



Doctoral Thesis

On principles of cortical computation

Author(s):

Krautz, Christoph

Publication Date:

2012

Permanent Link:

<https://doi.org/10.3929/ethz-a-007348196> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH No. 20207

On Principles of Cortical Computation

A dissertation submitted to
ETH ZURICH

for the degree of
DOCTOR OF SCIENCES

presented by
CHRISTOPH KRAUTZ
Dipl. Inf. Univ.
born 27.04.1980
citizen of Germany

accepted on the recommendation of
Prof. Dr. Angelika Steger, examiner
Prof. Dr. Rodney Douglas, co-examiner
Dr. Matthew Cook, co-examiner

2012

Abstract

In many areas, the brain outperforms the technologies that mankind has invented so far. For several decades, researchers have attempted to build computational systems that are inspired by the brain. However, all approaches have failed to create systems that would be comparable to the capabilities of the brain. Consequently, we still lack understanding of the fundamental principles of brain function. We still ask ourselves the same questions as the early pioneers of neuroscience did – how can this network of billions of nerve cells carry out perception, reasoning, or consciousness?

In this thesis, we focus our studies on four fundamental abilities of the brain that are crucial for building models of cortical computation: *(i)* coherent interpretation of information, *(ii)* consistent propagation of information, *(iii)* input normalization, and *(iv)* building a reliable system from unreliable neural components.

In the first part we study the coherent interpretation of sensory information. We explore computation that employs distributed representations of information and mutual influence between the representations. For this, we design a network to interpret input from a neuromorphic sensor by means of recurrently interconnected areas, each of which encodes a different aspect of the visual interpretation, such as light intensity or optic flow. As each area of the network tries to be consistent with the information in neighboring areas, the visual interpretation converges towards global mutual consistency. This work demonstrates how a biologically inspired recurrent architecture of mutually influencing modules can be designed to coherently interpret noisy or ambiguous data, even when there is no feed-forward method to generate the desired interpretation.

The second part is concerned with the consistent propagation of information in recurrent architectures. It is important that correct information about the environment is not distorted or turned into contradicting information, while reverberating throughout a recurrent network. This can be ensured by using a precisely aligned connectivity within these networks. Anatomical studies have shown that cortical neurons are able to create reciprocal topographical connections between areas. However, the morphology observed in experiments limits the precision of such anatomical reciprocity. To this day, the question of how precisely aligned reciprocal connections are obtained in the brain has been left open. We have found that reciprocal connectivity is obtainable through the adjustment of synaptic strengths. We present a model that employs a combination of biologically realistic learning mechanisms, which we refer to collectively as sharp learning. The mechanisms that we use to connect and train our networks are Hebbian learning at synapses, continuous winner-take-all circuitry within areas, and homeostatic activity regulation within single neurons. Our experiments show that by means of sharp learning it is possible to sharpen inter-area projections in a variety of network architectures.

In the third part, we discuss how the brain is able to cope with highly dynamical input. In neural systems, the reliable transmission of information can only be guaranteed if each area is able to process strongly varying amounts of input, without exceeding the limits of the processing units. Hence, we investigate how networks of spiking neural units can be connected so that they increase the dynamic range of a whole population of these units. We model a connectivity pattern known as feed-forward inhibition and explore its dynamics. Our results show that there exist multiple parameter regimes, in which reliable transmission of information is possible.

In the last part of this thesis we investigate how the brain is able to work reliably even though it is built from very unreliable components. Neurons in the brain exhibit a high variability in properties even if they belong to the same sub-type. Thus, we replace the reliable units in the networks of the third part by unreliable units. We show that it is possible to build reliable systems which consist of unreliable components.

Zusammenfassung

In vielen Bereichen ist das Gehirn den Technologien überlegen, welche die Menschheit bis heute erfunden hat. Seit mehreren Jahrzehnten versuchen Wissenschaftler, vom Gehirn inspirierte Rechensysteme zu bauen. Jedoch hat keiner der bisherigen Ansätze Systeme hervorgebracht, die mit den Fähigkeiten des Gehirns vergleichbar sind. Das heißt, uns fehlt noch immer das Verständnis der grundlegenden Prinzipien der Funktionsweise des Gehirns. Wir stellen uns die gleichen Fragen wie die frühen Pioniere der Neurowissenschaften – wie kann dieses Netzwerk aus Milliarden von Nervenzellen Sinneseindrücke wahrnehmen, aus Informationen Schlussfolgerungen ziehen oder gar ein Bewusstsein entwickeln?

In dieser Arbeit konzentrieren wir unsere Untersuchungen auf vier grundlegende Fähigkeiten des Gehirns, die entscheidend für den Bau von Modellen kortikaler Berechnung sind: auf *(i)* die kohärente Interpretation von Informationen, *(ii)* die konsistente Verbreitung von Informationen, *(iii)* die Normalisierung von Eingangssignalen und *(iv)* den Aufbau eines zuverlässigen Systems aus unzuverlässigen neuronalen Komponenten.

Im ersten Teil untersuchen wir die kohärente Interpretation von sensorischen Informationen. Wir erforschen Berechnungen, welche auf einer verteilten Repräsentation von Informationen und der wechselseitigen Beeinflussung dieser Informationen aufbauen. Dazu entwickeln wir ein Netzwerk, welches Signale eines neuromorphen Bildsensors interpretieren kann. Das Netzwerk besteht aus rekurrent verbundenen Bereichen, von denen jeder einen anderen Aspekt der visuellen Interpretation, wie zum Beispiel die Lichtintensität oder den optischen Fluss, kodiert. Jeder Bereich des Netzwerks versucht, im Einklang mit den Informationen in benachbarten Bereichen zu stehen. Daher konvergiert das Netzwerk zu einer global konsistenten Interpretation des Eingangssignals. Der erste Teil dieser Arbeit demonstriert den Aufbau eines biologisch inspirierten rekurrenten Systems sich gegenseitig beeinflussender Module, welches

mehrdeutige Daten kohärent interpretieren kann. Die Fähigkeit zur Interpretation ist dabei unabhängig von der Existenz eines direkten (feed-forward) Verfahrens zur Berechnung der gesuchten Interpretation.

Der zweite Teil beschäftigt sich mit der konsistenten Verbreitung von Informationen in rekurrenten Architekturen. Es ist wichtig, dass korrekte Informationen über die Umwelt nicht in widersprüchliche Informationen transformiert werden, während sie durch ein rekurrentes Netzwerk pulsieren. Dies kann durch eine exakt ausgerichtete Konnektivität innerhalb solcher Netzwerke gewährleistet werden. Anatomische Studien haben gezeigt, dass kortikale Neuronen in der Lage sind, reziproke topographische Verbindungen zwischen den verschiedenen Arealen des Kortex zu bilden. Allerdings wurde in Experimenten auch beobachtet, dass die Morphologie die Genauigkeit einer solchen anatomischen Reziprozität begrenzt. Bis heute ist es ungeklärt, wie exakt ausgerichtete reziproke Verbindungen im Gehirn geformt werden. Wir zeigen im zweiten Teil dieser Arbeit, dass sich solche reziproken Verbindungen durch eine Anpassung der synaptischen Verbindungsstärke herausbilden können. Dazu entwickeln wir ein Modell, das eine Kombination aus biologisch realistischen Lernmechanismen verwendet, die wir zusammen als Sharp Learning bezeichnen. Wir verwenden folgende Mechanismen, um unsere Netzwerke zu verbinden und zu trainieren: Hebb'sches Lernen an den Synapsen, kontinuierliche Winner-Take-All-Schaltungen innerhalb einzelner Netzwerkbereiche und homöostatische Regelung der Aktivität einzelner Neuronen. Unsere Experimente zeigen, dass es mittels Sharp Learning möglich ist, die Genauigkeit interarealer Projektionen in einer Vielzahl verschiedener Netzwerkarchitekturen drastisch zu erhöhen.

Im dritten Teil diskutieren wir, wie das Gehirn den großen Umfang an verschieden starken Eingangssignalen bewältigt. In neuronalen Systemen, kann eine zuverlässige Übertragung von Informationen nur dann gewährleistet werden, wenn jeder Bereich in der Lage ist, sowohl sehr schwache als auch sehr starke Eingabestärken zu verarbeiten, ohne dabei an die Leistungsgrenze der verarbeitenden Einheiten zu gelangen. Daher untersuchen wir, wie Netzwerke von feuernenden neuronalen Einheiten verbunden werden können, so dass der dynamische Bereich einer Population solcher Einheiten größer ist als der einer einzelnen Einheit. Wir modellieren ein Konnektivitätsmuster, das in der Literatur als Feed-Forward-Inhibition bekannt ist, und erforschen dessen Dynamik. Unsere Ergebnisse zeigen, dass es unterschiedliche Konfigurationen der

Netzwerkparameter gibt, in denen eine zuverlässige Übertragung von Informationen möglich ist.

Im letzten Teil dieser Arbeit untersuchen wir die Fähigkeit des Gehirns, zuverlässig zu arbeiten, obwohl es aus sehr unzuverlässigen Komponenten besteht. Neuronen im Gehirn weisen eine hohe Variabilität ihrer Eigenschaften auf, auch wenn sie zur gleichen Neuronenklasse gehören. Daher ersetzen wir die zuverlässigen Einheiten, die wir im dritten Teil dieser Arbeit zum Aufbau der Netzwerke verwendet haben, durch unzuverlässige Einheiten. Wir zeigen, dass es möglich ist, zuverlässige Systeme zu bauen, die aus unzuverlässigen Komponenten bestehen.