A Computational Model of Early Neurogenesis in *Drosophila*

George Marnellos and Eric Mjolsness

Technical Report Number CS97-523

January 1997

Computer Science and Engineering
University of California, San Diego

A Computational Model of Early Neurogenesis in *Drosophila*

George Marnellos and Eric Mjolsness

Technical Report Number CS97-523

January 1997

A Computational Model of Early Neurogenesis in *Drosophila*

George Marnellos[1] and Eric Mjolsness[2]

[1] Section of Neurobiology, Yale University, New Haven CT 06510, USA, and
Sloan Center for Theoretical Neurobiology, The Salk Institute,
La Jolla CA 92037, USA
[2] Department of Computer Science and Engineering and
Institute for Neural Computation, UCSD,
La Jolla CA 92093, USA

Abstract. We present a computational model that simulates how neuroblasts and SOP cells differentiate from proneural clusters of cells. Parameters of the model are optimized in order to fit the dynamical expression patterns of genes that are involved in this process of cell fate specification. We report on simulations in which we fit qualitative expression pattern datasets and draw conclusions about what mechanisms may be sufficient or necessary for neuroblast and SOP differentiation.

1 Introduction

In Drosophila, neuroblasts and sensory organ precursor (SOP) cells differentiate from epithelia to give rise to the central nervous system in the fly embryo and to epidermal sensory organs in the peripheral nervous system of the adult fly, respectively. Neuroblasts are neural precursor cells that divide to form neurons and glia; they segregate from the ventral neuroectoderm of the embryo in a regular segmental pattern. SOPs appear at stereotypical positions on imaginal discs of fly third instar larvae and divide to produce a neuron and three other cells that form Drosophila's sensory organs, like the bristles on its thorax (see [2] and [3]).

Neuroblasts and SOPs differentiate from apparently equivalent clusters of cells that express genes of the *achaete-scute* locus, so called *proneural* genes. Eventually only one cell from each proneural cluster in the neuroectoderm and only one or a few cells from each proneural cluster in imaginal discs (clusters in the discs are typically larger than those in the embryo) retain proneural gene expression and become neuroblasts and SOPs (we will be referring to this process as *cluster resolution*).

Several other genes that are involved in this specification of cell-fate are also expressed in characteristic spatial and temporal patterns during the process. Genetic, molecular, ablation and other studies have pointed to a lateral signalling

interaction between adjacent cells, through which the neural fate is promoted in the future neuroblasts and SOPs and suppressed in other cells (for reviews see [2] and [3]).

Despite the number of empirical observations that have been gathered, several features of this system remain unexplained: a precise characterization of lateral signalling is still lacking; we do not understand dynamical aspects of the system, for example, whether and how the shape and size of proneural clusters determine how cluster resolution proceeds; it is not clear what the role, if any, of cell delamination, which accompanies neuroblast differentiation in the fly embryo, is.

To address questions like these and guide further experiments, we have constructed a computational model to simulate neuroblast and SOP differentiation, based on the modelling framework described in [15]; the same framework has been used to simulate gene expression patterns in the *Drosophila* blastoderm (see [17] and [18]).

2 Model

In our model, cells are represented as cylinders in a 2-dimensional hexagonal lattice; the diameters of the cylinders may vary and the extent of their surface overlap determines the strength of interaction between neighbouring cells; cells in the model express a small number of genes corresponding to the genes that are involved in neuroblast and SOP differentiation (see Fig. 1). Following the regulatory gene circuit framework developed in [15], we model genes as nodes in fully connected neural nets, with connection weights depending on the kind of interaction: we allow two kinds of interaction, an intracellular and a lateral signalling one.

The concentration $v_a(t)$ of the product of gene a in a particular cell at time t changes as follows: gene a sums inputs from other genes in the same cell or in neighbouring cells according to

$$u_a(t) = \sum_b T_{ab} v_b(t) + \sum_{i \in N} \Lambda^i \sum_b \hat{T}_{ab} \hat{v}_b^i(t) \tag{1}$$

where T is the matrix of gene interactions and $\mathbf{v}(t)$ the vector of gene product concentrations within the cell, $\hat{T}$ is the matrix of gene interactions with neighbouring cells, $\hat{\mathbf{v}}^i(t)$ the vector of gene product concentrations in neighbouring cell i, N the set of neighbouring cells cells (the neighbourhood of a cell consists of the six surrounding cells) and Λ^i a factor depending on the surface overlap of the cell with neighbouring cell i.

Concentration $v_a(t)$ then changes by an amount

$$\Delta v_a(t) = R_a g(u_a(t) + h_a) - \lambda_a v_a(t) \tag{2}$$

where $u_a(t)$ is the linear sum of eq. 1, g is a bounded, monotonic, non-linear function, e.g. a sigmoid, R_a is the rate of production of gene a's product, h_a is the threshold for activation of gene a and λ_a is the rate of decay of gene a product.

We want to find the values of gene interaction strengths (and other parameters in the equations above) so as to optimally fit gene expression patterns; we use simulated annealing and genetic algorithms for this optimization, see [14].

3 Simulation Results

3.1 Design of optimization and simulation runs

In the runs presented here, we use circuits of two genes to fit two types of expression patterns, one representative of the proneural genes (which all have very similar expression patterns and promote the neuronal fate) and another corresponding to genes, like the *Enhancer of split* genes, that act to sustain the undifferentiated, epithelial state of a cell and which we refer to as *epithelial* genes.

The training datasets we use are based on experimentally described expression patterns from the literature ([19], [6], [10]); they specify the initial concentrations of the gene products (the proneural clusters) and the desired final pattern when the proneural clusters have resolved to single cells (neuroblasts or SOPs); it is left to the optimization to find the right model parameters so that the system develop from the initial state to the desired final one (see Fig. 1). All cells in a proneural cluster are initially equivalent in terms of gene expression levels. The size and cluster arrangement of the training datasets do not have any particular biological significance; the datasets have been designed in such a way as to keep the number of cells low while including as many 7-cell symmetrical clusters as possible, since the optimization is very expensive computationally and a large number of cells in the training dataset would take several days on an IBM PowerPC or an SGI Indigo.

Apart from the training dataset illustrated in Fig. 1, in runs with cell delamination we have used another identical dateset (not shown) which additionally contains information about the delamination of the cells at the desired final state (when the prospective neuroblasts have fully delaminated, while all the other cells not).

The test datasets specify only initial concentrations of gene products, are larger (since we do not optimize on them but simply run the optimization solutions obtained with the training set to see how these solutions perform on novel proneural clusters), and contain more clusters of various shapes and in various spatial arrangements (see top left panels in Figures 3 and 4).

We mainly focus on optimizing gene interaction strengths, i.e. the T and $\hat{T}$

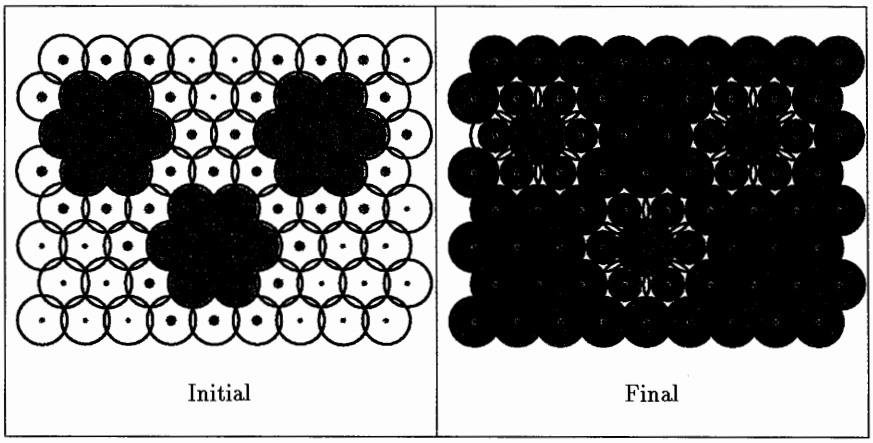

Fig. 1. Cells are modelled as cylinders in a hexagonal lattice (viewed here from above as circles). Gene expression is represented by coloured disks, proneural expression in brown and epithelial in green; disk radius is proportional to level of expression. This figure shows the training dataset: on the left, the initial concentrations of the gene products - there is only proneural gene expression in three symmetrical clusters; on the right, the desired final pattern of gene expression - proneural expression is retained only in the central cell of each cluster, corresponding to the future neuroblast or SOP, whereas all other cells express the epithelial gene.

matrices of eq. 2, and not so much other parameters of the model. An illustration of the designs we have used appears below; the gene interaction strengths optimized on are indicated with an "X", the empty boxes signify zero interaction strength (i.e. no interaction). Columns in these matrices are for input genes and rows for genes affected:

Intracellular Interactions

	Proneural	Epithelial
Proneural	X	X
Epithelial	X	X

Lateral Signalling Interactions

	Proneural	Epithelial
Proneural		X
Epithelial		

An example of the result of such an optimization run is presented below; there are five optimized parameters, four of them for gene interactions within a cell and one for a gene interaction across cells. A solution that the optimization has

come up with is the following:

Intracellular Interactions

	Proneural	Epithelial
Proneural	0.283	−0.015
Epithelial	−0.555	9.74

Lateral Signalling Interactions

	Proneural	Epithelial
Proneural		−10.1
Epithelial		

The simulation corresponding to the solution above appears in Fig. 2.

We have tried to limit the number of parameters we optimize on, so as not to overfit the data. Optimization on just five parameters as above, using the training dataset of Fig. 1, yields solutions that also work for larger test datasets with a greater number of 7-cell symmetrical clusters in various spatial arrangements (not shown). This indicates that the optimization does not just find parameter values that only work for the specific size and cluster arrangement of the training dataset, but rather extracts rules for resolving clusters.

3.2 Lateral interactions

Which lateral signalling interactions are allowed plays a crucial role in whether optimization can fit the training dataset in a satisfactory way or not. For example, if no Epithelial-to-Proneural or Epithelial-to-Epithelial lateral signalling interactions are allowed (i.e. if both entries in the right column of the of the lateral signalling matrix shown above are zero), then our optimization runs on the remaining interaction strengths have not produced values that lead to cluster resolution - and of course the same is true when no lateral interactions at all are allowed; whereas, if only the Epithelial-to-Proneural interaction of the lateral matrix is allowed, then optimization can readily yield values leading to cluster resolution (as mentioned above and shown in Fig. 2). This could be considered a prediction of the model, since it suggests that some lateral gene interactions are more important than others in producing cluster resolution. It also suggests that lateral signalling is sufficient and may also be necessary for cluster resolution.

If all four lateral signalling interactions are allowed, then, apart from fitting the training data and the 7-cell symmetrical cluster test sets, optimization solutions can successfully resolve bigger or smaller asymmetrical clusters as is illustrated in Fig. 3. This suggests that, although Proneural-to-Proneural and Proneural-to-Epithelial lateral interactions may not be sufficient on their own to bring about

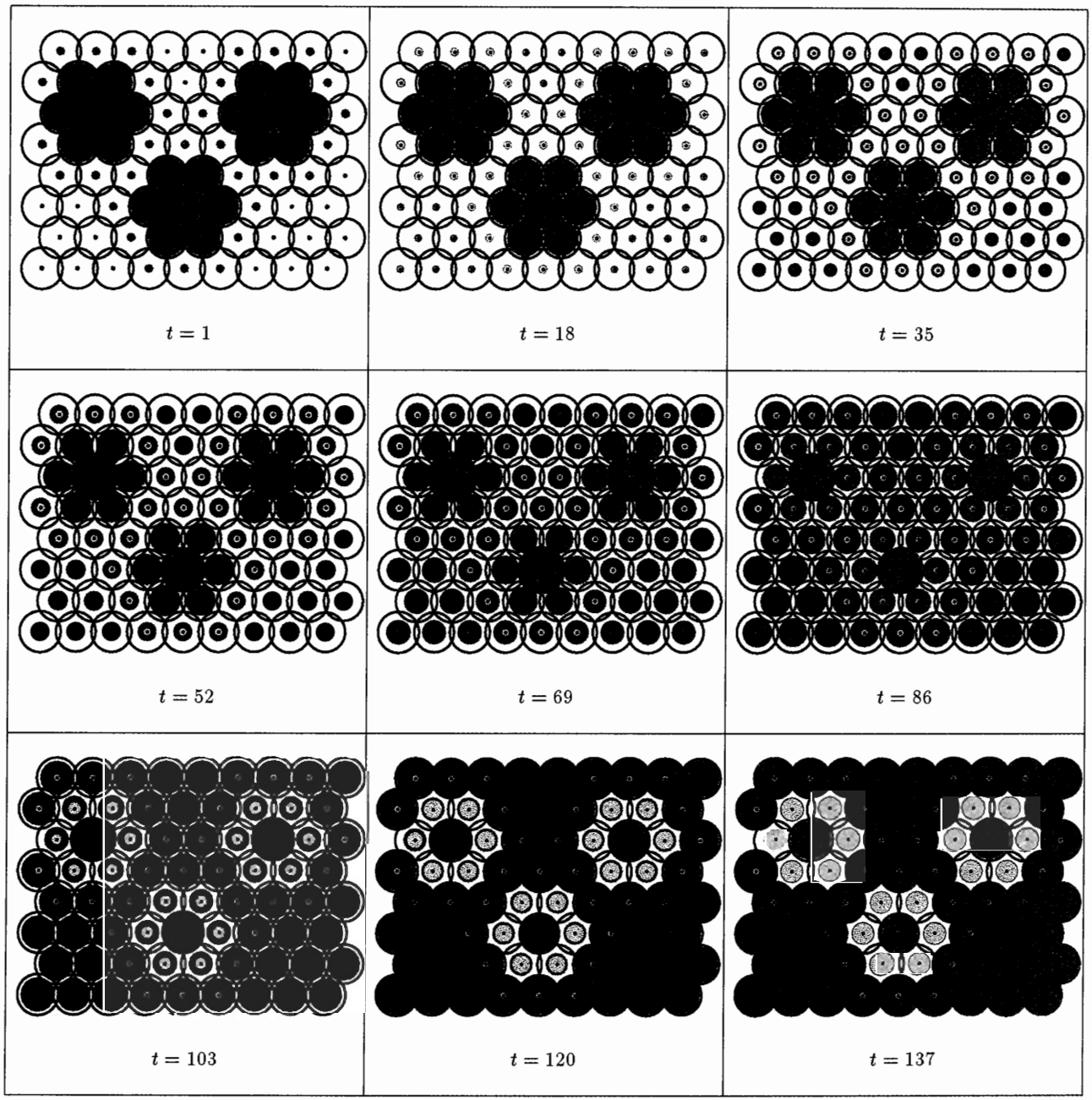

Fig. 2. Sequence of frames at different times of a neurogenesis model run: neuroblasts segregate from the middle of 3 proneural clusters. Proneural expression is depicted in brown, epithelial in green and their overlap in light green (other conventions as in Fig. 1). The parameters used in this simulation and their derivations are described in section 3.1.

cluster resolution, they may still enable lateral signalling to resolve a larger range of cluster shapes. It is interesting to note in Fig. 3 that 4-cell clusters do not produce a neuroblast but all cells in the cluster adopt the epithelial fate; this is the the case with all simulations we have run, irrespective of the number of optimized interaction strengths. Note also in Fig. 3 that the large cluster in the upper right corner does not resolve.

When sets of gene interaction strengths that include lateral interactions are obtained by optimization on the training data but these solutions are subsequently applied to the test data with the lateral interactions abolished, then the clusters do not resolve but all cells in them retain proneural gene expression. This parallels the effect of the so called *neurogenic* mutations in the real biological system; these mutations disrupt lateral communication between cells and lead to overproduction of neurons.

3.3 Cell delamination

Cell delamination has also been implemented in the model: it is directly controlled by genes within a cell and only indirectly by neighbouring cells; there are two parameters associated with the intracellular control of delamination, as well as another two that determine when delamination is initiated and how fast it proceeds. Delamination affects the surface overlap of neighbouring cells, i.e. changes the A^i factors in eq. 1. If the delamination parameters are optimized on together with four lateral signalling and four intracellular gene interaction parameters, then optimization solutions can resolve an even larger variety of shapes and sizes of clusters, that cannot be resolved readily without delamination (see Fig. 4).

With larger clusters it is easier to notice differences in the timing of cluster resolution: smaller clusters generally resolve faster than larger ones. It also becomes apparent that proneural and epithelial expression changes at different rates depending on cell position in the cluster, even for cells that adopt the same fate. Furthermore, from simulations both with and without cell delamination, it becomes apparent that lateral effects propagate further than one cell, as is indicated by the fact that gene expression in a group, for instance, of five cells changes in a particular way when this group of cells is a separate cluster (in this case one cell from the group becomes a neuroblast) and a different way when such a group is part of a bigger cluster (when all cells in the group revert to the epithelial fate); see figures 3 and 4. These observations are predictions of the model, since they have not been built into the model in any way.

3.4 Dynamics of cluster resolution

In order to study the dynamics of cluster resolution in our simulations we have also sketched the phase portraits, at successive points in time, of solutions obtained through optimization. For each point in time, we have plotted the di-

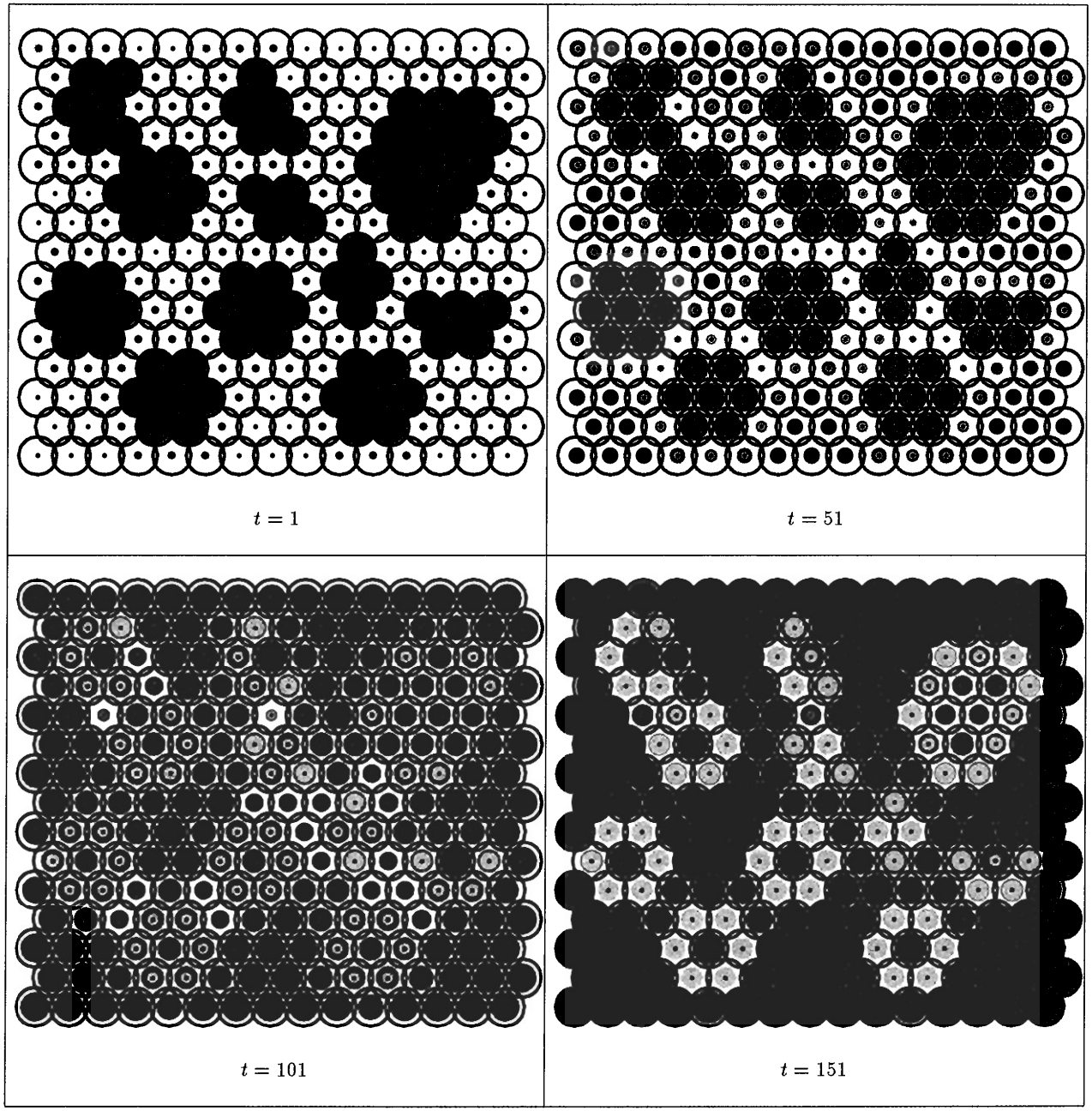

Fig. 3. Simulation with optimized parameters for all four lateral signalling interactions. Same conventions as in Fig. 2. Asymmetrical 5- and 6-cell clusters resolve, while all cells in the 4-cell clusters adopt the epithelial fate. Note that the large cluster in the top right corner is not resolved.

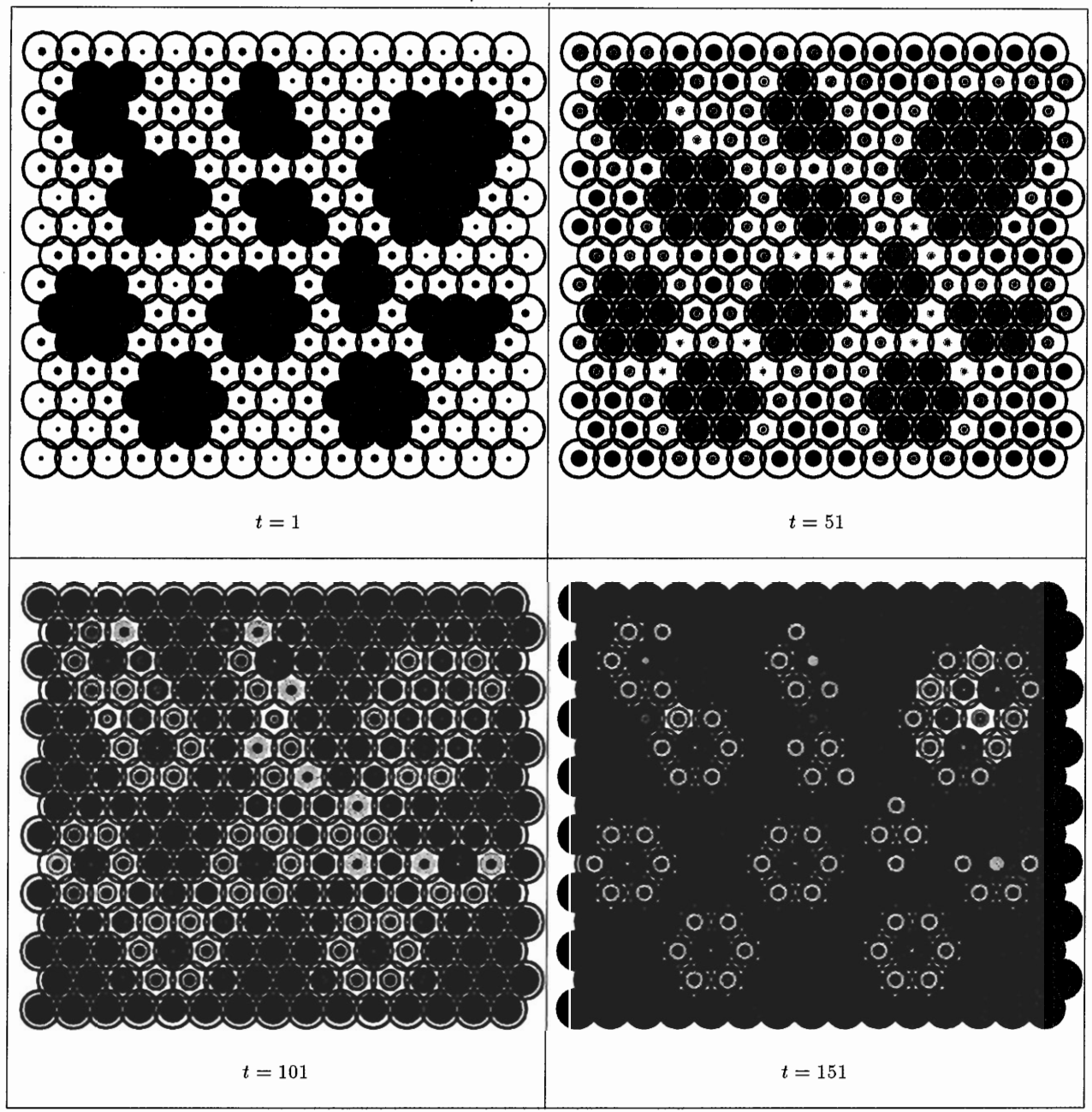

Fig. 4. Simulation with cell delamination and full lateral interactions. Same conventions as in Fig. 2, with the addition that cell delamination is represented by the thicker inner circle in the middle of the cells; the greater the radius of this circle, the further a cell has delaminated. Note that the large cluster on the right corner has resolved.

rection and magnitude of the change in gene product levels in a particular cell, given any value for the current product levels (see Fig. 5). This essentially shows how a cell would respond (in terms of modulation of its gene expression) if we altered (increased or decreased) the levels of its gene products.

If there are no lateral interactions the phase portrait of proneural versus epithelial gene expression levels does not change, of course, over time and is the same for all cells. By contrast, when there are lateral interactions, each cell has a different phase portrait that changes in time, depending on the cell's position (as in the example of Fig. 5). The phase portrait can be thought to represent the epigenetic landscape that each cell finds itself in; this is a dynamic landscape that changes depending on the strength of lateral interactions (as well as on the geometry of the tissue and the intracellular interactions).

4 Discussion

4.1 Implications of the model

Despite the simplicity of our model, the simulations described in this paper have allowed us to look closer at questions like the role of lateral signalling and cell delamination in neuroblast and SOP differentiation and have provided a tool to study the dynamics of gene expression during proneural cluster resolution.

Several conclusions can be drawn from these simulations which can be thought of as predictions of the model: lateral signalling, involving communication with just the immediate neighborhood of a cell, is sufficient for neuroblast and SOP segregation and may also be necessary. Abolishing lateral interactions prevents the resolution of proneural clusters. Not all lateral signalling interactions have the same effects, but some may be more important in bringing about cluster resolution, while the rest may enable the system to resolve clusters of a large variety of shapes and sizes. Lateral effects can propagate further than the cell's immediate neighborhood through a cell-to-cell relay.

Cell delamination produces sharper resolution for clusters of all sizes and facilitates the resolution of larger clusters; since in *Drosophila* neurogenesis delamination occurs only in the embryo CNS and not in imaginal discs, this raises the question of why this is the case: are there other mechanisms in discs, isomorphic to delamination, that facilitate the resolution of larger clusters, e.g. a specific pattern of subcellular localization of receptors and ligands on the cell membrane that modulates cell-cell communication the same way as delamination does? or is there a totally different mechanism with equivalent outcomes? In connection with this, it is interesting to note that Notch strongly localizes apically in imaginal disc cells (see [7]). More experimental and modelling work is needed to address questions like these.

The model makes further predictions about how cluster resolution proceeds:

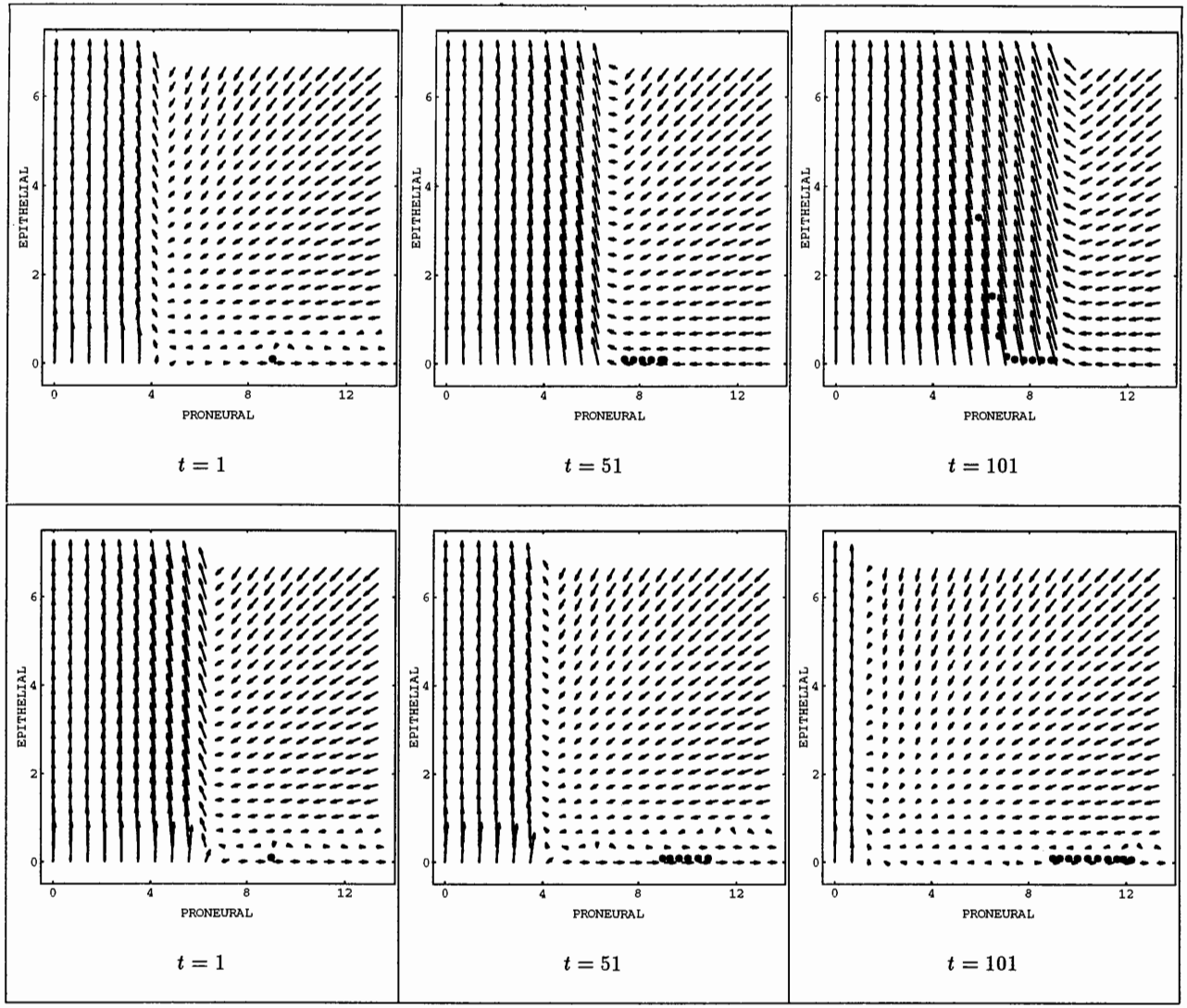

Fig. 5. Phase-portraits in the Epithelial-vs-Proneural plane for two cells from the same 7-cell symmetrical cluster at three successive stages in time; this is from a simulation with cell delamination and full lateral interactions. The dots indicate the trajectory of the cell. Top row: phase-portraits for a cell at the cluster periphery; the cell starts at high proneural and low epithelial levels in the left-most panel but soon the dynamic landscape changes and forces the cell to move to the epithelial attractor towards the top left in the following panels; note that by the time-point of the middle panel the cell seems to be committed to the epithelial fate despite its almost zero epithelial gene expression levels, since, even if its proneural gene expression level were almost doubled at this point, it would still move towards the epithelial attractor. Bottom row: phase-portraits for a cell at the cluster center; this cell starts at the same point in the phase plane as the peripheral cell in the top row, but the phase-portrait at the start already looks different from that of the peripheral cell and changes in a different way over time; note that even a small increase in epithelial expression would push this cell towards the epithelial attractor to the left, otherwise the cell moves towards the neural attractor to the bottom right.

smaller clusters generally resolve faster than larger ones; gene expression changes at different rates depending on cell position in the proneural cluster, even for cells that will eventually adopt the same fate; the degree of encirclement of a cell in a proneural cluster by other cells in the cluster can specify which cell becomes the neuroblast or SOP (namely the most encircled cell), especially in smaller clusters.

The analysis of gene interaction dynamics is also a source of very specific and quantitative predictions about how cells would respond to externally imposed changes in their gene product levels; for example, such analysis predicts that a cell in the proneural cluster periphery may be destined to the epithelial fate, even when its levels of epithelial gene expression are almost zero; such predictions are now testable in *Drosophila* since techniques have been developed that allow manipulations of gene expression in individual cells (see [9]).

4.2 Extensions

An obvious way to extend this model and make it even more biologically realistic is to include more genes in the regulatory circuit; this will perhaps make the analysis of the results more complicated, but such difficulties may be kept at a minimum if we use more extended and detailed datasets of gene expression derived directly from biological data, that will constrain and guide the optimization search for parameter values.

A specific extension of the model that would focus on how lateral signalling is mediated (and would therefore bear on most predictions of this model) is the following: as described in this paper, genes in the model can directly interact with genes in neighbouring cells; although this is a reasonable abstraction, introducing gene products on the cell membrane that gate gene interactions across cells is more biologically realistic and can produce more detailed predictions (in the real system lateral communication is mediated by the products of Notch, which is a receptor molecule, and Delta, its ligand, see [1]). We have already included Notch and Delta in our gene circuits and started running simulations with this additional feature. Questions surrounding the Notch-Delta interaction at the cell membrane are intimately related with cell delamination and subcellular localization of gene products, as alluded above, since delamination alters cell-cell communication through Notch and Delta in a way that depends upon how Notch and Delta are each localized on the cell membrane.

4.3 Further considerations

The kinds of questions that can be posed with the model described in this paper are not only of relevance to neurogenesis in *Drosophila* but are common to many developing organisms, especially in view of the fact that homologues to genes involved in *Drosophila* neurogenesis have been isolated in many species from worms to mammals participating in a variety of developmental processes.

In vertebrate neurogenesis such homologues act in ways similar to those of the *Drosophila* genes to regulate the number of neurons generated (see [5], [11], [12]); one would therefore expect that a theoretical and empirical understanding of *Drosophila* neurogenesis would provide insights into neurogenesis in higher vertebrates, for instance, into questions surrounding neuronal proliferation in the developing mammalian cortex (see [4], [16]).

Furthermore, the model of *Drosophila* neurogenesis described in this paper can be used to study quantitatively questions on genotype-phenotype mappings: the number of *Drosophila* bristles has long been a favourite quantitative phenotype of geneticists and evolutionary biologists (see [13]) and it will be useful to examine how such phenotypes are generated during development. This directly relates to the evolutionary questions of what variation such mappings can generate for selection to act upon and how evolvable they are (see [8], [20]).

References

1. Artavanis-Tsakonas, S., Matsuno, K. and Fortini, M.E.: Notch Signaling. Science **268** (1995) 225–232
2. Campos-Ortega, J.A.: Early neurogenesis in Drosophila melanogaster. In *The development of Drosophila melanogaster*, Bate, M. and Martinez-Arias, A. (eds.), Cold Spring Laboratory Press, 1993, 1091–1129
3. Campuzano, S. and Modollel, J.: Patterning of the *Drosophila* nervous system - the *achaete-scute* gene complex. Trends Genet. **8** (1992) 202–208
4. Caviness, V.S., Takahashi, T. and Nowakowski, R.S.: Numbers, time and neocortical neuronogenesis: a general developmental and evolutionary model. Trends Neurosci. **18** (1995) 379–383
5. Chitnis, A., Henrique, D., Lewis, J., Ish-Horowicz, D. and Kintner, C.: Primary neurogenesis in *Xenopus* embryos regulated by a homologue of the *Drosophila* neurogenic gene *Delta*. Nature **375** (1995) 761–766
6. Cubas, P., de Celis, J.-F., Campuzano, S. and Modolell, J.: Proneural clusters of *achaete-scute* expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. Genes and Dev. **5** (1991) 996–1008
7. Fehon, R.G., Johansen, K., Rebay, I. and Artavanis-Tsakonas, S.: Complex cellular and subcellular regulation of Notch expression during embryonic and imaginal development of *Drosophila*: implications for Notch function. J. Cell Biol. **113** (1991) 657–669
8. Goodwin, B.: Emergent form: evolving beyond Darwinism. Complexity **1** (1996) 11–15
9. Halfon, M.S., Kose, H., Chiba, A. and Keshishian, H.: Targeted gene expression without a tissue-specific promoter: using laser-induced single-cell heat shock to create mosaic embryos. In *Abstracts of the 37th Annual Drosophila Research Conference*, The Genetics Society of America, 1996, 46
10. Jennings, B., Preiss, A., Delidakis, C. and Bray, S.: The Notch signalling pathway is required for *Enhancer of split* bHLH protein expression during neurogenesis in the *Drosophila* embryo. Development **120** (1994) 3537–3548
11. Lewis, J.: Neurogenic genes and vertebrate neurogenesis. Curr. Opin. Neurobiol. **6** (1996) 3–10

12. Ma, Q., Kintner, C. and Anderson, D.J.: Identification of *neurogenin*, a vertebrate neuronal determination gene. Cell **87** (1996) 43–52
13. Mackay, T.F.C.: The genetic basis of quantitative variation: numbers of sensory bristles of *Drosophila melanogaster* as a model system. Trends Genet. **11** (1995) 464–470
14. Marnellos, G. and Mjolsness, E.: Optimization in biological models that use recurrent neural nets. To appear as a UCSD-CSE-TR, Oct. 1996.
15. Mjolsness, E., Sharp, D.H. and Reinitz, J.: A connectionist model of development. J.Theor.Biol. **152** (1991) 429–453
16. Rakic, P.: A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. Trends Neurosci. **18** (1995) 383–388
17. Reinitz, J., Mjolsness, E. and Sharp, D.H.: Model for cooperative control of positional information in *Drosophila* by Bicoid and maternal Hunchback. J.Exp.Zool. **271** (1995) 47–56
18. Reinitz, J. and Sharp, D.H.: Mechanism of *eve* stripe formation. Mech.Dev. **49** (1995) 133–158
19. Skeath, J.B. and Carroll, S.B.: Regulation of proneural gene expression and cell fate during neuroblast segregation in the *Drosophila* embryo. Development **114** (1992) 939–946
20. Wagner, G.P. and Altenberg, L.: Complex adaptations and the evolution of evolvability. Evolution **50** (1996) 967–976