Input Prediction and Autonomous Movement Analysis in Recurrent Circuits of Spiking Neurons[1]

Robert Legenstein, Wolfgang Maass[2], and Henry Markram

Robert Legenstein, Wolfgang Maass
Institute for Theoretical Computer Science
Technische Universitaet Graz
Inffeldgasse 16b, A-8010 Graz, Austria
E-mail: {legi, maass}@igi.tu-graz.ac.at
http://www.igi.tugraz.at/legi/
http://www.igi.tugraz.at/maass/

Henry Markram
Brain Mind Institute
Ecole Polytechnique Federale De Lausanne
CE-Ecublens
CH- 1015 Lausanne
E-mail: Henry.Markram@weizmann.ac.il
http://www.weizmann.ac.il/neurobiology/labs/markram/markram.html

[1] The work was partially supported by the Austrian Science Fond FWF, project # P15386, and the NeuroCOLT project of the EU.
[2] Corresponding author.

Abstract
We show that a generic recurrently connected circuit of integrate-and-fire neurons can be trained in an unsupervised manner to predict movements of different objects, that move within an unlimited number of combinations of speed, angle, and offset over a simulated sensor field. The autonomously trained neural circuit is also able to compute the direction of motion, which is a computationally difficult problem since it requires disambiguation of local sensor readings through the context of other sensor readings at the current and preceding moments. Finally it is shown that the trained neural circuit supports novelty detection and the generation of "imagined movements".

1. Introduction

Biological organisms are bombarded with a continuous stream of sensory input that delivers millions of bits per second. In order to survive, the organism needs to extract from this stream in real-time the most salient bits for higher-level processing and decision-making. For fast responses an organism also has to take into account transmission delays and the inertia of its actuators, and therefore needs to base its motor commands on the anticipated next state of the environment and itself, rather than on the current one. Incidentally, all these problems also arise in the control of mobile robots, and are largely unsolved. It has long been conjectured that unsupervised learning plays an important role for the way in which many organisms solve these problems. But it has remained difficult to explain how this could be achieved for realistically complex and diverse input streams, and how it could be implemented in generic neural microcircuits of the neocortex.

We present in this article results from computer simulations of biologically quite realistic recurrent circuits of integrate-and-fire neurons, which are presented with a continuous stream of inputs from 64 simulated sensors that are caused by movements of different objects at an unlimited number of different velocities and angles over a quadratic field covered by the sensors. This field could represent for example a map for a 2-dimensional visual, auditory or somatosensory field (e.g. barrel cortex in rats). Our strategy for modeling computation in recurrent circuits of integrate-and-fire neurons is based on the general scheme of the liquid state machine [Maass et al., 2001]. According to that scheme the recurrent neural circuit (the "liquid") is viewed as an unbiased "universal" analog memory device into which the input stream is fed. An unlimited number of readout neurons can extract and recombine from the current firing activity of this recurrent circuit (its current "liquid state") the information that is required for the solution of specific tasks. Whereas the efficacies of synapses within the recurrent circuit are left unchanged, thereby allowing that the same recurrent circuit is used simultaneously for completely different analyses of the same input stream, a simple unsupervised local learning algorithm is used for adapting the weights of neurons that read out information from the recurrent circuit. Different readout neurons are trained in this unsupervised manner to predict the next inputs for each of the 64 simulated sensors. In fact, for each sensor one can train several pools of readout neurons to predict future sensor activity for different time intervals into the future. No supervisor is needed for this training, since the environment itself provides the correct solution for each prediction task a moment later. Other readout neurons are trained in an unsupervised manner to "jump ahead" in this sequence prediction task by a context-dependent time span, and to predict where the moving object is going to leave the sensory field. These neurons obviously have to compute the *direction of movement* of an object, a well-known and demanding computational problem that requires integration of information from many local sensors and time slices due to the "aperture problem" (see [Mead, 1989]). This problem, which also is a key problem in computer vision,

arises from the fact that for extended (i.e., not point-like) moving objects one cannot infer the direction of its movement from the direction in which the edge of the object moves over a local receptive field (due to its limited "aperture"). Any solution of this problem requires substantial context-dependent integration of information over time and space. Our computer simulations demonstrate that this task can be solved through a simple unsupervised learning process, that can run all the time autonomously during the lifetime of a behaving organism or robot.

Through online prediction of sensory input an automatic preprocessing of the high-volume sensory input stream is accomplished, that makes it possible to surpress during normal sensory processing most of the incoming bits. Reports from sensors which had already been autonomously predicted by the system need not be transmitted to higher-level areas, since they provide no new information to the organism. This allows the higher-level areas to focus on the most far-reaching predictions and abstractions, such as the direction and velocity of movement (that have been extracted by lower level neural microcircuits from the raw input stream), and to use this sparse and temporally more stable information for higher level processing. Our simulations show that the same autonomously trained lower level circuits may also provide optimal support for higher-level areas if an unusual event takes place. If at any moment the movement of the object deviates from its predicted "usual" trajectory, an instant analysis of this unexpected event can take place, without any loss of information from sensors, since at this moment the system can make use of the full high bit rate of the incoming sensory input stream, and evaluate the complete pattern of deviations between actual and predicted input at all local sensors.

Since the recurrent "liquid" neural circuit has not been altered during this autonomous training of readout neurons, it can at the same time serve for completely different information processing tasks. To illustrate this point, we have simultaneously trained other readout neurons to extract from the same recurrent neural circuit information about the identity of the moving object. This simultaneous object classification can be carried out with high precision, thereby showing that it is no mystery how a generic neural microcircuit in a sensory cortical area can support completely different analyses of the sensory input (for example a "what stream" and a "where stream").

In view of the controversies about the actual biological implementation of synaptic plasticity, and in order to focus on general computational principles rather than implementation details, we have used in this article simple Hebbian local learning rules that appear to be in principle biologically plausible, without making assumptions about their precise biological implementation. But one could use here just as well the temporally asymmetric spike time dependent plasticity rule explored in [Rao and Sejnowski, 2000] in order to train the weights of the readout neurons (which should in this case receive additional synaptic input from those sensors which they are trained to predict). On a more abstract level, in the context of sigmoidal rather than spiking neurons, a corresponding temporally asymmetric learning rule had previously been explored in [Abbott and Blum, 1996] for sequence prediction tasks. The main difference between the computational models analyzed in [Abbott and Blum, 1996] and [Rao and Sejnowski, 2000], and the one investigated in this article, is that those earlier studies focused on neural circuits that had essentially "no hidden neurons", but rather consisted of arrays of neurons where each excitatory neuron in the network was directly targeted by exactly one "sensory" input, with excitatory or inhibitory (mediated by point-to-point interneurons) lateral connections between these directly input-driven neurons. Consequently these articles focused on predictions of input movements of pointwise objects whose direction

and speed was aligned with the specific pattern of connections and transmission delays of the lateral connections between the array of input-driven neurons. In contrast, we consider in this article generic randomly connected networks of integrate-and-fire neurons into which the input is mapped by random connections that favor a topographic map. Such circuit might for example represent a neural microcircuit in primary sensory cortex. We demonstrate that the computational advantage of this biologically more realistic architecture lies in its capability to support analyses of a much larger diversity of shapes, directions and speeds of moving objects that could possibly be anticipated in the pattern of connections and transmission delays of a circuit that has been explicitly constructed to support this family of tasks. For example, we show that the current activity of those 70 % "hidden" neurons in the randomly connected recurrent circuit that are not directly driven by sensory input, provides sufficient context information to disambiguate hundreds of different movements that all excite a very similar subset of sensors when they cross the mid-area of the sensory field. Since we show that it is not required that the architecture of a neural microcircuit is specialized for the specific spatial and temporal input patterns that it usually processes, we also provide a possible explanation for the large amount of common anatomical structure among neural microcircuits in different sensory areas, and also among microcircuits in sensory and higher cortical areas.

The possible role of hidden neurons in randomly connected recurrent neural networks for sequence prediction and disambiguation had previously already been investigated by Levy and his collaborators in the context of computational models for hippocampus (see [Levy, 1996] for a survey). Their recurrent neural networks consisted of McCulloch-Pitts neurons (with shunting inhibition), rather than integrate-and-fire neurons. It is not clear whether their results could also be achieved for biologically more realistic circuit models with integrate-and-fire neurons and dynamic synapses. The learning approach which they explored is complementary to the one considered in this article: they adapted weights of synaptic connections *within* the recurrent network, rather than just the synapses of neurons that read out information for specific tasks. As a consequence, they were only able to predict and disambiguate a set of input movements that is substantially smaller and simpler than the one considered in this article, with a number of neurons in the recurrent circuit that is comparable to ours. The idea to train just the neurons that read out information from a recurrent neural circuit had apparently first been proposed by [Buonomano and Merzenich, 1995]. The computational power of the resulting model has previously been analyzed in [Maass et al., 2001] and [Maass and Markram, 2002]. It was independently investigated for artificial neural networks by [Jaeger, 2001].

2. **Methods**

We employed a generic randomly drawn recurrent circuit consisting of 768 integrate-and-fire neurons, which in the terminology of [Maass et al., 2001] represented the "liquid" of the liquid state machine. 20% of these neurons were randomly chosen to be inhibitory. Connection probabilities and other parameters were chosen to reflect empirical data from neurobiology. Neuron parameters were chosen as in [Tsodyks et al., 2000]: membrane time constant 30ms, absolute refractory period 3ms (excitatory neurons), 2ms (inhibitory neurons). The threshold of each neuron was chosen to be -55 mV, the resting membrane potential and the reset voltage had a value of -56.5 mV. A value of 1 MΩ was chosen for the input resistance. In addition, each neuron received a background current of 13.5 nA. The probability of a synaptic connection from neuron a to neuron b (as well as that of a synaptic connection from neuron b to neuron a) was defined as $C \cdot e^{-(D(a,b)/l)^2}$, where l is a parameter which controls both the average number of connections and the average distance between neurons

that are synaptically connected. We assumed that the 768 neurons were located on the integer points of a $16 \times 16 \times 3$ cube in space, where $D(a,b)$ is the Euclidean distance between neurons *a* and *b*. Depending on whether a and b were excitatory (E) or inhibitory (I), the value of C was 0.4 (EE), 0.2 (EI), 0.5 (IE), 0.1 (II). In the case of a synaptic connection from a to b we modeled the synaptic dynamics according to the model proposed in [Markram et al., 1998], with the synaptic parameters U (use), D (time constant for depression), F (time constant for facilitation) randomly chosen from Gaussian distributions that were based on empirical data reported for such connections in [Gupta et al., 2000] and [Markram et al., 1998]. Depending on whether a,b were excitatory (E) or inhibitory (I), the mean values of these three parameters (with D,F expressed in second, s) were chosen to be 0.5, 1.1, 0.05 (EE), 0.05, 0.125, 1.2 (EI), 0.25, 0.7, 0.02 (IE), 0.32, 0.144, 0.06 (II). The maximal amplitude A (in nA) for postsynaptic currents was chosen to be 30 (EE), 60 (EI), -19 (IE), -19 (II). The SD of each synaptic parameter was chosen to be 50 % of its mean (with negative values replaced by values chosen from an appropriate uniform distribution). Synaptic connections from sensors (inputs) to neurons in the liquid were static, with their amplitude randomly drawn from a Gaussian distribution, the mean for their maximal amplitude *A* had a value of 3 nA, and the SD was 60% of the mean. The temporal evolution of postsynaptic currents was modeled by an exponential decay exp(-t/τs) with τs=3ms (τs=6ms) for excitatory (inhibitory) synapses. The transmission delays between liquid neurons were chosen uniformly to be 1.5 ms (EE), 0.7 ms (EI), and 0.8 for the other connections.

Input to this recurrent circuit was provided by 64 simulated sensors that were arranged in an 8×8 2D array. The receptive field of each sensor was modeled as a square of unit size. The sensor output (with range [0, 1]), sampled every 5 ms, reflected at any moment the fraction of the corresponding unit square that was currently covered by a simulated moving object. The outputs of the 64 sensors were projected as input to the liquid in a topographic manner: the $16 \times 16 \times 3$ neuronal sheet of the liquid was divided into 64 $2 \times 2 \times 3$ *input regions*, and each sensor from the 8×8 sensor array projected to one such input region in a topographic manner, i.e., neighboring sensors projected onto neighboring input regions. Each sensor output was injected into a randomly chosen subset of the neurons in the corresponding input region (selection probability 0.6) in the form of additional input current (added to their background input current). One could just as well provide this input in the form of Poisson spike trains with a corresponding time-varying firing rate, with a slight loss in performance of the system.
Neural readouts from the randomly connected recurrent circuit of integrate-and-fire neurons (the liquid) were simulated as in [Maass et al., 2001] by pools of 50 neurons (without lateral connections) that received postsynaptic currents from all neurons in the liquid, caused by their firing. For simplicity we assumed that each readout neuron outputs a value of 1 at time t if its membrane potential is above its threshold at that time t, and otherwise a value of 0. A piecewise linear squashing function applied to the fraction of these 50 perceptrons that currently output a 1 was interpreted as a time-varying analog output of that readout pool with values in [0, 1] (sampled every 25 ms). This corresponds to a space rate code if the perceptrons are replaced by integrate-and-fire neurons, which can be done with a 5-10% performance loss of the system. The synapses of the neurons in each readout pool were adapted according to the p-delta rule of [Auer et al., 2001]. But in contrast to [Maass et al., 2001], this learning rule was used here in an unsupervised mode, where target output values provided by a supervisor were replaced by the actual later activations of the sensors which they predicted (with prediction spans of 25 and 50 ms into the future). Other readouts were trained in the same unsupervised manner whether a sensor on the perimeter was going to be activated by more than 50 % when the moving object finally left the sensor field. These

neurons needed to predict farer into the future (100 – 150 ms, depending on the speed of the moving object, since they were trained to produce their prediction while the object was still in the mid-area of the sensor field). The latter readouts only needed to predict a binary variable, and therefore the corresponding readout pools could be replaced by a single perceptron (or a single integrate-and-fire neuron), at a cost of about 5 % in prediction accuracy (see [Haeusler et al., 2002] for related empirical data). Altogether there were 102 readout pools that simultaneously received their input from the same recurrent circuit consisting of 768 integrate-and-fire neurons. 36 of them were trained to predict subsequent sensor activation 25 ms later in the interior 6×6 subarray of the 8×8 sensor array, 36 other ones were trained for a 50 ms prediction of the same sensors (note that prediction for those sensors on the perimeter where the object enters the field is impossible, hence we have not tried to predict all 64 sensors). 28 readout pools were trained to predict which sensors on the perimeter of the 8×8 array were later going to be activated when the moving object left the sensor field. All these 100 readout pools were trained in an unsupervised manner through a large number of movements of two different objects, a ball and a bar, over the sensor field. In order to examine the claim of [Maass et al., 2001] that other readout pools could be trained simultaneously for completely different tasks, we trained one further readout pool in a supervised manner by the p-delta rule to classify the object that moved (ball or bar), and another readout pool to estimate the speed of the moving object.

3. **Results**

The general setup of the prediction task is illustrated in Figure 1. Moving objects, a ball or a bar, are presented to an 8×8 array of sensors (panel a), where the current activation of each sensor is proportional to the percentage of the corresponding unit square (the "receptive field" of that sensor) that is currently covered by the object. The time course of activations of 8 randomly selected sensors, resulting from a typical movement of the ball, is shown in panel b. Corresponding functions of time, but for all 64 sensors, are projected as 64 dimensional input by a topographic map into a generic recurrent circuit of spiking neurons (see Section 2). The resulting firing activity of all 768 integrate-and-fire neurons in the recurrent circuit is shown in panel c (firing times of inhibitory neurons are indicated as +). Panel d of Figure 1 shows the target output for 8 of the 102 readout pools. These 8 readout pools have the task to predict the output that the 8 sensors shown in panel b will produce 25 ms later. Hence their target output (dashed line) is formally the same function as shown in panel b, but shifted by 50 ms to the left. The solid lines in panel d show the actual output of the corresponding readout pools after unsupervised learning. Thus in each row of panel d the difference between the dashed and predicted line is the prediction error of the corresponding readout pool.

The diversity of object movements that are presented to the 64 sensors is indicated in Figure 2. Any straight line that crosses the marked horizontal or vertical line segments of length 4 in the middle of the 8x8 field may occur as trajectory for the center of an object. Training examples are drawn randomly from this – in principle infinite – set of trajectories, each with a movement speed that was drawn independently from a uniform distribution over the interval from 30 to 50 units per second (unit = side length of a unit square). Shown in Figure 2 are 20 trajectories that were randomly drawn from this distribution. Any such movement is carried out by an independently drawn object type (ball or bar), where bars were assumed to be oriented vertically to their direction of movement.

36 readout pools were trained to predict for any such object movement the sensor activations of the 6×6 sensors in the interior of the 8×8 array 25 ms into the future. Further 36 readout

pools were independently trained to predict their activation 50 ms into the future, showing that the prediction span can basically be chosen arbitrarily. One could just as well also train corresponding predictions for the 28 sensors on the perimeter of the 8×8 array, but for these sensors we investigated instead a more difficult and far reaching prediction task: a prediction for the moment when the object leaves the sensory field, hence with a varying prediction span ΔT that depends on the speed of the moving object, and which is substantially larger (100 to 150 ms) (see Figure 4 and discussion below). At any time t (sampled every 25 ms from 0 to 400 ms) one uses for each of the 72 readout pools that predict sensory input ΔT into the future the actual activation of the corresponding sensor at time $t + \Delta T$ as target value ("correction") for the learning rule. An illustration is provided in Figure 3. The current sensor activation at time $t = 200\,\text{ms}$ during one particular movement is shown in panel a. The prediction targets for 72 readout pools are shown in panel d. 36 of them are trained to predict the sensor activations of the inner 6×6 array of sensors $\Delta T = 25\,\text{ms}$ later at time t = 225 ms (shown on the l.h.s. of panel d), and 36 their activations $\Delta T = 50\,\text{ms}$ later at time $t = 250\,\text{ms}$ (shown on the r.h.s. of panel d). The normalized synaptic input to the neurons in these 72 readout pools (assuming all their synaptic weights had value 1) at time $t = 200\,\text{ms}$ is indicated by the horizontal bars on the right hand side of panel b. Each line indicates for this particular time t the contribution of one of the neurons in the recurrent circuit to the membrane potential of the readout neurons (modeled with a membrane time constant of 30 ms), with contributions of inhibitory neurons shown as bars going to the left. Technically, these bars simply represent the current output of a low pass filter that is applied to the spike train of the corresponding neuron shown on the same row in the raster plot to the left, with the most relevant time interval (due to the 30 ms membrane time constant of the simulated readout neurons) from $t - 30$ to t marked at the end. The array of these 768 bars represents the *liquid state* of the recurrent circuit at time $t = 200\,\text{ms}$ in the terminology of [Maass et al., 2001], and this is the common input at time t to all neurons in the 102 readout pools. Since after training the synapses to the neurons in the readout pools have all different weights, the actual contribution of the neurons in the "liquid" to the membrane potential of the readout neurons at time t is after training for each readout neuron a different weighted sum of the values indicated on the right hand side in panel b. The 72 readout pools for short-term movement prediction were trained by 1500 randomly drawn examples of object movements. More precisely, they were trained to predict future sensor activation at any time (sampled every 25 ms) during the 400 ms time interval while the object (ball or bar) moved over the sensory field, each with another trajectory and speed.

Panel c of Figure 3 shows after training the output of the 72 readout pools at time $t = 200\,\text{ms}$, where 36 readout pools were trained for the $\Delta T = 25\,\text{ms}$ prediction, and 36 for the $\Delta T = 50\,\text{ms}$ prediction. Perfect prediction would yield a perfect match to the corresponding sensor activation plots in panel d. Among the predictions of the 72 different readout pools on 300 novel test inputs there were for the 25 ms prediction 8.5 % false alarms (sensory activity erroneously predicted) and 4.8 % missed predictions of subsequent sensor activity. For those cases where a readout pool correctly predicted that a sensor will become active, the mean of the time period of its activation was predicted with an average error of 10.1 ms. For the 50 ms prediction there were for 300 novel test inputs 16.5 % false alarms, 4.6 % missed predictions of sensory activations, and an average 14.5 ms error in the prediction of the mean of the time interval of sensory activity.

One should keep in mind that movement prediction is actually a computationally quite difficult task, especially for a moving ball, that requires context-dependent integration of

information from past inputs over time and space. This computational problem is often referred to as the "aperture problem" [Mead, 1989]: from the perspective of a single sensor (or a small group of sensors) that is currently becoming partially activated because the moving ball is covering part of its associated unit square (i.e., its "receptive field") it is impossible to predict whether this sensor will become more or less activated at the next movement. In order to decide that question, one has to know whether the center of the ball is moving towards its receptive field, or is just passing it tangentially. To predict whether a sensor that is currently not even activated will be activated 25 or 50 ms later, poses an even more difficult problem that requires not only information about the direction of the moving object, but also about its speed and shape. Since there exists in this experiment no preprocessor that extracts these features, which are vital for a successful prediction, each readout pool that carries out predictions for a particular sensor has to extract on its own these relevant pieces of information from the raw and unfiltered information about the recent history of sensor activities, which are still reverberating in the recurrent circuit (the "liquid") and leave a trace in its current "liquid state".

28 further readout pools were trained in a similar unsupervised manner (with 1000 training examples) to predict *where* the moving object is going to leave the sensor field. More precisely, they predict which of the 28 sensors on the perimeter are going to be activated by more than 50 % when the moving object leaves the 8×8 sensor field. This requires a prediction for a context-dependent time span into the future that varies by 66 % between instances of the task, due to the varying speeds of moving objects. We arranged that this prediction had to be made while the object crossed the central region of the 8×8 field, hence at a time when the current position of the moving object provided hardly any information about the location where it will leave the field, since all movements go through the mid area of the field. Therefore the tasks of these 28 readout neurons require the computation of the direction of movement of the object, and hence a computationally difficult disambiguation of the current sensory input. We refer to the discussion of the disambiguation problem of sequence prediction in [Levy, 1996] and [Abbott and Blum, 1996]. The latter article discusses difficulties of disambiguation of movement prediction that arise already if one has just pointwise objects moving at a fixed speed, and just 2 of their possible trajectories cross. Obviously the disambiguation problem is substantially more severe in our case, since a virtually unlimited number of trajectories of different extended objects (see Figure 2), moving at different speeds, crosses in the mid area of the sensor field. The disambiguation is provided in our case simply through the "context" established inside the recurrent circuit through the traces (or "reverberations") left by preceding sensor activations. Figure 4 shows in panel a a typical current position of the moving ball, as well as the sensors on the perimeter that are going to be active by ≥ 50 % when the object will finally leave the sensory field. In panel b the predictions of the corresponding 28 readout neurons (at the time when the object crosses the mid-area of the sensory field) is also indicated (striped squares). The prediction performance of these 28 readout neurons was evaluated as follows. We considered for each movement the line from that point on the opposite part of the perimeter, where the center of the ball had entered the sensory field, to the midpoint of the group of those sensors on the perimeter that were activated when the ball left the sensory field (dashed line). We compared this line with the line that started at the same point, but went to the midpoint of those sensor positions which were predicted by the 28 readout neurons to be activated when the ball left the sensory field (solid line). The angle between these two lines had an average value of 4.9

degrees for 100 randomly drawn novel test movements of the ball (each with an independently drawn trajectory and speed)[3].

The readout pool that was independently trained in a supervised manner to classify the moving object (ball or bar) had an error of 0 % on 300 test examples of moving objects. The other readout pool that was trained in a supervised manner to estimate the speed of the moving bars and balls, which ranged from 30 to 50 units per second, made an average error of 1.48 units per second on 300 test examples. This shows that the same recurrent circuit that provides the input for the movement prediction can be used simultaneously by a basically unlimited number of other readouts, that are trained to extract completely different, or more abstract, information from the current "liquid state" of the recurrent circuit.

We also evaluated the 72 readout pools that had been trained in an unsupervised fashion (see Figure 1-3) to predict future sensor activity for completely different tasks, that were unrelated to those shown during training. We first considered the response to object movements when the object changed its direction of movement while crossing the sensor field. Figure 5 shows the prediction error of the 36 neurons that predict sensor activation $\Delta T = 50$ ms later when the object changes its direction of movement at time t = 250 ms. The weights of these readout neurons resulted from a training procedure described before (see Figure 2), where only linear movement paths were shown. Panel b indicates that at time t = 300 ms, thus 50 ms after the object has changed its direction of movement, the prediction error is very high, much higher than during any straight movement. Hence another neural circuit that compares in an online manner predictions of sensor activation with their actual later activation can easily discern that something unusual has happened. Furthermore the spatial distribution of prediction errors (r.h.s. of panel b in Figure 5) provides the most relevant input for analyzing this event. Note that the time point t = 250 ms where the moving object changed its direction had been picked arbitrarily in this experiment. Hence any-time novelty detection and analysis is supported by our neural circuit that had been trained in an unsupervised manner, although it had never been trained for novelty detection.

Finally, we show in Figure 6 what happens if from some arbitrarily chosen time point on (here $t = 125$ ms) the sensor input to the recurrent circuit is removed, and replaced by predictions of future inputs by the readout pools. More precisely, the time series of inputs (sampled every 5 ms) was replaced for each sensor after $t = 125$ ms by the preceding prediction of the corresponding readout pool (that had been trained for this prediction in an unsupervised manner as described before). Hence further predictions after time $t = 125$ ms are made based on an increasing portion of imagined rather than real inputs to the recurrent circuit. We found that such imagined continuations of movements require very small errors for short term predictions. Therefore the result shown in Figure 6 is from an experiment where the readouts were only trained to predict upwards movements for $\Delta T = 25$ ms, resulting in a substantially lower error for that prediction. The resulting autonomously "imagined" continuation of the object movement is shown in panels b – d. It turned out that this imagined movement proceeded by 87.5 % faster than the initial "real" part of the movement. Panel e shows the firing activity of 100 neurons in the recurrent circuit for the case where the input arises from the "real" object movement, and panel f shows the firing activity of the same neurons when the "real" input is replaced after $t = 125$ ms by imagined (predicted) inputs. One

[3] One can show that even more information about the precise movement direction is contained in the liquid state of the recurrent circuit when the moving ball crosses the mid area of the sensor field. Using the algorithm from section 10.2 in [Vapnik, 1998] to compute optimal hyperplanes for predicting which sensors on the perimeter were going to be activated reduced the average error from 4.9° to 1.5°.

sees from these raster plots that our model neural circuit may provide an explanation for the observed speed-up of spike-patterns in rat hippocampus from periods of sensory exploration to subsequent periods of sleep with presumed internal replay of preceding exploration phases (see [Nadasdy et al., 1999]).

4. **Discussion**

We have demonstrated that generic neural microcircuits are in principle capable of learning in an autonomous manner to augment and structure the complex high volume input stream that they receive: They can learn to predict individual components of the subsequent frames of typical "input movies", thereby allowing the system to focus both on more abstract and on surprising aspects of the input. For example, they can autonomously learn to extract the direction of movement of an object, which requires integration of information from many sensors ("pixels") and many frames of the input movie. Because of the diversity of moving objects, movement speeds, movement angles, and spatial offsets that occurred, it appears to be very difficult to construct explicitly any circuit of the same size that could achieve the same performance. Furthermore the prediction errors of our approach can be reduced by simply employing a larger generic recurrent circuit as liquid. On the other hand, given the complexity of this prediction task (for two different objects and a large diversity in movement directions and movement speeds), the recurrent circuit consisting of 768 neurons that we employed – which had not been constructed for this type of task – was doing already quite well. Its performance provides an interesting comparison to the analog VLSI circuit for motion analysis on a 7×7 array of sensors discussed in [Stocker and Douglas, 1999].

Whereas a circuit that would have been constructed for this particular task is likely to be specialized to a particular range of moving objects and movement speeds, the circuit that we have employed in our simulations is a completely generic circuit, consisting of randomly connected integrate-and-fire neurons, that has not at all been specialized for this task. We have demonstrated that it can at the same time be used by other readout neurons for completely different tasks, such as for example object classification. Furthermore we have shown that a generic neural circuit that has been trained in an unsupervised manner to predict future inputs automatically supports novelty detection when being exposed to new types of input movements. Finally we have demonstrated that if from some point on the circuit input is replaced by input predictions that are fed back from neural readouts, the emerging sequence of further predictions on the basis of preceding predictions generates an imagined continuation of a moving stimulus, triggered by the initial sequence of inputs from the beginning of that movement.

If such circuit would be a neural microcircuit in a particular sensory area of the neocortex, its task might become supported by more specialized neurons (such as for example simple or complex cells in visual cortex) that have been trained by evolution and/or individual experience to extract particular features from the types of input streams that usually arrive in that cortical area. In the very simple generic neural circuit that we have simulated, such specialization of the neural circuit would already occur if the output of some trained readout pools would contribute to the input of other readouts, or would be fed back as additional input into the recurrent circuit ("liquid"). This would improve the performance of the system for a whole category of tasks and inputs, but would also degrade – although not destroy – its flexibility and readiness to process completely different inputs (for example if the environment changes drastically, or if sensors are damaged). It would also degrade the capability of the liquid to serve as unbiased "universal" analog memory for readouts that

might be trained later for a completely new task, from whose perspective the feedback from previously trained readout neurons may just amplify special types of noise in the input stream.

The results that we have demonstrated in this article are very stable, and they work for a large variety of recurrent neural circuits. In particular they can be implemented on the most realistic computer models for neural microcircuits that are currently known. Since we now can perform complex computations on computer models of biologically realistic neural circuitry, these computer simulations can be employed as test platform for investigating the computational role of specific anatomic and physiological details of biological neural microcircuits, such as specific combinations of neuron- and synapse type, as well as specific details of their dynamic profile.

The only adaptive component in the simulated neural circuit that we have employed for prediction are the synaptic weights of readout neurons that are autonomously trained in an online manner to predict specific input components, and which will automatically adapt and optimize their prediction when the statistics of the input dynamics changes. One may hypothesize that similar predictive mechanisms take place simultaneously in all the millions of neural microcircuits of which a complex neural system is composed, thereby providing a powerful mechanism for filtering information in real time in order to report only the most salient and unexpected aspects of their input stream to other parts of the system. In fact, if all correctly predicted components of the input stream are surpressed before they can enter "higher" neural systems, the sensory input stream is recoded by such circuit into a neural code that is closer to a theoretically optimal code which reserves the shortest "words" – in this case the fewest spikes – for reporting the most frequently occurring events (comparable with the Huffman code, see e.g. [Cover and Thomas, 1991]). Our simulations demonstrate that simultaneously the same neural microcircuits that facilitate an efficient recoding of the input stream can also be used to augment the input stream by extrapolating it into the future, thereby allowing faster responses.

The autonomous adaptive mechanisms applied to an analog fading memory that were investigated in this article also provide new ideas for processing complex input streams in artificial behaving systems, such as mobile robots, in order to generate a drastically reduced and task-optimal state set for higher level reinforcement learning, and in order to achieve real-time responses in spite of the inertia of actuators by anticipating their subsequently required activation.

Acknowledgement:
We would like to thank Thomas Natschlaeger for helpful discussions.

5. References

[Abbott and Blum, 1996] Abbott, L. F., and Blum, K. I. (1996) Functional significance of long-term potentiation for sequence learning and prediction, *Cerebral Cortex*, vol. 6, 406-416.

[Auer et al., 2001] Auer, P., Burgsteiner, H. M., and Maass, W. (2001) *The p-delta learning rule for parallel perceptrons*, submitted for publication. Online available at http://www.igi.tugraz.at/maass/p_delta_learning.pdf.

[Buonomano and Merzenich, 1995] Buonomano, D. V., and Merzenich, M. M. (1995) *Temporal information transformed into a spatial code by a neural network with realistic properties*, Science, vol. 267, Feb. 1995, 1028-1030.

[Cover and Thomas, 1991] Cover, T. M., and Thomas, J. A. (1991) *Elements of Information Theory*, Wiley, New York.

[Gupta et al., 2000] Gupta, A., Wang, Y., and Markram, H. (2000) Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex, *Science* 287, 2000, 273-278.

[Haeusler et al., 2002] Haeusler, S., Markram, H., and Maass, W. (2002) *Low dimensional readout from high dimensional neural circuits*, submitted for publication.

[Jaeger, 2001] Jaeger, H. (2001) *The "echo state" approach to analyzing and training recurrent neural networks*, submitted for publication.

[Levy, 1996] Levy, W. B. (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks, *Hippocampus*, vol. 6, 579-590.

[Maass and Markram, 2002] Maass, W., and Markram, H. (2002) *On the computational power of recurrent circuits of spiking neurons*, submitted for publication.

[Maass et al., 2001] Maass, W., Natschlaeger, T., and Markram, H. (2001) *Real-time computing without stable states: A new framework for neural computation based on perturbations*, submitted for publication.

[Markram et al., 1998] Markram, H., Wang, Y., and Tsodyks, M. (1998) Differential signaling via the same axon of neocortical pyramidal neurons, *Proc. Natl. Acad. Sci.*, 95, 5323-5328.

[Mead, 1989] Mead, C. (1989) *Analog VLSI and Neural Systems*, Addison-Wesley, Reading, Mass.

[Nadasdy et al., 1999] Nadasdy, Z., Hirase, H., Czurko, A., Csicsvari, J., and Buzsaki, G. (1999) Replay and time compression of recurring spike sequences in the hippocampus, *J. Neuroscience*, 19 (21), 9497-9507.

[Rao and Sejnowski, 2000] Rao, R. P. N., and Sejnowski, T. J. (2000) Predictive sequence learning in recurrent neocortical circuits, *Advances in Neural Information Processing Systems* 12, (NIPS*99), 164-170, S. A. Solla, T. K. Leen, and K. R. Muller (Eds.), MIT Press.

[Stocker and Douglas, 1999] Stocker, A., and Douglas, R. (1999) Computation of smooth optical flow in a feedback connected analog network. *Advances in Neural Information Processing Systems* 11, (NIPS*98), 706-712.

[Tsodyks et al., 2000] Tsodyks, M., Uziel, A., and Markram, H. (2000) Synchrony generation in recurrent networks with frequency-dependent synapses, *J. Neuroscience*, Vol. 20 RC50.

[Vapnik, 1998] Vapnik, V. N. (1998) *Statistical Learning Theory*. John Wiley, New York.

Figure Captions

Figure 1: The prediction task.
a) Typical movements of objects over a 8 x 8 sensor field.
b) Time course of activation of 8 randomly selected sensors caused by the movement of the ball indicated on the l.h.s. of panel a.
c) Resulting firing times of 768 integrate-and-fire neurons in the recurrent circuit of integrate-and-fire neurons (firing of inhibitory neurons marked by +). The neurons in the 16 x 16 x 3 array were numbered layer by layer. Hence the 3 clusters in the spike raster result from concurrent activity in the 3 layers of the circuit.
d) Prediction targets (dashed lines) and actual predictions (solid lines) for the 8 sensors from panel b by 8 neural readouts, which were trained for this 50 ms prediction task by unsupervised learning. (Predictions were sampled every 25 ms, solid curves result from linear interpolation.)

Figure 2: 20 typical trajectories of movements of the center of an object (ball or bar).

Figure 3: Predictions by the autonomously trained neural circuit for a novel test example.
a) Current sensor activation at time t = 200 for this novel test movement of a ball.
b) Shown on the left hand side is the protocol of the firing activity in the recurrent circuit up to that time point. The resulting 768-dimensional "liquid state" at time $t = 200$ ms, consisting of the contributions of the 768 neurons in the recurrent circuit to the membrane potential of a hypothetical readout neuron with unit weights, is shown on the right (with contributions of excitatory neurons indicated by bars going to the right, contributions of inhibitory neurons by bars going to the left). Since we assumed a membrane time constant of 30 ms for neurons in the readout pools (see section 2), their current membrane potential depends primarily on the firing activity in the recurrent circuit during the preceding 30 ms time interval marked by vertical lines on the left hand side.
c) Outputs of the 72 readout pools, trained for 25 ms and 50 ms prediction, at time t = 200 for the input shown on the r.h.s. of panel b.
d) Prediction targets at time $t = 200$ ms for the 25 ms and the 50 ms predictions. The differences between panels c and d represent the prediction error for this novel test example.

Figure 4: Computation of movement direction. Dashed line is the trajectory of a moving ball. Sensors on the perimeter that will be activated by $\geq$ 50 % when the moving ball leaves the sensor field are marked in panel a. During unsupervised learning they are used for training the 28 readout pools that predict sensor activity on the perimeter. Sensors marked by stripes in panel b indicate a typical prediction by the trained circuit of the location of sensors on the perimeter that are going to be activated by $\geq$ 50 %, when the ball will leave the sensor field (time span into the future varies for this prediction between 100 and 150 ms, depending on the speed and angle of the object movement). Solid line in panel b represents the estimated direction of ball movement resulting from this prediction (its right end point is the average of sensors positions on the perimeter that are predicted to become $\geq$ 50 % activated). The angle between the dashed and solid line (average value 4.9° for test movements) is the error of this particular computation of movement direction by the simulated neural circuit.

Figure 5: Novelty detection. Shown in panel a is the 50 ms prediction of future sensor activity made at time $t = 200$ ms, the actual sensor activation at time $t = 250$ ms, and for each sensor the absolute value of the prediction error. The ball changed its direction of movement

at time $t = 250$ ms, resulting in a large error of the 50 ms prediction made at time 250 ms, which is shown on the l.h.s. of panel b. The comparison of this prediction with the actual sensor activation at time $t = 300$ ms (difference shown on the r.h.s. of panel b) provides all salient input needed for an analysis of this surprising event.

Figure 6: Imagined movement generated by the neural circuit. Panels a-d show the predictions of object positions at times $t = 130$ ms, 155 ms, 180 ms, 205 ms. Only the first prediction, shown in panel a, is based on sensor input. The predictions in panels b-d are primarily based on preceding input predictions, that were fed back as input into the recurrent neural circuit. This imagined movement happens to proceed faster than the actual movement which it continues, demonstrating that it is not constrained to have the same speed. Panel f shows the firing activity of a subset of 100 neurons in the recurrent neural circuit during this "imagined movement". Compared with the firing pattern of these neurons during a continuation of sensory input from the actual object movement after $t = 125$ ms (if it would continue on the same trajectory, with the same speed as at the beginning), shown in panel e, the firing pattern is very similar, but contracted in time.

Figure 1

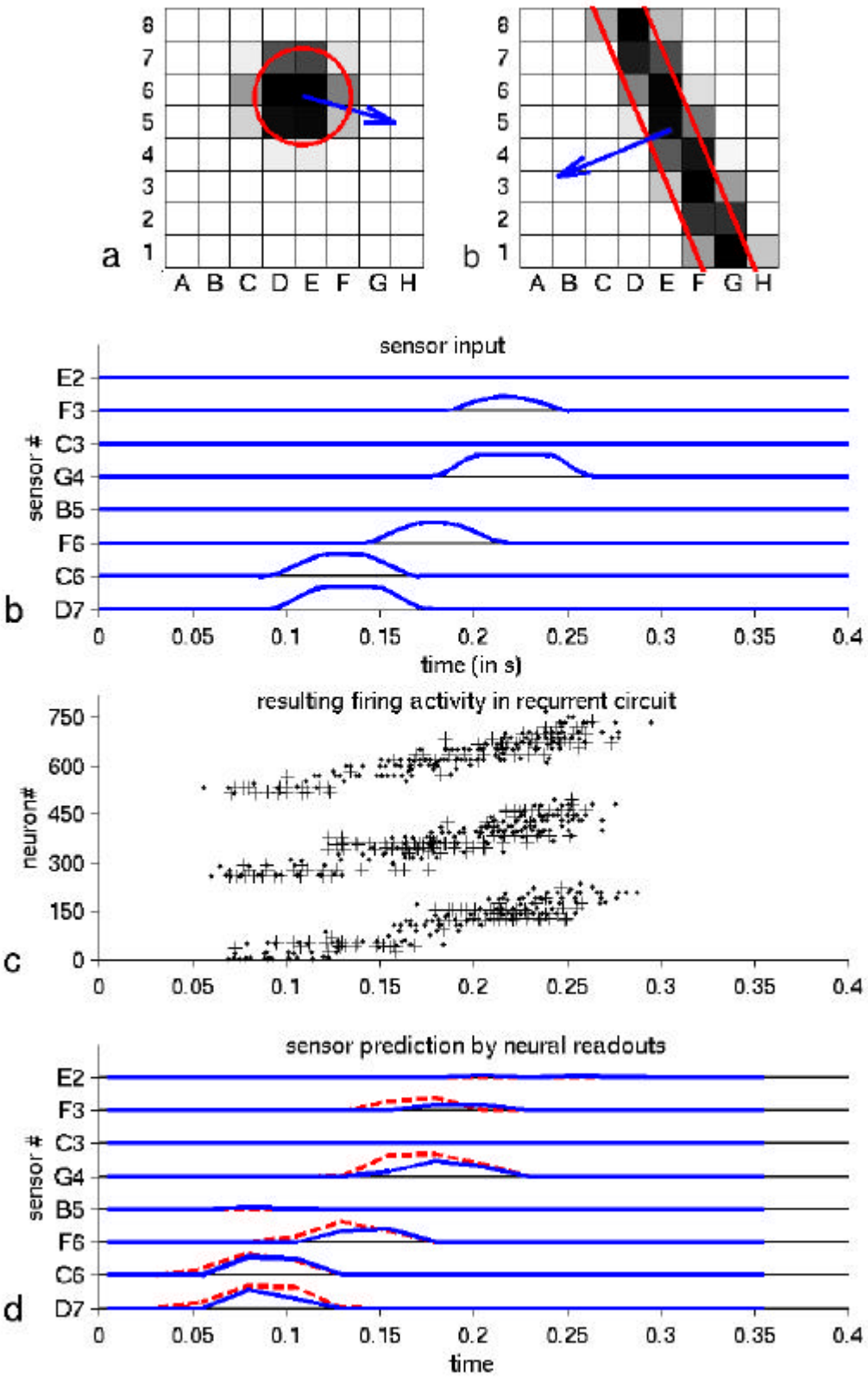

a
b
8
7
6
5
4
3
2
1
A B C D E F G H
sensor input
sensor #
E2
F3
C3
G4
B5
F6
C6
D7
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4
time (in s)
c
resulting firing activity in recurrent circuit
neuron#
750
600
450
300
150
0
d
sensor prediction by neural readouts
time

Figure 2

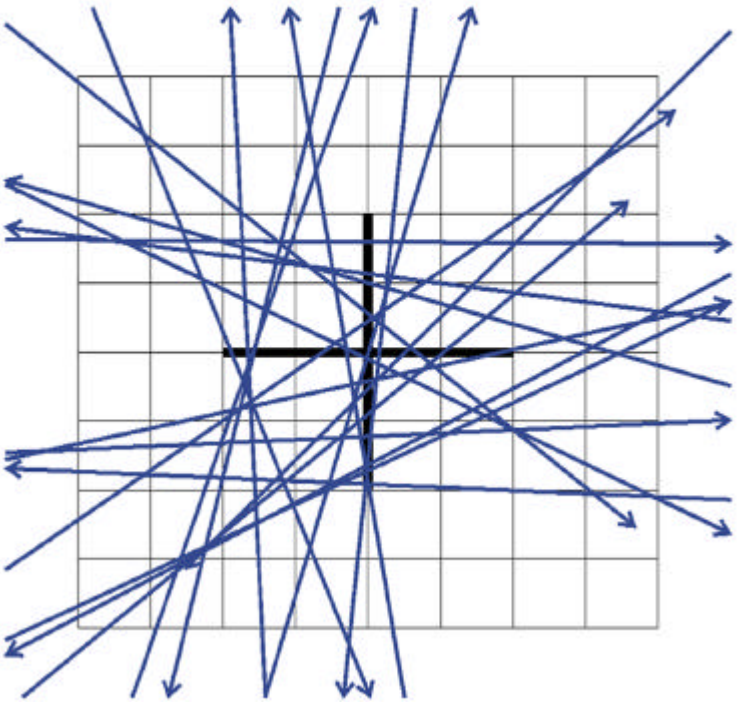


Figure 3

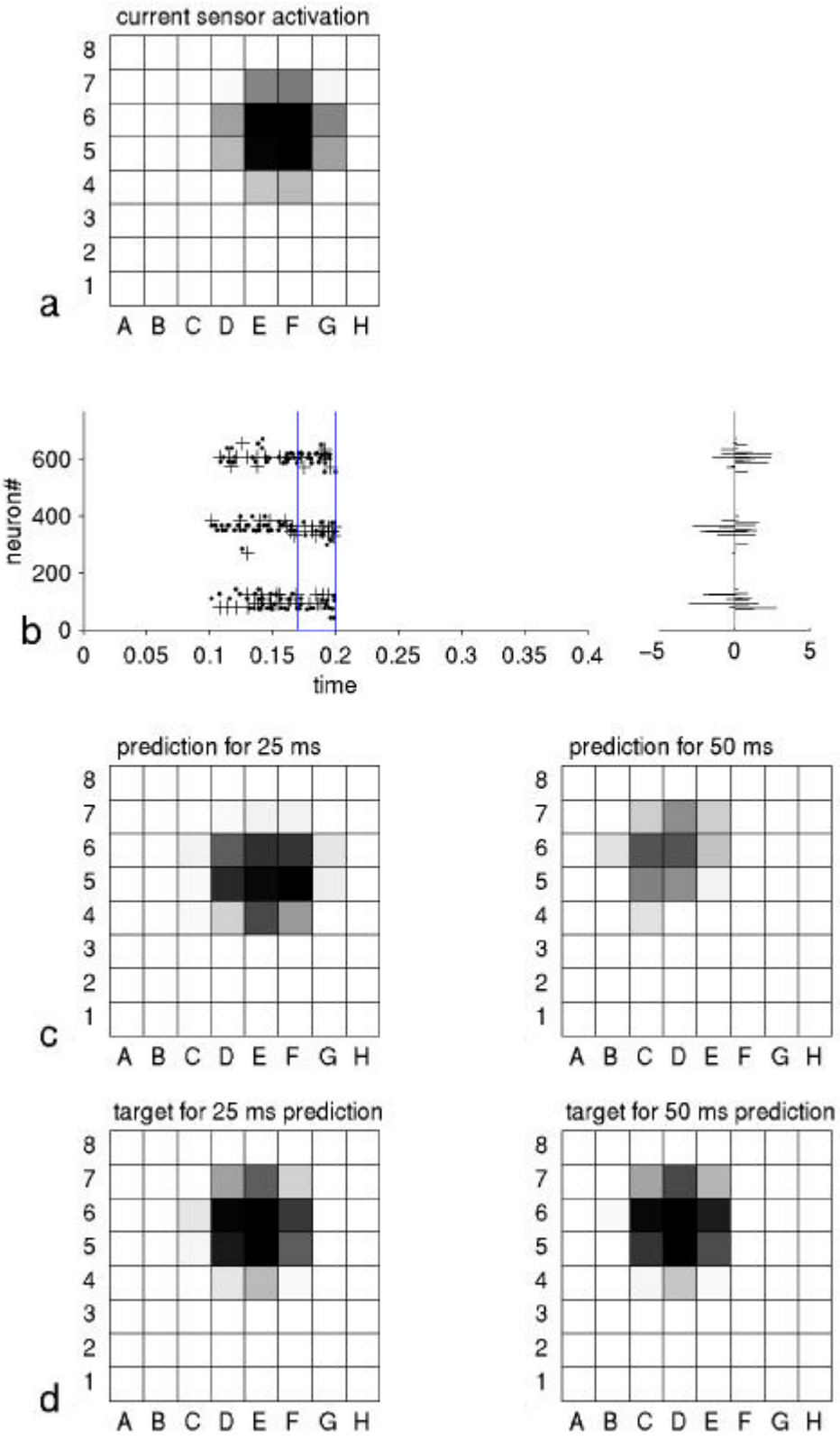

current sensor activation
a
8
7
6
5
4
3
2
1
A B C D E F G H
b
neuron#
600
400
200
0
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4
time
−5 0 5
prediction for 25 ms
c
prediction for 50 ms
target for 25 ms prediction
d
target for 50 ms prediction

Figure 4

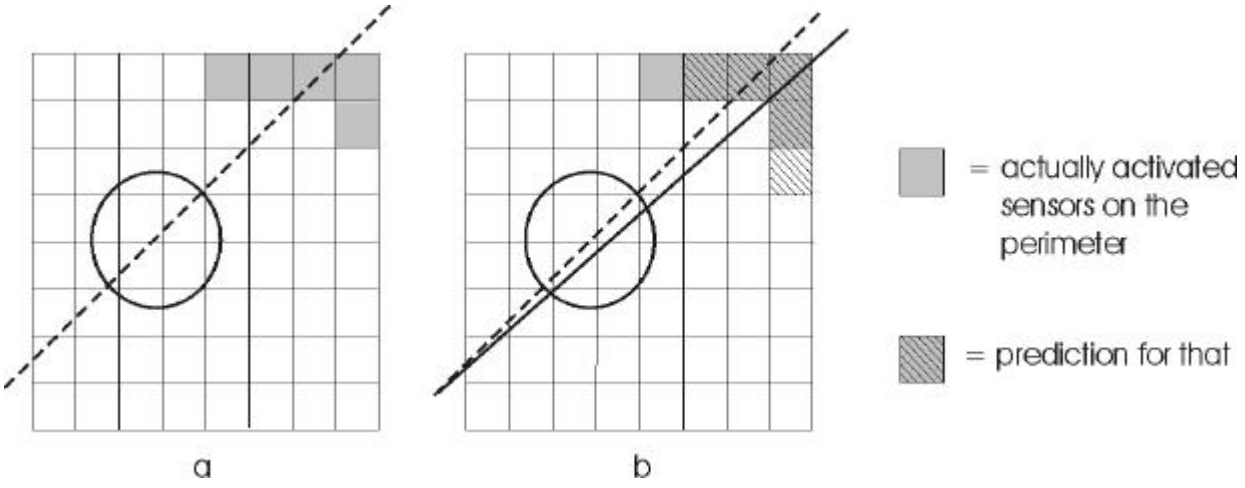
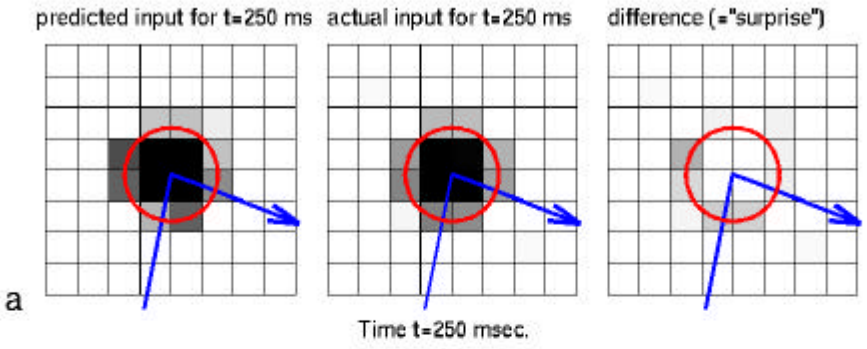

a
b
= actually activated sensors on the perimeter
= prediction for that

Figure 5


predicted input for t=250 ms
actual input for t=250 ms
difference (="surprise")
a
Time t=250 msec.

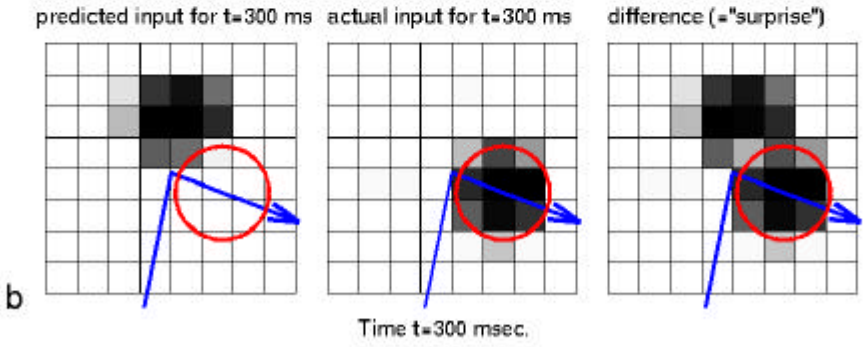

predicted input for t=300 ms
actual input for t=300 ms
difference (="surprise")
b
Time t=300 msec.

Figure 6