Synchronously parallel Boltzmann machines: a mathematical model
Zwietering, P.J.; Aarts, E.H.L.

Published: 01/01/1989

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal?

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Memorandum COSOR 89-21

Synchronously Parallel
Boltzmann Machines:
a Mathematical Model

Patrick Zwietering and Emile Aarts

Eindhoven, August 1989
The Netherlands
Synchronously Parallel Boltzmann Machines: 
a Mathematical Model

Patrick Zwietering¹ and Emile Aarts¹,²

1. Eindhoven University of Technology  
P.O. Box 513, 5600 MB Eindhoven, the Netherlands  
2. Philips Research Laboratories 
P.O. Box 80.000, 5600 JA Eindhoven, the Netherlands

Abstract

A mathematical model is presented for the description of synchronously parallel Boltzmann machines. The model is based on the theory of Markov chains and combines a number of previously known results into one generic model. It is argued that synchronously parallel Boltzmann machines maximize a function consisting of a weighted sum of the well-known consensus function and a pseudo consensus function. The weighting is determined by the amount of parallelism exploited in the Boltzmann machine and turns out to be problem dependent.

Keywords: Boltzmann machines, neural networks, synchronous parallelism, simulated annealing.

1 Introduction

Boltzmann machines, first introduced by Hinton, Sejnowsky & Ackley [1984], constitute a class of neural network models that can cope with difficult search, representation and learning problems [Aarts & Korst, 1989]. These models follow the basic paradigms of connectionist models [Fahlman & Hinton, 1987; Feldman & Ballard, 1982], assuming that information can be processed by massively parallel networks consisting of cooperating neuron-like computing elements that are highly interconnected by weighted links. Furthermore, A typical feature of these models is that the response of an individual computing element to other elements in the network is given by a non-linear scalar function of the weighted activity of the elements to which it is connected. For a review of neural network models see [Rumelhart, McClelland & the PDP Research Group, 1986]

More specifically, a Boltzmann machine is a symmetric network of simple 2-state computing elements interconnected by weighted links [Aarts & Korst, 1989]. A consensus function is used to quantify the "goodness" of a global configuration of the Boltzmann machine. Furthermore, the computing elements respond to their changing environment according to a stochastic transition function. Due to the stochastic nature of the response function
it is possible to model the behaviour of a Boltzmann machine exactly, and to prove that it converges asymptotically to globally optimal configurations. This is a unique feature of Boltzmann machines since most other neural networks follow the lines of Hopfield networks [Hopfield, 1982; 1984] in which a deterministic response function is used, performing a steepest descent, for which in general it is not possible to predict its performance.

In the book by Aarts & Korst [1989] an extensive treatment is given of the mathematical modelling of the convergence properties of Boltzmann machines. The treatment is primarily based on the assumption that Boltzmann machines operate sequentially, i.e. that only one element or unit may change its state at a time. Although the authors give some extensions to special parallel cases, no mathematical framework is presented that models parallel execution in general. Boltzmann machines, however, are intended to operate in parallel, and an adequate mathematical description is considered of great use for a proper understanding of their functioning.

This paper is intended as a first approach to the mathematical modelling of parallel Boltzmann machines. Our approach is limited to a description of synchronous parallelism applied to maximize the consensus function in a Boltzmann machine. We do not model the effect of parallelism on the learning algorithm as described for sequential Boltzmann machines (see for instance Aarts & Korst [1989] and Ackley, Hinton & Sejnowsky[1985]).

The organization of the paper is as follows. In Section 2 we briefly recall the most important aspects of Boltzmann machines within the framework of Markov chains. We then introduce a model for synchronously parallel Boltzmann machines (Section 3). Using this model a conjecture is formulated, concerning the general behaviour of synchronously parallel Boltzmann machines, and proven in a number of special cases (Section 4). Section 5 deals with the implications resulting from the modelling of Sections 3 and 4. In section 6 the paper is concluded with a discussion of the obtained results.

2 Boltzmann Machines: Premisses

A Boltzmann machine can be viewed as a network consisting of a number of two-state units, that are connected in some way. The network is represented by a pseudograph \( B = (\mathcal{U}, \mathcal{C}) \), where \( \mathcal{U} \) denotes the finite set of units and \( \mathcal{C} \) is a set of unordered pairs of elements of \( \mathcal{U} \) denoting the symmetric connections between the units. A connection \( \{u, v\} \in \mathcal{C} \) joins the units \( u \) and \( v \). The set of connections usually includes all loops or bias connections, i.e. \( \{\{u, u\} | u \in \mathcal{U}\} \subset \mathcal{C} \). If two units are connected, they are called adjacent. Furthermore, we say that the set \( \mathcal{C} \) specifies the connection pattern of a Boltzmann machine.

A unit \( u \) can be in one of two states: is either “on” or “off”. With each unit, a 0-1 variable is associated, denoting the state of the unit, where 0 and 1 correspond to “off” and “on”, respectively.

**Definition 1** A configuration \( k \) of a Boltzmann machine is given by a global state of the Boltzmann machine and is uniquely defined by a sequence of length \( |\mathcal{U}| \), whose \( u \)th component \( k(u) \) denotes the state of unit \( u \) in configuration \( k \). The configuration space \( \mathcal{R} \)
is given by the set of all possible configurations. Clearly, the cardinality of $\mathcal{R}$ equals $2^{1^u1}$.

**Definition 2** Let $\{u,v\} \in \mathcal{C}$ be a connection joining the units $u$ and $v$, then $\{u,v\}$ is activated in a given configuration $k$ if both units $u$ and $v$ are “on”, i.e. if

$$k(u) \cdot k(v) = 1.$$  \hspace{1cm} (1)

Otherwise the connection is not activated.

**Definition 3** With each connection $\{u,v\} \in \mathcal{C}$ a connection strength $s_{\{u,v\}} \in \mathbb{R}$ is associated as a quantitative measure for the desirability that the connection $\{u,v\}$ is activated. By definition, $s_{\{u,v\}} = s_{\{v,u\}}$. If $s_{\{u,v\}} > 0$ it is desirable that $\{u,v\}$ is activated; if $s_{\{u,v\}} < 0$ it is undesirable. Connections with a positive strength are called excitatory; connections with a negative strength are called inhibitory. Furthermore, the strength $s_{\{u,u\}}$ of the bias connection $\{u,u\}$ is called the bias of unit $u$.

**Definition 4** The consensus function $C : \mathcal{R} \rightarrow \mathbb{R}$ assigns to each configuration $k$ a real number, called the consensus, which equals the sum of the strengths of the activated connections, i.e.

$$C_k = \sum_{\{u,v\} \in \mathcal{C}} s_{\{u,v\}} k(u) k(v).$$  \hspace{1cm} (2)

In general the consensus will be large if many excitatory connections are activated, and it will be small if many inhibitory connections are activated. In fact, the consensus is a global measure indicating to what extent the units in a Boltzmann machine have reached a consensus about their individual states, subject to the desirabilities expressed by the individual connection strengths.

The objective of a Boltzmann machine is to reach a globally optimal configuration, i.e. a configuration with maximal consensus. To reach a maximal consensus, a state transition mechanism is introduced which allows the units to change their states (from “0” to “1” or vice versa), so as to adjust themselves to the states of the neighbouring units.

**Definition 5** Let a Boltzmann machine be in configuration $k$, then changing the state of unit $u$ results in a configuration $l$ for which $l(u) = 1 - k(u)$. Furthermore, Let $C_u$ denote the set of connections incident with unit $u$, excluding $\{u,u\}$, then the difference in consensus induced by changing the state of unit $u$ in configuration $k$ is given by

$$\Delta C_k(u) = r_k(u) h_k(u),$$  \hspace{1cm} (3)

where

$$r_k(u) = 1 - 2k(u)$$  \hspace{1cm} (4)

and

$$h_k(u) = \sum_{\{u,v\} \in C_u} s_{\{u,v\}} k(v) + s_{\{u,u\}}.$$  \hspace{1cm} (5)

From (3)-(5) it is apparent that the effect on the consensus, resulting from changing the state of unit $u$, is completely determined by the states of its neighbouring units and the corresponding connection strengths. Consequently, each unit can locally evaluate its own
state transition since no global calculations are required. This is a very important property since it means that there is potential for parallel execution.

Typical of a Boltzmann machine is the probabilistic response $A_{k}^{(\tau)}(u)$ of an individual unit $u$ to its neighbouring units given configuration $k$, which takes the following form [Aarts & Korst, 1989; Ackley, Hinton & Sejnowski, 1984]

$$A_{k}^{(\tau)}(u) = \frac{1}{1 + \exp(-\tau \Delta C_k(u))},$$

(6)

where $\Delta C_k(u)$ is given by (3) and $\tau \in \mathbb{R}^+$ denotes a control parameter. The probabilistic response function given above allows a Boltzmann machine to escape from locally optimal configurations. In the next section we elaborate on this in greater detail.

To model the state transitions of the units in a Boltzmann machine, we use the theory of Markov chains. Before we can do this we need some definitions and basic results.

**Definition 6** A Markov chain is a sequence of trials, where the probability of the outcome of a given trial depends only on the outcome of the previous trial. Let $X(i)$ be a stochastic variable denoting the outcome of the $i^{th}$ trial, then the transition probability at the $i^{th}$ trial for each pair $k,l$ of outcomes is defined as

$$P_{kl}(i) = \mathbb{P}\{X(i) = l \mid X(i-1) = k\}.$$  

(7)

The matrix $P(i)$ whose elements are given by (7) is called the transition matrix. Furthermore, the Markov chain is called homogeneous if $P(i)$ is independent of $i$; it is called finite if the set of outcomes is finite.

**Definition 7** A Markov chain with transition matrix $P$ is irreducible, if for each pair of outcomes $k,l$ there is a positive probability of reaching $l$ from $k$ in a finite number of trials, i.e. if $\forall k,l \exists n \geq 1 : (P^n)_{kl} > 0$. Moreover, the chain is aperiodic if $\exists k : P_{kk} > 0$.

**Definition 8** [Feller, 1950] The stationary distribution of a finite homogeneous Markov chain with transition matrix $P$ is the probability distribution of the outcomes after an infinite number of trials, and can be defined as the vector $q$, whose $k^{th}$ component is given by

$$q_k = \lim_{i \to \infty} \mathbb{P}\{X(i) = k\}.$$  

(8)

**Property 1** [Feller, 1950] Let $P$ be the stochastic transition matrix associated with a finite homogeneous Markov chain and let the Markov chain be irreducible and aperiodic. Then there exists a stochastic vector $q$ whose components $q_i$ are uniquely determined by the following equation

$$\forall k : \sum_l q_l P_{lk} = q_k.$$  

(9)

The vector $q$ is the stationary distribution of the Markov chain, because it satisfies (8). Furthermore, it can be easily proven that the following so-called detailed balance equation

\footnote{We usually use $e^{-1}$ instead of $\tau$, c.f. [Aarts & Korst, 1989]. Here the inverse is chosen for convenience sake. Furthermore, $\tau$ is also denoted as the inverse temperature [Ackley, Hinton & Sejnowski, 1984].}
is sufficient for the verification of the stationary distribution

\[ \forall k, l : q_k P_{kl} = q_l P_{kl}. \]  

(10)

3 Synchronously Parallel Boltzmann Machines

There are a number of different approaches that can be pursued to realize parallel state transitions in a Boltzmann machine. Here, we distinguish between the following two modes of parallelism.

**Synchronous parallelism** Sets of state transitions are scheduled in successive trials, each trial consisting of a number of individual state transitions. After each trial the accepted state transitions are communicated through the network so that all units have up-to-date information about the states of their neighbours before the next trial is initiated. During each trial a unit is allowed to propose a state transition only once. Evidently, synchronous parallelism requires a global clocking scheme to control the synchronization.

**Asynchronous parallelism** State transitions are evaluated simultaneously and independently. Units continuously generate state transitions and accept or reject them on the basis of information that is not necessarily up-to-date, since the states of its neighbours may have changed in the meantime. Clearly, asynchronous parallelism does not require a global scheme, which is of great advantage in the hardware implementations of the Boltzmann machine.

In this paper we only consider the *synchronous* approach. Asynchronous parallelism in a Boltzmann machine is much harder to model since in that case no time discretization can be introduced which is essential for the description based on the theory of Markov chains.

The state transition mechanism in a synchronously parallel Boltzmann machine is thought to consist of two steps: firstly a subset of units is generated that may propose a state transition, and secondly each unit in the subset evaluates a state transition based on the current states of all other units. With this mechanism we can associate a transition matrix in the following way.

**Definition 9** Let, for each pair of configurations \( k, l \in \mathcal{R} \), \( U_{kl} \) be the subset of units that should change their states in order to transform \( k \) into \( l \), i.e.

\[ U_{kl} = \{ u \in \mathcal{U} \mid k(u) \neq l(u) \}. \]  

(11)

Then the *transition probability* \( P_{kl}^{(\tau)} \) of transforming configuration \( k \) into \( l \) is defined by

\[ P_{kl}^{(\tau)} = \sum_{U \subseteq U_{kl}} G(U_s) A_{kl}^{(\tau)}(U_s), \]  

(12)

where \( G(U_s) \) denotes the *generation probability*, i.e. the probability of generating the set \( U_s \subseteq \mathcal{U} \), and \( A_{kl}^{(\tau)}(U_s) \) the *acceptance probability*, i.e. the probability of accepting the state
transitions required to transform \( k \) into \( l \), given that the set of units \( \mathcal{U}_s \) is allowed to make a transition; \( \tau \) denotes the control parameter mentioned before.

The choices for the generation and acceptance probabilities play an important role in the modelling of parallel Boltzmann machines, and the discussion in the remainder of this section is centered on these two items.

**The Generation Probability**

The generation probability determines the amount parallelism in a Boltzmann machine and the way the units are synchronized. Furthermore, the outcome \( \mathcal{U}_s \) of the generation process defines for each configuration \( k \) a *neighbourhood* \( \mathcal{R}_k(\mathcal{U}_s) \) consisting of all configurations \( l \in \mathcal{R} \) that can be reached from \( k \) by a state transition of one or more units in \( \mathcal{U}_s \), i.e.

\[
\mathcal{R}_k(\mathcal{U}_s) = \{ m \in \mathcal{R} | \mathcal{U}_{km} \subseteq \mathcal{U}_s \}. \tag{13}
\]

Note that \( k \in \mathcal{R}_k(\mathcal{U}_s) \) and \( l \in \mathcal{R}_k(\mathcal{U}_s) \Leftrightarrow k \in \mathcal{R}_l(\mathcal{U}_s) \).

We impose the following conditions on \( G \)

\[
\forall \mathcal{U}_s \subseteq \mathcal{U} : G(\mathcal{U}_s) \geq 0, \tag{14}
\]

\[
\sum_{\mathcal{U}_s \subseteq \mathcal{U}} G(\mathcal{U}_s) = 1, \tag{15}
\]

\[
\bigcup_{G(\mathcal{U}_s) > 0} \mathcal{U}_s = \mathcal{U}. \tag{16}
\]

The first two conditions ensure that \( G \) is a correctly defined probability function. The last condition guarantees that each unit can be in the set of units that is allowed to make a transition.

**The Acceptance Probability**

As a direct consequence of the probabilistic response function of the individual units given in (6), the acceptance probability takes the following form:

\[
A_{\mathcal{U}_s}^{(r)}(\mathcal{U}_s) = \prod_{u \in \mathcal{U}_s} A_k^{(r)}(u) \prod_{u \in \mathcal{U} \backslash \mathcal{U}_s} \left( 1 - A_k^{(r)}(u) \right) \\
\prod_{u \in \mathcal{U}_s} [1 + \exp(-\tau \Delta C_k(u))]^{-1} \prod_{u \in \mathcal{U} \backslash \mathcal{U}_s} [1 + \exp(+\tau \Delta C_k(u))]^{-1} \\
\prod_{u \in \mathcal{U}_s} [1 + \exp(\tau r_i(u) h_k(u))]^{-1} \\
\exp\left( -\frac{\tau}{2} \sum_{u \in \mathcal{U}_s} r_i(u) h_k(u) \right) \\
= \frac{\prod_{u \in \mathcal{U}_s} 2 \cosh\left( \frac{\tau}{2} h_k(u) \right)}{\prod_{u \in \mathcal{U}_s}}. \tag{17}
\]

The factor \( r_i(u) \in \{-1, +1\} \) in the denominator of (17) can be omitted because \( \cosh(x) \) is an even function.
Using the following identity (see [Peretto, 1984])
\[
\prod_{u \in \mathcal{U}_s} 2 \cosh\left( \frac{r}{2} h_k(u) \right) = \sum_{m \in \mathcal{M}(\mathcal{U}_s)} \exp\left( -\frac{r}{2} \sum_{u \in \mathcal{U}_s} r_m(u) h_k(u) \right),
\]
and, denoting
\[
F_{kl}(\mathcal{U}_s) = -\frac{1}{2} \sum_{u \in \mathcal{U}_s} r_l(u) h_k(u),
\]
the expression of (17) can be rewritten as
\[
A_{kl}^{(\tau)}(\mathcal{U}_s) = \frac{\exp(\tau F_{kl}(\mathcal{U}_s))}{\sum_{m \in \mathcal{M}(\mathcal{U}_s)} \exp(\tau F_{km}(\mathcal{U}_s))}.
\]
From (20) it is clear that \(A_{kl}^{(\tau)}(\mathcal{U}_s)\) defines a stochastic matrix, i.e. \(\sum_{l \in \mathcal{M}(\mathcal{U}_s)} A_{kl}^{(\tau)}(\mathcal{U}_s) = 1\). Together with condition (15) this implies that \(P^{(\tau)}\) given by (12) is a stochastic matrix.

It can be easily verified that the conditions of (14)-(16) together with the definition of the acceptance probability of (17) ensure that the Markov chain induced by (12) is irreducible and aperiodic (see Definition 7). This is an important result since it states that the conditions of Property 1 are satisfied and that, consequently, there exists a unique stationary distribution of the corresponding Markov chain.

In Section 4 we derive closed forms of the stationary distribution for a number of cases with different additional conditions on the generation probability. These expressions are then used to prove Conjecture 1 for these cases, which is the main issue of this paper.

In the remainder of this section we introduce some new notions which will be used in the rest of this paper. We derive a few elementary properties of these notions which makes it possible to rewrite the acceptance probability of (20) into a more comprehensive form. The expressions also serve as a basis for the results derived in later sections. The following shorthand notation will be of use in the forthcoming definitions.

\[
\forall u, v \in \mathcal{U} : B(u, v) = \begin{cases} 
\frac{1}{2} s_{\{u, v\}} & \text{if } \{u, v\} \in C_u, \\
0 & \text{if } \{u, v\} \notin C_u.
\end{cases}
\]

\[
\forall u \in \mathcal{U} : a(u) = s_{\{u, u\}},
\]

\[
\forall u \in \mathcal{U} : (B \cdot k)(u) = \sum_{v \in \mathcal{U}} B(u, v) k(v),
\]

\[
k \cdot a = \sum_{u \in \mathcal{U}} k(u) a(u).
\]

With these notations, the expressions of \(C_k\) and \(h_k\) given by (2) and (5), respectively, can be rewritten as

\[
C_k = k \cdot B \cdot k + k \cdot a,
\]

\[
h_k = 2B \cdot k + a.
\]
Definition 10 The auxiliary functions $D_{kl}, E_{kl}$ and $J_{kl}(U_s)$ are defined as

$$D_{kl} = k \cdot a + 2k \cdot B \cdot l + l \cdot a,$$  \hspace{1cm} (24)  

$$E_{kl} = -(k - l) \cdot B \cdot (k - l),$$  \hspace{1cm} (25)  

$$J_{kl}(U_s) = D_{kl} - F_{kl}(U_s).$$  \hspace{1cm} (26)  

Note that (22) and (24) imply that $C_k = \frac{1}{2} D_{kk}.$

Lemma 1 Let the functions $D_{kl}$ and $E_{kl}$ be given by (24) and (25), respectively. Then

$$D_{kl} = \frac{1}{2} (D_{kk} + D_{ll}) + E_{kl}. \hspace{1cm} (27)$$

Proof

From (24) it follows that

$$D_{kl} - \frac{1}{2} (D_{kk} + D_{ll}) = l \cdot a + 2l \cdot B \cdot k + k \cdot a - k \cdot B \cdot k - k \cdot a - l \cdot B \cdot l - l \cdot a$$

$$= k \cdot B \cdot l - k \cdot B \cdot k - l \cdot B \cdot l + l \cdot B \cdot k$$

$$= -(k - l) \cdot B \cdot (k - l).$$

Using (25) then completes the proof of the lemma. ■

Lemma 2 Let $U_{kl} \subseteq U_s$. Then $J_{kl}(U_s) = J_{kk}(U_s)$.

Proof $U_{kl} \subseteq U_s$ implies $k(u) = l(u)$ for $u \not\in U_s$. Hence

$$F_{kk}(U_s) - F_{kl}(U_s) \equiv \sum_{u \in U_s} \frac{1}{2} (r_l(u) - r_k(u)) h_k(u)$$

$$\equiv \sum_{u \in U_s} (k(u) - l(u)) h_k(u)$$

$$\equiv \sum (k(u) - l(u)) h_k(u)$$

$$\equiv (k - l) \cdot h_k$$

$$\equiv (k - l) \cdot (2B \cdot k + a)$$

$$\equiv D_{kk} - D_{kl}.$$  \hspace{1cm} (28)

Combining this result with the expression of (26) completes the proof. ■

The following theorem gives an alternative expression for the acceptance probability $A_{kl}^{(r)}(U_s)$ of (20). The expression consists of a factorization into two terms, one term depending only on $l$, the other one depending on both $k$ and $l$.

Theorem 1 Let the acceptance probability of a synchronously parallel Boltzmann machine be given by (20). Then

$$\forall k, l \in R \forall U_s \ni U_{kl} : A_{kl}^{(r)}(U_s) = \frac{\exp(\tau C_l) \exp(\tau E_{kl})}{\sum_{m \in R_k(U_s)} \exp(\tau C_m) \exp(\tau E_{km})}.$$  \hspace{1cm} (28)
Proof
From Lemmas 1 and 2 it follows that
\[ \forall k, l \in \mathcal{R} : D_{kl} = C_k + C_l + E_{kl} \tag{29} \]
and
\[ \forall k, l \in \mathcal{R} \forall U_s : F_{kl}(U_s) = D_{kl} - J_{kk}(U_s). \tag{30} \]
Hence, we obtain
\[
\begin{aligned}
A_{kl}^{(r)}(U_s) & \equiv \frac{\exp(\tau F_{kl}(U_s))}{\sum_{m \in \mathcal{R}_k(U_s)} \exp(\tau F_{km}(U_s))} \\
& \overset{(29)}{=} \frac{\exp(\tau D_{kl} - \tau J_{kk}(U_s))}{\sum_{m \in \mathcal{R}_k(U_s)} \exp(\tau D_{km} - \tau J_{kk}(U_s))} \\
& = \frac{\exp(\tau D_{kl})}{\sum_{m \in \mathcal{R}_k(U_s)} \exp(\tau D_{km})} \\
& \overset{(29)}{=} \frac{\exp(\tau C_k) \exp(\tau E_{kl})}{\sum_{m \in \mathcal{R}_k(U_s)} \exp(\tau C_m) \exp(\tau E_{km})},
\end{aligned}
\]
which completes the proof of the theorem.

The resulting expression for the acceptance probability given by (28), reveals two important aspects.

1. If \( E_{km} = 0 \) for all \( m \in \mathcal{R}_k(U_s) \), then the desired drift towards configurations with larger consensus is guaranteed.

2. If \( E_{km} \neq 0 \) for some \( m \in \mathcal{R}_k(U_s) \), then it is not (yet) clear what the behaviour of the synchronously parallel Boltzmann will be.

In the following section we elaborate on these aspects in more detail and partly answer the question posed in the second item.

4 Main Result

From the previous section we know that the Markov chain corresponding to the transition probability matrix \( P^{(r)} \) of a synchronously parallel Boltzmann machine with rather general choices for the generation and acceptance probability, has a stationary distribution which we shall denote by \( q^{(r)} \), \( r \) again denoting the control parameter.

In general it is desirable to have an analytical expression for \( q^{(r)} \), in order to study the asymptotic behaviour of the Boltzmann machine. Although Theorem 1 provides a promising first step in this direction, we can in fact only use it to obtain the desired expression for
a special subclass of synchronously parallel Boltzmann machines for which the expression for \( q^{(r)} \) is already known (see Section 4.1, [Aarts & Korst, 1989]).

The next step was to do without \( q^{(r)} \). Using a different approach we were able to conjecture the outcome of \( \lim_{T \to \infty} q^{(r)} \) in our most general model. This is the main issue of this paper. Furthermore we show that our conjecture is true in those cases for which we can calculate \( q^{(r)} \) (including all previously known results) and hence, one can view the conjecture as a generic model for these cases. We also give motivation for the conjecture being true in the yet uncovered situations. The next section deals with the implications of the conjecture and gives some applications.

Before we can formulate our main result we need the following definitions.

**Definition 11** For every configuration \( k \in \mathcal{R} \) and \( U_s \subseteq U \), a configuration \( k_{U_s} \in \mathcal{R}_k(U_s) \) is defined as

\[
    k_{U_s}(u) = \begin{cases} 
    1 & \text{if } h_k(u) > 0 \text{ and } u \in U_s \\
    0 & \text{if } h_k(u) < 0 \text{ and } u \in U_s \\
    k(u) & \text{if } h_k(u) = 0 \text{ or } u \notin U_s,
    \end{cases}
\]

where \( h_k \) is the function defined in equation 5.

**Definition 12** For every configuration \( k \in \mathcal{R} \) and subset \( U_s \subseteq U \) the generalized pseudo consensus function \( S_k(U_s) \) is defined by

\[
    S_k(U_s) = D_{kk}^* U_s.
\]

Note that \( k_{\emptyset}^* = k \) and hence \( S_k(\emptyset) = D_{kk} = 2C_k \). Furthermore, we have that \( S_k(U) = S_k \), where \( S_k \) denotes the pseudo consensus function introduced by Aarts & Korst [1989].

**Definition 13** For every configuration \( k \in \mathcal{R} \) the extended consensus function \( C_k \) is defined by

\[
    C_k = \sum_{U_s \subseteq U} G(U_s) S_k(U_s).
\]

We now come to the main result.

**Conjecture 1** Let \( B \) be a synchronously parallel Boltzmann machine with transition matrix \( P^{(r)} \) given by (12). Then the stationary distribution \( q^{(r)} \) of the Markov chain induced by \( P^{(r)} \) converges as \( r \to \infty \) to a uniform distribution over the set \( \mathcal{R}_{opt} \), where \( \mathcal{R}_{opt} \) denotes the set of configurations \( k \in \mathcal{R} \) for which the extended consensus function \( C_k \) is maximal.

In the next two sections we prove the correctness of the conjecture for some special subclasses of synchronously parallel Boltzmann machines. We therefore distinguish between the following two cases of synchronous parallelism.

**Limited parallelism** Units may change their states in parallel only if they are not adjacent.

**Unlimited parallelism** Units may change their states in parallel whether or not they are adjacent.
4.1 Limited Parallelism

In this section we consider synchronously parallel Boltzmann machines applying limited parallelism. In this case the stationary distribution \( q^{(r)} \) can be expressed explicitly, see Lemma 4, which allows us to prove Conjecture 1, see theorem 2.

The concept of limited parallelism is formalized in the next definition.

**Definition 14** A synchronous parallel Boltzmann machine is said to apply limited parallelism if the generation if the generation probability satisfies

\[
\forall U_s \subseteq U : G(U_s) > 0 \implies \forall u, v \in U_s, u \neq v : \{u, v\} \not\in \mathcal{C}.
\]

(34)

This definition yields the following result.

**Lemma 3** Let \( B \) be a synchronously parallel Boltzmann machine applying limited parallelism. Then

\[
\forall U_s \subseteq U : G(U_s) > 0 \implies \forall k \in \mathcal{R}, l \in \mathcal{R}_k(U_s) : E_{kl} = 0.
\]

(35)

**Proof**
If \( U_{kl} \subseteq U_s \), then it follows directly from the definitions of \( E_{kl} \) (25) and \( U_{kl} \) (11) that

\[
|E_{kl}| \leq \sum_{\{u,v\} \in \mathcal{C}} |s_{\{u,v\}}| \leq \sum_{\{u,v\} \in \mathcal{C}} |s_{\{u,v\}}|.
\]

(36)

Furthermore if \( U_s \subseteq U \) with \( G(U_s) > 0 \), then (34) implies that the summation in the right-hand side of (36) runs over an empty set. This implies that \( E_{kl} = 0 \) which completes the proof.

**Lemma 4** Let \( B \) be a synchronously parallel Boltzmann machine applying limited parallelism with transition matrix \( P^{(r)} \) given by (12). Then

(i) there exists a stationary distribution of the finite homogeneous Markov chain induced by \( P^{(r)} \) on \( \mathcal{R} \), whose components are given by

\[
\forall k \in \mathcal{R} : q^{(r)}_k = \frac{\exp(\tau C_k)}{\sum_{m \in \mathcal{R}} \exp(\tau C_m)},
\]

(37)

and

(ii) the stationary distribution converges as \( r \to \infty \) to a uniform distribution over the set of optimal configurations, i.e. \( \lim_{r \to \infty} q^{(r)} = r \), where

\[
r_k = \begin{cases} 
|\mathcal{R}_{\text{opt}}|^{-1} & k \in \mathcal{R}_{\text{opt}} \\
0 & k \not\in \mathcal{R}_{\text{opt}},
\end{cases}
\]

(38)

and \( \mathcal{R}_{\text{opt}} \) denotes the set of configurations \( k \in \mathcal{R} \) with maximal consensus.
Proof

(i) To prove part (i) of the lemma we use the fact that the Markov chain induced by $P^{(r)}$ is irreducible and aperiodic, see Section 3, and hence from Property 1 we know that there exists a unique stationary distribution. The proof is then completed by showing that the components of the stationary distribution of (37) satisfy the detailed balance equation (10). From Lemma 3 we know that for all $k, l \in \mathcal{R}$ and $U_s \supseteq U_{kl}$ with $G(U_s) > 0$, we have that $E_{kl} = 0$, and thus from Theorem 1 we obtain

\[ A_{kl}^{(r)}(U_s) = \frac{\exp(\tau C_l)}{\sum_{m \in \mathcal{R}_k(U_s)} \exp(\tau C_m)}. \]  

Furthermore, it is not difficult to show that $U_s \supseteq U_{kl}$ implies that $\mathcal{R}_k(U_s) = \mathcal{R}_l(U_s)$. From (37) and (39) it then follows directly that $q_k^{(r)} P_k^{(r)} = q_l^{(r)} P_l^{(r)}$ which validates the detailed balance equation (10), and hence the proof of part (i) is completed.

(ii) Follows directly from (37).

Next, from Lemma 4, we obtain the following result.

**Theorem 2** Conjecture 1 holds for a synchronously parallel Boltzmann machine applying limited parallelism.

Proof

Following Lemma 4 it suffices to show that $\tilde{C}_{\text{opt}} = \mathcal{R}_{\text{opt}}$.

Using the definition of the generalized pseudo consensus function we can rewrite the extended consensus function as

\[ \tilde{C}_k = \sum_{U_s \supseteq U} G(U_s) D_{kk} U_s. \]

\[ \tilde{C}_k \overset{(20)}{=} \sum_{U_s \supseteq U} G(U_s)(C_k + C_k U_s) + E_{kk} U_s, \]

\[ = \sum_{U_s \supseteq U} G(U_s)(C_k + C_k U_s), \]  

because $G(U_s) > 0$ implies $E_{kk} U_s = 0$, since limited parallelism is applied and $k U_s \in \mathcal{R}_k(U_s)$.

Let $C_{\text{opt}} = C_k$ for some $k \in \mathcal{R}_{\text{opt}}$ and $k \in \mathcal{R}$, if $C_k < C_{\text{opt}}$, then we find with (40)

\[ \tilde{C}_k \overset{(15)}{=} 2 C_{\text{opt}} \sum_{U_s \supseteq U} G(U_s) \]

Hence

\[ C_k < C_{\text{opt}} \Rightarrow \tilde{C}_k < 2 C_{\text{opt}}. \]
If $C_k = C_{opt}$, then $\Delta C_k(u) = C_{k(u)} - C_k \leq 0$ for all $u \in \mathcal{U}$. Hence, using
\[
\Delta C_k(u) = r_k(u)h_k(u) = (1 - 2k(u))h_k(u),
\]
we have
\[
k(u) = \begin{cases} 
1 & \text{if } h_k(u) > 0 \\
0 & \text{if } h_k(u) < 0,
\end{cases}
\]
i.e. $k(u) = k_{U_s}^*(u)$ for all $u \in \mathcal{U}$ and $U_s \subseteq \mathcal{U}$. Thus $C_k = C_{opt}$ implies that $C_{k_{U_s}} = C_k = C_{opt}$. Using (40) it follows directly that
\[
C_k = C_{opt} \Rightarrow \tilde{C}_k = 2C_{opt}. 
\]
Combining (41) and (42) then completes the proof.

4.2 Unlimited Parallelism

In the case that a Boltzmann machine applies unlimited parallelism, we cannot calculate the stationary distribution like we did for synchronously parallel Boltzmann machines applying limited parallelism (see previous section). Therefore we pursue a different approach consisting of the following two steps.

First, we consider the isolated process in which the matrix $A^{(r)}(U_s)$ induces a finite homogeneous Markov chain on the subspace $\mathcal{R}_{k_0}(U_s)$ of the configuration space $\mathcal{R}$, where $k_0 \in \mathcal{R}$ is an arbitrary but fixed configuration. This is a valid Markov process since $\mathcal{R}_{k_0}(U_s)$ is a recurrent subspace and $A^{(r)}(U_s)$ is a stochastic matrix.

The process described above corresponds to a Boltzmann machine in which the same "isolated" subset of units $U_s$ is generated all the time, allowing all units in the subset to change their states simultaneously, i.e. we consider the case in which the outcome of the generation mechanism yields the same subset $U_s$ all the time. This is the subject of Lemma 6.

Next, in the second step of our approach, we return to the general model of a synchronously parallel Boltzmann machine as given by equation 12, where the generation mechanism may yield a different subset each time it is applied. We explain how this approach supports Conjecture 1 and show where a possible proof may start.

Finally, for a special kind of unlimited parallelism, we show that the results can be used to prove Conjecture 1 in this special case. This is the subject of Theorem 3.

We start by introducing an auxiliary function and proving a lemma, which provides a relation between this function and the generalized pseudo consensus function.

**Definition 15** The function $f_k^{(r)}(U_s)$ is for all $k \in \mathcal{R}$, $U_s \subseteq \mathcal{U}$ and $r \geq 0$ defined by
\[
f_k^{(r)}(U_s) = rJ_{kk}(U_s) + \sum_{u \in U_s} \ln(2 \cosh(\frac{r}{2}h_k(u))). 
\]
Lemma 5 For all $k,l \in \mathcal{R}$

$$\lim_{r \to \infty} \left( f_k^r(U_s) - f_l^r(U_s) \right) = \begin{cases} +\infty & \text{if } S_k(U_s) > S_l(U_s), \\ 0 & \text{if } S_k(U_s) = S_l(U_s), \\ -\infty & \text{if } S_k(U_s) < S_l(U_s). \end{cases}$$

(44)

Proof

The first step is to prove that the generalized pseudo consensus function given by (32) satisfies

$$S_k(U_s) = J_{kk}(U_s) + \frac{1}{2} \sum_{u \in U_s} |h_k(u)|. \quad (45)$$

One can directly verify that the definition of $k^*_{Us}$, (31) implies that

$$\frac{1}{2} \sum_{u \in U_s} |h_k(u)| \overset{(4)}{=} -\frac{1}{2} \sum_{u \in U_s} T_{kk^*}(u)h_k(u) \overset{(19)}{=} F_{kk^*}(U_s).$$

Hence, using that $U_{kk^*} \subseteq U_s$, we find (45) from

$$S_k(U_s) \overset{(32)}{=} D_{kk^*} \overset{(30)}{=} J_{kk}(U_s) + F_{kk^*}(U_s).$$

Now using

$$\ln(2 \cosh(x)) = \ln(e^{|x|} + e^{-|x|})$$

$$= |x| + \ln(1 + e^{-2|x|})$$

$$= |x| + O(e^{-2|x|}), \quad (|x| \to \infty),$$

it follows directly from (43) and (45) that for all $k \in \mathcal{R}$ we have

$$\lim_{r \to \infty} \left( f_k^r(U_s) - rS_k(U_s) \right) = \lim_{r \to \infty} \sum_{u \in U_s} \left\{ \ln(2 \cosh(\frac{r}{2}h_k(u))) - |\frac{r}{2}h_k(u)| \right\}$$

$$= 0.$$

This combined with some standard calculus then completes the proof.

The next lemma describes the behaviour of the Markov chain on the subspace $\mathcal{R}_{k_0}(U_s)$ induced by the matrix $A^r(U_s)$.

Lemma 6 Let $k_0 \in \mathcal{R}$ be a configuration and $U_s \subseteq U$ a subset of units. Consider the subspace $\mathcal{R}_{k_0}(U_s)$ and transition matrix $P = A^r(U_s)$. Then

(i) there exists a unique stationary distribution $q_k^r(U_s)$ of the finite homogeneous Markov chain induced by $P$ on $\mathcal{R}_{k_0}(U_s)$, whose components are given by

$$q_k^r(U_s) = \begin{cases} \frac{1}{N_k^r(U_s)} \exp(f_k^r(U_s)) & k \in \mathcal{R}_{k_0}(U_s) \\ 0 & k \notin \mathcal{R}_{k_0}(U_s), \end{cases}$$

(46)
where

\[ N_{k_0}^{(\tau)}(U_s) = \sum_{m \in R_{k_0}(U_s)} \exp(f_m^{(\tau)}(U_s)), \]

and

(ii) the stationary distribution converges as \( \tau \to \infty \) to a uniform distribution over the set \( R_{k_0}(U_s) \), where \( R_{k_0}(U_s) \) denotes the set of configurations \( k \in R_{k_0}(U_s) \) for which the generalized pseudo consensus function \( S_k(U_s) \) is maximal.

Proof

(i) The proof is similar to the proof of Lemma 4. Again it is easily verified that the Markov chain induced by \( P \) is irreducible and aperiodic and hence from Property 1 there exists a unique stationary distribution, whose shape can be verified with the detailed balance equation (10). To verify this equation we first rewrite the expression for the acceptance probability given by (17). This is done following the lines of the proof of Theorem 1, with the substitutions only applied to the denominator, i.e. for \( k, l \in R, U_s \supseteq U_{kl} \) we write using (19) and (30)

\[ A_{kl}^{(\tau)}(U_s) = \frac{\exp(\tau D_{kl})}{\prod_{u \in U_s} 2 \cosh(\frac{x}{2} h_k(u))} \cdot \]

(47)

So, let \( k \in R_{k_0}(U_s) \) (otherwise the proof is trivial) then, using the fact that \( D_{kl} = D_{lk} \)

\[ N_{k_0}^{(\tau)}(U_s) q_k^{(\tau)}(U_s) A_{kl}^{(\tau)}(U_s) \]

(46)

\[ \exp(f_k^{(\tau)}(U_s)) A_{kl}^{(\tau)}(U_s) \]

(43)

\[ \exp(\tau J_{kl}(U_s)) \prod_{u \in U_s} 2 \cosh(\frac{x}{2} h_k(u)) A_{kl}^{(\tau)}(U_s) \]

(47)

\[ \exp(\tau J_{li}(U_s)) \prod_{u \in U_s} 2 \cosh(\frac{x}{2} h_i(u)) A_{lk}^{(\tau)}(U_s) \]

which completes the proof of part (i).

(ii) Using Lemma 5 we find for \( k \in R_{k_0}(U_s) \)

\[ \lim_{\tau \to \infty} q_k^{(\tau)}(U_s) = \lim_{\tau \to \infty} \sum_{m \in R_{k_0}(U_s)} \frac{\exp(f_m^{(\tau)}(U_s))}{\sum_{m' \in R_{k_0}(U_s)} \exp(f_m^{(\tau)}(U_s))} \cdot \]

(44)

\[ r_k(U_s), \]

where

\[ r_k(U_s) = \left\{ \begin{array}{ll} |R_{k_0}(U_s)|^{-1} & k \in R_{k_0}(U_s) \\ 0 & k \notin R_{k_0}(U_s) \end{array} \right. \]

(48)

and

\[ R_{k_0}(U_s) = \{ k \in R_{k_0}(U_s) \mid \forall i \in R_{k_0}(U_s) : S_i(U_s) \leq S_k(U_s) \}, \]

(49)

which completes the proof of the lemma.
We now can use Lemma 6 to obtain an alternative interpretation of the general model which is based on (12) and provide some evidence for the correctness of Conjecture 1. Instead of viewing $P(\tau)$ as a matrix that induces a single Markov chain on the configuration space $\mathcal{R}$, we can consider $P(\tau)$ as a collection of matrices $A(\tau)(\mathcal{U}_s)$, each of which induces a Markov chain on its corresponding subspace $\mathcal{R}_k(\mathcal{U}_s)$. In this perspective, we view $G$ as a selecting mechanism between the different Markov chains; the Markov chain induced by $A(\tau)(\mathcal{U}_s)$ is selected with probability $G(\mathcal{U}_s)$.

Lemma 6 states that the selected Markov chain induced by $A(\tau)(\mathcal{U}_s)$ maximizes $S_k(\mathcal{U}_s)$ over the subspace $\mathcal{R}_k(\mathcal{U}_s)$. Due to the fact that the generation may yield a different subset each time it is applied (where because of condition (16) each unit has a positive probability of being in that subset), the maximization becomes effective over the whole configuration space. Hence, $P(\tau)$ (with $\tau \to \infty$) models a process where with probability $G(\mathcal{U}_s)$ a function $S_k(\mathcal{U}_s)$ is maximized. In other words, $P(\tau)$ describes the process of maximizing $\tilde{C}_k$.

So far we have not been able to prove Conjecture 1 in a formal way, which is therefore left for further research. However, in a special case of unlimited parallelism which we define below, the proof follows directly from Lemma 6 (c.f. Little's model in [Peretto, 1984]).

**Definition 16** A synchronous parallel Boltzmann machine is said to apply full parallelism if the generation probability satisfies $G(\mathcal{U}_s) = 1 \iff \mathcal{U}_s = \mathcal{U}$.

**Theorem 3** Conjecture 1 holds for a synchronously parallel Boltzmann machine applying full parallelism.

**Proof**
In the situation of full parallelism we have $P(\tau) = A(\tau)(\mathcal{U})$ and $\tilde{C}_k = S_k(\mathcal{U})$. Hence, using $\mathcal{R}_{k_0}(\mathcal{U}) = \mathcal{R}$ for all $k_0 \in \mathcal{R}$, we find that Conjecture 1 and the second part of Lemma 6 are identical.

### 5 Implications

In this section we study the implications of Conjecture 1 for a large class of synchronously parallel Boltzmann machines, defined in the following way.

**Definition 17** A synchronously parallel Boltzmann machine is called fair if the probability that a unit is allowed to make a transition is equal for all units, i.e. there exists a constant $\alpha \in [0,1]$ such that

$$\forall u \in \mathcal{U} : \sum_{\mathcal{U}_s \subseteq \mathcal{U}} G(\mathcal{U}_s) \chi_{\mathcal{U}_s}(u) = \alpha, \quad (50)$$

where $\chi_{\mathcal{U}_s}$ denotes the characteristic function on the set $\mathcal{U}_s$.

The definition of fair Boltzmann machines formally describes a very natural condition for the generation mechanism. In practice one always deals with fair Boltzmann machines and it is in fact hard to implement an unfair scheme. The following property gives an alternative interpretation of $\alpha$. Since the proof is trivial we leave it for the reader.
Property 2 Let $B$ be a fair synchronously parallel Boltzmann machine. Then the constant $\alpha$ defined by (50) equals the expected fraction of the units that is allowed to make a transition, each time the generation mechanism is applied.

In the next theorem we show that the extended consensus function of a fair Boltzmann machine can be rewritten as a weighted sum of the consensus function and the pseudo consensus function $S_k = S_k(U)$. We also introduce a critical value $\alpha_c$ which yields an upper bound on the amount of parallelism that can be exploited in a Boltzmann machine if one wants to guarantee consensus maximization.

**Theorem 4** Let $B$ be a fair synchronously parallel Boltzmann machine. Then

(i) the extended consensus function $\tilde{C}$ defined by (33) satisfies for all $k \in \mathcal{R}$

$$\tilde{C}_k = (1 - \alpha)2C_k + \alpha S_k,$$

where $S_k = S_k(U)$ and $\alpha$ is the value defined in (50), and

(ii) if $\alpha_c$ is defined by

$$\alpha_c = \min_{k \in \mathcal{R} \setminus \mathcal{R}_{opt}} \frac{2C_{opt} - 2C_k}{S_k - 2C_k},$$

then $\tilde{R}_{opt} = R_{opt}$ if and only if $\alpha < \alpha_c$.

**Proof**

(i) Let $k \in \mathcal{R}$. Using the expression of (45) in combination with the expressions of (26) and (19), one can straightforwardly verify that there exists a not specific function $z_k(u)$ such that for all $U_s \subseteq U$

$$S_k(U_s) = 2C_k + \sum_{u \in U_s} z_k(u).$$

Now exploiting the fairness of the Boltzmann machine it follows that

$$\tilde{C}_k - 2C_k \overset{(33)}{=} -2C_k + \sum_{U_s \subseteq U_s} G(U_s)S_k(U_s)$$

$$\overset{(53)}{=} -2C_k + \sum_{U_s \subseteq U_s} G(U_s)(2C_k + \sum_{u \in U_s} z_k(u))$$

$$\overset{(15)}{=} \sum_{U_s \subseteq U_s} G(U_s) \sum_{u \in U_s} z_k(u)$$

$$= \sum_{U_s \subseteq U_s} G(U_s) \sum_{u \in U_s} z_k(u)X_{U_s}(u)$$

$$= \sum_{u \in U_s} \sum_{U_s \subseteq U_s} G(U_s)X_{U_s}(u)$$

$$\overset{(50)}{=} \sum_{u \in U_s} z_k(u)\alpha$$

$$\overset{(53)}{=} (S_k(U) - 2C_k)\alpha,$$
from which the final result is directly obtained.

(ii) Suppose \( \alpha < \alpha_c \).

It is easily verified that (42) also holds for the unlimited case, since \( k = k^*_k \) implies \( E_{kk^*u} = 0 \). Thus \( k \in \mathcal{R}_{opt} \) gives \( \tilde{C}_k = 2C_{opt} \). Now let \( k \in \mathcal{R} \setminus \mathcal{R}_{opt} \), we then show that \( \tilde{C}_k < 2C_{opt} \).

If \( S_k < 2C_{opt} \) then, using the fact that \( \alpha \leq 1 \), we obtain

\[
\tilde{C}_k = 2C_k + \alpha(S_k - 2C_k) < 2C_k + \alpha(2C_{opt} - 2C_k) \leq 2C_{opt}.
\]

If \( S_k \geq 2C_{opt} \) then, using the definition of \( \alpha_c \), we get

\[
\tilde{C}_k = 2C_k + \alpha(S_k - 2C_k) < 2C_k + \alpha_c(S_k - 2C_k) \leq 2C_{opt},
\]

Suppose \( \alpha \geq \alpha_c \).

If \( \tilde{C}_k > 2C_{opt} \) for some \( k \in \mathcal{R} \), then \( \tilde{R}_{opt} \setminus \mathcal{R}_{opt} \neq \emptyset \) which proves our assertion. Now assume \( \tilde{C}_k \leq 2C_{opt} \) for all \( k \in \mathcal{R} \).

From \( \alpha_c \leq \alpha \leq 1 \) we obtain that there exists a \( k_0 \in \mathcal{R} \setminus \mathcal{R}_{opt} \) with \( S_{k_0} \geq 2C_{opt} > 2C_{k_0} \) and \( \alpha_c = (2C_{opt} - 2C_{k_0})/(S_{k_0} - 2C_{k_0}) \). Hence,

\[
\tilde{C}_{k_0} = 2C_{k_0} + \alpha(S_{k_0} - 2C_{k_0}) \geq 2C_{k_0} + \alpha_c(S_{k_0} - 2C_{k_0}) = 2C_{opt}.
\]

This, combined with our assumption, gives again \( \tilde{R}_{opt} \setminus \mathcal{R}_{opt} \neq \emptyset \), which completes the proof.

Remark. We can rewrite the expression of (51) for the extended consensus function of a fair Boltzmann machine as

\[
\tilde{C}_k = (1 - \alpha)C_k^0 + \alpha C_k^1,
\]

where \( C_k^0 = D_{kk^*} \) and \( C_k^1 = D_{kk^*} \), obtaining an expression which is easily interpreted.

It is not hard to calculate lower and upper bounds for the critical value \( \alpha_c \). Let \( 0 \) denote the configuration with all units "off". Suppose \( 0 \notin \mathcal{R}_{opt} \) and \( S_0 \geq 2C_{opt} \), then we obtain directly from the definition of \( \alpha_c \) (52)

\[
\alpha_c \leq \frac{2C_{opt}}{S_0}.
\]

Furthermore, \( S_0 \) can be easily calculated using \( S_0 = \sum_{u \in \mathcal{U}} s^*_t(u,u) \).

To obtain an lower bound, we define \( C_{nop} = \max \{ C_k \mid k \in \mathcal{R} \setminus \mathcal{R}_{opt} \} \) and \( S_{opt} = \max \{ S_k \mid k \in \mathcal{R} \} \). Suppose \( S_{opt} \geq 2C_{opt} \), then we have for all \( k \in \mathcal{R} \setminus \mathcal{R}_{opt} \) with \( S_k \geq 2C_{opt} \)

\[
\frac{2C_{opt} - 2C_k}{S_k - 2C_k} \geq \frac{2C_{opt} - 2C_k}{S_{opt} - 2C_k} = \frac{2C_{opt} - S_{opt}}{S_{opt} - 2C_k} + 1 \geq \frac{2C_{opt} - S_{opt}}{S_{opt} - 2C_{nop}} + 1.
\]
which implies that
\[
\alpha_c \geq \frac{2C_{\text{opt}} - 2C_{\text{nop}}}{S_{\text{opt}} - 2C_{\text{nop}}}. \tag{56}
\]

5.1 Applications

In order to demonstrate how to use the results obtained so far, we apply them to some of the Boltzmann machines described in the book by Aarts & Korst [1989], designed to solve combinatorial optimization problems.

First we consider the problem of finding the largest independent set of a graph. It is straightforward to construct a Boltzmann machine that solves this problem for a given graph [Aarts & Korst, 1989]. The resulting Boltzmann machine consists of the given graph extended with the set of bias connections which obtain a connection strength +1. The connections in the original graph are given a strength -2. If we let \( V \) denote the largest independent set, one can verify that \( C_{\text{opt}} = |V| \), \( C_{\text{nop}} = |V| - 1 \) and \( S_{\text{opt}} = S_0 = |U| \).

We exclude disconnected graphs, which trivially implies that \( 2|V| \leq |U| \). Now, using (56) and (55) we obtain
\[
\frac{2}{|V| + 2 - 2|V|} \leq \alpha_c \leq \frac{2|V|}{|U|}. \tag{57}
\]

The bounds in (57) coincide if and only if \( |V| = 1 \) or \( 2|V| = |U| \).

In [Aarts & Korst, 1989] the maximum independent set problem is solved with the type of Boltzmann machine described above, for a class of random graphs constructed in the following way. A fixed set of vertices \( U \) (≡ units) is taken (\(|U| = 50, 100, 150, 200, 250\)) and the edges are chosen randomly with a probability \( \frac{10}{|U|-1} \) of being present. This implies that the expected degree equals 10 for all vertices. It turned out that for these kind of graphs the maximum independent set was equal to \( |V| \approx \frac{3}{10}|U| \), yielding \( \alpha_c \leq \frac{6}{10} \).

The second problem we consider is the well-known max cut problem. Given a graph with positive weights on the edges, one wants to find a partition of the vertices into two subsets, such that the sum of the weights of the edges with endpoints in different subsets is maximal. The construction of the Boltzmann machine is identical to the one for the first problem. The connection strength of non-bias connections is taken minus two times the weight of the edge it corresponds to, whereas the bias of a unit will equal the total weight of all the edges incident with this vertex (see [Aarts & Korst, 1989] for a detailed description).

The above construction leads to \( C_{\text{opt}} = \text{max-cut} \), the value corresponding to the optimal cut, and \( S_0 = 2W \), with \( W \) denoting the sum of all the weights in the graph. The same random graphs as in the first problem were used, with edges given an integer weight randomly chosen in \( \{1, \ldots, 10\} \). For these kind of graphs the expected number of edges is \( 5|U| \), and hence the expected value of \( S_0 \) is \( 2 \cdot \frac{11}{2} \cdot 5|U| = 55|U| \). The resulting value of \( C_{\text{opt}} \) turned out to be approximately \( 20|U| \), which gives \( \alpha_c \leq \frac{18}{11} \).
Numerical experiments with simulations of Boltzmann machines for both problems revealed that $\alpha$ could be chosen smaller then or equal to $\frac{2}{3}$, a value which is comparable to both estimated upper bounds. Values larger then this bound yielded very poor convergence.

6 Conclusions

In this paper a mathematical model is presented for the description of synchronously parallel Boltzmann machines. The main result is a conjecture about the general behaviour of this type of Boltzmann machines. It states that synchronously parallel Boltzmann machines maximize an extended consensus function, which — for all "standard" choices of the generation mechanism — consists of a weighted sum of two functions. The first function denotes the function that is maximized by the Boltzmann machine if it operates sequentially (the consensus function); the second one denotes the function that is maximized if the Boltzmann machine operates fully parallel (the pseudo consensus function).

The weighting in the extended consensus function is determined by a single parameter, denoting the average fraction of units in the Boltzmann machine that is operating. If one wants to guarantee that a given synchronously parallel Boltzmann machine maximizes the genuine consensus function, the amount of parallelism which can be exploited effectively is bounded by an upperlimit. This upper bound can be easily estimated for a given problem and compares fairly with values found in practice.

Bibliography