglichen Größen sind.

Nach (4.12) stellt sich das interne Feld damit als Summe aus einer gaußverteilten Zufallsvariable  $\tilde{z}_t$  und einem Gedächtnisterm  $\sum_{\tau < t} \tilde{K}_{t\tau} S_\tau$  dar, in dem die Zustände des Neurons zu allen früheren Zeiten auftreten. Die Dynamik des Neuronalen Netzwerkes ist also ein deutlich nicht-markovscher Prozeß. Trotz der Einfachheit von (4.12) ergibt sich im allgemeinen eine komplizierte Feldverteilung für  $P_t(h)$ , da die  $S_\tau$ 's sowohl untereinander als auch mit dem gaußschen Rauschen  $\tilde{z}_t$  korreliert sind.

Die einzelnen Terme in (4.12) können mit der ursprünglich für Spinsysteme entwickelten Cavity-Methode von Parisi and Mézard [ME87] einfach interpretiert werden. Danach wird ein Teil des inneren Feldes eines zu einem  $N$ -Spinsystem neu hinzugefügten Spins  $S_0$  durch die Magnetisierungen des ungestörten Systems erzeugt; dies ist das gaußsche Feld  $\tilde{z}_t$ . Der Rest, der Antwortterm  $\sum_{\tau < t} \tilde{K}_{t\tau} S_\tau$ , entsteht durch die Polarisation der  $N$  Spins des alten Systems durch die Anwesenheit des neuen Spins zu allen früheren Zeiten  $\tau$ . Da die Cavity-Interpretation nicht an unsere spezielle Kopplungsmatrix gebunden ist, sollte die Struktur von (4.12) für eine große Klasse von Netzwerken gültig sein. Wir werden deshalb (4.12) zum Ausgangspunkt für unsere approximative dynamische Theorie machen, welche in Abschnitt 4.2 entwickelt werden wird.

\* Ab dieser Stelle werden die Ortsindices weggelassen, da alle Neuronen durch die Mittelung über die Kopplungen äquivalent geworden sind. Stattdessen wird im folgenden die Zeit als Index verwendet.

#### 4.1.4 Die ersten zwei Zeitschritte

Mit jedem neuen Zeitschritt kommen in (4.12) neue Rauschterme  $z_t$  einschließlich ihrer Korrelationen zu den Variablen früherer Zeitpunkte hinzu. Zudem sind in jedem Zeitschritt neue  $\widehat{K}_{t\tau}$  und  $K_{t\tau}$  zu berechnen, sodaß die Zahl der zu lösenden Gleichungen schnell anwächst. Diese Situation, die schon bei anderen Netzwerken auftrat [GA87], verhindert auch hier eine exakte Analyse der Dynamik über längere Zeiten. Wir werden deshalb lediglich Resultate für die ersten zwei Zeitschritte anführen; der dritte Zeitschritt wird in einem der nächsten Abschnitte diskutiert werden.

Wie in Anhang A dargestellt, ist  $P_0(h)$ , die Feldverteilung im ersten Zeitschritt, eine Gaußverteilung mit dem Mittelwert

$$\overline{h_0} = (1 - \alpha) m_0$$

und der Varianz

$$\overline{\Delta h_0^2} = \alpha(1 - \alpha) \cdot (1 - m_0^2) .$$

Im zweiten Zeitschritt, also für  $t = 1$ , folgt für das interne Feld  $h$  die Struktur

$$h_1 = z + d_1 \cdot S_0 \tag{4.13}$$

mit der gaußverteilten Zufallsvariable  $z$ . Damit tritt hier erstmals ein Erinnerungsterm, allerdings in seiner einfachsten Form, auf. Das gaußsche Rauschen  $z$  ist noch unkorreliert zur diskret verteilten Zufallsvariable  $S_0$ . Die Feldverteilung  $P_1(h)$  wird damit eine Summe zweier Gaußverteilungen gleicher Breite  $\overline{\Delta z^2}$ , die jeweils bei  $(1 - \alpha) \overline{z} \pm d_1$  zentriert sind. Die relativen Amplituden der beiden Gaußverteilungen sind dabei  $(1 + \overline{S_0})/2$  bzw.  $(1 - \overline{S_0})/2$  (vergl. Formel (4.17)).

Der Mittelwert des internen Feldes zum zweiten Zeitschritt ist durch

$$\overline{h_1} = (1 - \alpha) m_1 + d_1 m_0$$

gegeben. Für die Varianz des internen Feldes zum zweiten Zeitschritt,  $\overline{\Delta h_1^2}$ , erhält man

$$\begin{aligned} \overline{\Delta h_1^2} = & \alpha(1 - \alpha) \cdot (1 - m_1^2) - (1 - 2\alpha) \left\{ 2(1 - \alpha) K_{10} m_0 m_1 \right. \\ & \left. + (1 - 2\alpha) \frac{(1 - \alpha)^3}{\alpha} K_{10}^2 (1 - m_0^2) \right\} . \end{aligned}$$

Die Separation der beiden Gaußverteilungen,  $d_1$ , folgt aus

$$d_1 = -(1 - \alpha)^2 K_{10} \tag{4.14}$$

mit

$$K_{10} = -2 \frac{\alpha}{(1 - \alpha)} P_0(0) .$$

Das hier erhaltene Resultat für die Feldverteilung zum zweiten Zeitschritt stimmt mit den Resultaten der in Abschnitt 3.2.2 geführten, allgemeineren Cavity-Diskussion überein. Dort war  $d_1$  als

$$d_1 = 2P_{h_i}(0) \sum_j J_{ij} J_{ji}$$

hergeleitet worden.

Die doppelte Gaußverteilung zum zweiten Zeitschritt ist, daran sei nocheinmal erinnert, ein allgemein gültiges Resultat. Wesentlich sind für das Auftreten des Gedächtnisterms im zweiten Zeitschritt sind die Schleifen über zwei Kopplungen, also beispielsweise über  $J_{ij}J_{ji}$ , nur wenn sie fehlen, bleibt das Feld im zweiten Zeitschritt gaußisch. Approximative Beschreibungen der Dynamik von Neuronalen Netzwerken, die auf gaußverteilten internen Feldern beruhen, sind daher nur bedingt aussagekräftig. Im allgemeinen liefern sie schon für den zweiten Zeitschritt falsche Ergebnisse.

## 4.2 Die Näherung für die inneren Felder — die Double-Peak-Dynamik

Alle interessanten Fragen, wie Attraktionsgebiete, Remanenzeffekte usw., erfordern die Behandlung der Dynamik über längere Zeitabschnitte, typischerweise, wie numerische Simulationen andeuten, in der Größenordnung von 10 bis 20 Zeitschritten. Eine exakte analytische Beschreibung so langer Zeiträume scheint schwierig oder sogar unmöglich zu sein. Wir benötigen also eine approximative Theorie der Dynamik, die es ermöglicht, über längere Zeiträume einigermaßen korrekte Voraussagen zu machen. Um anwendbar zu sein, sollte diese approximative Theorie gleichzeitig mit einer drastisch reduzierten Anzahl von dynamischen Parametern auskommen.

### 4.2.1 Der Double-Peak-Ansatz

Wir wissen von der exakten Theorie des vorherigen Abschnittes, daß das innere Feld die Struktur

$$h_t = \tilde{z}_t + \sum_{\tau < t} \tilde{K}_{t\tau} S_\tau \quad (4.15)$$

hat. Es besteht aus einem gaußschen Rauschterm und einem Gedächtnisterm, in dem die Werte der Spins zu früheren Zeiten auftauchen. Dieser Gedächtnisterm führt, wie numerische Simulationen zeigen (vergl. Abbildung 4.2), zu einer ausgeprägten Doppelstruktur der Feldverteilung; der Gedächtnisterm ist die Quelle aller Remanenzeffekte, welche man bei der Dynamik Neuronaler Netze findet.

Wir führen jetzt die einfachste nichtgaußsche Approximation — im folgenden als *Double Peak Dynamic (DPD)* bezeichnet — ein, die in der Lage ist, diese Doppelstruktur der Feldverteilung zu approximieren [HE90, HE91]. Wir versuchen dabei gleichzeitig, den physikalischen Grundgehalt von (4.15) in unsere Approximation zu übertragen. Daraus resultiert der DPD-Ansatz für die internen Felder:

$$h_t = u_t + d_t \cdot S_{t-1}. \quad (4.16)$$

Hierbei ist  $S_{t-1}$  der Wert des Neurons zum Zeitpunkt  $t - 1$  und die Variable  $u_t$  ein effektives gaußsches Rauschen, das wir als unkorreliert zu  $S_{t-1}$  voraussetzen. Wir haben also den Erinnerungsterm  $\sum_{\tau < t} \tilde{K}_{t\tau} S_\tau$  durch den Term  $d_t \cdot S_{t-1}$  ersetzt.

Der DPD-Ansatz scheint eine starke Näherung zur exakten Feldverteilung (4.15) zu sein, wird doch der komplette Erinnerungsterm auf einen einzigen elementaren Beitrag reduziert. Doch ist (4.16) für die ersten zwei Zeitschritte noch exakt; zu späteren Zeitpunkten wird in (4.15) der Term  $\tilde{K}_{t,t-1} S_{t-1}$  den dominanten Beitrag liefern, der Beitrag der übrigen Terme kann durch eine entsprechende Renormierung von  $d_t$  realisiert werden.

Die aus dem DPD-Ansatz (4.16) resultierende Wahrscheinlichkeitsverteilung  $P_t(h)$  besteht aus zwei Gaußverteilungen

$$P_t(h) = \frac{1}{\sqrt{2\pi \overline{\Delta u_t^2}}} \left\{ \frac{1 + m_{t-1}}{2} \exp\left(-\frac{(h - \overline{u_t} - d_t)^2}{2 \overline{\Delta u_t^2}}\right) + \frac{1 - m_{t-1}}{2} \exp\left(-\frac{(h - \overline{u_t} + d_t)^2}{2 \overline{\Delta u_t^2}}\right) \right\}, \quad (4.17)$$

deren Mittelpunkte im Abstand  $2d_t$  liegen. Sie zeigt damit dieselbe Struktur wie die exakte Feldverteilung zum zweiten Zeitschritt. Es gilt hierbei

$$\overline{u_t} = \overline{h_t} - d_t m_{t-1}, \quad (4.18)$$

$$\overline{\Delta u_t^2} = \overline{\Delta h_t^2} - d_t^2 (1 - m_{t-1}^2). \quad (4.19)$$

Unser Ansatz erfüllt die Forderung nach einer drastisch reduzierten Anzahl von dynamischen Parametern: lediglich  $\overline{h_t}$ ,  $\overline{\Delta h_t^2}$  und  $d_t$  sind zu jedem Zeitpunkt neu zu bestimmen. Wie wir im folgenden durch Vergleich mit numerischen Simulationen zeigen werden, ist ein Gedächtnisterm in dieser einfachsten Form ausreichend, um die Dynamik eines Neuronalen Netzes weitgehend korrekt zu beschreiben.

#### 4.2.2 Die Rekursionsgleichungen der DPD

Die drei abzuleitenden Rekursionsgleichungen der DPD teilen sich auf in eine allgemein gültige, von der jeweiligen Kopplungsmatrix unabhängigen Iterationsgleichung für  $d_t$  und in die Iterationsgleichungen für Mittelwert und Varianz der Feldverteilung, welche von der speziellen Kopplungsmatrix des Netzwerkes abhängen.

##### Iterationsgleichung für $d_t$

Die Rekursionsrelation für  $d_t$  kann ganz allgemein für den Fall symmetrischer Kopplungen, also für  $J_{ij} = J_{ji}$ , hergeleitet werden. Sie folgt direkt aus unserem Ansatz (4.16). Multiplizieren wir Gleichung (4.16) mit  $S_{t-1}$  und mitteln, erhalten wir zunächst

$$\begin{aligned} d_t &= \overline{S_i(t-1) h_i(t)} - \overline{u_i(t) S_i(t-1)} \\ &= N^{-1} \sum_{i,j} \overline{S_i(t) J_{ij} S_j(t-1)} - \overline{u_t} m_{t-1}. \end{aligned} \quad (4.20)$$

Wir nutzen nun die Symmetrie der Matrix aus, summieren den ersten Term in (4.20) über  $j$  und erhalten

$$\begin{aligned} N^{-1} \sum_i \overline{S_i(t) h_i(t-1)} &= N^{-1} \sum_i \overline{\text{sign}(h_i(t-1)) h_i(t-1)} \\ &= \int P_{t-1}(h) |h| dh. \end{aligned}$$

Damit folgt schließlich für  $d_t$  die Rekursionsgleichung

$$\begin{aligned} (1 - m_{t-1}^2) d_t &= \int P_{t-1}(h) |h| dh - \overline{h_t} m_{t-1} \\ &= \overline{|h_{t-1}|} - \overline{h_t} m_{t-1}. \end{aligned} \quad (4.21)$$

### Iterationsgleichungen für $\overline{h}_t$ und $\overline{\Delta h}_t^2$

Die Rekursionsgleichungen für die beiden verbleibenden Parameter der DPD werden nicht direkt durch den DPD—Ansatz geliefert, vielmehr müssen die ersten zwei Momente der Feldverteilung aus der Kenntnis der Feldverteilung bzw. der dynamischen Parameter zu früheren Zeitpunkten bestimmt werden. Hier kommen explizit die speziellen Eigenschaften der jeweiligen Kopplungsmatrix des Neuronalen Netzwerkes zum tragen. In unserem Falle, der pseudoinversen Kopplungsmatrix, sind die beiden Momente durch

$$\overline{h}_t = (1 - \alpha) m_t \quad (4.22)$$

und

$$\begin{aligned} \overline{\Delta h}_t^2 = & \alpha(1 - \alpha) (1 - m_t^2) + \frac{d_t^2 (1 - m_{t-1}^2)^2}{\overline{\Delta h}_{t-1}^2} \cdot \left\{ (1 - 2\alpha)^2 \right. \\ & \left. + \alpha(1 - \alpha) \left( 1 - \frac{\alpha(1 - \alpha)}{\overline{\Delta h}_{t-1}^2} (1 - m_{t-1}^2) \right) \right\} \quad (4.23) \end{aligned}$$

gegeben. Hierbei gilt Gleichung (4.22) exakt, während Gleichung (4.23) eine Approximation darstellt. Sie wird aber für  $\alpha = 1/2$  exakt; dann reduziert sie sich auf

$$\overline{\Delta h}_t^2 = \frac{1}{4} (1 - m_t^2) . \quad (4.24)$$

Die Gleichungen (4.17) bis (4.19) und (4.21) bis (4.23) stellen, zusammen mit Gleichung (4.2), ein geschlossenes System von Rekursionsrelationen dar. Wir haben damit die exakte Dynamik des Neuronalen Netzwerkes mit pseudoinverser Kopplungsmatrix auf eine approximative Dynamik weniger dynamischer Parameter reduziert. Im folgenden werden wir die Ergebnisse dieser DPD—Approximation diskutieren.

#### 4.2.3 Ein erläuterndes Beispiel: Die Feldverteilung zum 3. Zeitschritt

Die DPD besitzt lediglich drei dynamische Parameter und, wegen der durch den DPD—Ansatz festgelegten Form von (4.17), nur beschränkte Möglichkeiten, eine allgemeine, exakte Feldverteilung zu approximieren. Es stellt sich damit die Frage, wie gut die Approximation der DPD in nichttrivialen Fällen ist. Zur Illustration wollen wir deshalb die Resultate von exakter Theorie und DPD für den dritten Zeitschritt vergleichen.

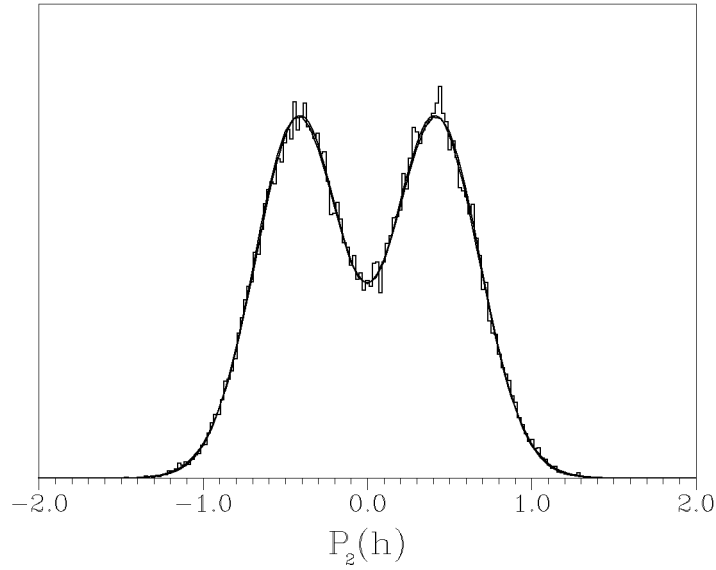
Der Einfachheit halber beschränken wir uns dabei auf  $\alpha = 1/2$  and  $m_0 = 0$ . Mit den Gleichungen (4.22) und (4.23) finden wir dann zunächst

$$\begin{aligned} \overline{h}_t &= 0 \\ \overline{\Delta h}_t^2 &= 1/4 . \end{aligned}$$

Also sind bereits zwei der drei Parameter des DPD—Ansatzes festgelegt. Es bleibt nur noch ein Parameter, die Selbstkopplung  $d_t$ , der sich justieren kann, um mit dem Doppelgauß der DPD die exakte Feldverteilung zu nähern.

Gleichung (4.11) der exakten Theorie liefert zunächst für den dritten Zeitschritt die folgende Struktur des inneren Feldes:

$$\begin{aligned} h_2 &= (1 - \alpha) w_2 \\ &+ (1 - \alpha)^2 K_{21} w_1 \end{aligned}$$



**Abbildung 4.1:** Die exakte Feldverteilung zum 3. Zeitschritt, sowie, zum Vergleich, die DPD-Approximation und eine numerisch bestimmte Feldverteilung. Die exakte Feldverteilung und die DPD-Approximation liegen dabei so eng beieinander, daß beide Verteilungen in der Grafik nicht voneinander zu trennen sind.

$$\begin{aligned}
& + (1 - \alpha)^2 \left\{ (1 - \alpha) K_{21} K_{10} + K_{20} \right\} w_0 \\
& - (1 - \alpha)^2 K_{21} S_1 \\
& - (1 - \alpha)^2 \left\{ (1 - \alpha) K_{21} K_{10} + K_{20} \right\} S_0
\end{aligned}$$

Dies ist die Summe eines gaußschen Rauschterms

$$\begin{aligned}
\tilde{w} & = (1 - \alpha) w_2 + (1 - \alpha)^2 K_{21} w_1 \\
& + (1 - \alpha)^2 \left\{ (1 - \alpha) K_{21} K_{10} + K_{20} \right\} w_0
\end{aligned}$$

mit zwei diskret verteilten Zufallsvariablen  $S_1$  und  $S_0$ . Da  $S_1 = \text{sign}((1 - \alpha)w_0)$  gilt, ist eine dieser Variablen,  $S_1$ , mit dem Rauschterm  $\tilde{w}$  korreliert.

Man kann nun zeigen, daß

$$(1 - \alpha) K_{21} K_{10} + K_{20} = 2K_{10} P_1(0) \{1 - 2\alpha\}$$

gilt. Also verschwindet, wegen  $\alpha = \frac{1}{2}$ , in  $h_2$  der Term mit  $S_0$ . Wir können  $h_2$  damit wieder als Summe eines Gaußterms mit einer diskret verteilten Zufallsvariablen schreiben:

$$h_2 = \tilde{w} + \tilde{d} \cdot S_1 . \quad (4.25)$$

Hierbei ist  $\tilde{d}$  durch

$$\begin{aligned}
\tilde{d} & = -(1 - \alpha)^2 K_{21} \\
& = \frac{1}{2} \sqrt{\frac{2}{\pi - 2}} \exp \left\{ \frac{1}{\pi - 2} \right\} .
\end{aligned}$$



gegeben. Die Struktur von (4.25) ist identisch mit unserem Ansatz (4.16), nun aber sind  $S_1$  und  $\tilde{w}$  miteinander korreliert. Die Korrelation führt zu einer nicht-gaußschen Feldverteilung für den dritten Zeitschritt:

$$P_2(h) = \frac{1}{2} \frac{1}{\sqrt{2\pi \Delta \tilde{w}^2}} \left\{ \left[ 1 + \operatorname{erf}(\Delta_+(h)) \right] \exp\left(-\frac{1}{2} \frac{(h - \tilde{d})^2}{\Delta \tilde{w}^2}\right) + \left[ 1 + \operatorname{erf}(\Delta_-(h)) \right] \exp\left(-\frac{1}{2} \frac{(h + \tilde{d})^2}{\Delta \tilde{w}^2}\right) \right\}. \quad (4.26)$$

Hierbei ist

$$\Delta_{\pm}(h) = \frac{1}{\sqrt{2(1 - \rho^2)}} \frac{\rho(h \mp \tilde{d})}{\sqrt{\Delta \tilde{w}^2}}.$$

Die Parameter dieser Verteilung sind durch

$$\begin{aligned} \tilde{d} &= 0.27576, \\ \rho &= 0.61854, \\ \frac{\tilde{d}}{\Delta \tilde{w}^2} &= 0.09167, \end{aligned}$$

gegeben.

In Abbildung 4.1 ist die exakte Feldverteilung (4.26) im Vergleich mit der korrespondierenden DPD-Approximation dargestellt. Wie bereits bemerkt, sind  $\tilde{h}_i$  und  $\Delta \tilde{h}_i^2$  durch die Wahl der Netzwerkparameter  $\alpha = 1/2$  und  $m_0 = 0$  festgelegt und lediglich die Selbstkopplung  $d_i$  kann sich justieren, um die exakte Feldverteilung (4.26) zu approximieren. Wir erhalten als Parameter für die DPD-Verteilung zur Zeit  $t = 2$ ,

$$\begin{aligned} d_2 &= 0.42504, \\ \frac{d_2}{\Delta u_2^2} &= 0.06934. \end{aligned}$$

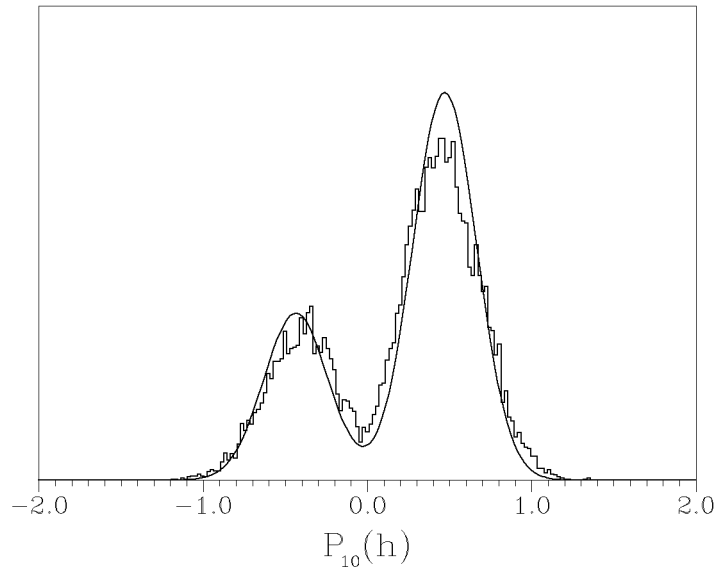
Natürlich ist  $d_2 \neq \tilde{d}$ , vielmehr ist  $d_2$  durch Gleichung (4.21) renormiert worden, um die fehlende Korrelation zwischen  $u_2$  und  $S_1$  im DPD-Ansatz zu korrigieren. Zusätzlich ist die Breite des gaußschen Rauschens  $u_2$  durch die DPD-Gleichungen verändert worden. Obwohl das exakte  $P_2(h)$  nicht die Struktur der DPD-Gleichung (4.17) besitzt, stimmen die wirkliche Feldverteilung und die approximierende DPD-Verteilung gut überein.

### 4.3 Analytische Resultate und numerische Simulationen

Die Iteration der DPD-Gleichungen erzeugt die Parameter der Feldverteilung  $P_i(h)$ , aus welcher wiederum diverse andere dynamische Größen berechnet werden können. In diesem Abschnitt werden u.a die Dynamik des Überlapps, Attraktionsgebiete und einige zeitliche Korrelationsfunktionen diskutiert. Alle Vorhersagen der DPD werden dabei mit umfangreichen numerischen Untersuchungen verglichen.

#### 4.3.1 Test der DPD

Bevor das allgemeine Netzwerkverhalten im Rahmen der DPD diskutiert wird, wollen wir in diesem Abschnitt einige kritische Annahmen der DPD anhand numerischer Simulationen überprüfen.



**Abbildung 4.2:** Die Feldverteilung für  $t = 10$ , ein Vergleich zwischen der DPD-Approximation und einer numerisch bestimmten Feldverteilung. Man beachte die unterschiedliche Höhe der beiden Maxima der Feldverteilung, was durch den nichtverschwindenden Musterüberlapp  $m_t \neq 0$  verursacht wird.

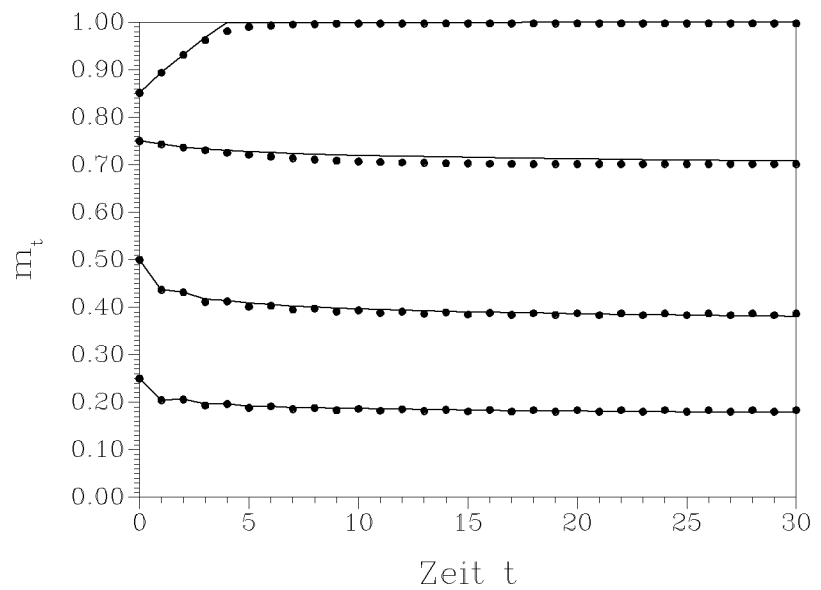
#### Feldverteilung für $t = 10$

Zunächst wollen wir die Feldverteilung des Neuronalen Netzes zu einem späten Zeitpunkt untersuchen. Die Form der DPD-Verteilung wird, wie bereits diskutiert, durch den Ansatz (4.16) relativ stark eingeschränkt. A priori ist nicht klar, ob die wirkliche Feldverteilung, insbesondere zu späten Zeitpunkten, durch die DPD-Verteilung korrekt wiedergegeben werden kann.

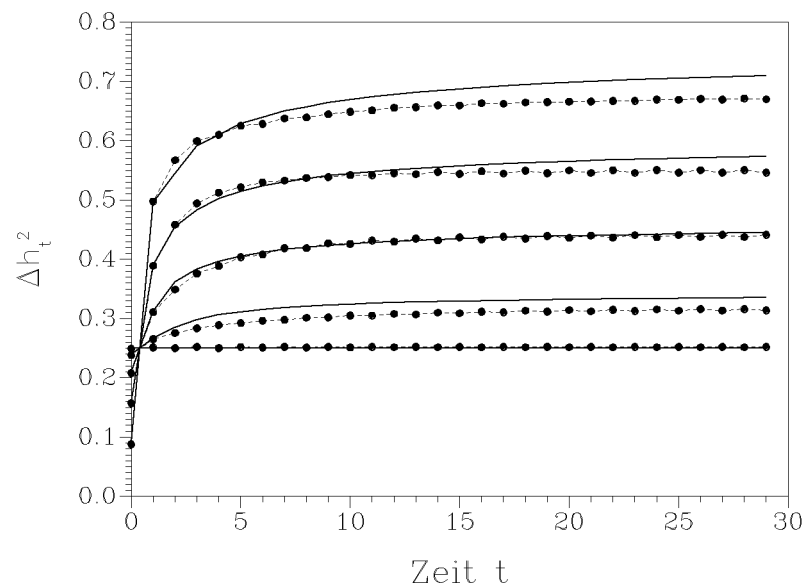
Da es praktisch nicht möglich ist, die exakte Feldverteilung für späte Zeiten analytisch auszurechnen, wird in Abbildung 4.2 die DPD-Feldverteilung für  $t = 10$  mit einer numerisch bestimmten Feldverteilung verglichen. Auch zu diesem späten Zeitpunkt ist eine gute Übereinstimmung zwischen Analytik und Simulation festzustellen. Die Separation der beiden Gaußterme, sowie deren Breite, werden durch die DPD korrekt wiedergegeben. Auch die unterschiedliche Höhe der beiden Gaußverteilungen, die, wie gesagt, innerhalb der DPD durch den Musterüberlapp zum vorhergehenden Zeitpunkt festgelegt wird, zeigt die numerisch bestimmte Feldverteilung deutlich. Der Ansatz (4.16) und die dort gemachten Annahmen über die Struktur des inneren Feldes werden also bestätigt.

#### Netzwerkverhalten bei $\alpha = \frac{1}{2}$

Als einen weiteren Test des DPD-Ansatzes kann die Dynamik des Neuronalen Netzwerkes bei  $\alpha = \frac{1}{2}$  interpretiert werden. Bei dieser Speicherdichte wird die Iterationsgleichung (4.23), welche die Varianz  $\overline{\Delta h_i^2}$  der Feldverteilung normalerweise nur approximativ liefert, exakt, sodaß als einzige Näherung gegenüber der exakten Theorie nur der DPD-Ansatz selbst übrigbleibt.



**Abbildung 4.3:** Das dynamische Verhalten des Musterüberlapps bei der Speicherdichte  $\alpha = \frac{1}{2}$ . Für Startüberlappungen  $m_0 < 0.8$  ergeben sich starke Remanenzeffekte.



**Abbildung 4.4:** Die zeitliche Entwicklung der Varianz der inneren Felder für die Speicherdichten  $\alpha = 0.5, 0.6, 0.7, 0.8$  und  $0.9$  (von unten nach oben). Die Linien sind die Vorhersage der DPD, die Punkte numerische Daten ( $N = 256$ , Mittelung über 1024 Systeme). Bei allen Trajektorien war der Musterüberlapp identisch null, weshalb die  $\alpha = 0.5$ -Kurve exakt auf  $\frac{1}{4}$  liegt (vergl. Formel 4.24).

Abbildung 4.3 zeigt den Fluß der Netzwerkdynamik bei diesem  $\alpha$  und verschiedenen Startüberlappungen  $m_0$ . Zunächst ist festzustellen, daß DPD-Vorhersage und Simulationsdaten sehr gut übereinstimmen. Wir beobachten dabei, daß die Endüberlappungen, falls keine vollständige Restaurierung des verrauschten Eingangsmusters stattfindet, stark mit den jeweiligen Startüberlappungen korreliert sind. Dies ist das erste Auftreten eines Remanenzeffektes im Rahmen der DPD.

### Zeitliches Verhalten der Varianz der inneren Felder

Eine weitere, nicht mit dem DPD-Ansatz zusammenhängende Approximation in den dynamischen Gleichungen steckt in Formel (4.23), welche die Varianz des internen Feldes liefert. Für  $\alpha \neq \frac{1}{2}$  gilt diese Gleichung nur näherungsweise. Um die Güte dieser Approximation zu überprüfen, wurden Vorhersagen der DPD für  $\overline{\Delta h_t^2}$  mit numerischen Ergebnissen verglichen (Abbildung 4.4). Die festzustellenden Abweichungen sind relativ klein und zum Teil wohl auch durch die endliche Größe der bei der numerischen Simulation verwendeten Netzwerke bedingt.

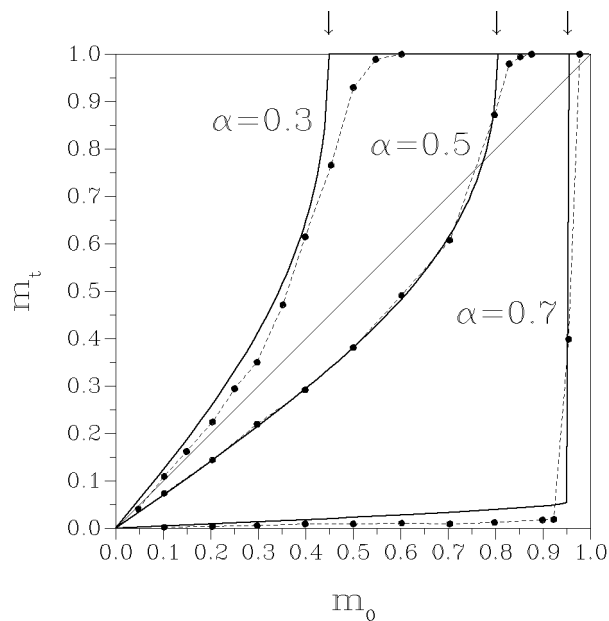
### 4.3.2 Dynamisches Verhalten der Pseudoinversen

Das beim Test der DPD gefundene dynamische Verhalten des Netzwerkes bei  $\alpha = \frac{1}{2}$  ist keineswegs typisch für alle  $\alpha$ -Werte. Das dynamische Verhalten des Neuronalen Netzes unter synchroner Dynamik stellt sich vielmehr als stark  $\alpha$ -abhängig heraus. Zur Übersicht geeignet sind  $m_t/m_0$ -Diagramme, bei denen der Überlapp zu einem späten Zeitpunkt als Funktion der Anfangsüberlappungen dargestellt wird. Abbildung 4.5 zeigt Kurven für  $\alpha = 0.3, 0.5$  und  $0.7$ , als Testzeitpunkt wurde  $t = 30$  gewählt. Zu diesem Zeitpunkt haben fast alle numerisch simulierten Systeme gestoppt (vergl. auch die Diskussion in Abschnitt 4.3.4), der Überlapp der DPD verändert sich nur noch marginal.

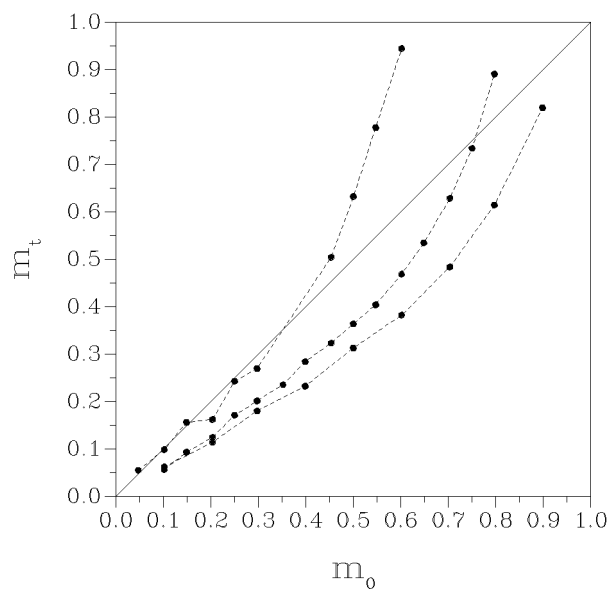
Wir finden zunächst für jeden  $\alpha$ -Werte einen kritischen Überlapp  $m_c$ , oberhalb dem eine vollständige Restaurierung des verrauschten Eingangsmusters stattfindet. Dieser kritische Überlapp zeigt den Rand des Attraktionsgebietes an. Wie erwartet, liegt  $m_c$  für höhere  $\alpha$ -Werte näher an  $m = 1$ , das Attraktionsgebiet der Muster wird also kleiner. Für  $m_0 \rightarrow m_c$  divergiert die Zeit, die zur vollständigen Restaurierung eines Bildes benötigt wird. Eine Gesetzmäßigkeit dieser Divergenz konnte weder im Rahmen der DPD noch durch numerische Simulationen des Netzwerkes festgestellt werden. Die DPD zeigt eine starke Divergenz, die leicht  $\alpha$ -abhängig ist. Numerische Untersuchungen zur Relaxationsgeschwindigkeit in's Muster liegen — bis auf sehr kurze Relaxationszeiten — jenseits der Leistungsgrenze heutiger Computer. Kanter [KA89] hat durch Simulationen mit sehr großen Systemgrößen (bis zu 2000 Neuronen waren nötig) die Relaxationszeit für die Parameterwerte  $\alpha = 0.1, m_0 = 0.3$  zu  $t_{recall} = 5$  bestimmt. Dies stimmt mit der Vorhersage der DPD überein (Abbildung 4.7).

Für Startüberlappungen unterhalb des kritischen Überlapps  $m_c$  hängt das dynamische Verhalten des Netzwerkes stark von der Speicherdichte  $\alpha$  ab. Bei  $\alpha \approx \frac{1}{2}$  bleibt der Überlapp in der Nähe der Startüberlappungen, man findet starke Remanenzeffekte. Bei niedrigeren  $\alpha$ -Werten hingegen ( $\alpha = 0.3$  in Abbildung 4.5) findet *immer* eine teilweise Restaurierung des verrauschten Bildes statt. Das Netzwerk bleibt aber während des Assoziationsvorgangs in den dynamischen Fallen, also metastabilen Zuständen oder Zweierzyklen, hängen.

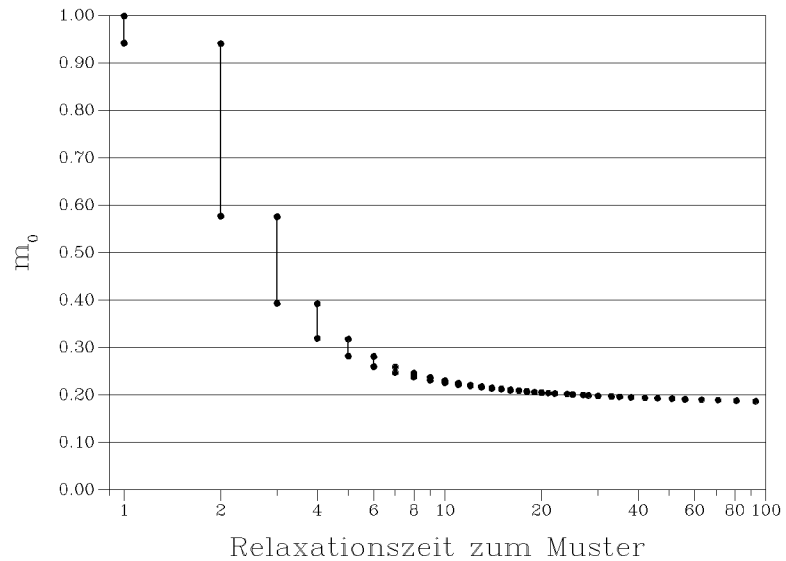
Bei  $\alpha$ -Werten größer als 0.5 unterscheidet sich das dynamische Verhalten deutlich von dem Verhalten bei kleinen Speicherdichten (in Abbildung 4.5 wird  $\alpha = 0.7$  gezeigt). Im wesentlichen finden wir nur  $m_t \rightarrow 0$  oder 1. Die Rela-



**Abbildung 4.5:** Die synchrone Dynamik des Netzwerkes mit pseudoinverser Koppelungsmatrix bei verschiedenen  $\alpha$ -Werten. Für jeden  $\alpha$ -Wert existiert ein kritischer Überlapp  $m_c$  (die Pfeile '↓' in der Graphik), oberhalb dem eine vollständige Restaurierung des verrauschten Eingangsmusters stattfindet. Die Punkte sind die Resultate über 100 Systeme pro D



**Abbildung 4.6:** Die Dynamik des Netzwerkes unter asynchroner Dynamik (von links nach rechts  $\alpha = 0.3, 0.5$  und  $0.7$ ). Man beachte die starken Remanenzeffekte auch für hohe Speicherdichten; dies ist ein deutlicher Unterschied zur synchronen Dynamik. Die Parameter der numerischen Simulationen entsprechen denen der Abbildung 4.5.



**Abbildung 4.7:** Die Relaxationsgeschwindigkeit in das Muster für eine Speicherdichte von  $\alpha = 0.1$  und verschiedenen Startüberlappungen  $m_0$ . Die von Kanter [KA89] numerisch gefundene Relaxationszeit von  $t_{recall} = 5$  bei  $m_0 = 0.3$  und  $\alpha = 0.1$  wird durch die DPD bestätigt.

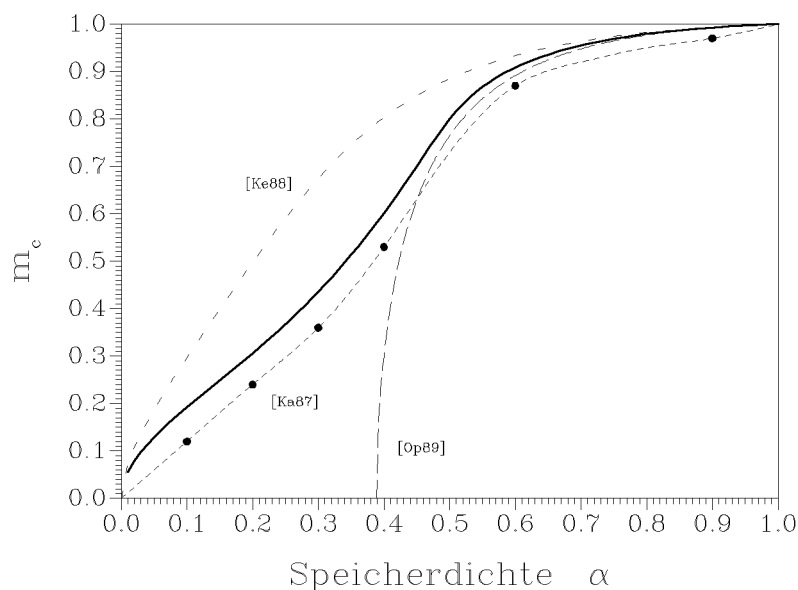
xation erfolgt dabei sehr schnell, dynamische Fallen, die zu Remanenzeffekten führen könnten, scheinen in diesem  $\alpha$ -Bereich nicht vorhanden zu sein.

Die starke Abhängigkeit des dynamischen Verhaltens des Netzwerkes von der Speicherdichte  $\alpha$  ist eine Eigenheit der synchronen Dynamik. Unter asynchroner Dynamik, für die in Abbildung 4.6 einige numerische Daten gezeigt werden, findet man für alle  $\alpha$ -Werte, insbesondere auch für  $\alpha > \frac{1}{2}$ , remanenten Überlapp. Interessanterweise liefert die asynchrone Dynamik in etwa die selben kritischen  $m_c$ 's wie die synchrone Dynamik.

### 4.3.3 Das Attraktionsgebiet der Pseudoinversen

Als Attraktionsgebiet des Neuronalen Netzwerkes bezeichnen wir den kritischen Überlapp  $m_c$  als Funktion von  $\alpha$ . Wie Abbildung 4.8 zeigt, finden wir für  $\alpha < \frac{1}{2}$  recht große Attraktionsgebiete, aber für  $\alpha > \frac{1}{2}$  schrumpft das Attraktionsgebiet schnell zu einem schmalen Streifen um das jeweilige Muster zusammen. Ein volles Attraktionsgebiet wird nur im Limes  $\alpha \rightarrow 0$  erzielt. Würde man allerdings als Attraktionsgebiet alle Startüberlappungen  $m_0$  definieren, bei denen wenigstens eine partielle Restaurierung des verrauschten Bildes stattfindet, würde sich bis fast  $\alpha = \frac{1}{2}$  ein volles Attraktionsgebiet ergeben.

Es sei angemerkt, daß die angeführten Daten im wesentlichen auch für das Netzwerk mit Perceptron-Matrix gelten, da sich pseudoinverse Kopplungsmatrix und Perceptronmatrix im niedrigen  $\alpha$ -Bereich nur unwesentlich voneinander unterscheiden (siehe Anhang C). Numerische Simulationen [KR90] lassen auch im hohen  $\alpha$ -Bereich keine wesentliche Verbesserung der Eigenschaften des Neuronalen Netzwerkes bei Verwendung der Perceptronmatrix erwarten. Da die Pseudoinverse aber sehr viel einfacher und schneller zu berechnen ist als das Perceptron,



**Abbildung 4.8:** Das Attraktionsgebiet des Netzwerkes mit pseudoinverser Kopplungsmatrix als Funktion von  $\alpha$  (durchgezogene Linie). Ebenfalls eingetragen sind zwei im Text diskutierte, heuristische Regel und numerische Daten für das Attraktionsgebiet dieses Netzwerkes aus [KA87].

sollte in praktischen Anwendungen der ersteren der Vorzug gegeben werden.

In Abbildung 4.8 sind noch zwei heuristische Regel zur Bestimmung des Attraktionsgebietes eines Neuronalen Netzwerkes eingezeichnet. Ähnlich wie im Falle der Hopfield-Kopplungsmatrix liefert der 'verdünnte Netzwerk'-Ansatz [OP89] nur in unmittelbarer Nachbarschaft der Muster korrekte Resultate, für kleine  $\alpha$ -Werte wird ein volles Attraktionsgebiet vorhergesagt, was offensichtlich falsch ist. Der Grund liegt in der Vernachlässigung des Erinnerungsterms. Sich aufbauende Korrelationen, und damit die Tatsache, daß das Neuronale Netz in dynamischen Fallen stoppen könnte, werden innerhalb dieses Ansatzes nicht behandelt.

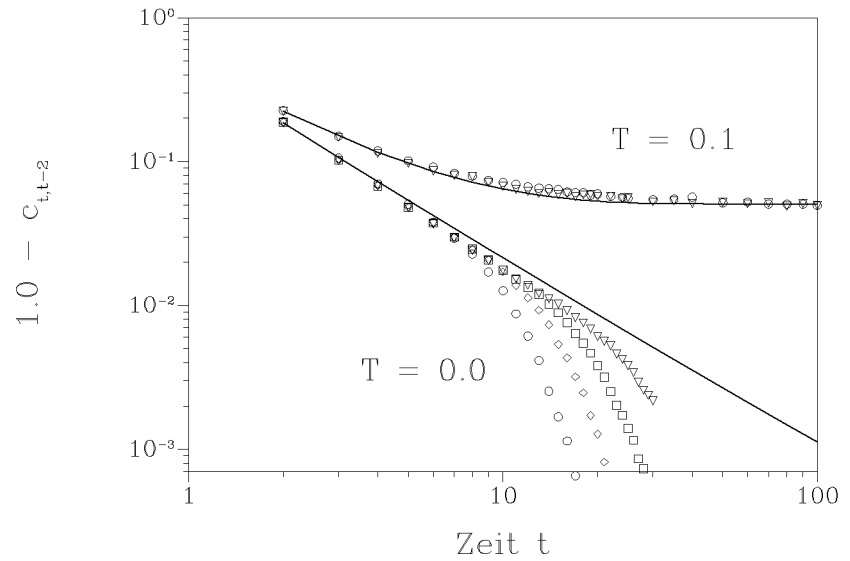
Kepler und Abbott [KE88] versuchten diesen Nachteil der verdünnten Theorie dadurch zu beheben, daß sie eine bestimmte Konvergenzgeschwindigkeit im ersten Zeitschritt in Richtung des Musters fordern. Dies behebt zwar das Problem des vollen Attraktionsgebietes bei niedrigen  $\alpha$ -Werten, liefert aber trotzdem, im Falle der Pseudoinversen, keine qualitativ korrekten Resultate.

#### 4.3.4 Die Korrelationsfunktion $\overline{S_t S_{t-2}}$

Wird die Startkonfiguration  $S_i(0)$  aus dem Raum orthogonal zu den Mustern gewählt, verschwindet der Startüberlapp  $m_0$ , und das Netzwerk bewegt sich für alle Zeiten im Orthogonalraum der Muster; der Musterüberlapp  $m_t$  bleibt null und ist kein guter dynamischer Parameter mehr.

Einblicke in die Dynamik des Neuronalen Netzes gibt in diesem Fall die zeitliche Korrelationsfunktion  $c_{t,t-2} = \overline{S_t S_{t-2}}$ . Sie ist im Rahmen der DPD durch

$$\begin{aligned} c_{t,t-2} &= \overline{S_t \cdot S_{t-2}} \\ &= \overline{\text{sign}(u_{t-1} + d_{t-1} S_{t-2}) \cdot S_{t-2}} \end{aligned}$$



**Abbildung 4.9:** Das zeitliche Verhalten der Korrelationsfunktion  $1 - c_{t,t-2}$  für verschiedene Systemgrößen. Die numerischen Daten wurden mit Systemgrößen von 128 ( $\circ$ ), 256 ( $\diamond$ ), 512 ( $\square$ ) und 1024 ( $\nabla$ ) Neuronen erzielt. Jeder Datenpunkt wurde über 100 ( $T = 0.1$ ) oder 1000 ( $T = 0.0$ ) Systeme gemittelt.

$$\begin{aligned}
 &= \overline{\text{sign}(u_{t-1}S_{t-2} + d_{t-1})} \\
 &= \text{erf} \left( \frac{d_{t-1}}{\sqrt{2\Delta u_{t-1}^2}} \right)
 \end{aligned}$$

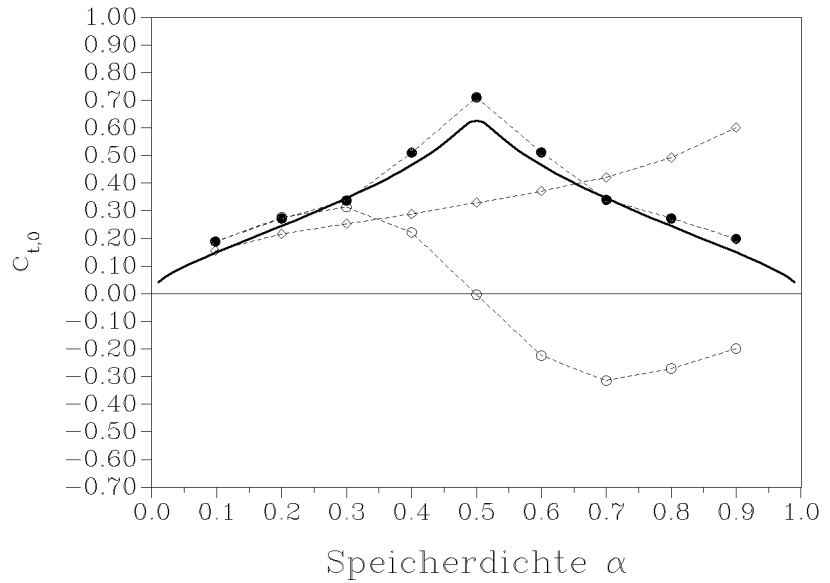
gegeben. Man findet einen zeitlichen Zerfall der Korrelationsfunktion, der sich wie  $1 - c_{t,t-2} \sim t^{-\lambda}$  verhält (Abbildung 4.9). Der Exponent  $\lambda$  folgt aus der DPD zu  $\approx -1.37$ , während die numerischen Simulationen  $\lambda \approx -1.47$  liefern. Ein ähnliches Verhalten der Funktion  $1 - c_{t,t-2}$  wurde auch in numerischen Simulationen der synchronen Dynamik des SK-Modells der Spingläser gefunden [GA87]. Hier wurde ein Exponent von  $\lambda \approx -\frac{3}{2}$  gefunden.

Die numerischen Daten von Abbildung 4.9 zeigen nach einer bestimmten Zeit eine deutliche Abweichung von der DPD-Vorhersage. Diese Zeit ist eng mit der jeweiligen Systemgröße korreliert und kann als guter Indikator für das Auftreten von 'finite size'-Effekten in numerischen Simulationen angesehen werden. Interessanterweise liegt der Zeitpunkt, an dem solche 'finite size'-Effekte wirksam werden, selbst bei großen Systemen ( $N = 1024$ ) relativ früh ( $t \approx 20$ ). Ergebnisse für das Langzeitverhalten Neuronaler Netze, die aus numerischen Simulationen gewonnen wurden, sind deshalb äußerst kritisch zu bewerten.

#### 4.3.5 Die remanente Magnetisierung

Eine andere Größe, die Auskunft über das dynamische Verhalten des Neuronalen Netzes bei verschwindendem Musterüberlapp gibt, ist die remanente Magnetisierung  $c_{t,0} = \overline{S_t S_0}$ . Es ist die Projektion des Zustandsvektors  $S_i(t)$  zum Zeitpunkt  $t$  auf den Anfangszustand  $S_i(0)$ . Sie kann im Rahmen der DPD zunächst nicht direkt ausgerechnet werden, da die DPD ja nur den Musterüberlapp auf





**Abbildung 4.10:** Die remanente Magnetisierung in Abhängigkeit von  $\alpha$ . Die durchgezogene Linie ist die Vorhersage der DPD. Die numerischen Daten für die synchrone Dynamik ('•' = gerade Zeitschritte, '◦' = ungerade Zeitschritte) zeigen eine deutliche Symmetrie um  $\alpha = \frac{1}{2}$ . Eine solche Symmetrie fehlt bei den Daten für die asynchrone Dynamik ('◊'). Die Simulationen wurden für  $N = 256$  durchgeführt, gemittelt über 1024 Systeme pro Datenpunkt.

das erste Muster, mit  $\xi_i^1 = 1$ , liefert. Wir können aber die Eichtransformation der Dynamik (vergl. Anhang C) benutzen, um das erste Muster auf  $S_i(0)$  zu transformieren:

$$\begin{aligned}\xi_i^1 &\rightarrow \xi_i^1 S_i(0), \\ J_{ij} &\rightarrow S_i(0) J_{ij} S_j(0).\end{aligned}$$

Damit wird der Musterüberlapp  $m_t$  auf die remanente Magnetisierung  $c_{t,0}$  abgebildet und die DPD kann zur Analyse des zeitlichen Verhaltens dieser Größe herangezogen werden.

Durch die Eichtransformation werden die dynamischen Parameter  $d_t$  und  $\overline{h_t^2}$  nicht verändert, da in ihnen die transformierten Größen immer paarweise auftreten.  $d_t$  also ist immer noch durch Gleichung (4.20) und  $\overline{h_t^2}$  durch Gleichung (4.23) mit  $m_t = 0$  gegeben. Da nun aber — bedingt durch die Eichtransformation — das eine Muster mit kondensiertem Musterüberlapp,  $\xi_i^1$ , nicht mehr identisch dem Eins-Muster  $\xi_i = 1$  ist, kann der noch verbleibende Parameter der DPD, nämlich  $\overline{h_t}$ , nicht mehr mittels Gleichung (4.22) bestimmt werden. Wir machen stattdessen die heuristische Annahme  $\overline{u_t} = 0$ , was zumindest mit den Verhältnissen der ersten zwei Zeitschritte konsistent ist. Mit diesem heuristischen Ansatz faktorisiert die remanente Magnetisierung in der Zeit und wir erhalten

$$c_{2t,0} = \prod_{\tau=1}^t c_{2\tau,2\tau-2}.$$

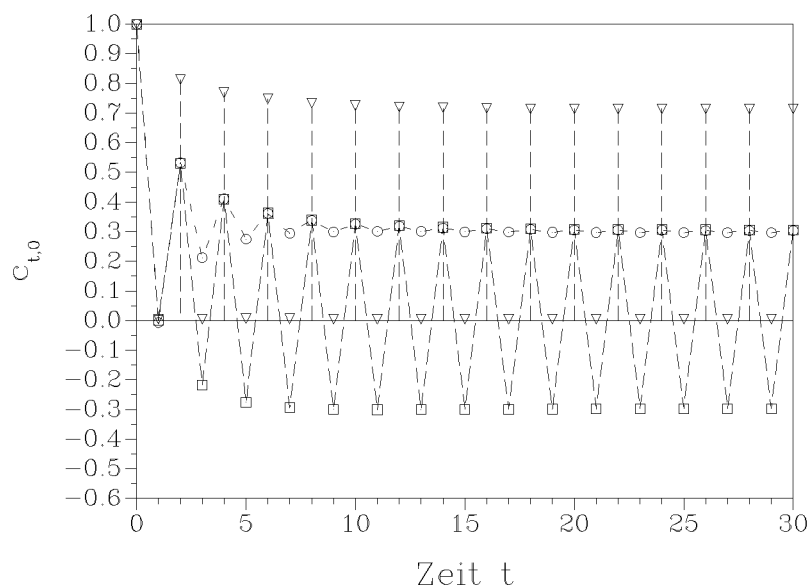
Die remanente Magnetisierung ist damit auf ein Produkt von Korrelationsfunktionen zurückgeführt.

Die Iteration der DPD-Gleichungen liefert jetzt folgendes Ergebnis (Abbildung 4.10): Die remanente Magnetisierung hat zu geraden Zeitschritten ein Maximum bei  $\alpha = \frac{1}{2}$  und fällt nach hohen und niedrigen  $\alpha$ -Werten hin zu null ab. Bei ungeraden Zeitschritten verschwindet die remanente Magnetisierung, es ergibt sich damit ein stark oszillatives Verhalten. Dies ist sehr ähnlich den Verhältnissen, welche bei der numerischen Simulation des SK-Modells gefunden werden.

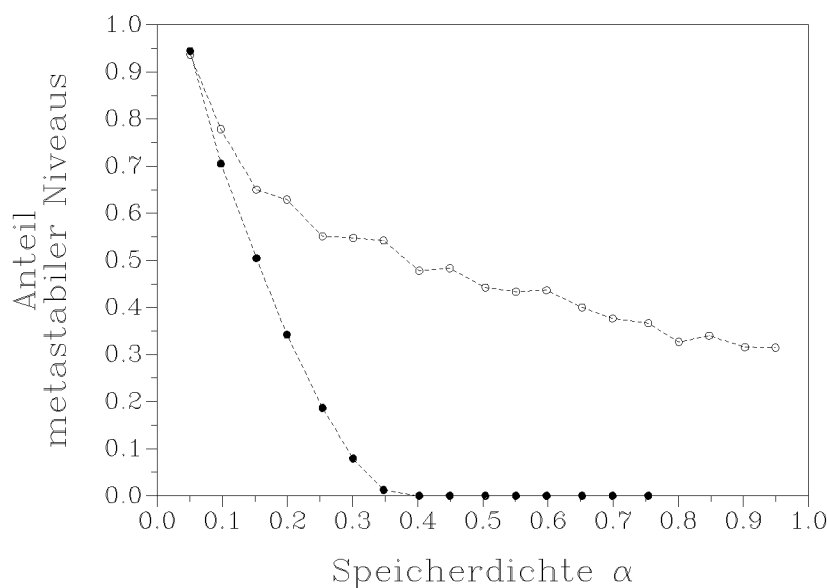
Die numerischen Simulationen des Netzwerkes mit pseudoinverser Kopplungsmatrix liefern allerdings das vorhergesagte oszillative Verhalten der remanenten Magnetisierung in Reinform nur bei  $\alpha = \frac{1}{2}$ . Für niedrigere  $\alpha$ -Werte verschwindet es hingegen. Für hohe  $\alpha$ -Werte treten aber dafür in den numerischen Simulationen sogar negative Werte der remanenten Magnetisierung auf! Besonders deutlich wird dieses stark  $\alpha$ -abhängige dynamische Verhalten von  $c_{t,0}$  in Abbildung 4.11, wo die zeitliche Entwicklung der remanenten Magnetisierung für die Speicherdichten  $\alpha = \frac{1}{4}$ ,  $\frac{1}{2}$  und  $\frac{3}{4}$  dargestellt ist. Die starke Oszillation zwischen positiven und negative Werten bei der  $\alpha = \frac{3}{4}$ -Kurve ist deutlich zu sehen.

Die hier beobachtete Oszillation der remanenten Magnetisierung eines Neuronalen Netzes ist ein neuer, interessanter Effekt, der technologisch vielleicht als 'Novelty detector' benutzt werden könnte. Normalerweise muß nämlich bei einem assoziativen Netzwerk von einem übergeordneten Netzwerk überprüft werden, ob eines der Muster erkannt worden ist oder ob das Netzwerk in irgendeinem der metastabilen Niveau gestoppt hat. Dies kann durch sogenannte 'Großmutter'-Neuronen geschehen, die im wesentlichen den Überlapp zu jeweils einem fest vorgegebenen Muster berechnen. Dazu sind jedoch, bei  $p = \alpha N$  Mustern, ebenso viele Skalarprodukte zu bilden. Es bietet sich nun folgendes, wesentlich ein-

re-  
ein



**Abbildung 4.11:** Die zeitliche Entwicklung der remanenten Magnetisierung für die  $\alpha$ -Werte  $\frac{1}{4}$  ('o'),  $\frac{1}{2}$  ('▽') und  $\frac{3}{4}$  ('□'). Deutlich zu sehen ist die starke Oszillation für hohe  $\alpha$ -Werte. Die Daten stammen von Neuronalen Netzen mit 256 Neuronen, gemittelt wurde über je 1024 Systeme.



**Abbildung 4.12:** Der Anteil der metastabilen Niveaus bei den dynamischen Fällen der synchronen Dynamik. Im hohen  $\alpha$ -Bereich spielen bei der pseudoinversen Kopplungsmatrix (‘•’) nur noch die Zweierzyklen eine Rolle. Für  $\alpha \rightarrow 0$  dominieren bei beiden Kopplungsmatrizen die metastabilen Niveaus als dynamische Fallen. Die Oszillationen der numerischen Daten beim Hopfield-Modell (‘◦’) werden durch die diskrete Verteilung der internen Felder dieser Kopplungsmatrix verursacht (256 Neuronen, 900 Systeme).

oszillatives Verhalten, ist das gerade getestete Eingangsmuster sicher nicht in der Menge der eingespeicherten Muster zu finden. Nur wenn die remanente Magnetisierung einen konstanten Wert annimmt, wurde das Eingangsmuster erfolgreich erkannt. Dieses Verfahren zur Erkennung neuer Muster setzt lediglich die synchrone Dynamik des Netzwerkes mit pseudoinverser Kopplungsmatrix und eine Speicherdichte  $\alpha > \frac{1}{2}$  voraus.

Das stark  $\alpha$ -abhängige Verhalten der remanenten Magnetisierung findet sich nicht bei der asynchronen Dynamik des Netzwerkes (siehe ebenfalls Abbildung 4.10). Hier treten keine Oszillationen auf, und auch die bei der synchronen Dynamik gefundene Symmetrie der Werte um  $\alpha = \frac{1}{2}$  fehlt. Stattdessen steigt die remanente Magnetisierung für asynchrone Dynamik mit wachsendem  $\alpha$  kontinuierlich an, was eine erhöhte Dichte dynamischer Fallen im hohen  $\alpha$ -Bereich nahelegt. Diese Vermutung wird auch durch die im Musterüberlapp gefundenen starken Remanenzeffekte unterstützt (Abbildung 4.6). In der Tat nimmt die Zahl der metastabilen Niveaus mit wachsendem  $\alpha$  zu [Ku90], sie zeigt sogar funktionell das gleiche Verhalten wie die remanente Magnetisierung der asynchronen Dynamik.

Die synchrone Dynamik besitzt zusätzlich zu den metastabilen Niveaus der asynchronen Dynamik noch die Zweierzyklen als dynamische Fallen, aber trotzdem führt dies im hohen  $\alpha$ -Bereich nicht zu erhöhten Remanenzeffekten bei dieser Dynamik. Einen Hinweis, warum dies so ist, liefern die numerischen Daten der Abbildung 4.12, bei denen der Prozentsatz der Fälle dargestellt ist, in welchen die synchrone Dynamik in metastabilen Niveaus endete. Für das Netzwerk mit

pseudoinverser Kopplungsmatrix treten oberhalb von  $\alpha \approx 0.4$  gar keine metastabilen Niveaus mehr als dynamische Fallen auf. Die Dynamik verläuft für diese  $\alpha$ -Werte offensichtlich durch Bereiche, die frei von Punktattraktoren sind. Die numerischen Daten von Abbildung 4.12 zeigen übrigens auch, daß, wie in Abschnitt 2.6 diskutiert, die Zweierzyklen für  $\alpha \rightarrow 0$  bei beiden Kopplungsmatrizen keine Rolle mehr spielen.

## 4.4 Dynamik unter externem Rauschen

### 4.4.1 Die Modifikation der dynamischen Gleichungen

Die DPD kann leicht auf eine Dynamik unter externem Rauschen erweitert werden. Die Übertragungsfunktion der Neuronen ist unter dem Einfluß externem Rauschens durch

$$S_i(t+1) = \text{sign}(h_i(t) + r_i(t))$$

definiert, wobei die Zufallsvariablen  $r_i(t)$  das externe Rauschen darstellen. Wir wählen jetzt

$$r_i(t) = \frac{1}{2\beta} \ln \left( \frac{1 - x_i(t)}{x_i(t)} \right)$$

mit  $x_i(t)$  in  $(0, 1)$  gleichverteilt. Damit erhalten wir nämlich

$$S_i(t+1) = \pm 1 \quad \text{mit der Wahrscheinlichkeit} \quad \frac{1}{1 + e^{\mp 2\beta h_i(t)}}.$$

was mit der üblichen Glauber-Dynamik übereinstimmt ([GL63], vergl. Abschnitt 1.3.4). Der Rauschparameter  $\beta$  kann damit als inverse Temperatur  $T$  angesehen werden:  $\beta = 1/T$ . Die synchrone Dynamik erfüllt 'detailed balance' mit dem Hamiltonian

$$H[\{S_i\}] = -\frac{1}{\beta} \sum_i \ln \left\{ 2 \cosh \left( \beta \sum_j J_{ij} S_j \right) \right\}$$

Eine Analyse der zu  $H[\{S_i\}]$  korrespondierenden Freien Energie wurde bislang nicht durchgeführt, und wir werden deshalb die DPD verwenden, um das thermodynamische Verhalten dieses Systems zu studieren. Die erhaltenen Resultate können aber mit Resultaten von Kanter und Sompolinsky [KA87], welche die Gleichgewichtszustände des Netzwerkes unter asynchroner Glauberdynamik in replika-symmetrischer Näherung untersucht haben, verglichen werden.

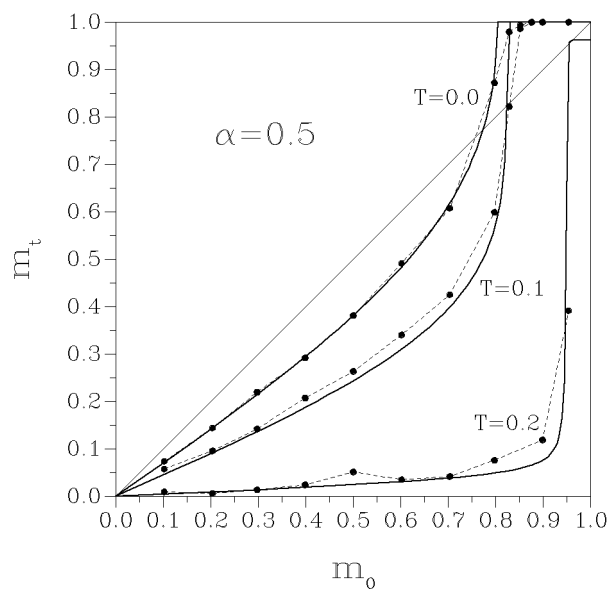
Die Erweiterungen der DPD auf die Dynamik unter externem Rauschen ist einfach. Die Feldverteilung  $P_t(h)$  enthält nun noch eine zusätzliche Mittelung über das synaptische Rauschen zu früheren Zeiten  $\tau < t$ . Da das synaptische Rauschen mit allen anderen auftretenden Zufallsprozessen unkorreliert ist, sind die sich ergebenden Änderungen der DPD klein. Der Überlapp ist nun

$$m_t = \overline{\tanh(\beta h_{t-1})}$$

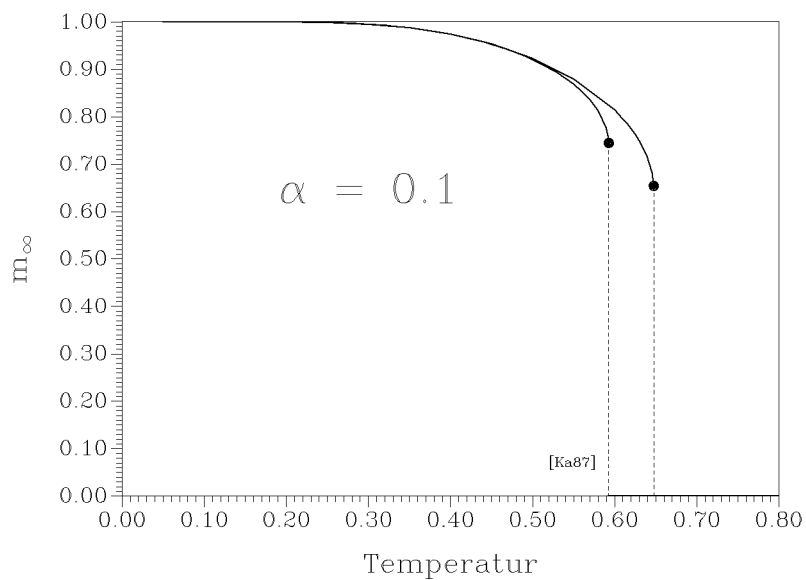
gegeben und die Separation der beiden Gaußverteilungen berechnet sich zu

$$(1 - m_{t-1}^2) d_t = \overline{h_{t-1} \tanh(\beta h_{t-1})} - \overline{h_t} m_{t-1}$$

Die übrigen Gleichungen bleiben unverändert. Natürlich werden für  $\beta \rightarrow \infty$  wieder die ursprünglichen DPD-Gleichungen erhalten.



**Abbildung 4.13:** Das Verhalten des Neuronalen Netzwerkes unter Temperatur. Die durchgezogenen Linien stellen die Vorhersage der DPD dar, während die gestrichelten Linien Resultate numerischer Simulationen sind (256 Neuronen, gemittelt über 100 Symmetrie



**Abbildung 4.14:** Der Überlapp der dynamischen Attraktoren mit den gelernten Mustern als Funktion der Temperatur  $T$  bei  $\alpha = 0.1$  (rechts). Ebenfalls eingetragen ist das Ergebnis der replika-symmetrischen Näherung von Kanter/Sompolinsky [KA87] für diesen  $\alpha$ -Wert.

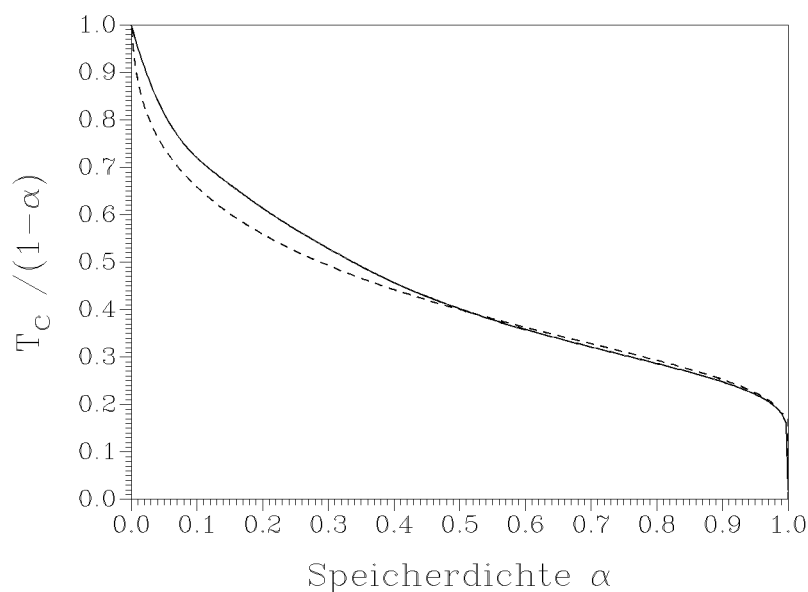
#### 4.4.2 Netzwerk-Dynamik unter externem Rauschen

In Abbildung 4.13 wird das Verhalten des Neuronalen Netzwerkes für  $\alpha = 0.5$  und bei Temperaturen  $T$  von 0.0, 0.1 und 0.2 dargestellt. Offensichtlich schrumpft das Attraktionsgebiet der Muster mit ansteigendem Rauschen. Gleichzeitig entwickelt sich dabei das  $m_i/m_0$ -Diagramm zu einer Stufenfunktion. Letzteres legt nahe, daß die dynamischen Fallen der deterministischen  $T = 0$ -Dynamik instabil gegenüber thermischen Fluktuationen sind (ein auch aus der thermodynamischen Analyse bekanntes Resultat). DPD-Vorhersage und numerische Simulation stimmen wiederum gut überein.

#### 4.4.3 1. Ordnungsphasenübergang

Mit steigendem Rauschpegel entfernen sich die dynamischen Attraktoren von den eingespeicherten Mustern. Dies ist schön in der  $T = 0.2$ -Kurve von Abbildung 4.13 zu sehen. Das Plateau des Attraktors hat nur noch einen Überlapp von etwa 0.96 mit dem Muster. Die Temperatur  $T = 0.2$  liegt, bei  $\alpha = \frac{1}{2}$ , sehr nahe an der kritischen Temperatur  $T = 0.201$ , bei der die dynamischen Attraktoren in der Nähe der Muster selbst nicht mehr stabil gegenüber dem thermischen Rauschen sind. Der Übergang stellt sich als starker 1. Ordnungsphasenübergang heraus, wie Abbildung 4.14 am Beispiel  $\alpha = 0.1$  zeigt. Ein ähnliches Verhalten

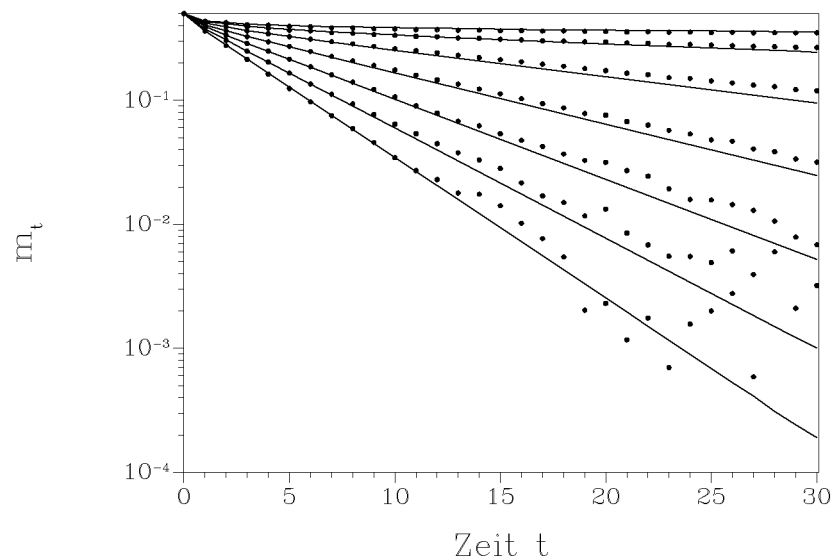
1. Das  
ildung



**Abbildung 4.15:** Die kritische Temperatur  $T_c$ , bei der die dynamischen Attraktoren in der Nähe der Muster instabil werden, als Funktion von  $\alpha$  (durchgezogene Linie). Die gestrichelte Linie stellt das Resultat der statischen Analyse von Kanter/Sompolinsky dar [KA87].

#### 4.4.4 Die kritische Temperatur $T_c$

Die kritische Temperatur  $T_c$ , bei der die dynamischen Attraktoren in der Nähe der Muster verschwinden, ist eine Funktion der Speicherauslastung  $\alpha$  (Abbil-



**Abbildung 4.16:** Unter externem Rauschen zerfällt der Musterüberlapp exponentiell schnell, falls das Muster nicht erkannt wird. Die Linien sind die Vorhersage der DPD, während die Punkte Daten aus numerischen Simulationen darstellen (256 Neuronen, 1024 Systeme). Die Temperaturen waren  $T = 0.05, 0.10, 0.15, 0.20, 0.25, 0.30$  und  $0.35$  (von oben nach unten).

dung 4.15). Auch hier finden wir eine Übereinstimmung mit den statischen Resultaten von Kanter/Sompolinsky, obwohl ihr Hamiltonian eigentlich nur für die asynchrone Dynamik gültig ist. Eine solche Übereinstimmung — von Gleichgewichtsergebnissen — konnte zwischen synchroner und asynchroner Dynamik auch im Falle der Hopfield-Kopplungsmatrix beobachtet werden [Fo87].

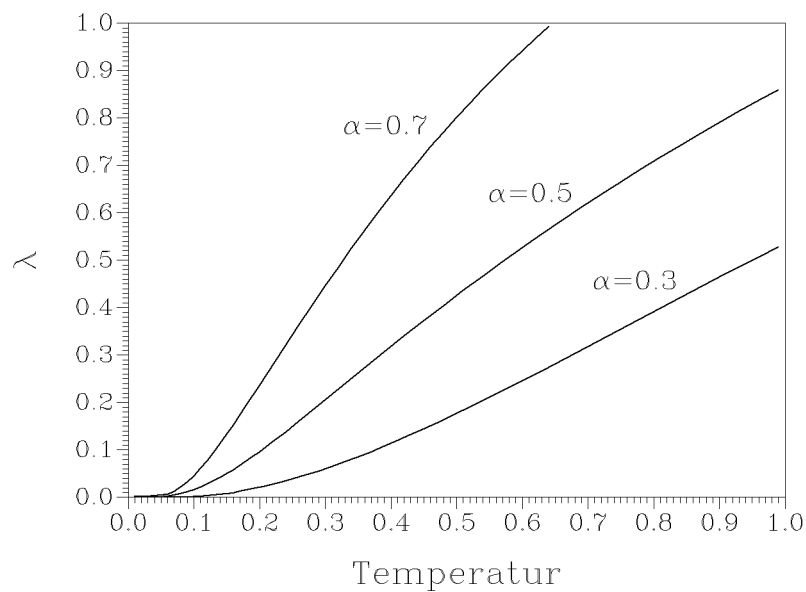
#### 4.4.5 Der exponentielle Zerfall von Überlapp und remanenter Magnetisierung

Wir untersuchen nun die Dynamik des Neuronalen Netzwerkes, falls das Netzwerk nicht in das Muster relaxiert. Wie Abbildung 4.16 für verschiedene Temperaturen zeigt, zerfällt der Musterüberlapp dann exponentiell schnell gegen null. Die Zeitkonstante des Zerfalls stellt sich als unabhängig von  $m_0$  heraus und ist als Funktion von  $\alpha$  und  $T$  in Abbildung 4.17 dargestellt. Für niedrige Temperaturen werden recht lange Relaxationszeiten gefunden.

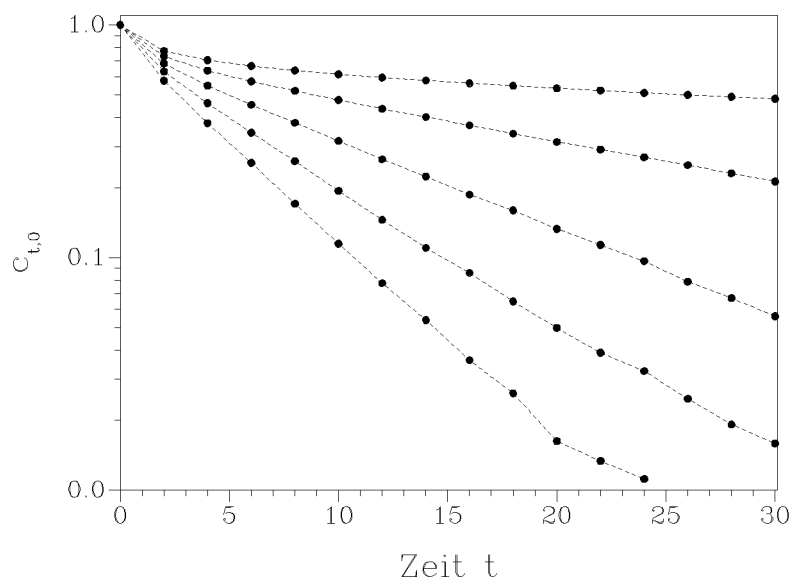
Die numerische Untersuchungen zur remanenten Magnetisierung  $c_{t,0}$  zeigen für diese Größe ebenfalls einen exponentiellen Zerfall (Abbildung 4.18). Dies ist ein deutlicher Unterschied zur synchronen Dynamik des SK-Spinalgases, bei der ein Potenzgesetz zu finden ist.

#### 4.4.6 Die Korrelationsfunktion $c_{t,t-2}$ bei stochastischer Dynamik

Für die Parameterwerte  $\alpha = 0.5$  und  $T = 0.1$  sind in Abbildung 4.9 (Seite 71) Daten für die zeitliche Korrelationsfunktion  $1 - c_{t,t-2}$  eingetragen. Die in der Abbildung ebenfalls angeführten, numerischen Daten für Systemgrößen von 128 und 1024 Neuronen stimmen beide mit der DPD-Vorhersage überein. Die prominenten 'finite-size'-Effekte der deterministischen Dynamik treten hier offen-

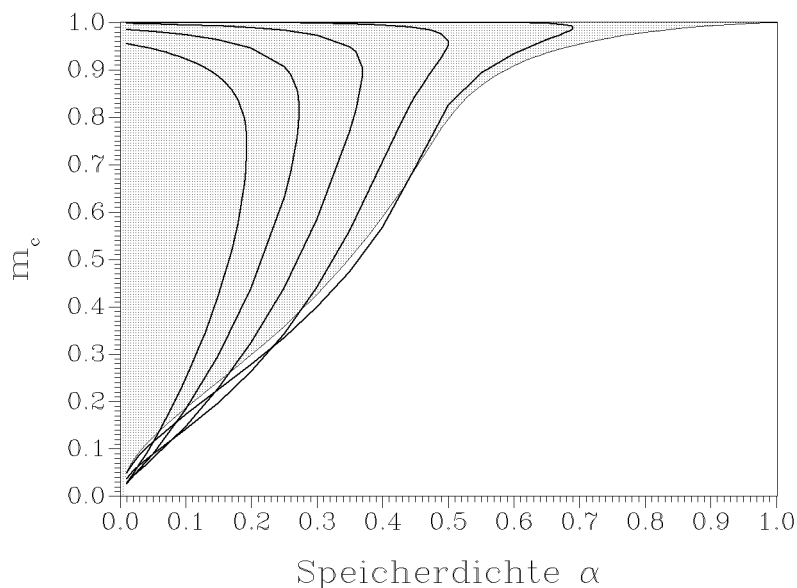


**Abbildung 4.17:** Die Zerfallskonstante des Überlapps  $m_t$ , als Funktion der Temperatur  $T$ , für drei verschiedene Speicherdichten  $\alpha$ . Das Verschwinden der Musterattraktoren bei den jeweiligen  $T$ 's hat offensichtlich keinen Einfluß auf die Zerfallsgeschwin-



**Abbildung 4.18:** Der Zerfall der remanenten Magnetisierung beim Neuronalen Netzwerk mit pseudoinverser Kopplungsmatrix unter synchroner Dynamik. Gezeigt sind, von oben nach unten, die Temperaturen 0.1, 0.15, 0.20, 0.25 und 0.3 bei einem  $\alpha$ -Wert von 0.5.





**Abbildung 4.19:** Das Attraktionsgebiet des Netzwerkes mit pseudoinverser Kopplungsmatrix bei Temperatur. Gezeigt sind die Attraktionsgebiete für die Temperaturen  $T = 0.0, 0.1, 0.2, 0.3, 0.4$  und  $0.5$  (von rechts nach links). Das Attraktionsgebiet für die deterministische  $T = 0$ -Dynamik ist grau unterlegt. Eine Vergrößerung der Attraktionsgebiete der Muster bei höherer Temperatur findet praktisch nicht statt.

sichtlich nicht auf. Man beachte dabei, daß  $T = 0.1$  immer noch weit unterhalb der kritischen Temperatur  $T = 0.201$  für diesen  $\alpha$ -Wert liegt.

#### 4.4.7 Das temperaturabhängige Attraktionsgebiet

Zum Abschluß ist in Abbildung 4.19 noch das temperaturabhängige Attraktionsgebiet des Netzwerkes mit pseudoinverser Kopplungsmatrix dargestellt. Für jede der angeführten Temperaturen gibt der untere Zweig der Kurven den kritischen Überlapp  $m_c$  an, während der obere Zweig den Überlapp der dynamischen Attraktoren in der Nähe der Muster darstellt. Für alle  $m_0$ -Werte zwischen 1 und  $m_c$  fließt das System zu den dynamischen Attraktoren. Startüberlapp unterhalb von  $m_c$  führen hingegen, für  $t \rightarrow \infty$ , zu verschwindendem Musterüberlapp,  $m_\infty = 0$ .

Wie die Abbildung 4.19 zeigt, verkleinern sich die Attraktionsgebiete mit steigender Temperatur. Es ist vermutet worden, daß bei mittleren Temperaturen das Netzwerkverhalten gegenüber dem Verhalten der deterministischen Dynamik werden kann, da alle metastabilen Niveaus durch das thermische Rauschen destabilisiert werden. Die Muster würden dann als einzige noch vorhandene, thermodynamisch stabile Attraktoren ihre Attraktionsgebiete ausdehnen [AM89]. Dies ist jedoch zumindest bei der Pseudoinversen nicht der Fall. Lediglich bei kleinen  $\alpha$ -Werten ist eine marginale Vergrößerung der Attraktionsgebiete festzustellen.