Phase Diagrams of Self-Organizing Maps

H.-U. Bauer, M. Riesenhuber*, T. Geisel
Institut für Theoretische Physik und SFB Nichtlineare Dynamik,
Universität Frankfurt, D-60054 Frankfurt/Main, Germany
(May 15, 1996)

We present a method which allows the analytic determination of phase diagrams in the Self-Organizing Map (SOM), a model for the formation of topographic projection patterns in the brain and in signal processing applications. The method only requires an ansatz for the tesselation of the data space induced by the map, not for the explicit state of the map. We analytically obtain phase diagrams for various examples, including models for the development of orientation and ocular dominance maps. The latter phase diagram exhibits transitions to broadening ocular dominance patterns as observed in a recent experiment.

PACS numbers: 87.10.+e, 89.70.+c, 05.90.+m

Topographic maps occur in many areas of the brain where sensory and other information is represented topographically, as well as in signal processing applications where data points are projected from one space to another in a neighborhood preserving fashion. An archetypical example involves the projection of oriented edge elements to the visual cortex where neighboring neurons respond to edges of similar orientation, at neighboring positions in the visual field [1]. Topographic maps were found to be most often generated or refined by externally driven self-organization processes [2]. Among the numerous models [3–6] for these pattern formation phenomena, Kohonen's Self-Organizing Map (SOM) [7–9] has found particularly wide distribution. In the domain of technical signal processing, the low-dimensional "Feature Map" variant of the SOM is utilized for neighborhood preserving vector quantization, motor control [10,11], data visualization [12], or speech data preprocessing (for many further examples see [8,9]). Studies of self-organization in the brain are often based on the slightly different high-dimensional SOM version. Here stimuli and receptive fields are described not in terms of prespecified "features", but in terms of (high-dimensional) activity resp. weight distributions [13–15]. This allows for a simultaneous self-organization not only of the map topography, but also of the shapes of individual receptive fields.

The popularity of the SOM is based on its simple formulation, its numerical robustness and the empirical success of its applications. However, due to a strong nonlinearity of this model, a general analytical treatment of the corresponding pattern formation process has been lacking. In particular, the conditions on map and data set parameters for patterns to occur are found most often only empirically, an unsatisfactory and numerically costly procedure. In this letter we present a method to analytically relate map and data set parameters to specific states of SOMs, i.e. to calculate phase diagrams of SOMs. The method is based on a comparison of the distortions of different data space tesselations, i.e. of different ways to distribute the data points among the map elements. Even though the method is applicable to the low-dimensional as well as to the high-dimensional variant of the SOM, it achieves its full potential in the latter case, where an ansatz for the tesselation is comparatively easy, but an ansatz for the map itself is unfeasible. We first apply our method to a tutorial mapping example and then solve two models for map formation in the visual cortex which previously could be investigated only numerically.

A Self-Organizing Map (SOM) consists of nodes (neurons) characterized by a position $\mathbf{r}$ in the map output space lattice and a weight vector (receptive field) $\mathbf{w_r}$ in the map input space, the data space. A data point $\mathbf{v}$ is mapped onto that node $\mathbf{s}$ whose weight vector $\mathbf{w_s}$ matches $\mathbf{v}$ best. This amounts to a winner-take-all rule, a strong nonlinearity which in a biological context is explained as a consequence of lateral inhibition [8]. In the context of technical applications this projection rule is identical to that of a regular vector quantizer [16]. The map results as a stationary state of a self-organization process, which successively changes all vectors $\mathbf{w_r}$,

$$\Delta\mathbf{w_r} = \epsilon h_{\mathbf{rs}}(\mathbf{v} - \mathbf{w_r}), \qquad h_{\mathbf{rs}} = e^{-\frac{||\mathbf{r}-\mathbf{s}||^2}{2\sigma^2}}, \tag{1}$$

following the presentation of stimuli $\mathbf{v}$. ϵ controls the size of learning steps. The neighborhood function $h_{\mathbf{rs}}$ enforces neighboring neurons to align their receptive fields, imposing the property of topography on the SOM.

In the general case, data points $\mathbf{v}$ and receptive field vectors $\mathbf{w_r}$ are activity and weight distributions across M input channels, respectively (normalized to a constant total activity, $\sum_i^M v_i = S$). The winner $\mathbf{s}$ is determined by

$$\mathbf{s} = \arg\max_{\mathbf{r}}(\mathbf{w_r}\cdot\mathbf{v}). \tag{2}$$

The input channels correspond, e.g., to the sensors in a sensory layer, like the retinal ganglion cells as input channels to a visual map. The typically large number of such channels warrants the notion of a "high-dimensional" map. In case the distributions $\mathbf{v}$ and $\mathbf{w_r}$

are replaced by (a small number of) features $\tilde{\mathbf{v}}$, $\tilde{\mathbf{w}}_{\mathbf{r}}$, like the centers of gravity of $\mathbf{v}$ and $\mathbf{w}_{\mathbf{r}}$, one arrives at the low-dimensional SOM-variant. This replacement precludes a self-organization of an internal shape of the $\mathbf{w}_{\mathbf{r}}$, but has the advantage of a drastically reduced numerical expense.

A theoretical analysis of pattern formation in SOMs has only been performed as yet for the low-dimensional version of the model [17–20]. There, a linear approximation of the map dynamics about an equilibrium state led to conditions for this equilibrium state to become unstable. This analysis exploited that in particular mapping geometries equilibrium values of the $\tilde{\mathbf{w}}_{\mathbf{r}}$ trivially are known. In the high-dimensional SOM the equilibrium values of the $\mathbf{w}_{\mathbf{r}}$ depend on stimulus parameters as well as on the map parameter σ in a non-trivial fashion even if the map performs a trivial projection. Since the $\mathbf{w}_{\mathbf{r}}$ are not analytically accessible in this case, an analysis based on the $\mathbf{w}_{\mathbf{r}}$ is not feasible.

Here, we take a different approach which focuses on the data space tesselation, i.e. the distribution of data points among map neurons. Even if small changes of stimulus or map parameters change the $\mathbf{w}_{\mathbf{r}}$ slightly, the tesselation can remain unaffected. We consider it to be a change of the state of the map, only if the tesselation is changed, corresponding, e.g., to a break of symmetry in the receptive fields. In many cases (cf. examples), an ansatz for the tesselations corresponding to these different map states is easy to make.

To evaluate different possible states, we note that the SOM approximately minimizes the distortion function

$$E_{\mathbf{w}} = \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} (\mathbf{v}' - \mathbf{w}_{\mathbf{r}})^2 e^{\left(-\frac{||\mathbf{r}-\mathbf{r}'||^2}{2\sigma^2}\right)}. \qquad (3)$$

$\Omega_{\mathbf{r}}$, the so-called Voronoi cell of neuron $\mathbf{r}$, denotes the set of data points $\mathbf{v}$ which are mapped onto node $\mathbf{r}$ via (2). Even though the SOM learning dynamics does not proceed along the gradient of this function (or any other energy function), the deviations become small in the limit of an ordered map with large values for σ [21]. It is also known that modification of the SOM winner rule lead to a map formation algorithm following exactly the gradient of an energy function [22,23]. Therefore, a sensible strategy to determine the final state of a SOM is to compare distortion functions for different map states. To avoid the $\mathbf{w}_{\mathbf{r}}$-problem inherent to an evaluation of $E_{\mathbf{w}}$, we replace $E_{\mathbf{w}}$ by the related distortion function

$$E_{\mathbf{v}} = \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} \sum_{\mathbf{v} \in \Omega_{\mathbf{r}}} (\mathbf{v}' - \mathbf{v})^2 e^{\left(-\frac{||\mathbf{r}-\mathbf{r}'||^2}{2\sigma^2}\right)}, \qquad (4)$$

which requires knowledge of the data space tesselations $\Omega_{\mathbf{r}}$ only. Under quite general assumptions [24], $E_{\mathbf{v}}$ is approximately related to $E_{\mathbf{w}}$ by a multiplicative factor (to compensate for the additional summation in $E_{\mathbf{v}}$), hence $E_{\mathbf{w}}$ and $E_{\mathbf{v}}$ become minimal for the same sets of parameters.

To elucidate the application of our method, and to illustrate the intricacies of high-dimensional SOMs, we first discuss a simple, tutorial example. Consider the map of a rectangular input space, extensions $4 \times 2s$, $0 < s < 2$, with periodic boundaries in the first and open boundaries in the second direction, onto a ring of 4 neurons. The stimuli are activity distributions of Gaussian shape, with small but finite width. This input space is discretized with $M \times M$ channels per unit square. Hence, the $\mathbf{v}$ and $\mathbf{w}_{\mathbf{r}}$ are sum-normalized $4M \times 2sM$-dimensional vectors, not just two-dimensional pointers as they would be in the feature map version of this example.

Even for simple projection patterns, the explicit values of the $\mathbf{w}_{\mathbf{r}}$ depend on the shape of the stimuli, and on σ, and cannot be determined easily. What are sensible data space tesselations in this example? Since the SOM algorithm tries to group similar stimuli into common Voronoi cells, stimuli will be grouped such that they form connected, simple regions in the input rectangle. One possible tesselation is such that half of the stimuli will be mapped to one neuron, and the rest of the stimuli to the first neuron's next nearest neighbor (state a, see Fig. 1a). The remaining two neurons are never best-matching. Contrary to the low-dimensional feature map, this tesselation is realizable for high-dimensional SOMs (for a numerical example, see Fig. 1e). Other tesselations distribute the stimuli evenly to all four neurons (states b-d, see Fig. 1 b-d, f-h). A tesselation with all stimuli going to just one neuron would not be stable.

When are these states attained? Let us first consider the transition between the two-neuron state a and the four-neuron states b,c,d. In the limit of a small spatial stimulus extension, most pairs $\mathbf{v}, \mathbf{v}'$ do not overlap, and their squared difference takes on a value $(\mathbf{v} - \mathbf{v}')^2 = c_0$. Then $E_{\mathbf{v}}$ can be evaluated approximately,

$$\begin{aligned} E_{\mathbf{v},a} &= \sum_{\mathbf{r}=1,3} \sum_{\mathbf{r}'=1,3} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} \sum_{\mathbf{v} \in \Omega_{\mathbf{r}}} (\mathbf{v} - \mathbf{v}')^2 e^{-\frac{||\mathbf{r}-\mathbf{r}'||^2}{2\sigma^2}} \\ &\approx 2c_0 \left((2M \cdot 2sM)^2 e^0 + (2M \cdot 2sM)^2 e^{\frac{-4}{2\sigma^2}} \right) \qquad (5) \end{aligned}$$

$$\begin{aligned} E_{\mathbf{v},b,c,d} &= \sum_{\mathbf{r}=1}^{4} \sum_{\mathbf{r}'=1}^{4} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} \sum_{\mathbf{v} \in \Omega_{\mathbf{r}}} (\mathbf{v} - \mathbf{v}')^2 e^{-\frac{||\mathbf{r}-\mathbf{r}'||^2}{2\sigma^2}} \\ &\approx 16c_0 \left(s^2M^4 + 2s^2M^4 e^{\frac{-1}{2\sigma^2}} + s^2M^4 e^{\frac{-4}{2\sigma^2}} \right) \qquad (6) \end{aligned}$$

The transition between the two states occurs for

$$e^0 + e^{\frac{-4}{2\sigma^2}} = 2e^{\frac{-1}{2\sigma^2}}, \qquad \text{i.e. for} \quad \sigma \approx 0.91. \qquad (7)$$

For a subsequent distinction between the four-neuron states b-d, the previously neglected impact of stimulus pairs overlapping across boundaries of $\Omega_{\mathbf{r}}$ has to be taken into account (for a more detailed presentation of these calculations, see [24]). Then the condition of minimal

$E_{\mathbf{v}}$ leads to the phase diagram depicted in Fig. 2a. It coincides very well with the results of simulations of the SOM-algorithm for this mapping example (Fig. 2b). The simulations also showed that all the tesselations we considered do indeed occur, but no others.

We now analyze two SOM-models for the development of topographic maps in the visual cortex [13,15], which were previously accessible only numerically. In these models a sensory input space with one resp. two layers of $N \times N$ retinal channels is mapped onto an $N \times N$-neuron output layer, the cortical area. The first model is concerned with the development of orientation maps where neurons respond best to stimuli of a particular orientation and location. Using ellipsoidal Gaussian activity distributions as stimuli (minor axis σ_1, major axis $\sigma_2 > \sigma_1$), simulations led to maps with oriented receptive fields for substantially elongated stimuli [13]. Using rather circular stimuli ($\sigma_2 \approx \sigma_1$) non-oriented receptive fields were also observed. The tesselations corresponding to these states have simple characteristics: in maps with non-oriented receptive fields stimuli of different orientation but same retinal position are grouped into one Voronoi cell. Maps with oriented receptive fields have stimuli of the same orientation but different positions in one Voronoi cell. Using a finite set of stimuli (i.e. discrete orientations and positions), we evaluated and equated the corresponding distortions $E_{\mathbf{v},ori}$ and $E_{\mathbf{v},non-ori}$. This led to an analytic relation for the transition point between the stimulus parameters σ_1, σ_2 and the network parameter σ,

$$\sigma_{2,crit} \approx \sigma_1 + \sqrt{3}\sigma. \qquad (8)$$

Eq. (8) for the break of receptive field symmetry is very well corroborated by numerical simulations using the finite stimulus set employed in the calculations (Fig. 3), as well as by additional simulations with the full stimulus set (all orientations and positions). The additive relation between σ_1 and the critical σ_2 deviates from the multiplicative relation found for a corresponding model in feature map approximation [18].

The second model is concerned with the development of ocular dominance (OD) maps in the visual cortex. As in [15], the input space consists of two retinal layers. Stimuli occur simultaneously at the same position in both retinae, correlated by a factor of c. For large c, neurons do not develop a preference for one retina over the other (no OD). For decreasing values of c, a transition takes place to solutions where each neuron has a preference for one retina (OD). As in the previous example, there is no obvious direct way to obtain the weight vectors $\mathbf{w_r}$ analytically. But again the corresponding stimulus space tesselations are simple: either stimuli from the same positions but from both retinae, or stimuli from the same retina but different position are grouped together in a Voronoi cell. In the latter case, we can make further assumptions about the local spatial arrangement of neurons which prefer the same retina. Among the exhaustively many arrangements evaluated, only three turned out to be of relevance: bands of width 1, bands of width 2, or bands of "infinite" width (i.e. half the system size). Evaluating $E_{\mathbf{v}}(c)$ for the different cases (Fig. 4) yields a transition point $c \approx 0.64$ for the change from OD to non-OD, depending only very slightly on the ansatz for the spatial OD layout. Simulating maps for different values of c we find a steep increase of OD at $c \approx 0.66$, very close to our analytical result.

Evaluating $E_{\mathbf{w}}$ and $E_{\mathbf{v}}$ for these maps yields two very similar curves, indicating the close relationship of the two distortion measures (3) and (4). The deviation between the numerical and analytical values of $E_{\mathbf{v}}$ is a consequence of simplifying assumptions about the tesselations made in order to obtain the analytical values.

Comparison of the curves for the different OD band layouts reveals that with decreasing c layouts of increasing band width are preferred. Even though we consider our ansatz for the different layouts as too crude to be realized precisely in simulated maps, the tendency of increasing band width with decreasing correlation is nevertheless established. This result coincides with the findings of a recent neuroanatomical experiment involving strabismic cats and normal-sighted cats (corresponding to smaller and larger values of c, resp.) [25].

Our method also turned out to be helpful in the derivation and solution of a SOM-based model for the development of orientation maps for the competition of On-center and Off-center cells [26]. This demonstrates the universality of the analytic method in the investigation of any high-dimensional SOM-model, including more complicated mapping problems, which as yet were handicapped by the need for very costly numerical simulations.

This work was supported by the Deutsche Forschungsgemeinschaft.

[1] D. H. Hubel, T. N. Wiesel, J. Comp. Neur. **158**, 267 (1974).

[2] C. Shatz, Neuron **5**, 645 (1990).

[3] C. von der Malsburg, Kybernetik **14**, 85 (1973); Biol. Cyb. **32**, 49 (1979).

[4] R. Durbin, D. J. Willshaw, Nature **326**, 689 (1987).

[5] K. D. Miller, J. B. Keller, M. P. Stryker, Science **245**, 605 (1989).

[6] M. Stetter, A. Müller, E. W. Lang, Phys. Rev. E **50**, 4167 (1994).

[7] T. Kohonen, Biol. Cyb. **43**, 59 (1982);

[8] T. Kohonen, *Self-Organizing Maps*, Springer (1995).

[9] H. Ritter, Th. Martinetz, K. Schulten, *Neural Computation and Self-Organizing Maps*, Addison Wesley (1992).

[10] H. Ritter, T. Martinetz, K. Schulten, Neur. Netw. **2**, 159

(1989).
[11] J. A. Walter, K. Schulten, IEEE Trans. Neur. Netw. **4**, 86 (1993).
[12] A. Dekker, Network **5**, 351 (1994).
[13] K. Obermayer, H. Ritter, K. Schulten, Proc. Nat. Acad. Sci. USA **87**, 8345 (1990).
[14] K. Obermayer, H. Ritter, K. Schulten, Par. Comp. **14**, 381 (1990).
[15] G. J. Goodhill, Biol. Cyb. **69**, 109 (1993).
[16] Y. Linde, A. Buzo, R. Gray, IEEE Trans. Comm. **28**, 84 (1980).
[17] H. Ritter, K. Schulten, Biol. Cyb. **60**, 59 (1988).
[18] K. Obermayer, G. G. Blasdel, K. Schulten, Phys. Rev. A **45**, 7568 (1992).
[19] G. A. van Velzen, J. Phys. A **27**, 1665 (1994).
[20] H.-U. Bauer, Neur. Comp. **7**, 36 (1995).
[21] E. Erwin, K. Obermayer, K. Schulten, Biol. Cyb. **67**, 47 (1992).
[22] T. M. Heskes, B. Kappen, Proc. IEEE ICNN 1993, IEEE Press, 1219 (1993).
[23] S. Luttrell, Neur. Comp. **6**, 767 (1994).
[24] M. Riesenhuber, H.-U. Bauer, T. Geisel, to be submitted.
[25] S. Löwel, J. Neurosci. **14**, 7451 (1994).
[26] M. Riesenhuber, H.-U. Bauer, T. Geisel, submitted to CNS 96 (1996).

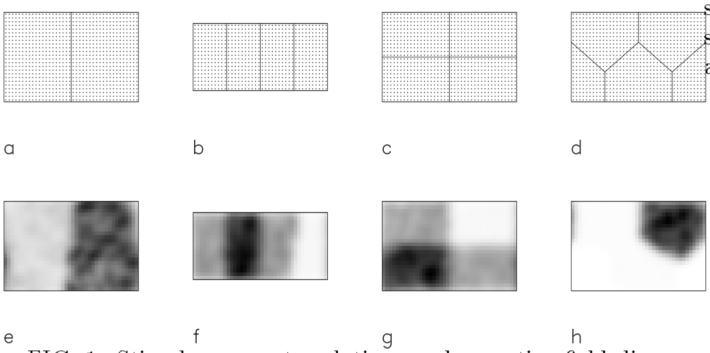

FIG. 1. Stimulus space tesselations and receptive field distributions $\mathbf{w_r}$ for the 4-neuron example (see text). (a-d): input space regions, such that all stimuli centered in one region are mapped to the same neuron. Note the division into only two such regions in (a), and into four regions of different shape in (b-d). Depending on s and σ, maps corresponding to these tesselations are observed. (e-h): gray value images of corresponding $\mathbf{w_r}$ for one exemplary neuron, resp. (e: $s = 1.3$, $\sigma = 1.13$, f: $s = 0.8$, $\sigma = 0.78$, g: $s = 1.2$, $\sigma = 0.78$, h: $s = 1.2$, $\sigma = 0.35$). Note the coincidence of the shapes of the black regions in (e-h) with the tesselation regions in (a-d).

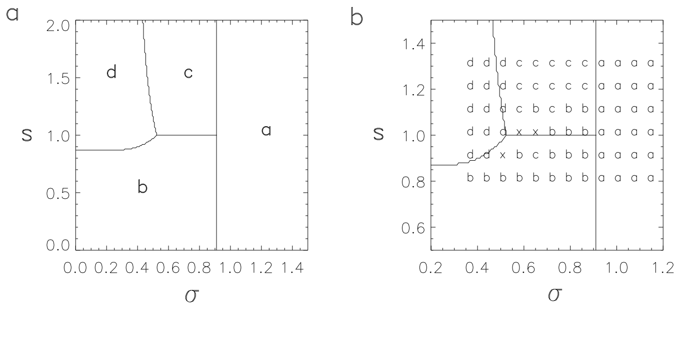

FIG. 2. Phase diagrams for the 4-neuron example. a: analytical, b: numerical results (letters) superimposed on the analytic phase diagram. The letters indicate the solution types of Fig. 1, x denotes non-classifiable maps.

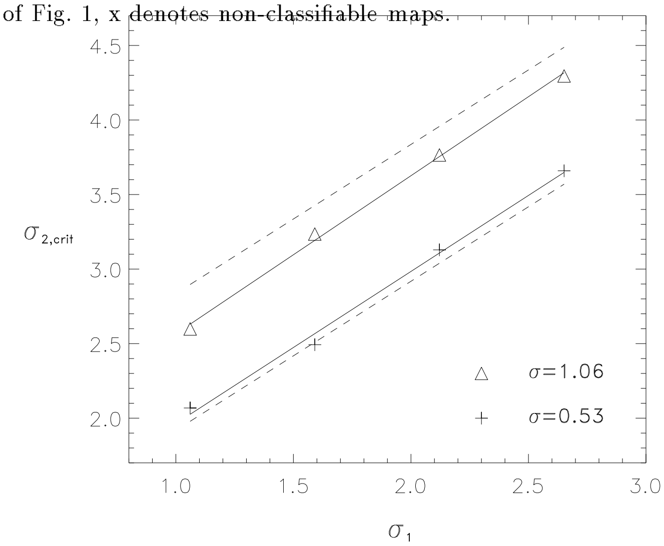

FIG. 3. Critical value $\sigma_{2,crit}$ of the longer half-axis of elliptic stimuli for the occurrence of an orientation map, as a function of the shorter half-axis σ_1, for two exemplary widths σ of the map neighborhood function. Symbols: values from simulations of SOMs (15×15 nodes, $\epsilon = 0.1 \rightarrow 0.001$, 2×10^5 steps), solid line: fit to these points, resp. Dashed lines: analytic result (8).

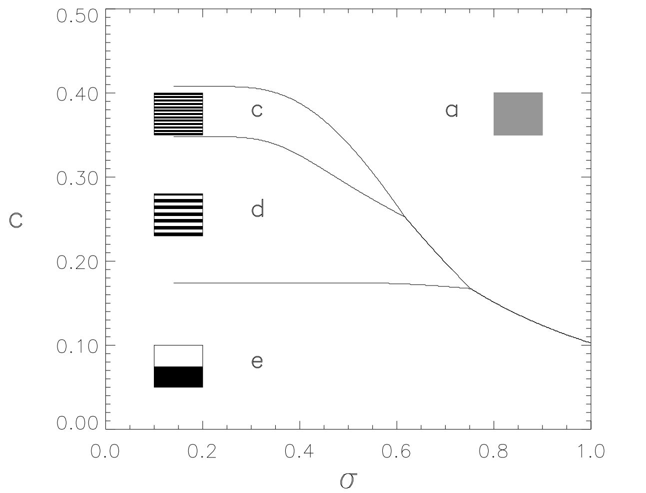

FIG. 4. Distortions $E_{\mathbf{v}}$ and $E_{\mathbf{w}}$ as a function of stimulus correlations c for SOM ocular-dominance model. Dashed line: analytical $E_{\mathbf{v}}$ of non-OD solution, dotted line: analytical $E_{\mathbf{v}}$ of OD solution (width 1), solid line: mean ocularity $O =< ||\mathbf{w}_{\mathbf{r},lefteye} - \mathbf{w}_{\mathbf{r},righteye}|| >_{\mathbf{r}}$, indicates state transition at $c = 0.66$, solid line with stars: numerical $E_{\mathbf{v}}$, solid line with crosses: numerical $E_{\mathbf{w}}$. The $E_{\mathbf{v}}$-values were scaled in order to compensate for the additional summation as compared to $E_{\mathbf{w}}$. Upper right inset: enlargement of the cross-over region. Lower left inset: analytical $E_{\mathbf{v}}$ of non-OD solution (dashed line), together with three analytical $E_{\mathbf{v}}$ of OD solutions with increasing width. Note that for decreasing c, the minimal value for $E_{\mathbf{v}}$ is attained for solutions with increasing band width.