

Quantifying Input and Output Spike Statistics of a Winner-Take-All Network in a Vision System

Matthias Oster, Rodney Douglas, and Shih-Chii Liu
Institute of Neuroinformatics, Uni-ETH Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
Email: {mao,rjd,shih}@ini.phys.ethz.ch

Abstract—Event-driven spike-based processing systems offer new possibilities for real-time vision. Signals are encoded asynchronously in time thus preserving the time information of the occurrence of an event. We examine this form of coding using experimental data from a multi-layered multi-chip system which consists of an artificial retina, a convolution filterbank and a winner-take-all network which detect the position of a moving object. The spike outputs of the convolution stage can be described by an inhomogeneous Poisson distribution of Gaussian profile, although the underlying building blocks are completely deterministic and exhibit only a small amount of variation. We discuss a method for measuring the accuracy of the asynchronous spiking representation in both time and value, thereby quantifying the performance of the winner-take-all network in determining the position of a ball rotating in front of the system.

I. INTRODUCTION

Event-driven processing hardware architectures inspired by biological computational principles have evolved to a state in which large-scale multi-chip systems can be assembled, e.g. [1], [2], [3]. These systems offer new possibilities for applications that require short response times and large processing power, such as in vision processing. The key features of these new systems are the use of spikes (or short digital pulses) to encode signals and to trigger event-based computation. With spikes, signals can be represented with the same principle across all stages of a processing chain. To successfully build spike-based systems that are complex enough to solve real-world tasks, we need a thorough understanding of the encoding of spike-based signals, since their properties differ largely from that of conventional synchronous architectures.

In this work we analyze how signals are encoded in the form of spikes at different levels of a large-scale processing system. The CAVIAR (Convolution AER Vision Architecture for Real-time processing) project is a multi-lab EU-funded research project that explores spike-based computation in a multi-chip vision architecture [1]. The system consists of an artificial retina, a convolution filterbank, a winner-take-all decision network and a learning module, that together detect and classify the different trajectories of a moving object.

We show that the spike outputs of the convolution stage in response to a moving object can be described by a wave of activity that travels along the neurons of the winner-take-all network. The wave can be described by an inhomogeneous Poisson distribution with a Gaussian profile. The winner-take-all (WTA) network transforms these inputs into a sparse

representation of the object position. This asynchronous encoding is subject to variation in both time and value, which translates to jitter in the output spikes and errors in the classified position. We discuss how we measure the accuracy of this output representation in terms of the object location. Understanding the statistics of the input and output spikes of the WTA network allows us to quantify the performance of the spike-based WTA operation in the context of a real-world application.

We first introduce the CAVIAR system and the experiment in Section II. We then introduce our analysis method of the spike outputs of the convolution stage in Section III. From the output spikes of the WTA network we quantify the object position in Section IV before we discuss the results in Section V.

II. EXPERIMENTS

The modules in the CAVIAR system communicate using an asynchronous spike-based communication protocol, the address-event-representation (AER). In principle, the building blocks are interchangeable and can be assembled into different architectures. The CAVIAR system specifies a system architecture that serves as a demonstrator.

Figure 1 describes the functionality of the system. Input to the system is first processed by a contrast-sensitive spike-based retina which codes the temporal contrast edges in ON (positive contrast edges) and OFF spikes (negative temporal contrast edges) [4] [5]. The spikes are transmitted to a set of convolution chips [6], whose output indicates if the convolution operation exceeds a positive threshold ('positive spikes') or a negative threshold ('negative spikes'). The object chip detects the current object position from these output spikes using the WTA operation [7]. The position of the detected object is expanded over time in the delay line chip, and the resulting trajectories are learned by the learning chip. The output of the learning chip is one of its 32 neurons that span the presented input space.

In the experiment considered here, the stimulus is a constant-speed rotating disc with a set of different-sized circles on a white background. Each of the four convolution chips contains a matched-filter circular kernel of a particular size. The winner-take-all selects the strongest input signal and suppresses the output of all other neurons. We do not consider the subsequent learning stage in this work.

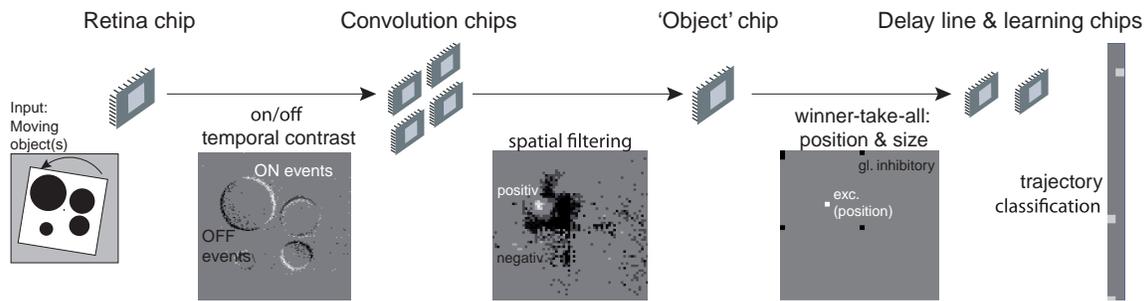


Fig. 1. Overview of the spike-based CAVIAR demonstrator: Input to the transient retina in this case was 4 rotating black disks on a white background moving with constant speed. The retina converts temporal contrast edges into spikes (left inlay): black dots represent 'OFF' events (responses to negative temporal contrast edges), white dots 'ON' ones (responses to positive contrast edges). The spikes are transmitted to a group of convolution chips. Each spike-based convolution chip detects the center of the disk that best matches the programmed convolution kernel (middle inlay). Spikes indicating positive outputs of the convolution operation are shown as white dots, spikes indicating negative outputs as black dots. Four convolution chips are tiled to increase the resolution; in other configurations they can be programmed to detect different objects. The output of the spatial filtering is cleaned by the 'object' chip, which performs a winner-take-all process to decide the best object position (right inlay). The white dot marks the spike output of the object chip; the black dot represents the spikes of the inhibitory neurons involved in the computation. Object position and size (in case of the convolution kernels programmed for different ball sizes) are then expanded over time in a delay line chip and the resulting trajectories are classified by the learning chip. Additional logic modules between the chips can be used to monitor the spike trains in the system, to map the synaptic connections, and to inject artificial spike trains into the system.

III. OUTPUT OF CONVOLUTION STAGE

For this analysis we focus on the ON-spikes of the convolution stage that indicate the presence of an object matching the programmed kernel in both space and time with a certain probability (the negative spikes are not taken into account here). Depending on the parameters of the convolution, the number of spikes and their relevance will change, resulting in a different shape of the average spike waveform that forms the input to the WTA network.

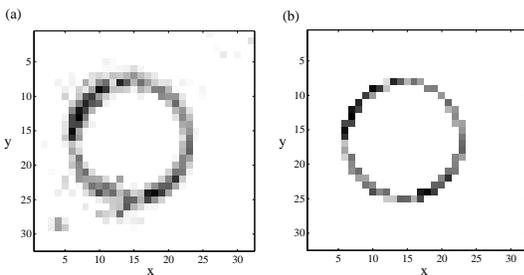


Fig. 2. Spatial trajectory of the stimulus center. The stimulus is a disc that rotates with constant velocity in front of the retina. The convolution stage contains a matched-filter kernel; its output is a smoothed version of the center of the object (a). The gray level gives the spike count for one revolution of the stimulus. For the analysis, we consider only pixels that fall onto the trajectory of the stimulus center (b). We masked these pixels with a manually defined region of interest.

We consider the neurons to be arranged along a one-dimensional space. We mapped the two-dimensional convolution output onto this one-dimensional space by considering only neurons along the stimulus trajectory (Fig. 2). This is possible since we know the trajectory in this simple problem. The transformation discards activity from neurons outside the trajectory of the stimulus center. These outliers receive less input than neurons on the trajectory and do not evoke output

spikes from the WTA. Our analysis focuses on the spatio-temporal estimation of the stimulus position, for which only the neurons with a significant spike input are relevant.

To assess how well the input data to the WTA can be described by a travelling Gaussian wave with Poisson statistics, we examine the average input to each neuron of the one-dimensional WTA. Since the neurons are aligned to a certain stimulus position, the input to each neuron corresponds to a movement step of the stimulus. We can describe the input in response to this step by the Peri-Stimulus Time Histogram (PSTH), a commonly used method to describe spike train responses. In the PSTH, spike trains from different trials are aligned to the stimulus onset. The histogram is obtained by binning all spikes using a fixed bin time. The resulting histogram is normalized to the number of trials, resulting in the average input response of each neuron to the stimulus.

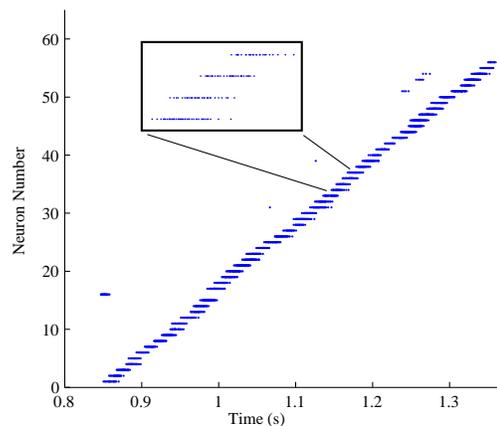


Fig. 3. Raster plot of the input spike trains. Input channels along the trajectory are sorted in the order of the stimulus movement. Each point marks one spike (every spike train contains about 20-40 spikes). Data from one revolution of the stimulus disc are shown.

In our case the stimulus is rotation symmetric, so the input to different neurons of the one-dimensional WTA can be seen as different trials of the same experiment. We obtain the PSTH by averaging over all neurons and over multiple revolutions of the stimulus.

Figure 3 shows the spike trains before alignment, but sorted by their mean spike time. From this representation the average travelling time d from one neuron to the next can be calculated, by averaging the difference in the mean time between each pair of neighboring input channels. The resulting PSTH is shown in Fig. 4, together with a Gaussian fit.

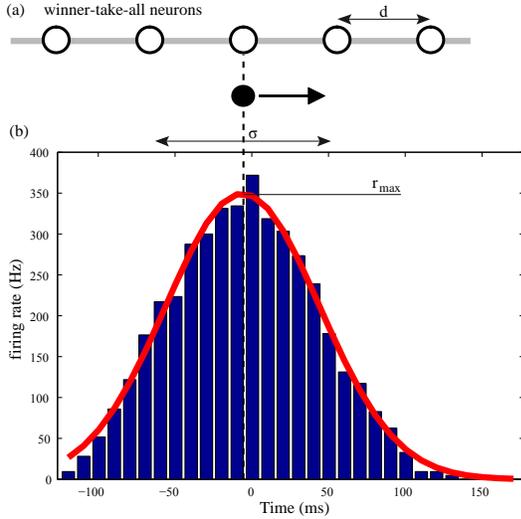


Fig. 4. Peri-Stimulus Time Histogram (PSTH) obtained by averaging the convolution spikes over about 2500 trials (60 neurons and multiple revolutions of the stimulus). The red line shows a Gaussian fit.

From the fit, the parameters σ (standard deviation) and peak spike rate r_{\max} are extracted. To quantify the goodness of the fit, we extracted the parameters separately for each input channel and then averaged over the results of each spike train. All parameters are listed in Table 1.

Parameter	CAVIAR	simulated Poisson
d	95ms \pm 31%	95ms \pm 9.5%
σ	46ms \pm 11%	46ms \pm 4.5%
r_{\max}	373Hz \pm 37%	372Hz \pm 17.6%
CV	0.83 \pm 23%	0.96 \pm 16%

TABLE I

COMPARISON OF PARAMETER VALUES BETWEEN THE CAVIAR DATA SET FOR A GAUSSIAN-SHAPE TRAVELLING WAVE STIMULUS WITH AN ARTIFICIALLY CREATED DATA SET.

To determine the variation in the spike times, we have transform the non-stationary spike trains of the convolution output into stationary ones. We use the time-rescaling theorem, which rescales the spike times according to an average rate function that is obtained by smoothing the PSTH [8]. The resulting spike train is homogeneous which allows us to test for Poisson statistics. We calculate the coefficient of variation

(CV) as the standard deviation of the inter-spike intervals (ISIs) by their mean (see Table 1). As a sufficient test condition for Poisson statistics, the ISIs have to follow an exponential distribution. Figure 5 shows the distribution of the inter-spike intervals.

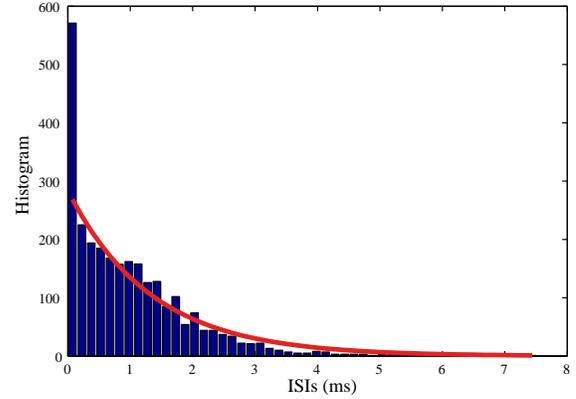


Fig. 5. Inter-spike interval distribution after transforming the spike trains into a homogeneous process. Data from one stimulus revolution. The continuous line shows an exponential fit excluding the peak at the minimum inter-spike interval that originates from subsampling in the system.

The distribution has a large peak at minimum inter-spike intervals that is induced by subsampling. From the retina resolution of 128x128 the input channels are sub-sampled to 64x64 for the convolution stage, and 32x32 for the winner-take-all stage. Sub-sampling combines neighboring input channels. Since these channels originate from spatially related stimulus input, neighboring spike trains exhibit correlations that result in minimum inter-spike intervals in the distribution.

The resulting distribution of inter-spike intervals can be well approximated by an exponential function (see Fig. 5, continuous line), which implies a Poisson distribution of the output spike trains of the convolution chip.

IV. OUTPUT OF WINNER-TAKE-ALL

The WTA network integrates the output spikes of the convolution stage. The winner, that is the neuron that reaches threshold first, makes an output spike and suppresses the other neurons through strong inhibition. The output spikes of the WTA network indicate the object position, discretized in value, that is, to the neurons that indicate the object position; and in time, that is, to the time occurrence of the output spikes.

Figure 6 shows how the stimulus position is reconstructed from this asynchronous representation. The ideal output of the network is an update of the object position as soon as the object is aligned to a new neuron, resulting in a staircase function. Since the object in our experiment moves with constant speed, the transition times from each neuron to the next are constant.

The object position reconstructed from the output spikes of the WTA network deviates from this ideal case in both time and position. We call an error in the position 'classification error', if a neuron other than the one aligned to the object

position makes an output spike. Spiking before or after the ideal object position results in jitter. Both errors are induced by the Poisson statistics of the input and the variation in its parameters.

Jitter and classification error are not independent. For example, an earlier spike can also be seen as the classification of a different position. We combine both errors by defining an area error e , which quantifies the area difference between the reconstructed object location and the ideal case, normalized to one neuron.

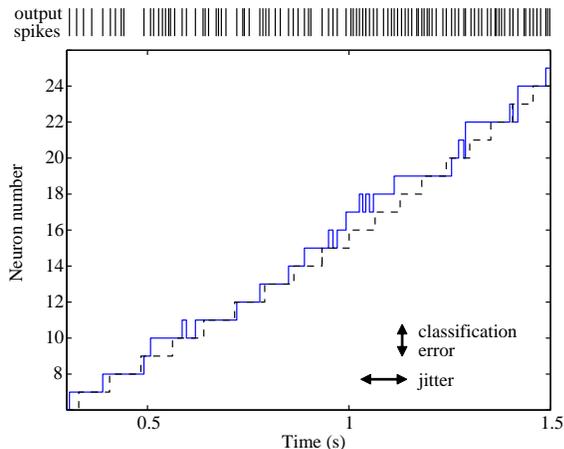


Fig. 6. Reconstruction of object position from the winner-take-all output spikes. The ideal output of the network is an update of the object position as soon as the object is aligned to a new neuron (dashed staircase). In the sparsest representation the network would elicit one spike at each vertical line of the dashed staircase function. In the CAVIAR data (continuous staircase), the WTA network elicits more spikes than one spike per position, as illustrated by the output spike train of the network at the top. This leads to switching in the predicted object position, for example between neurons 17 and 18. In addition, the output sometimes indicates an incorrect position, for example at neurons 18 and 19, or the spike times are jittered, for example at neurons 13 and 14. The area error e quantifies these errors by considering the area between the predicted ball position (dashed staircase) and the actual prediction from the data (continuous staircase), normalized to one neuron. The area error from this data is $e \approx 0.64$.

V. DISCUSSION

Our analysis of the output of the CAVIAR convolution stage show that the output spike trains of the convolution stage can be approximated by a Gaussian waveform of Poisson statistics that travels along the neurons. The building blocks of the CAVIAR system are completely deterministic and have only a small amount of variation across the neurons. A spiking system exhibits two types of variation. First, there is variation in the spike rates, for example, in response to a fixed contrast change seen by the retina pixels. Second, there are variations in the spikes latencies, resulting in spike jitter. Mismatch in the input circuits of for example, the sensor, will affect both timing and the output spike rate, since it varies how fast the neurons reach threshold.

We were astonished to find that these small variations add up to Poisson statistics already at the output of the first post-processing stage after the sensor. Further analyses are neces-

sary to determine if this phenomenon is due to the hardware of the system or due to the nature of the problem, that is, the law of large numbers leading to a Poisson distribution. Interestingly, the Poisson distribution is also one of the most commonly used distributions used to model biological spike train statistics.

We then quantified the output of the winner-take-all implementation, a sparse representation in which each output spike indicates a new stimulus position. Although the area measure as the difference between estimated and measured position does not capture all properties of the output spike trains, it allows a quantification of the classified stimulus position and therefore the performance of the network. We had developed a theoretical model of the winner-take-all performance in dependence of the Poisson parameters of the input published in [9]. This shows that the output of winner-take-all in CAVIAR system as we analyzed it here comes close to the performance limit that is induced by the Poisson statistics of the input spike trains.

VI. ACKNOWLEDGMENTS

We acknowledge members of the INI hardware group for the development of the software and hardware infrastructure. This work was mainly funded by EU-grant IST-2001-34124.

REFERENCES

- [1] R. Serrano-Gotarredona, M. Oster, P. Lichtsteiner, A. Linares-Barranco, R. Paz-Vincent, F. Gómez-Rodríguez, H. Kolle Riis, T. Delbrück, S. C. Liu, S. Zahnd, A. M. Whatley, R. Douglas, P. Häfziger, G. Jimenez-Moreno, A. Civit, T. Serrano-Gotarredona, A. Acosta-Jiménez, and B. Linares-Barranco, "AER building blocks for multi-layer multi-chip neuromorphic vision systems," in *Advances in Neural Information Processing Systems (NIPS)*, Y. Weiss, B. Schölkopf, and J. Platt, Eds., vol. 18, Neural Information Processing Systems Foundation. Cambridge, MA: MIT Press, Dec 2005, pp. 1217–1224.
- [2] T. Choi, B. Shi, and K. Boahen, "A multi-chip implementation of cortical orientation hypercolumns," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, vol. 3, 2004, pp. 13–16.
- [3] R. Vogelstein, U. Mallik, E. Culurciello, G. Cauwenberghs, and R. Etienne-Cummings, "Saliency-Driven Image Acuity Modulation on a Reconfigurable Silicon Array of Spiking Neurons," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 16, 2005, pp. 1457–1464.
- [4] J. Kramer, "An integrated optical transient sensor," *IEEE Transactions on Circuits and Systems II, Analog and Digital Signal Processing*, vol. 49, no. 9, pp. 612–628, Sep 2002.
- [5] P. Lichtsteiner, C. Posch, and T. Delbrück, "A 128x128 120dB 30mW asynchronous vision sensor that responds to relative intensity change," in *ISSCC Digest of Technical Papers*, 2006, pp. 508–509.
- [6] R. Serrano-Gotarredona, T. Serrano-Gotarredona, A. Acosta-Jimenez, and B. Linares-Barranco, "A neuromorphic cortical layer microchip for spike based event processing vision systems," *IEEE Transactions on Circuits and Systems I*, 2006.
- [7] S.-C. Liu and M. Oster, "Feature competition in a spike-based winner-take-all VLSI network," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2006, pp. 3634–3637.
- [8] E. Brown, R. Barbieri, V. Ventura, R. Kass, and L. Frank, "The time-rescaling theorem and its application to neural spike train data analysis," *Neural Computation*, vol. 14, no. 2, pp. 325–346, 2002.
- [9] M. Oster and S. Liu, "Spiking inputs to a winner-take-all network," in *Advances in Neural Information Processing Systems (NIPS)*, Y. Weiss, B. Schölkopf, and J. Platt, Eds., vol. 18, 2005, pp. 1051–1058.