Predictive Coding Light: learning compact visual codes by combining excitatory and inhibitory spike timing-dependent plasticity[*]

Antony W. N'dri, Thomas Barbier, Céline Teulière
Université Clermont Auvergne, Clermont Auvergne INP, CNRS, Institut Pascal,
F-63000 Clermont-Ferrand, France
antony.n_dri@doctorant.uca.fr, {thomas.barbier,celine.teuliere}@uca.fr

Jochen Triesch
Frankfurt Institute for Advanced Studies,
Frankfurt am Main, Germany
triesch@fias.uni-frankfurt.de

Abstract

It is widely believed that the brain uses predictive coding schemes to represent sensory inputs in an efficient manner. However, it is still debated how networks of spiking neurons can learn such coding schemes in an unsupervised fashion. Here we present a hierarchical spiking neural network architecture that learns an efficient encoding of visual input from an event-based vision sensor by combining excitatory and inhibitory spike timing-dependent plasticity (STDP). The network develops receptive fields and exhibits surround suppression effects reminiscent of biological findings. We show that inhibitory STDP which aims to suppress predictable (and therefore redundant) spikes in neurons strongly reduces neural activity (and therefore energy costs) with only moderate reductions in coding fidelity.

1. Introduction

Biological vision systems have been optimized by millions of years of evolution to be highly energy efficient and they can learn without external supervision. Designing similarly efficient artificial vision systems capable of unsupervised learning is still a grand challenge. At the sensing front-end, much progress has been made with event-based cameras that mimic information processing in the retina, providing high energy efficiency, low latency, and high dynamic range [14]. However, it is much less clear how unsupervised learning processes in later visual processing stages can be understood and replicated in artificial vision systems.

A prominent conceptual framework for understanding such unsupervised learning is the efficient coding hypothesis, which states that sensory representations and sensory processing are adapted to the statistics of naturally occurring sensory signals. In particular, Barlow argued that visual inputs may be represented through an encoding scheme that removes redundancy inherent in sensory signals to make best use of the available neural resources [5]. A special variant of efficient coding schemes are predictive coding approaches [13, 24, 25, 29]. How such predictive coding ideas can be realized in spiking neural networks and how these may be set up by neuronal and synaptic plasticity mechanisms is still subject of much debate [3, 7, 22].

To address this challenge, we here propose a hierarchical spiking neural network that learns to encode visual input from an event-based camera in an efficient manner. To do so, it combines different forms of excitatory and inhibitory spike timing-dependent plasticity (eSTDP and iSTDP, respectively) together with homeostatic mechanisms. In contrast to most predictive coding approaches, there are no explicit neurons for detecting errors between bottom-up inputs and top-down predictions. Instead, the network simply uses iSTDP to learn to inhibit the most predictable spikes via lateral and top-down connections. The idea is that since these spikes are easy to predict from the network's ongoing activity, they only represent redundant information. Suppressing these redundant spikes therefore saves energy without much loss of information, since spiking is a major source of energy consumption in both biological brains [2, 19] and neuromorphic chips such as Loihi [11]. Our network also does not distinguish excitatory and inhibitory neurons, but any neuron can both excite and inhibit other neurons in the

[*] We thank UCA, Clermont Auvergne INP and the SFRI ANR for their financial support throughout this research. Without their funding, this work would not have been possible. JT acknowledges support from the Johanna Quandt foundation. We thank Fleur Zeldenrust for discussions.

network. Due to the simplicity of the approach compared to previous predictive coding schemes, we call it *Predictive Coding Light* (PCL).

We summarize our contributions as follows: 1) We propose PCL, a simple predictive coding scheme for spiking neural networks. 2) We train and test a PCL network on simulated and real data from an event-based camera. 3) We show that the network forms synaptic connectivity patterns that account for several biological observations including simple and complex cell-like receptive fields and surround suppression effects. 4) We demonstrate that the learnt inhibitory connections permit the network to encode its sensory inputs in an (energy) efficient manner: it represents inputs with much fewer spikes with only moderate information loss. Code and datasets are available from the authors upon request.

2. Related Work

The current surge of interest in predictive coding goes back to the seminal work of Rao and Ballard [25], although its intellectual origins can be traced back to Hermann von Helmholtz [32]. We focus our discussion on spiking neural network implementations of efficient and predictive coding schemes and models of surround suppression, which is considered a signature of predictive coding in visual information processing.

Models of efficient and predictive coding with spiking neural networks. How spiking neural networks can learn efficient and/or predictive representations using unsupervised learning approaches has been addressed by various works. Several authors have attempted to construct spiking neural network implementations of classic unsupervised learning schemes. For example, Zylberberg et al. propose an implementation of sparse coding with spiking neural networks [34]. Savin et al. consider independent component analysis in spiking neurons [27]. Burbank constructs a spiking implementation of an autoencoder [8]. This latter work uses different types of STDP for feedforward and feedback connections in the network, which we also use in PCL. Chauhan et al. show that visual receptive fields closely matching biological findings develop in a network using rank-based STDP [10]. Barbier et al. learn simple and complex cell-like receptive fields from input of an event-based camera [4].

Realizations of explicit predictive coding schemes with spiking neural networks are still quite rare. Brendel et al. [7] derive voltage-based local learning rules for spiking neurons from an objective function for coding efficiency. Their networks learn to represent signals in a spike by spike fashion to achieve codes that are highly efficient yet still robust to noise, delays, or constraints on connectivity. This work makes a number of very interesting biological predictions. An open question seems to be the biological plausibility of some of the derived learning rules, in particular a predicted "reverse STDP" for recurrent excitatory to excitatory connections [7]. Our work shares several features with [7]. In particular, we also assign different roles and learning rules to feedforward vs. feedback connections. However, we only consider learning rules for which there is some experimental support.

Most recently, the work of Mikulasch et al. [22] proposes to implement predictive coding in SNNs using multi-compartment neurons. In this approach, prediction errors are computed by the local voltage dynamics in the dendrites. Neurons compete via lateral and top-down connections in order to reduce prediction errors. Similar to [7], this scheme also leads to a tight balance between excitation and inhibition that matches biological observations. PCL proposes a simpler architecture without the need for multi-compartment neurons.

Models of surround suppression. The term *surround suppression* describes the neurophysiological observation that neurons in visual cortex can reduce their activity when a visual stimulus extends beyond its so-called classical receptive field. The origins of this effect have been a topic of neuroscientifc investigation [23, 26]. Rao and Ballard demonstrated that such behavior arises naturally in their predictive coding approach [25], while earlier approaches, e.g., [28, 30], and also more recent ones [9, 20] mostly tried to give mechanistic explanations of the phenomenon. Following-up on the functional perspective of Rao and Ballard, Zhu and Rozell [33] have shown that a broad range of nonclassical receptive field effects including end-stopping, surround suppression, contrast invariance of orientation tuning, and cross-orientation suppression emerge in a recurrent network that learns a sparse encoding of its visual input. Note that all the works above have used rate-coding model neurons. A recent study has considered surround suppression in an SNN [21], but the surround suppression was "hard-wired" into the network and not a consequence of learning an efficient/predictive code as is the case for our PCL network.

3. Methods

3.1. Network architecture

Our network is inspired by simple and complex cells in the primary visual cortex of mammals. It is composed of a simple cell layer, which receives inputs from an event-based vision sensor, and a complex cell layer, which is driven by the simple cell layer (Fig. 1).

The event-based vision sensor generates two kinds of events: 'ON' events signalling brightness increases and 'OFF' events signalling brightness decreases. These events

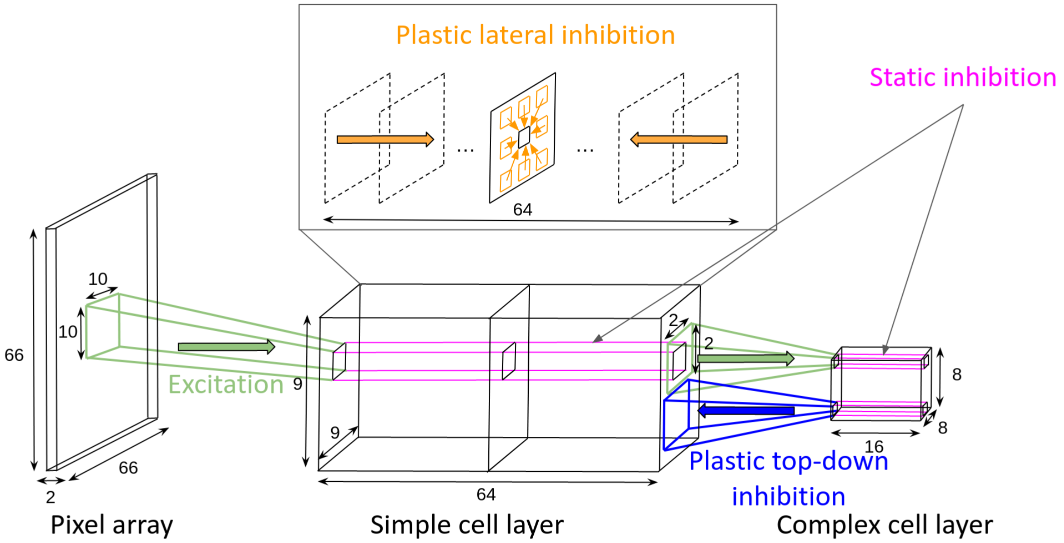


Figure 1. Network architecture.

are received by the simple cells via adjustable excitatory synaptic weights. The simple cells in turn excite complex cells in the complex cell layer via another set of adjustable excitatory synaptic weights.

We do not distinguish excitatory and inhibitory neurons in our network. Instead, units can both excite and inhibit other neurons. The network comprises three kinds of inhibition. First, there is a 'static inhibition' among simple cells with identical receptive field location that helps to decorrelate their responses. Second, there is an additional 'plastic lateral inhibition' among simple cells with neighboring receptive fields. Third, there is a 'plastic top-down inhibition' from complex cells to simple cells. These lateral and top-down inhibitory connections learn to cancel highly predictable simple cell spikes, thereby reducing the total amount of spiking without much information loss. Table 1 summarizes the different connectivity parameters.

3.2. Neuron model and static lateral inhibition

We use a standard leaky integrate-and-fire (LIF) neuron model, as it offers a good compromise between computational cost and realism. When a LIF neuron receives no inputs, its membrane potential decays exponentially to a resting value V_{rest}, that is defined as 0 mV in this work. Our simulation is event-based. The membrane potential of a neuron is only updated whenever a synaptic input arrives. In this case, the weight w_i of the corresponding synaptic connection is added to the neuron's membrane potential. The membrane potential of the neuron is updated according to:

$$\tilde{V}(t+\Delta t) = \max\{V_{\min}, V(t)e^{\frac{-\Delta t}{\tau_{\text{m}}}} + w_i(t)\}\,, \qquad (1)$$

where Δt is the time since the membrane potential was last updated, τ_{m} is the membrane time constant and $V(t)$ is the membrane potential. V_{min} is a minimum value of the membrane potential, here defined as -80 mV.

When the membrane potential reaches a threshold V_θ, the neuron generates an action potential that will be transmitted to other neurons. The membrane potential of the neuron is then reset to a value $V_{\text{reset}} = -20$ mV:

$$V(t+\Delta t) = \begin{cases} \tilde{V}(t+\Delta t) & : \tilde{V}(t+\Delta t) < V_\theta \\ V_{\text{reset}} & : \tilde{V}(t+\Delta t) \geq V_\theta\,. \end{cases} \qquad (2)$$

We incorporate a relative refractory period in the neuron model, limiting its ability to spike for a short time period after a previous spike. This is implemented via a decaying trace that is subtracted from the potential at each update:

$$\tilde{V}(t+\Delta t) = \max\{V_{\min}, V(t)e^{\frac{-\Delta t}{\tau_{\text{m}}}} + w_i(t) - \eta_{\text{RP}} e^{-\frac{t+\Delta t - t_s}{\tau_{\text{RP}}}}\}\,, \qquad (3)$$

with t_s the time since the last spike and η_{RP} and τ_{RP} the amplitude and decay rate of the refractory mechanism, respectively. This Eq. 3, replaces Eq. 1 for excitatory inputs.

Static lateral inhibition. Simple/complex cells inhibit other simple/complex cells at the same retinal location via a 'static inhibition'. When a cell spikes, the other cells receive a strong inhibitory input preventing them from spiking for a short time span:

$$\tilde{V}(t+\Delta t) = \max\{V_{\min}, V(t)e^{\frac{-\Delta t}{\tau_{\text{m}}}} - w_{\text{I}}\}\,, \qquad (4)$$

where w_{I} sets the strength of this inhibition. Eq. 4 replaces Eq. 3 in the case where an inhibitory input is received due to the static inhibition.

Simple cells:

Retina size	Input zone	$\mathrm{N}_{\mathrm{cells}}$
$(346 \times 260 \times 2)$	$(66 \times 66 \times 2)$	$(9 \times 9 \times 64)$
Receptive field	Overlap	Plastic lateral inhibition range
$(10 \times 10 \times 2)$	$(3 \times 3 \times 2)$	$(4 \times 4 \times 64)$

Complex cells:

Input size	Input zone	$\mathrm{N}_{\mathrm{cells}}$
$(9 \times 9 \times 64)$	$(9 \times 9 \times 64)$	$(8 \times 8 \times 16)$
Receptive field	Overlap	Plastic top-down inhibition range
$(2 \times 2 \times 64)$	$(1 \times 1 \times 64)$	$(2 \times 2 \times 64)$

Table 1. Connectivity parameters. The values indicate dimensions of the following form: $(x \times y \times z)$ cells.

3.3. Synaptic plasticity mechanisms

Spike-Timing Dependent Plasticity. Excitatory synaptic connections to the simple cells are learnt via a standard form of spike timing-dependent plasticity (STDP) with exponential windows. In our event-based simulation, we distinguish potentiating and depressing events (long term potentiation/depression, LTP/LTD). LTP is triggered whenever the post-synaptic neuron spikes. At this time the weight of the i-th synaptic input to this neuron is updated according to:

$$\Delta w_i^{\mathrm{LTP}} = \eta_{\mathrm{LTP}} e^{\frac{t_i - t_\mathrm{s}}{\tau_{\mathrm{LTP,s}}}} , \tag{5}$$

where t_i is the time when the last spike from input i was received, t_s is the spike time of the postsynaptic neuron, and $\eta_{\mathrm{LTP,s}}$ and $\tau_{\mathrm{LTP,s}}$ set the scale and time constant of the temporal window.

LTD is also triggered whenever the post-synaptic neuron spikes according to:

$$\Delta w_i^{\mathrm{LTD}} = -\eta_{\mathrm{LTP,s}} e^{\frac{t_{s-1} - t_i}{\tau_{\mathrm{LTD,s}}}} , \tag{6}$$

where t_i is again the time when the signal from input i was received, t_{s-1} is the previous spike time of the postsynaptic neuron, and $\eta_{\mathrm{LTD,s}}$ and $\tau_{\mathrm{LTD,s}}$ again set the scale and time constant of the temporal window.

For the excitatory connections from simple cells to complex cells we use slightly different STDP rules with relatively wide rectangular temporal windows to foster invariance:

$$\Delta w_i^{\mathrm{LTP}} = \begin{cases} \eta_{LTP} & : |t_i - t_s| \leq \tau_{\mathrm{LTP,c}} \\ 0 & : |t_i - t_s| > \tau_{\mathrm{LTP,c}} \end{cases} \tag{7}$$

$$\Delta w_i^{\mathrm{LTD}} = \begin{cases} \eta_{\mathrm{LTD}} & : |t_{s-1} - t_i| \leq \tau_{\mathrm{LTD,c}} \\ 0 & : |t_{s-1} - t_i| > \tau_{\mathrm{LTD,c}} \, . \end{cases} \tag{8}$$

We have experimented with different window functions and the results described below do not seem to critically depend on the precise functional form of the window, as long as it is sufficiently wide. A detailed discussion of this is beyond the scope of the current manuscript.

Weight normalization. The STDP rule proposed above can lead to a great disparity between weights and instability due to unbounded growth. We introduce a weight normalization mechanism in order to avoid this and to control the strengths of different connection types in the network. This aims to model a limited supply of synaptic building blocks in a simple fashion [31]. Let $\tilde{W}$ be a non normalized weight vector of a neuron, obtained after a weight update via STDP. The weights W are then updated according to:

$$W \leftarrow \lambda \frac{\tilde{W}}{\|\tilde{W}\|} , \tag{9}$$

where λ is a scaling factor. The aforementioned equation is applied to all weights (excitatory and inhibitory) of our network for all neurons. The value of λ varies according to the type of connection (see Tab. 2 for details).

Weight sharing. In order to extract an identical set of features with groups of simple cells representing different retinal locations, simple cells share their excitatory weights as in convolutional neural networks. Note that the rest of the weights in the network are not shared but individual. In particular, we did not apply any weight sharing mechanism for complex cells, so that the features extracted become tuned to the specific inputs at that spatial location.

Plastic lateral inhibition. Simple cells can also dynamically inhibit each other. Rather than using a fixed weight to subtract from the membrane potential as for the static inhibition, here inhibitory connections are learnt using the STDP rule from Eqs. *5* and *6*. A spiking simple cell will inhibit all other cells in a 3-dimensional range (excluding those at the same retinal location) to prevent them from sending the same information to deeper layers. The range of the inhibition is given in Tab. 1 and the hyperparameters used for training with STDP are given in Tab. 2. The update of the membrane potential when a plastic lateral inhibitory event is received follows Eq. *1*.

Plastic top-down inhibition. Complex cells inhibit simple cells via plastic top-down inhibitory connections. These connections are also learnt using the STDP rule from Eqs. *5* and *6*. The spatial range of these inhibitory connections and parameters used for training are detailed in Tabs. 1 and 2, respectively. The update of the membrane potential of a simple cell receiving an inhibitory top-down input also follows Eq. *1*, just like for the plastic lateral inhibition.

Simple cells										
$\mathrm{V}_{\mathrm{reset}}$ (mV)	$\mathrm{V}_{\mathrm{thresh}}$ (mV)	τ_{m} (ms)	τ_{RP} (ms)	η_{LTP} (mV)	η_{LTD} (mV)	w_{I} (mV)	η_{RP} (mV)	τ_{LTP} (ms)	τ_{LTD} (ms)	λ
-20	30	20	30	excit. = $77 \cdot 10^{-6}$ inhib. = $77 \cdot 10^{-2}$	excit. = $21 \cdot 10^{-6}$ inhib. = $21 \cdot 10^{-2}$	20	1	14	7	excit. = 4 lat. inhib. = 40 t-d. inhib. = 200
Complex cells										
$\mathrm{V}_{\mathrm{reset}}$ (mV)	$\mathrm{V}_{\mathrm{thresh}}$ (mV)	τ_{m} (ms)	τ_{RP} (ms)	η_{LTP} (mV)	η_{LTD} (mV)	w_{I} (mV)	η_{RP} (mV)	τ_{LTP} (ms)	τ_{LTD} (ms)	λ
-20	3	200	30	excit. = 0.02	excit. = 0.02	15	1	20	20	excit. = 10

Table 2. Network hyperparameters.

Bars per sequence		Speeds (pixels/ms)		Length (pixels)		Orientations (deg)	Widths (pixels)
Min	Max	Min	Max	Min	Max		
1	50	0.1	0.5	1	150	0, 23, 45, 68 90, 113, 135 and 158	1, 2 and 3

Table 3. Synthetic bars training dataset parameters.

3.4. Event-based datasets generation

For synthetic training sets we generate events from RGB image sequences using the PIX2NVS simulator [6]. This method is fast and gives us perfect control over the stimulus properties.

Training sets. We generated a training set of moving bars of various orientations, motion directions (orthogonal to their orientations), speeds, and widths, and also controlled the polarity of the events of our bars (white and black bars on a grey background). Our dataset comprises 144 sequences lasting from 1 to 4 seconds. The bar orientations covered 180°. Table 3 gives the parameters of the dataset.

To verify that the network can learn proper receptive fields from natural images, we create event sequences from natural images of the Van Hateren database [15]. Concretely, we move images with a fixed speed of 0.1 pixels/ms in a set of 32 possible directions (following the orientations listed in Tab. 3) to generate events. Our natural images dataset comprises 2000 sequences of 800 ms duration.

Figure 2 (left) shows bar and natural images stimuli.

Surround suppression test sets. For testing surround suppression effects, we create bar stimuli that increase in length from 1 to 75 pixels and have orientations between 0–180° as in the training set. Bars move in the two directions orthogonal to their orientation. In order to account for the variability of the bars training set, the test set also contains bars of the same widths listed in Table 3 with speeds of 0.1, 0.3 and 0.5 pixels/ms.

To evaluate our network on real events, we also recorded event sequences with a DVS event camera (DAVIS 346). We capture real moving bars by manually translating vertical bars drawn on a sheet of paper in front of the sensor with different speeds (estimated visually to be fast or slow). We obtain different orientations and bar lengths by rotating and cropping these event sequences.

Parameter reconstruction datasets. To assess the quality of stimulus encoding by the PCL network, we create training and test sequences containing bars of defined position, width, length, orientation, and movement direction. We want to evaluate the capacity of PCL to estimate them, and do that from the network's activity as described below. We use 200,000 samples as our training set and 50,000 samples as our test set.

3.5. Parameter reconstruction

The parameter reconstruction considers 30 ms intervals of the network's spiking activity to estimate the parameters of a moving bar (position, width, length, orientation, movement direction). We construct a feature descriptor by filtering the simple cell spike trains with a Gaussian kernel. For each neuron we calculate:

$$G(t) = \frac{1}{n} \sum_{i=1}^{n} e^{\frac{(t_i - t)^2}{2\sigma^2}} , \tag{10}$$

where n is the number of spikes in the sample, t_i is the time of the i-th spike, and σ is the standard deviation of the Gaussian kernel, which is equal to 5 ms here. We combine the G-values of each neuron into a vector of length $9 \times 9 \times 64$. We also perform a min-max normalization of the vector as it improves the results. The duration of the sequences was arbitrarily chosen to cover a long time scale even if the contribution of far-away spikes is marginal.

A single vector of G-values serves as input to a 3-layered fully connected network to estimate the bar parameters. We use 10 hidden units for the estimation of the width and 100 hidden units for the rest of the parameters. These values were found empirically to give good results while keeping computational costs low. We use the ReLU activation function for all layers. We use an MSE loss function for the length and the position of a bar, and a cross-entropy loss for the remaining parameters. The optimization is done with the ADAM [17] optimizer with standard PyTorch parameters. Training lasts 10 epochs using a batch size of 32.

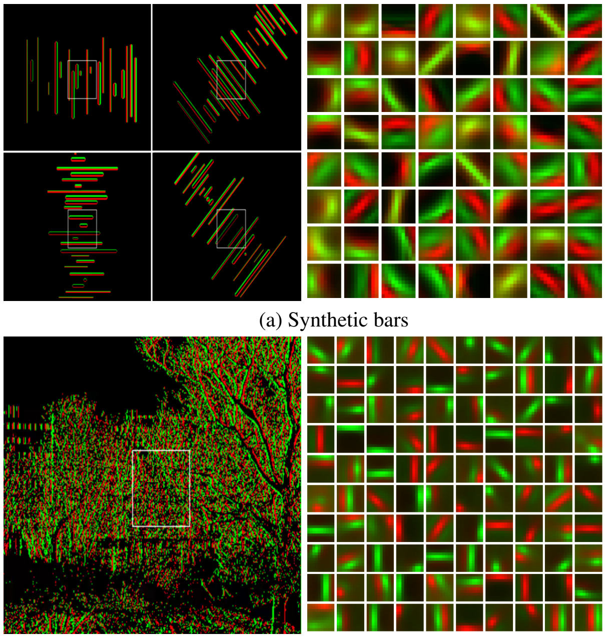
The neural network is trained to perform a regression for the position and the length of the bar and a classification for the remaining parameters. For the sake of convenience when presenting results, we convert the errors during estimation of bar position and length also into an accuracy score via a linear transformation. Specifically, we define an estimation error of 0 pixel to correspond to 100% and an error of 20 pixels (2 receptive field widths) to correspond to 0% accuracy.

4. Results

Previous predictive coding models have aimed to explain visual receptive field properties and surround suppression effects. The latter are typically verified by observing the responses of individual neurons to sequences of moving bar stimuli of different lengths that match the neuron's preferred orientation. Here we proceed in a similar way. We first show that our network learns simple and complex cell receptive fields. We then demonstrate surround suppression effects. Finally, for different versions of the network with and without different kinds of inhibition, we estimate stimulus parameters such as the position, width, length, orientation, and direction of a moving bar from the network's simple cells' spikes.

4.1. Learning receptive fields

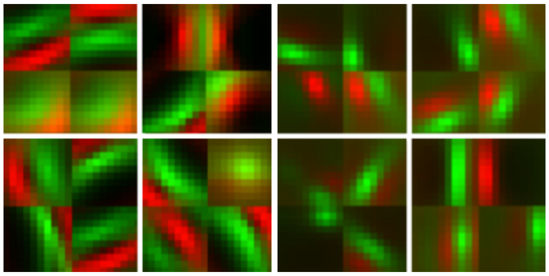
When training our network, all synaptic connections are learning at the same time. Thus, the simple and complex cells' excitatory feedforward connectivity develop at the same time as the recurrent inhibition. The neuron parameters are summarized in Tab. 2. Figure 2 shows a sample of the datasets and the learnt simple cell receptive fields. They resemble Gabor functions and exhibit a preference towards a particular orientation. For natural images, we observe a strong dominance of vertical and horizontal receptive fields. When we increased the number of simple cells to 100 we obtained a more diverse set of receptive fields covering many different orientations. As buildings, trees trunks, branches and wild grass are numerous in the dataset and contain many horizontally and vertically oriented patterns, the cells become more tuned to those orientations overall. This is reminiscent of the well-known oblique effect [1]. Increasing the number of simple cells helps learning a more diverse set of receptive field orientations. Another difference in the natural image condition is that many receptive fields only show a single elongated lobe of one polarity (green or red, corresponding to on or off). We speculate that this may be due to larger structures in the natural images, that mostly generate only one kind of event on their leading edge and the other kind of event on their (relatively distant) trailing edge.



(a) Synthetic bars

(b) Natural images

Figure 2. Learning of simple cell-like receptive fields for synthetic bar stimuli (a, top) and natural images (b, bottom). Left: Examples of event-based visual input sequences used during training. The rectangles in the middle show the input zones of the simple cells. Right: Learnt simple cell-like receptive fields.



(a) Synthetic bars network (b) Natural images network

Figure 3. Visualization of complex cell receptive fields for the network trained with synthetic bar stimuli (a, left) and natural images (b, right). In each case we visualize four example complex cell receptive fields by showing the 4 simple cell receptive fields that had the strongest excitatory connections to these complex cells.

Visualizing the receptive fields of complex cells is more difficult, since each complex cell combines responses of many simple cells with different receptive field locations in a nonlinear fashion. Figure 3 visualizes some complex cell receptive fields by showing the four simple cell receptive fields (one for each of the four spatial locations from where the complex cell receives simple cell input) that have the strongest connections to the example complex cells. Complex cells often tend to receive strong input from simple cells with similar orientation but different phase, making the complex cell responses more invariant to input polarity as can be seen in the example complex cell in the top left corner of panel (a) of Fig. 3.

4.2. Surround suppression

Having characterized the receptive fields in our simple and complex cell layers, we test whether PCL gives rise to surround suppression effects, which are considered a signature of predictive coding. We begin by testing for surround suppression effects in the network that is trained and tested with synthetic bar stimuli.

Synthetic bars. Figure 4a shows the results obtained with synthetic bars. With both types of plastic inhibition, we observe a suppressive effect around 50%, calculated from finding the percentage decrease from the peak to the response at maximum length. Both the lateral and the top-down inhibition contribute to the suppression. The plastic lateral inhibition gives the highest suppression effect of 40% while the plastic top-down inhibition leads to only 17%. This may be due to the relatively small number of complex cells in the network. However, the top-down inhibition significantly reduces the maximum spiking response compared to a network without any plastic inhibition. In contrast, the lateral inhibition has a stronger effect in terms of decreasing spiking activity with growing bar length. Their combination induces an overall strong suppressive effect with a distinct decrease in spiking with bar length.

Real bars. Next, we validate the surround suppression results by testing the network with real bar stimuli recorded with the event-based camera. Figure 4b shows the results. With both types of plastic inhibition, the network still exhibits surround suppression with a suppressive effect of 57%. For bars longer than the classical receptive field size, the simple cells' responses decrease with bar length, although the effect is not as pronounced as for the synthetic stimuli, on which the network was trained. Overall, the effects of plastic lateral and top-down inhibition are similar to those observed for synthetic stimuli. However, the results are somewhat more variable (note the larger error bars) because the process of event generation is less controlled and noisier for the real bar input.

Taken together, these results demonstrate that the PCL network learns simple and complex cell-like receptive fields and shows surround suppression effects as observed biologically. These biological features are considered a consequence of the brain attempting to learn an efficient code for its visual inputs, suggesting that the PCL network also manages to learn an efficient code.

4.3. Parameter reconstruction

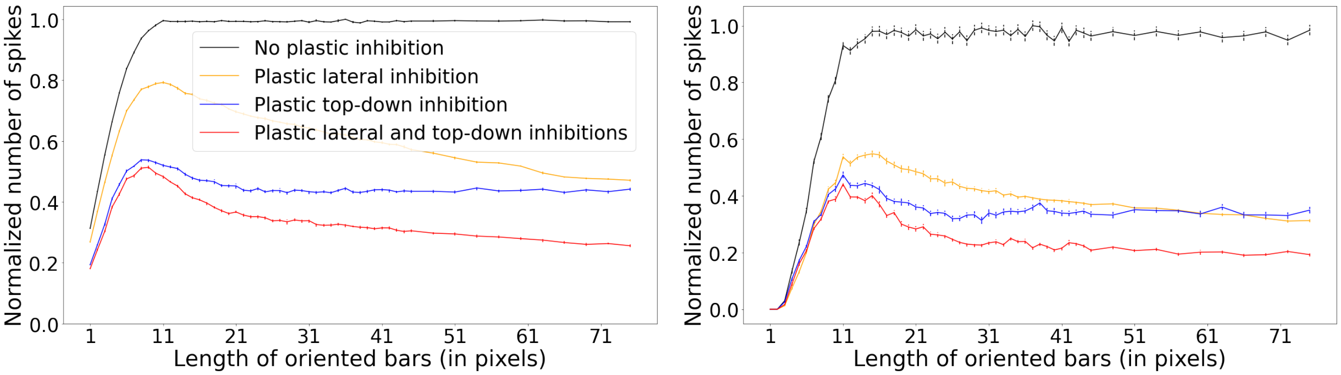
In this section, we try to test more directly if the PCL network learns an efficient code. Our reasoning is as follows: If the PCL network encodes its visual inputs efficiently (in terms of energy use) by removing spikes that are highly predictable, then it should retain much information about these inputs with a comparatively small number of spikes. In order to test this, we train a neural network to reconstruct the parameters (position, width, length, direction, orientation) of our bar stimuli from the simple cells' spikes. If these input parameters can be well reconstructed, this implies that the network has retained a high amount of information about the stimuli. Furthermore, we can compare networks with the learned plastic inhibition to networks without inhibition where we remove a similar number of spikes *randomly*. Our prediction is that in the latter case the network will perform worse, because the random removal of spikes will target both redundant and highly informative spikes equally.

To test this hypothesis we evaluate a network with random inhibition, learnt inhibition, and no inhibition. We simulate the random inhibition mechanism by randomly deleting some spikes in the simple cell spike trains to match the activity level in the network with plastic inhibition. This is analogous to lowering the black curve in Fig. 4a to the niveau of the red curve for each bar length by removing randomly selected spikes.

Figure 5 shows the results. In all cases, the learnt inhibition performs better than the random inhibition. The difference between the baseline classification results (without inhibition and thus, with more spikes) and the learnt inhibition is on average equal to 5%. On the other hand, the performance drop with the random inhibition is on average 2.5 times larger. These results show that the total number of spikes is important for the coding of information. Furthermore, they also suggest that the inhibitory STDP succeeds in preferentially removing more redundant spikes, saving energy with only moderate information loss.

5. Discussion

We have proposed *Predictive Coding Light* (PCL), a hierarchical spiking neural network architecture that learns to efficiently encode signals from an event-based camera. PCL relies on excitatory STDP to learn its feedforward connectivity and inhibitory STDP to learn its recurrent and top-down connectivity. The only exception to this is the *static inhibition*. It is an interesting open question if this could also be replaced by a form of plastic inhibition based on a local learning rule such as a form of iSTDP. The different



(a) Synthetic bars

(b) Real bars

Figure 4. Demonstration of surround suppression. We plot the normalized average number of spikes per sequence (about 5 s for synthetic bars and 1 s for real bars) produced by simple cells in different versions of the network when stimulated with moving oriented bars of different lengths. The networks with plastic inhibition exhibit surround suppression: Beyond a certain bar length corresponding to the size of the simple cells' classical receptive field, the spiking activity is reduced, matching biological findings. If plastic inhibitory connections are removed (black curve), the effect disappears. Results have been averaged over 64 simple cells in a single network with activation and inactivation of the plastic inhibition mechanisms. Each simple cell was stimulated with bars matching the preferred orientation of the cell. Error bars indicate the standard deviation across 5 (4) different sequences for synthetic (real) bars.

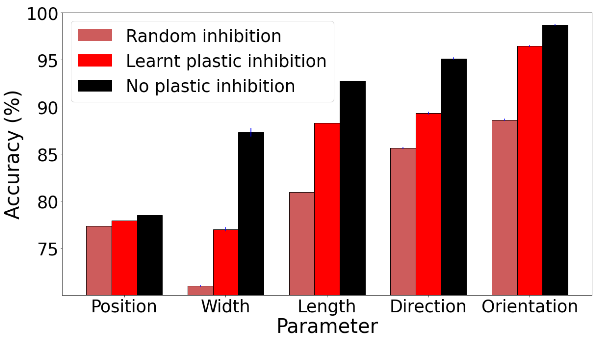


Figure 5. Comparison of parameter reconstruction. Error bars indicate the standard deviation across 3 simulations on separately trained classifier networks.

forms of inhibition allow the network to more than halve the number of spikes used to encode visual inputs (cf. Fig. 4) with only moderate loss of coding fidelity (5% drop in accuracy during parameter reconstruction, cf. Fig. 5). Such savings are a key advantage of efficient coding schemes [5, 7]. It should be noted, however, that the simple visual input consisting of oriented bars is highly compressible. Although we have mainly evaluated PCL on synthetic data in order to have better control over the input, we have also validated the approach with a real event-based camera observing similar reductions in spike rates.

While we took an engineering stance in this work, we hope that PCL may also help to better understand efficient information coding in visual cortex. In particular, we have demonstrated the learning of simple and complex cell-like receptive fields and demonstrated surround suppression effects similar to what is observed in primary visual cortex. The network also exhibits so-called cross-orientation suppression, which we hope to cover in a future publication. However, it must also be noted that we have made some gross simplifications in PCL compared to biological reality. This includes the absence of a distinction between excitatory and inhibitory neurons and a lack of recurrent excitatory connectivity. Adding recurrent excitation could help to explain phenomena such as a fading memory for recent visual inputs and the propagation of evidence during perceptual inference.

In the future, it will also be interesting to construct deep PCL architectures and explore their potential for problems such as energy efficient object recognition [12, 16]. Furthermore, even though we focused on vision here, PCL is generic and could be applied to other event-based data, e.g., from silicon cochleas, or even multimodal inputs. Furthermore, PCL uses a simple, fully spiking neural network architecture and only local “Hebbian” learning rules. This makes it suitable for implementation on neuromorphic hardware, a topic we would also like to explore in the future. Similarly, we would like to apply PCL on robotic platforms to build spiking implementations of fully self-calibrating active vision systems [18].

References

[1] Stuart Appelle. Perception and discrimination as a function of stimulus orientation: the” oblique effect” in man and animals. *Psychological bulletin*, 78(4):266, 1972. 6

[2] David Attwell and Simon B. Laughlin. An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral*

Blood Flow & Metabolism, 21(10):1133–1145, 2001. PMID: 11598490. 1

[3] Dana H Ballard, Rajesh PN Rao, and Zuohua Zhang. A single-spike model of predictive coding. *Neurocomputing*, 32:17–23, 2000. 1

[4] Thomas Barbier, Céline Teulière, and Jochen Triesch. Spike timing-based unsupervised learning of orientation, disparity, and motion representations in a spiking neural network. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1377–1386, 2021. 2

[5] Horace B Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1(01):217–233, 1961. 1, 8

[6] Yin Bi and Yiannis Andreopoulos. PIX2NVS: Parameterized conversion of pixel-domain video frames to neuromorphic vision streams. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1990–1994, 2017. 5

[7] Wieland Brendel, Ralph Bourdoukan, Pietro Vertechi, Christian K Machens, and Sophie Denéve. Learning to represent signals spike by spike. *PLoS computational biology*, 16(3):e1007692, 2020. 1, 2, 8

[8] Kendra S Burbank. Mirrored STDP implements autoencoder learning in a network of spiking neurons. *PLoS computational biology*, 11(12):e1004566, 2015. 2

[9] Matteo Carandini and David Heeger. Normalization as a canonical neural computation. nat. *Nature reviews. Neuroscience*, 13:51–62, 11 2011. 2

[10] Tushar Chauhan, Timothée Masquelier, and Benoit R Cottereau. Sub-optimality of the early visual system explained through biologically plausible plasticity. *Frontiers in Neuroscience*, 15:727448, 2021. 2

[11] Mike Davies, Narayan Srinivasa, Tsung-Han Lin, Gautham Chinya, Yongqiang Cao, Sri Harsha Choday, Georgios Dimou, Prasad Joshi, Nabil Imam, Shweta Jain, Yuyun Liao, Chit-Kwan Lin, Andrew Lines, Ruokun Liu, Deepak Mathaikutty, Steven McCoy, Arnab Paul, Jonathan Tse, Guruguhanathan Venkataramanan, Yi-Hsin Weng, Andreas Wild, Yoonseok Yang, and Hong Wang. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro*, 38(1):82–99, 2018. 1

[12] Pierre Falez, Pierre Tirilly, Ioan Marius Bilasco, Philippe Devienne, and Pierre Boulet. Multi-layered spiking neural network with target timestamp threshold adaptation and STDP. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2019. 8

[13] Karl Friston and Stefan Kiebel. Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences*, 364(1521):1211–1221, 2009. 1

[14] Guillermo Gallego, Tobi Delbrück, G. Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J. Davison, Jörg Conradt, Kostas Daniilidis, and Davide Scaramuzza. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:154–180, 2019. 1

[15] J. H. van Hateren and A. van der Schaaf. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings: Biological Sciences*, 265(1394):359–366, Mar 1998. 5

[16] Saeed Reza Kheradpisheh, Mohammad Ganjtabesh, Simon J. Thorpe, and Timothée Masquelier. STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, 99:56–67, 2018. 8

[17] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014. 6

[18] Alexander Lelais, Jonas Mahn, Vikram Narayan, Chong Zhang, Bertram E Shi, and Jochen Triesch. Autonomous development of active binocular and motion vision through active efficient coding. *Frontiers in neurorobotics*, 13:49, 2019. 8

[19] Peter Lennie. The cost of cortical computation. *Current Biology*, 13(6):493–497, 2003. 1

[20] Yao Li and Lai-Sang Young. Unraveling the mechanisms of surround suppression in early visual processing. *PLoS computational biology*, 17(4):e1008916, 2021. 2

[21] Haihua Liu, Na Shu, Qiling Tang, and Wensheng Zhang. Computational model based on neural network of visual cortex for human action recognition. *IEEE transactions on neural networks and learning systems*, 29(5):1427–1440, 2017. 2

[22] Fabian A. Mikulasch, Lucas Rudelt, Michael Wibral, and Viola Priesemann. Where is the error? hierarchical predictive coding through dendritic error computation. *Trends in Neurosciences*, 46(1):45–59, Jan 2023. 1, 2

[23] Hirofumi Ozeki, Ian M Finn, Evan S Schaffer, Kenneth D Miller, and David Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–592, 2009. 2

[24] Byron H. Price and Jeffrey P. Gavornik. Efficient temporal coding in the early visual system: Existing evidence and future directions. *Frontiers in Computational Neuroscience*, 16, 2022. 1

[25] Rajesh Rao and Dana Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2:79–87, 02 1999. 1, 2

[26] Robert Sachdev, Matthew Krause, and James Mazer. Surround suppression and sparse coding in visual and barrel cortices. *Frontiers in Neural Circuits*, 6, 2012. 2

[27] Cristina Savin, Prashant Joshi, and Jochen Triesch. Independent component analysis in spiking neurons. *PLoS computational biology*, 6(4):e1000757, 2010. 2

[28] David Somers, E Todorov, Athanassios Siapas, L Toth, D Kim, and M Sur. A local circuit approach to understanding integration of long-range inputs in primary visual cortex. *Cerebral cortex (New York, N.Y. : 1991)*, 8:204–17, 04 1998. 2

[29] Michael Spratling. Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, 2, 2008. 1

[30] Martin Stemmler, Marius Usher, and Ernst Niebur. Lateral interactions in primary visual cortex: A model bridging physiology and psychophysics. *Science*, 269(5232):1877–1880, 1995. 2

[31] Jochen Triesch, Anh Duong Vo, and Anne-Sophie Hafner. Competition for synaptic building blocks shapes synaptic plasticity. *Elife*, 7:e37836, 2018. 4

[32] Hermann Von Helmholtz. *Handbuch der physiologischen Optik*, volume 9. Voss, 1867. 2

[33] Mengchen Zhu and Christopher Rozell. Visual nonclassical receptive field effects emerge from sparse coding in a dynamical system. *PLoS computational biology*, 9:e1003191, 08 2013. 2

[34] Joel Zylberberg, Jason Timothy Murphy, and Michael Robert DeWeese. A sparse coding model with synaptically local plasticity and spiking neurons can account for the diverse shapes of v1 simple cell receptive fields. *PLoS computational biology*, 7(10):e1002250, 2011. 2