A Marker-based Model for the Ontogenesis of Routing Circuits

Philipp Wolfrum[1] and Christoph von der Malsburg[1,2]

[1] Frankfurt Institute for Advanced Studies, D-60438 Frankfurt am Main, Germany
[2] Computer Science Dept., University of Southern California, LA 90089-2520, USA

Abstract. We present a model for the ontogenesis of information routing architectures in the brain based on chemical markers guiding axon growth. The model produces all-to-all connectivity between given populations of input and output nodes using a minimum of cortical resources (links and intermediate nodes). The resulting structures are similar to architectures proposed in the literature, but with interesting qualitative differences making them biologically more plausible.
Keywords: information routing, shifter circuits, dynamic links, visual cortex

1 Introduction

An important part of brain function is the routing of information between different areas. The routes along which information flows cannot be static, but must be adaptable to the current requirements. The most prominent example for this necessity is visual attention, where a certain mechanism ensures that only a selected portion of the visual input reaches higher visual "target areas" like inferotemporal cortex (IT). Other examples in which information routing may be very useful include pitch-invariant recognition of melodies, or our ability to combine arbitrary words into grammatically correct sentences. Such abilities require routing structures that provide physical connections between all locations in a certain input region and all locations of a target area.

The necessity for dynamic information routing was appreciated early on [1], and models for its use in object recognition [2] and for frame-of-reference transforms [3] have been put forward. All-to-all routing between large cortical areas has to happen via intermediate stages to be biologically plausible (see problem definition in Sect. 2 and [4] for a detailed discussion). Several architectures for such a multi-stage routing have been proposed, like Shifter Circuits [5], the SCAN model [6], or the minimal architecture of [4]. What has been missing so far are models explaining the ontogenetic development of routing structures in the brain.

2 Routing structures

Let us pose the information routing problem as follows:

- Given are an input layer and an output layer both consisting of N feature nodes (or simply *nodes*). We are looking for a routing network that establishes all-to-all connectivity between those layers.

– The routing happens via $K-1$ intermediate layers of N nodes each.
– Nodes of adjacent feature layers can be connected by *links*. For connecting the $K+1$ feature layers, K stages of links are required, every stage containing N^2 potential links.

Several anatomically plausible architectures have been proposed that meet these requirements. The most prominent one is the so called *Shifter Circuit* [5]. While Shifter Circuits implement a redundant connectivity between input and output, in [4] we propose an architecture that provides full connectivity while requiring the minimally possible number of feature nodes and links. A one-dimensional version of this connectivity is shown in Fig. 1.

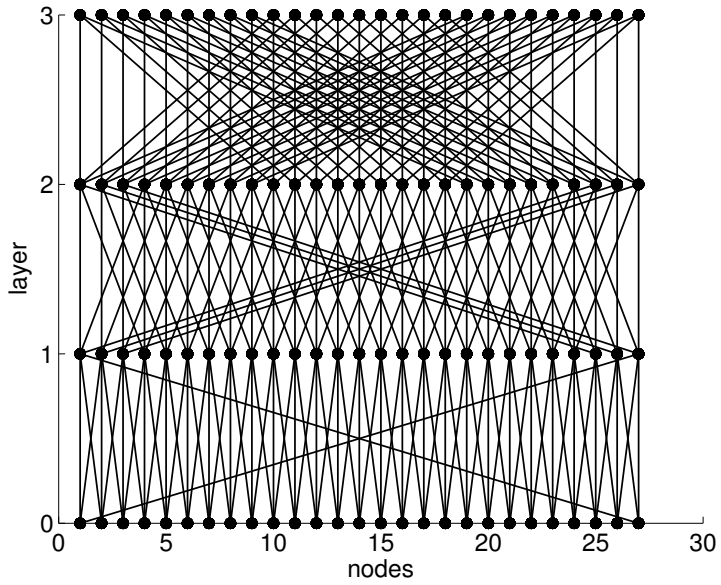


Fig. 1. Routing architecture from [4]. The $N = 27$ nodes of the input layer 0 are connected to all 27 nodes of output layer 3 via 2 intermediate layers and $K = 3$ stages of links. Note that this connectivity requires "wrap-around" links, i.e. links between one end of the presynaptic layer and the opposite end of the postsynaptic layer.

3 Ontogenetic Dynamics

How can ontogeny produce such routing circuits in the brain? Especially the large gaps necessary between links on higher stages are difficult to explain with traditional learning rules. Here we will investigate whether chemical markers can help forming such structures.

It is well known that axonal growth follows chemical gradients [7]. It was hypothesized early that chemical markers could help forming the point-to-point retinotopic mapping that exists between retina and the tectum [8], and Willshaw and von der Malsburg [9] presented a model that realizes the retinotectal mapping on the basis of a limited number of chemical markers present in the retina. Recent studies [10] have shown that the mechanisms by which axons detect these gradients are much more sensitive than previously assumed, allowing the question whether even more complicated patterns than the retinotectal map can arise from chemical marker interaction.

We here present a model that explains the development of routing structures based on chemical markers. For simplicity and ease of visualization we restrict ourselves here to the case of one-dimensional feature layers. Let $C^k_{i,j}$ denote the strength of the link between node i in layer k to node j of layer $k+1$. $C^k_{i,j}$ can vary between 0 and 1, with 0 representing an absent link and 1 a fully grown one. We will refer to all links of one stage by the $N \times N$ matrix C^k. When we make a statement that refers to the links of all K stages we will leave out the superscript k (this also applies for other variables introduced below).

We describe the growth of the links not directly in terms of C but of an unbounded variable U, which codes for the real links via the sigmoid function

$$C = \frac{1}{1 + e^{-sU}} \quad , \tag{1}$$

where s defines the steepness of the sigmoid. We let U start out at a homogeneous negative value with some noise added (see Sect. 4 for a discussion of robustness to noise), so that all links C are initially close to 0. The growth of U then follows the differential equation

$$\dot{U} = F^{\text{norm}} \times F^{\text{marker}} \times F^{\text{top}}, \tag{2}$$

where $\times$ denotes elementwise multiplication. The three terms have the roles of restraining local growth of connections (F^{norm}), keeping similarity of chemical markers on both sides of a link low (F^{marker}), and introducing topologic interactions (F^{top}). Thanks to the multiplicative combination, no "tuning" of the relative contributions of the terms is required; the mechanism works for different network sizes without need for adjusting many parameters.

The term

$$F^{\text{norm}}_{i,j} = d - \sum_{\tilde{j}} C_{i,\tilde{j}} \tag{3}$$

is a factor that tends to keep the sum of all efferent links from any position i close to a desired value d. Once the combined link strengths exceed d, F^{norm} turns negative, thus letting the respective link shrink.

The term F^{marker} makes a link's change sensitive to the similarity of chemical markers in the two nodes it connects. These markers are channeled from the input layer to higher levels by the very connectivity C whose growth in turn they influence. We assume each node of the input layer to contain a different type of chemical marker t_i (for a discussion of the plausibility of this and possible alternatives, see Sect. 5). In

matrix notation this means that the marker distribution in layer 0 is the identity matrix, $M^0 = I_{N\times N}$, with the marker types on the 1st and the node location on the 2nd dimension. Markers are then transported to higher layers via the existing links C:

$$M^{k+1} = M^k C^k \ . \tag{4}$$

To calculate F^{marker}, we first define a similarity term

$$F^{\text{sim},k}_{i,j} = \sum_t M^k_{t,i}(M^{k+1}_{t,j} - C^k_{i,j} M^k_{t,i}) \ , \tag{5}$$

which is the similarity (dot product) of the marker vector on the presynaptic side with that portion of the marker vector on the postsynaptic side that was not carried there by the link itself. Therefore, the similarity term signals to the link how well the routes between the part of input space it "sees" and its target node are already being served by other links (see Fig. 2). The role of F^{marker} is to let a link grow only if its similarity term is not too large. We therefore set

$$F^{\text{marker}}_{i,j} = 1 - H(F^{\text{sim}}_{i,j} - \alpha) \ , \tag{6}$$

with $H(\cdot)$ denoting the Heaviside function and a fixed parameter α.

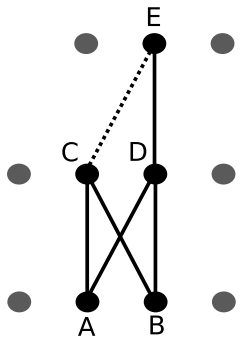


Fig. 2. Role of the similarity term. Already well-established links (solid lines) carry markers from input nodes A and B to intermediate nodes C and D, and from D to E. Therefore, a weak link C-E (dotted line) finds a marker distribution at its target E that is similar to the one at its origin C. This similarity keeps it from growing. Functionally, this mechanism prevents formation of redundant alternative routes between two points.

The term

$$F^{\text{top}}_{i,j} = \beta(C_{i-1,j-1} + C_{i+1,j+1}) + G_{i,j} \tag{7}$$

combines two different topological influences, their relative strength weighted by the parameter β. The first part adds cooperation between parallel neighboring links. The second term G favors the growth of links to the corresponding position in the next layer (i.e. $i = j$) over links to faraway positions. We assume it here to be a bounded

hyperbolic function of position difference of the two end nodes: $G_{i,j} = \frac{\gamma}{|i-j|+\gamma}$, with γ defining the steepness (see Fig. 3). G is necessary to tell the ontogenetic mechanism how to align the coordinate systems of the layers it is connecting. A possible way of implementing this term is to first allow development of a point-to-point mapping (e.g. through the mechanism presented in [9]), which then serves as a guidance for the growth of a routing connectivity.

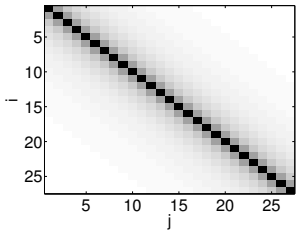


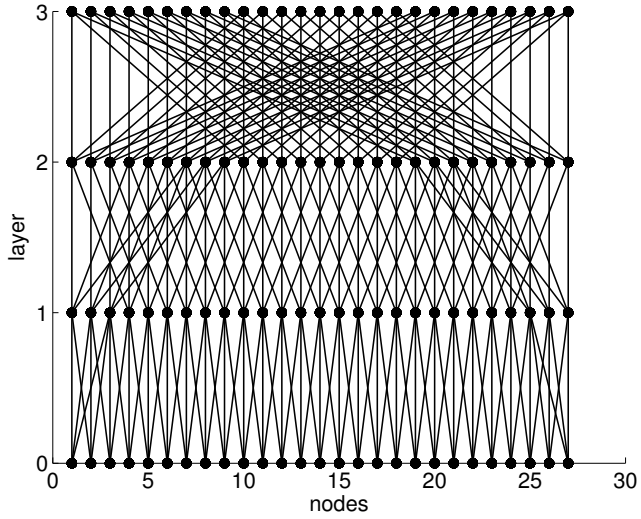
Fig. 3. The term G helps to align the coordinate systems of subsequent layers by favoring links between nodes with corresponding positions (middle diagonal) over links between distant nodes. $\gamma = 0.6$ like in the simulations of Sect. 4.

4 Results

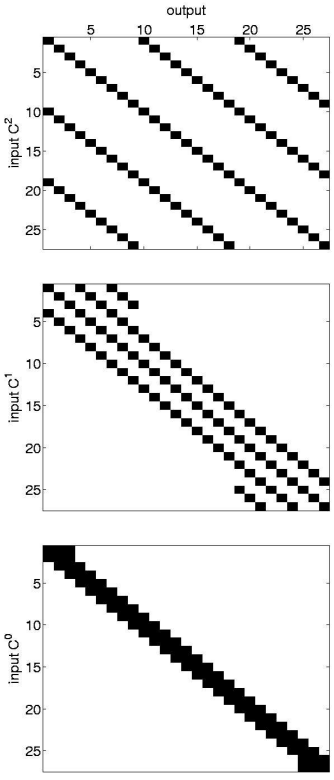
We chose to investigate the growth of networks containing $K = 3$ link stages. Equation (2) was integrated using the Euler method and the following parameter settings: $s = 30$ (steepness of sigmoid), $\alpha = 0.5$ (threshold for marker similarity), $\beta = 0.6$ (strength of neighbor interaction), $\gamma = 0.6$ (steepness of the hyperbolic term G). A delayed onset of growth at higher stages improves the final results. We chose a delay of 15% and 30% of overall simulation time for the middle and the highest stage, respectively.

First, we assumed $d = 3$ as target number of links per node (see (3)). With $K = 3$ link stages, this means that $N = d^K = 27$ input and output nodes can be connected. The network resulting from the ontogenetic mechanism for these parameter settings is shown in Fig. 4. Note how the distance between links increases from 1 to 3 to 9 from bottom to top, thus producing non-redundant full connectivity. We can see in Fig. 4(a) that the resulting network differs qualitatively from the manually produced one of Fig. 1: There are no wrap-around links (i.e. links from a node on one side of the feature layer to the opposite side of the next layer). Instead, these links appear on the other side of the central link (cf. Fig. 4(b)). Interestingly, this new structure produces the same perfect all-to-all connectivity as the one arising from theoretical considerations in [4], while being biologically more plausible.

The mechanism can also grow routing structures between larger feature layers. For this we only have to adjust the target number of links per node d, without changing any of the other parameters. Fig. 5 shows simulation results for $d = 5$, i.e. $N = d^3 = 125$ nodes per layer. We see that qualitatively the resulting structure is similar to the one



(a) Connection structure of the full network.



(b) Matrices C^k of the full network of (a) shown separately.

Fig. 4. Results for $N = 27$ nodes per layer and a target number of links $d = 3$.

obtained for $d = 3$, except that now each node makes 5 connections to the next layer, with appropriate spacings of 1, 5, and 25 nodes.

However, we also see that the structure in Fig. 5 is not as clean as the one in Fig. 4, with several links not going to the "correct" targets. This results in an overall input-output connectivity that is not perfectly homogeneous, i.e. some input-output pairs are connected by two different routes, while others have no connection. For the structure shown in Fig. 5, the strengths of the input-output connections have mean value and standard deviation of $\mu \approx 1$ and $\sigma \approx 0.15$.

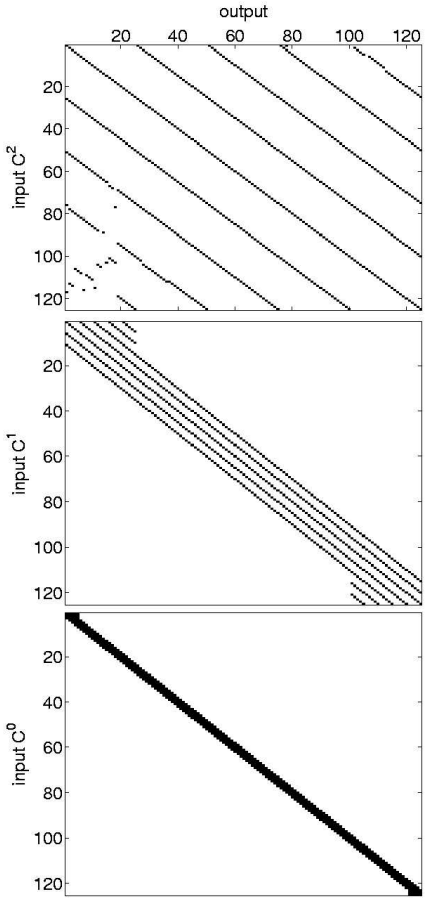


Fig. 5. Resulting connection matrices C^k for $N = 125$ nodes per layer and a target number of links $d = 5$. The initial values of U contained 10% of additive noise.

The reason for the uneven final structure lies in the noise that was introduced to the initial link strengths: We chose the initial values of U randomly from the interval

$[-16.5.. -15]$, which means that they contain 10% of additive uniformly distributed noise. Further simulations have shown that the mechanism generally results in a flawless connectivity only if the initial conditions contain less than $\approx 5\%$ of noise. The growth of smaller networks is far less sensitive to noise: For $N = 27$, up to 20% of additive noise in the initial conditions practically always results in the correct final connectivity.

5 Conclusions

One assumption crucial for the model as presented here is that there be a unique chemical marker for every input node. This is very unlikely to be the case in the brain. Future work will address the question how the N different chemical markers can be replaced by a small number of marker gradients spread evenly over the input layer. Also, it is possible to replace the unique markers assumed here by stochastic, uncorrelated signals produced by the different input nodes. Since these signals are orthogonal, their linear superpositions arising at higher layers could still be decomposed and, most notably, the scalar product of such superpositions would give exactly the same similarity term as assumed in the model here. This would yield an activity-based instead of a marker-based mechanism following the same mathematical model.

We have presented a neurally plausible ontogenetic mechanism modeling the formation of routing circuits in the brain. The mechanism requires only signals that are available locally at source and/or target of the respective connection. While the mechanism may be important for understanding the development of certain wiring structures of the brain, it may also turn out to have technological applications like the automatic wiring of computer networks.

Acknowledgments

This work was supported by the European Union through project FP6-2005-015803 ("Daisy") and by the Hertie Foundation.

References

1. Pitts, W., McCulloch, W.S.: How we know universals: the perception of auditory and visual forms. Bulletin of Mathematical Biophysics **9** (1947) 127–147
2. Lades, M., Vorbrüggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on computers **42** (1993) 300–311
3. Weber, C., Wermter, S.: A self-organizing map of Sigma-Pi units. Neurocomputing (2007) (in press)
4. Wolfrum, P., von der Malsburg, C.: What is the optimal architecture for visual information routing? Neural Computation (2007) (in press).
5. Olshausen, B.A., Anderson, C.H., van Essen, D.C.: A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. Journal of Neuroscience **13**(11) (1993) 4700–4719

6. Postma, E., van den Herik, H., Hudson, P.: SCAN: A Scalable Model of Attentional Selection. Neural Netw **10**(6) (1997) 993–1015
7. Yamamoto, N., Tamada, A., Murakami, F.: Wiring of the brain by a range of guidance cues. Prog Neurobiol **68**(6) (2002) 393–407
8. Sperry, R.W.: Chemoaffinity in the orderly growth of nerve fiber patterns and connections. Proc Natl Acad Sci U S A **50** (1963) 703–710
9. Willshaw, D.J., von der Malsburg, C.: A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. Philos Trans R Soc Lond B Biol Sci **287**(1021) (1979) 203–243
10. Rosoff, W.J., Urbach, J.S., Esrick, M.A., McAllister, R.G., Richards, L.J., Goodhill, G.J.: A new chemotaxis assay shows the extreme sensitivity of axons to molecular gradients. Nature Neuroscience **7** (2004) 678 – 682