SELF-ORGANIZING NEUROMORPHIC SYSTEMS WITH SILICON GROWTH CONES

Brian Seisho Taba

A DISSERTATION

in

Bioengineering

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2005

Kwabena Boahen, Ph.D.,	John Schotland, M.D., Ph.D.,
Supervisor of Dissertation	Graduate Group Chair

Acknowledgements

I would like to thank all of the people who supported me during my time here at Penn.

In particular, I would like to thank my advisor, Kwabena Boahen, for his enthusiastic motivation, generous support and patient mentorship over the past few years.

I would also like to thank my committee members – Rita Balice-Gordon, Leif Finkel, Daniel Lee, and Bertram Shi – for their time and valuable advice.

Finally, I would like to thank my fellow lab members – Kareem Zaghloul, Kai Hynna, John Arthur, Paul Merolla, Bo Wen, John Wittig, Thomas Choi, Rodrigo Alvarez, Zhijun Liu, and Joseph Lin – for their steady comradeship, both in and out of lab, which made graduate study more interesting and much more fun.

Abstract

SELF-ORGANIZING NEUROMORPHIC SYSTEMS WITH SILICON GROWTH CONES

Brian Seisho Taba

Supervisor: Kwabena Boahen, Ph.D.

Neuromorphic engineers have achieved considerable success in devising silicon implementations of progressively more complex neural architectures. However, the effort required to design a successful neuromorphic system grows dramatically as the scope of these projects expands to encompass multiple neuromorphic subsystems. This design process could be eased by automating difficult design tasks.

In this thesis I introduce a novel technique for automatically rewiring connectivity between spiking neurons based on a model of activity-dependent axonal growth cone navigation during neural development, and illustrate its performance with a silicon implementation of a model growth cone population whose migration is driven and directed by patterned neural activity. I develop a stochastic model of silicon growth cone motion to explain and characterize population behavior, and discover that performance is limited by an optimality criterion whose existence is implied by the fundamental physicality of the system.

Contents

List of Figures

Chapter 1

Introduction

In recent decades, experimental neuroscientists have managed to tease out many of the computational principles hidden in the inner workings of the brain. Neuromorphic engineers attempt to migrate these hard-won insights into practical engineering applications by using standard electrical engineering methods to design silicon chips that faithfully represent neural architectures. The typical design strategy is to carefully trace the wiring patterns in the physiological tissue under study and then instantiate an equivalent connectivity for the chip. While this approach has been very successful in modeling systems like the retina [67, 6, 107], the complexity of the wiring diagram increases combinatorially as systems become larger and more ambitious in their scope. Automation of some of the more arduous wiring tasks would greatly ease the burden of the neuromorphic system designer.

The human brain is the ultimate wiring problem, involving over 10^{15} synapses connecting 10^{12} neurons, yet somehow every newborn infant manages to automatically solve it from only the 10^{9} bits in the genetic code, relying on emergent properties of cellular populations that are guided by very simple programs. For example, in the mammalian

visual system, neural circuits wire themselves up according to schematics encoded in the statistical structure of their spontaneous input activity. The goal of this thesis is to harness analogous behavior to self-organize axonal connections between disparate populations of spiking silicon neurons.

Inspired by an examination of self-organizing feature maps in the brain, our basic wiring principle will be to translate temporal activity coincidence into spatial position coincidence, so that "cells that fire together, wire together." To implement this rule, we will design a silicon model of axonal growth cones, motile sensory appendages that steer the tips of elongating axons based on their assessment of the extracellular chemical environment. Growth cones can be guided by chemotropic gradients of diffusible substances like BDNF, whose release and uptake depend on patterned neural activity. BDNF is a good candidate for a retrograde messenger that communicates postsynaptic activity to a presynaptic afferent, which can use this message as feedback to determine the relative fitness of its current location. An activity-dependent gradient of BDNF would be sufficient to spatially congregate coactive growth cones, satisfying the desired navigation rule.

There are a number of neuromorphic potentiation circuits intended to model adult synaptic plasticity, but scant attention has been paid to developmental morphogenesis. More traditionally engineered neural clustering chips are based on vector quantization algorithms such as Kohonen's self-organizing map [59, 69, 11, 7, 82, 49] that rely on the synchronous presentation of digital datawords, and are not equipped to process the stochastic spike trains of neuromorphic chips. Furthermore, these devices are typically limited in size by their use of a global winner-take-all computation that collates information across the entire chip.

The Neurotrope1 system described in this thesis is the first neuromorphic model of

structural neural plasticity, and its behavior highlights many of the performance issues associated with any physical implementation of a neural algorithm, be it a silicon chip or the brain itself. The outline of the the thesis is as follows:

Chapter 2 reviews neural map formation from biological, theoretical, and engineering perspectives.

Chapter 3 presents Neurotrope1, the first neuromorphic self-organizing map chip. The Neurotrope1 system implements a simplified model of growth cone migration under the guidance of an activity-dependent diffusive factor called neurotropin.

Chapter 4 introduces a transition matrix analysis that allows us to characterize neurotropic growth cone guidance in the Neurotrope1 system, both theoretically and experimentally.

Chapter 5 applies the transition matrix analysis to the simplest nontrivial case of neurotropic growth cone guidance, supervised pair attraction. We discover that an intermediate value of the neurotropin spreading range optimizes equilibrium performance, as measured by the ability of an active growth cone to move to a coactive target.

Chapter 6 elucidates the origin of the optimal spreading range and the resource constraints that ultimately limit performance for supervised pair attraction.

Chapter 7 presents the general case of unsupervised self-organization. The Neurotrope1 system robustly refines a retinotopic map when driven by an ideal stimulus generator and a severely nonideal silicon retina.

Chapter 2

Neural self-organization

In this chapter we review the experimental and theoretical basis of self-organizing neural maps.

2.1 Structural plasticity

During development, neurons project axons to distant targets. Axon pathfinding is guided by labile structures called growth cones, which sense local chemotropic gradients that steer migrating axons to their targets. Upon arrival, axons elaborate highly dynamic arbors, continually sprouting and retracting transient branches that form nascent synaptic contacts with neighboring dendritic filopodia. Synapses and their corresponding branches are stabilized by correlated activity, and the final pattern of permanent synapses defines the neural circuit. Directed circuit formation is mediated by activity-dependent regulation of filopodial dynamics.

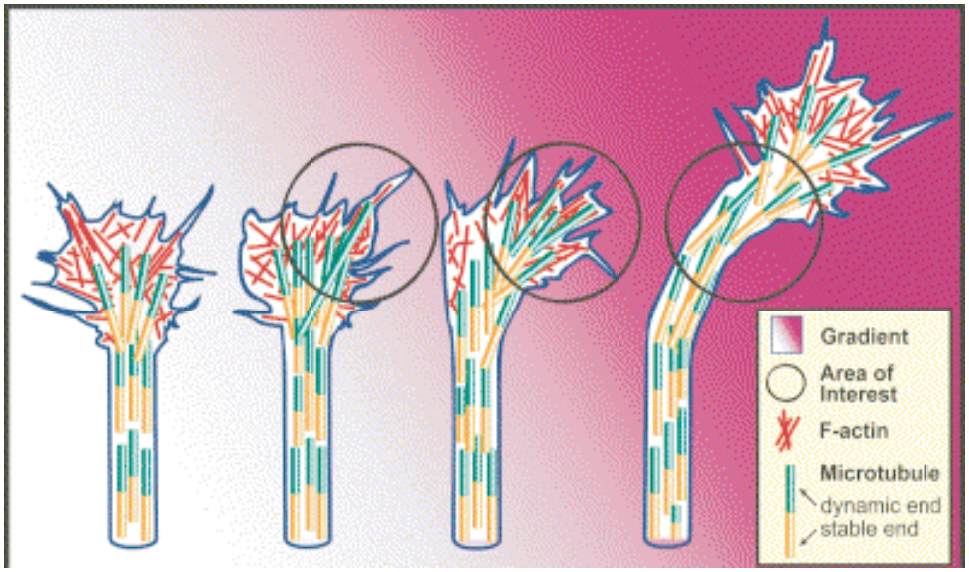


Figure 2.1: Growth cones steer axons

Growth cone filopodia project and retract, driven by F-actin polymerization and disassembly. Microtubules grow from the axon shaft into the growth cone body to anchor stable filopodia and extend the axon. Extracellular guidance cues bias growth cone dynamics to stabilize more filopodia on one side, steering the axon up the local gradient [23].

2.1.1 Growth cones

Elongating axons are tipped by amoeboid structures called *growth cones*. Growth cones consist of *filopodia*, finger-like membrane protrusions that probe the local environment and sense chemical cues, and *lamellipodia*, networks of cross-linked actin filaments that form the webbing between adjacent filopodia. Growth cones are highly dynamic structures, constantly sprouting and retracting filopodia. The relative rates at which filopodia extend and retract determine the net motion of the growth cone. If more filopodia sprout on one side of the growth cone, the growth cone turns toward that side. If more filopodia retract, the growth cone turns away, towing the axon behind it. Axon guidance is thus mediated by the cytoskeletal dynamics of the growth cone and its filopodia (Figure 2.1).

The growth cone cytoskeleton is constructed from polymers of the building-block pro-

teins tubulin and actin. Tubulin dimers polymerize into tubular arrays called *microtubules*, which offer structural support to the axon and transport nutrients and signals down their length. Actin monomers polymerize into helical filaments called *F-actin* or microfilaments, which provide structural support within the growth cone. The lamellipodial cytoskeleton consists of a disordered meshwork of F-actin, while the filopodia are built out of bundles of aligned F-actin filaments.

Axon elongation is a three-stage process. First, filopodia and lamellipodia protrude from the growth cone by constructing new F-actin bundles. Next, microtubules grow out of the axon into the growth cone body, engorging it with organelles and vesicles. Finally, the F-actin in the growth cone body disassembles, permitting the membrane to collapse into a tight bundle around the microtubules to consolidate into a new axon segment [23].

F-actin is polarized in the sense that monomers are preferentially added to one end of the polymer and removed from the other end. In filopodia, the polymerizing end of the F-actin bundle faces the leading edge and the depolymerizing end faces the growth cone body. In addition, F-actin bundles are constantly retracted toward the growth cone by a motor-driven transport process called retrograde flow. Thus, an F-actin bundle grows toward the tip of the filopodium, but is constantly pulled backward by retrograde flow into the growth cone body, where it dissociates. The balance between polymerization, retrograde flow, and depolymerization determines whether the F-actin bundle grows or shrinks, extending or retracting its filopodium.

2.1.1.1 Axon guidance cues

Filopodial dynamics are controlled by the relative rates of several processes including actin nucleation of new bundles, actin polymerization at the distal end of the filopodium, ret-

rograde F-actin flow, actin depolymerization at the proximal end of the filopodium, and nonmuscle myosin activity. All of these processes are regulated by the activity of a family of small GTP-binding proteins called Rho GTPases. Agents that influence Rho GTPase activity to selectively modulate one or more of these processes therefore represent potential axon guidance cues.

For example, an agent that inhibits actin polymerization would promote filopodial collapse by permitting retrograde actin flow and depolymerization to dominate. A growth cone moving in a gradient of this chemorepulsive agent would tend to retract more filopodia from areas of high concentration and the net effect would be to guide the growth cone toward regions of low concentration.

A number of axon guidance cues have been identified that direct filopodial growth by regulating Rho GTPase activity, either directly or through intermediate signaling cascades [53]. Some of these guidance cues are secreted, such as netrin, Slit, neurotrophins, and class 3 semaphorins; others are membrane-bound, such as class 1 and 4 semaphorins, ephrins, ligands for receptor protein tyrosine phosphatases, cell-adhesion receptors, and myelin-associated inhibitors. Individual guidance cues may be chemoattractive or chemorepulsive or both, depending on context.

For example, the same guidance cue can simultaneously activate separate signaling pathways via different receptors. Netrin binding to the DCC receptor triggers cAMP-dependent attraction in the growth cones of cultured *Xenopus* spinal neurons, while binding to a DCC-UNC5 receptor complex results in cGMP-dependent repulsion [78]. The balance between these opposing pathways determines the net effect of netrin on the growth cone. In fact, sufficiently reducing intracellular cAMP levels actually switches the polarity of this effect from attraction to repulsion [74]. Conversely, elevating cAMP enhances the strength

of netrin-induced attraction.

Interestingly, brief electrical stimulation can trigger this cAMP-dependent enhancement, implying that growth cone guidance can be modulated by neural activity [73]. Activity plays a major role in the morphogenetic development of synaptic circuits.

2.1.1.2 Arborization

As an axon approaches its target, it elaborates an arbor of motile branches that test and discard provisional synapses onto overlapping dendrites in a flurry of structural remodeling. Axons and dendrites each probe their vicinity for suitable synaptic partners by continually sprouting exploratory filopodia that initiate synaptogenesis at points of contact. Nascent synapses can form and disassemble in less than two hours, and their filopodia exhibit similarly rapid turnover [1]. Most filopodia are transient, retracting unless anchored by persistent synapses. Stable filopodia mature into branches of the final arbor.

Synapse turnover continues well past arbor assembly. In neonatal mice, for example, each geniculate cell initially receives over twenty separate retinal inputs immediately prior to eye opening, when axon arbors have already segregated into distinct eye-specific layers. Over the next three weeks, redistribution of presynaptic release sites reduces this number to one to three synapses whose strengths are increased 50-fold [15]. Early in development, therefore, weak preliminary synapses may serve to stabilize the arbor, which then provides a scaffold for more developed synapses in the final circuit.

Dendritic arborization appears to be synaptotropic, meaning it is guided by synapse formation (Figure 2.2). In zebrafish tectum, all nascent synapses appear at the tips of dendritic filopodia. A fraction of these synapses are actively maintained and are able to stabilize their

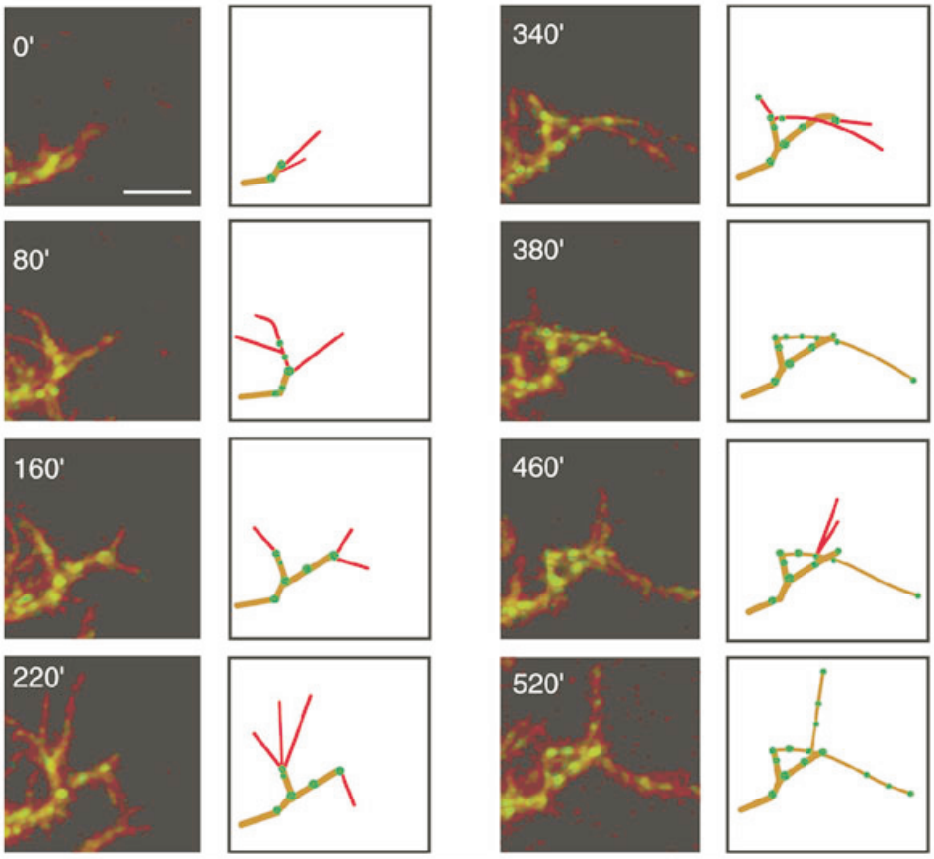


Figure 2.2: Synaptotropic arborization

Dendritic arborization is guided by synapse formation. Time lapse image sequence of dendritic filopodia *in vivo*. Red lines are newly formed and possibly transient branches; green dots are synapses. New branches sprout from synaptic sites [77].

filopodia, either by providing trophic support or even simple membrane adhesion. Their dendritic filopodia mature into permanent arbor branches that sprout new filopodia from their synaptic sites. Dendrites advance overall by selectively stabilizing branch segments between permanent synapses, whose spatial pattern defines the dendritic arbor [77].

Axon arborization is similarly synaptotropic, with most new branches originating at synaptic sites [1]. Branch dynamics are activity-dependent. Presynaptic activity blockade significantly increases branch addition and elimination rates, escalating arbor complex-

ity and interfering with branch stabilization [18]. The pattern of branch stabilization is instructed by correlations in activity. Retinotectal axons innervating *Xenopus* binocular tectum selectively remodel their arbors to reflect differences in activity correlations within and between eyes. Axon branches stabilize in territory dominated by afferents from the same eye and disappear from territory dominated by afferents from the opposite eye [85].

Branch stabilization is mediated by NMDA channels. The NMDA receptor is an ionotropic glutamate channel whose voltage dependence permits ion flow only during the coincidence of presynaptic glutamate release and postsynaptic depolarization, a conductance profile that can communicate correlations between presynaptic and postsynaptic activity to the postsynaptic cell [94]. Blocking NMDA channels in binocular tectum reduces the axon branch retraction rate in opposite eye territory but not in same eye territory, consistent with the idea that poorly correlated activity triggers an active elimination mechanism that prunes mistargeted branches [85].

New axon branches sprout preferentially into regions with low afferent density. Branch addition is not directly instructed by correlated activity, but selective branch elimination would automatically reduce afferent density, indirectly encouraging new sprouting to fill the emptied regions [85]. Axonal filopodia might measure local afferent density by using their own glutamate release as an active sensor for extracellular space. Extracellular space promotes axonal filopodial motility in the hippocampal mossy fibers of neonatal mice, as mediated by activity-dependent autocrine glutamate signaling. Glutamate regulation of motility is bidirectional; motility is upregulated by the relatively dilute concentrations that might result if most of the released glutamate diffused away into a large extracellular volume, and downregulated by the more intense concentrations that might result from confining extracellular glutamate to a small volume like a synaptic cleft [98]. Thus, afferent activity could permit axonal filopodia to increase their motility in empty space, improving

the chance of intercepting a dendrite and inducing synaptogenesis.

Alternatively, the predilection for new branches to sprout into unoccupied territory might simply reflect a competition between afferents for space. Selective branch elimination is presumably driven by a similar competition, which could be over a presynaptic resource such as neurotransmitter vesicles, a postsynaptic resource such as a target-derived survival factor, or both.

2.1.2 BDNF

The family of proteins called neurotrophins take their name from the neurotrophic hypothesis, which postulates that innervating neurons compete for a limited amount of a survival factor secreted by the target organ. Neurons supplied with insufficient survival factor withdraw or die. Neurons supplied with sufficient survival factor are maintained, and are said to receive trophic support from their target. Beyond this eponymous regulation of neuronal survival, neurotrophins control a host of developmental processes including cell fate decisions, phenotypic expression, axonal and dendritic arborization and pruning, and patterns of innervation. In the mature system, they modulate synaptic plasticity and function.

The first neurotrophin, nerve growth factor (NGF), was originally purified as a survival factor for sensory and sympathetic spinal neurons in culture. Subsequent work isolated additional members of the neurotrophin family, whose current roster comprises NGF, brain-derived neurotrophic factor (BDNF), NT-3, NT-4/5, NT-6, and NT-7. Neurotrophins all bind to the shared p75NTR receptor and to specific members of the tropomyosin-related kinase (Trk) receptor tyrosine kinase family. NGF binds specifically to TrkA, while BDNF and NT-4 are specific to TrkB. NT-3 binds to TrkC with high affinity, and less efficiently to the other Trk receptors [52].

Among the neurotrophins, BDNF is particularly well-studied because its activity-dependent dynamics offer a plausible substrate for synaptic plasticity and growth cone guidance. It is commonly believed that BDNF mediates synaptic Hebbian learning by acting as a retrograde messenger that communicates postsynaptic activity to the presynaptic afferent. Under this hypothesis, BDNF would need to be released by postsynaptic activity and taken up by presynaptic activity [56].

2.1.2.1 BDNF as retrograde messenger

Studies of cultured hippocampal neurons demonstrated that electrical activity can indeed evoke postsynaptic release of BDNF from CNS dendrites. BDNF secretion requires high levels of intracellular Ca^{2+}, which can be achieved through presynaptic depolarization of the postsynaptic membrane that triggers extracellular Ca^{2+} entry via voltage-gated calcium channels [46]. The onset of BDNF release lags the start of the electrical stimulus by tens of seconds, a delay attributed in part to the absence of a readily releasable pool of predocked peptide vesicles at the membrane, which requires vesicle diffusion to precede neuropeptide exocytosis. Once triggered, BDNF release can be sustained for minutes beyond the end of the stimulus through Ca^{2+}-induced Ca^{2+} release from intracellular stores and autocrine activation of TrkB receptors [60]. The slow onset and offset of BDNF release tend to smooth out the discrete burst stimuli, so that exogenous BDNF levels reflect a more persistent identification of recently active postsynaptic cells.

Specific patterns of activity can trigger greater BDNF release. Bursts of high frequency postsynaptic spikes enhance BDNF release, presumably because transient Ca^{2+} spikes sum more quickly to attain higher cumulative intracellular Ca^{2+} levels. Nonphasic and low frequency spike train stimuli are less effective, as is constant depolarization [5]. BDNF release depends on action potential generation, possibly to regenerate inward Ca^{2+} currents

through post-spike hyperpolarization relief of Ca^{2+} channel inactivation [38].

Presynaptic BDNF uptake is modulated by similar patterns of presynaptic activity. High frequency burst stimuli such as standard LTP-inducing tetanus protocols enhance TrkB insertion into the cell surface and internalization of the bound BDNF-TrkB complex, while low frequency spike trains and constant depolarization have little effect [24, 25].

Activity-dependent regulation of BDNF secretion, TrkB insertion, and BDNF-TrkB internalization all help to confine BDNF-TrkB signaling to active synapses. BDNF can diffuse up to 60 μm from its release site [73], so for BDNF to act as a synapse-specific retrograde messenger, mechanisms must be in place to prevent spillover from affecting neighboring synapses. Other observed localizing mechanisms include activity-dependent BDNF mRNA transcription, which limits BDNF expression to active postsynaptic cells [60], and local protein synthesis of synaptic potentiation agents, which permits internalized BDNF-TrkB complexes to selectively promote specific synapses instead of broadcasting a general potentiation signal by synthesizing agents at the cell body [109]. The diffusive spread of secreted BDNF might also be restricted by scavenging exogenous BDNF with an inactive receptor such as truncated TrkB located in the extrasynaptic membrane [64].

2.1.2.2 BDNF as tropic agent

On the other hand, the diffusibility of BDNF suggests that this versatile protein could also play a chemotropic role in guiding active growth cones over short distances to active targets. Growth cones are attracted to BDNF *in vitro*, and preserve their sensitivity to BDNF gradients by adapting to increasing basal concentrations near a secretion source [75].

On a larger scale, BDNF increases axon arbor growth through branch addition and

elongation, as well as the number of synapses per axon terminal in *Xenopus* tectum [20, 1]. Activity blockade interferes with branch stabilization but does not prevent BDNF from enhancing axon growth [18]. BDNF and glutamate might therefore represent complementary halves of correlation-mediated stabilization. BDNF acts presynaptically to generate new synaptic guesses through exploratory branch growth and synaptogenesis, while glutamate acts postsynaptically to preserve good guesses and prune bad ones. In its putative tropic role, BDNF would also encourage better guesses by attracting axonal filopodia to coactive dendrites.

As a retrograde messenger, BDNF could also play a role in synapse stabilization. Synapse assembly and disassembly require action on the part of both afferent and target [19, 40]. The amount of BDNF taken up by the afferent is a measure of the coincidence of the postsynaptic activity required for BDNF release and the presynaptic activity that facilitates BDNF uptake, so BDNF conveys activity correlation strength to the afferent. Strong correlations encourage the afferent to reinforce the presynaptic terminal, increasing its ability to drive the target cell, which responds by strengthening the postsynaptic terminal.

Blocking this retrograde correlation signal impairs normal circuit formation. Infusing an excess of BDNF or NT-4/5 into cat visual cortex during the critical period prevents ocular dominance maps from forming [9], as does competitive binding of TrkB ligands [10]. Occluding one eye during this period causes cortical cells to lose their response to that eye. NT-4/5 infusion can restore this loss at the cost of orientation selectivity and ocular dominance, implying that NT-4/5 promotes globally promiscuous connectivity [39]. TrkB ligands thus play an organizing role in the development of neural maps.

2.2 Neural maps

One of the organizing principles of the brain is that spatially adjacent cells in one region tend to receive input from spatially adjacent cells in another region. These *topographic maps* are ubiquitous in sensory cortex. Examples include retinotopic maps in the visual system, tonotopic maps in the auditory system, and receptor maps in the olfactory system. Each of these maps serves as an internal representation of some feature that varies continuously in the physical world, be it position, frequency, or odorant composition.

2.2.1 Map development

Sensory feature map development is controlled by a combination of activity-independent and activity-dependent processes. Typically, coarse global structure is laid down by intrinsic molecular markers. Precise local connectivity is then refined and maintained by patterned neural activity. The relative contributions of activity-independent and activity-dependent developmental processes have been extensively studied in the retinal projection to the midbrain, which for amphibians is the tectum and for mammals is the superior colliculus.

2.2.1.1 Activity-independent map formation

Retinal ganglion cell (RGC) axons navigate almost the entire path from retina to tectum using external guideposts whose instructions are intrinsic to each axon. For example, *Xenopus* tadpole eyes are located laterally, with no binocular overlap. Premetamorphic RGCs project uniformly to contralateral tectum. During metamorphosis, eye position rotates to create a partial binocular field and a new small population of new RGCs is generated in the

ventrotemporal retina that projects axons to ipsilateral tectum to form a binocular map with the preexisting contralateral projection. Upon reaching the midline, therefore, each RGC axon must decide which tectum to innervate.

An axon's decision to cross the midline at the optic chiasm is based on a chemorepulsive interaction between a membrane-bound receptor on the axon surface and a diffusible ligand secreted by the surrounding tissue. All RGC axons express the receptor EphB, but the optic chiasm does not express the ligand ephrin-B until metamorphosis, when all of the initial RGC axons have already crossed. The presence of ephrin-B repels ventrotemporal RGC axons, forcing them to project to ipsilateral tectum [76]. Ephrin-B thus acts as a guidance cue that is regulated both spatially and temporally. Its guidance is global, providing the same instruction to every EphB-expressing axon. Similar global guidance cues escort the axon tract to its target tissue [86].

Upon arrival at the tectum, axons sort themselves into loose retinotopy using matched gradients of molecular markers expressed by tectal cells and the corresponding receptors on the RGC axon terminals. Retinal expression of the receptor EphA decreases along the temporal-nasal axis, while tectal expression of the ligand ephrin-A increases along the rostral-caudal axis. The axons of temporal RGCs express more EphA and are therefore sensitive to lower ephrin-A concentrations than axons of nasal RGCs. Since ephrin-A binding to EphA mediates axon repulsion, temporal RGC axons migrate away from high caudal ephrin-A concentrations to the low-ephrin-A rostral tectum. Axons compete for space, so nasal RGC axons are displaced into caudal tectum. Axon attraction mechanisms mediated by EphB/ephrin-B interactions similarly guide dorsal RGC axons to the ventral tectum. The global topography of the initial retinotectal projection is thus organized by a combination of activity-independent chemoaffinity gradients and competition for space [57].

2.2.1.2 Activity-dependent map refinement

Matched gradients alone suffice to construct a bare-bones map structure, but fleshing it out requires electrical activity. Map formation is followed by a morphogenetic frenzy in which axon arbors rapidly sprout and retract transient branches which stabilize only in correctly targeted regions. Sparse, widely overlapping axon arbors contract into dense, tightly targeted termination zones, reinforced by a proliferation of precisely placed synapses, while stray collaterals are pruned. The net effect is to refine a coarse global axon ordering into precise local synaptic circuitry. This directed morphogenesis overlaps with the onset of sensory experience, and is strongly activity-dependent. Disrupting normal visual activity during this critical period of development can freeze or even unravel existing neural maps [21].

Activity can play one of two organizational roles in morphogenetic map refinement. First, activity might merely permit refinement, which would be directed by separate wiring instructions. For example, axons might require activity to read out a matched gradient. In this case, activity would be necessary but not sufficient for map maturation.

Alternatively, the pattern of activity might itself contain enough structure to instruct wiring directly. The canonical model for activity-instructed neural wiring posits a competition for synaptic territory in which connection strength is promoted by activity. Synchronous inputs to the same postsynaptic neuron strengthen their connections at the expense of less active or less synchronous inputs. Circuits are organized through cooperation between correlated inputs and competition between uncorrelated inputs [56].

2.2.1.3 Activity-dependent cooperation

In the retinotopy problem, each RGC occupies a unique retinal position and therefore no two RGCs have exactly the same spatiotemporal activity. However, correlations in natural images encourage visual processing elements to draw input from adjacent points in the visual field, so adjacent elements typically share many inputs and can cooperate to drive downstream targets. In a population of equally active RGC axons, each competitor is equally fit, so the preferred configuration is the one that permits the most cooperation at the target, which is achieved when adjacent RGCs project axons to adjacent targets, maximizing receptive field overlap. Retinotopy can therefore be generated from a cooperative process instructed by the statistics of visual computation.

During development, these statistics could be supplied externally by natural sensory experience [3, 71] or internally by correlated spontaneous activity, in the case of structures that develop prior to the onset of sensory input. Locally correlated spontaneous activity has been observed in several developing neural systems, including the spinal cord, cortex, and retina [56, 16, 36].

In the retina, spatiotemporally correlated patterns of rhythmic bursting activity called *retinal waves* appear in immature RGCs prior to the onset of photoreceptor-evoked stimulation [105]. A patterned event begins as a burst of spontaneous spiking in a retinal ganglion cell that recruits the activity of adjacent cells via a network of cholinergic amacrine cells, creating a wavefront of synchronized firing that sweeps across spatially contiguous domains of retina before fading (Figure 2.3). Domain boundaries are not repeated and no particular region is favored, although a refractory period prevents recently active domains from participating in a new wave [37]. Their restricted spatial extent and uniformly distributed initiation sites suggest that retinal waves could supply downstream axons with

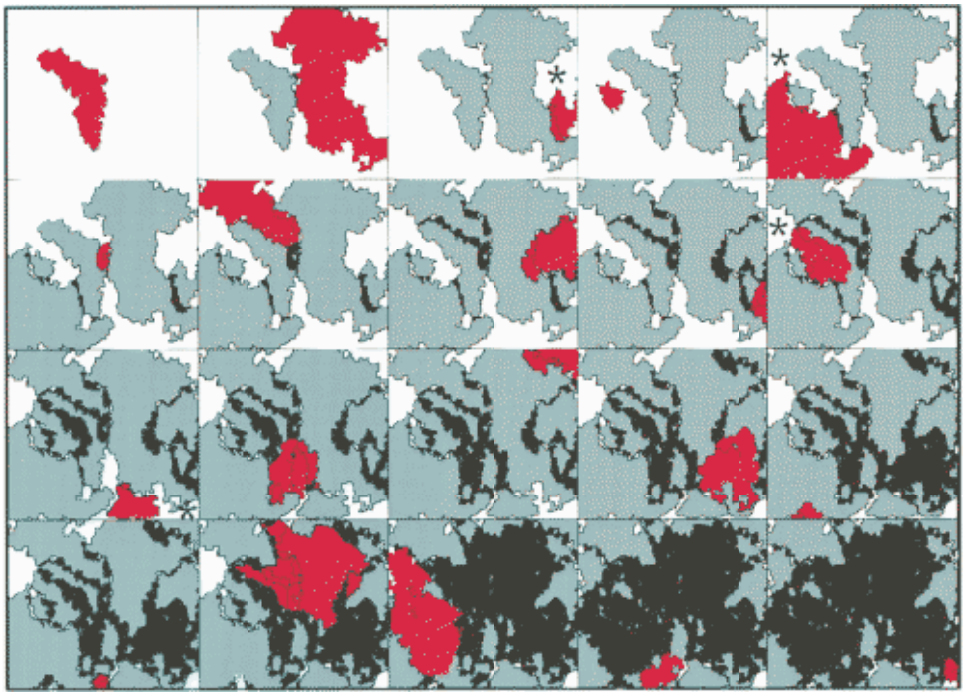


Figure 2.3: Retinal waves

Fluorescence imaging of spontaneous retinal wave domains in P2 ferret retina. Red domains denote the spatial extent of the current retinal wave; previous domains are colored blue. Overlap between domains is colored black. Waves typically propagate unless blocked by refractory participants in recent waves (exceptions are marked with asterisks). The field of view is 1.2mm $\times$ 1.4mm [37].

sufficient information to instruct retinotopic self-organization. An RGC axon could recognize an afferent projected by a retinal neighbor based on the coincident presynaptic activity evoked by participation in the same retinal wave.

Retinal wave correlations play an instructive role in retinotopic map refinement during a critical period of development in the mouse retinocollicular projection. Disabling retinal wave propagation through the network of cholinergic amacrine cells decorrelates the activity of neighboring RGCs during the first postnatal week while leaving absolute firing rates intact. Without these correlations, the diffuse initial projection fails to refine, despite the continued presence of activity. Retinotopy cannot be rescued by the subsequent onset of activity correlations due to glutamatergic retinal waves or visually evoked activity [68].

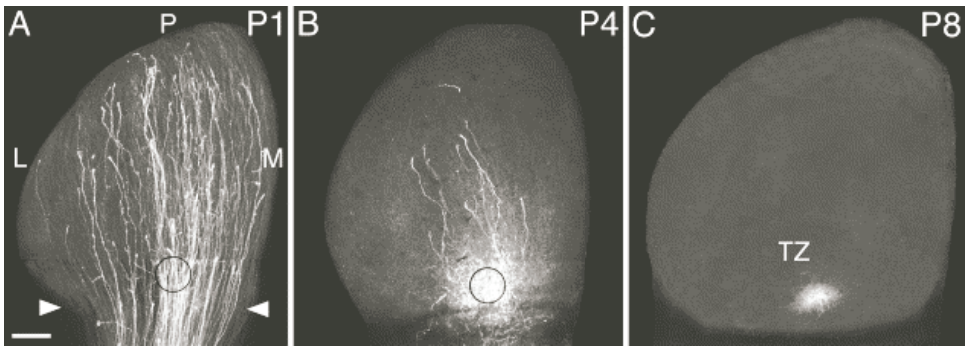


Figure 2.4: Retinotopic refinement

RGC axons innervating the mouse superior colliculus (SC) project to dense termination zones (TZ) that form a retinotopic map. (a) At P1, RGC axons have entered the anterior SC (arrowheads) and project well beyond their final TZ (circle) in the lateral (L), medial (M), and posterior (P) directions. (b) At P4, many axons have eliminated much of their overshoot and branches near the TZ begin to arborize. (c) At P8, the retinocollicular map approaches maturity as all RGC axons resolve to their retinotopic TZ [68].

2.2.1.4 Activity-dependent competition

In the mammalian visual system, RGCs from each eye project axons to the lateral geniculate nucleus (LGN), where they segregate into eye-specific layers, a process that requires postsynaptic activity [91]. Segregation is a competition-driven process in which an afferent's fitness is gated by its level of activity relative to its peers. Varying the level of spontaneous activity in one eye relative to the other causes the layer of the more active eye to expand into the territory of the less active eye [80, 93].

Compared to the problem of retinotopy in which every retinal position defines its own RGC subpopulation, eye-specific segregation is almost trivial, requiring only that within-eye correlations be higher than between-eye correlations [72]. Blocking cholinergic retinal wave propagation does not disturb segregation as long as the level of spontaneous activity in each eye remains intact [54]. A similar effect is observed in the olfactory bulb of transgenic mice, in which afferents segregate onto individual glomeruli according to their

odorant receptor identity. Olfactory maps form normally in the absence of odorant-evoked correlated activity, implying that uncorrelated spontaneous activity is sufficient to permit map formation. However, when presynaptic activity is only blocked in a subpopulation of olfactory neurons, the inactive fibers initially converge onto the correct glomerulus but subsequently vanish from the sensory epithelium, apparently expelled by more active neurons [106].

The ability to distinguish between separate afferent populations is critical for competition-driven segregation. At the rodent neuromuscular junction, each muscle fiber initially receives synaptic input from several inputs. Stronger synaptic inputs annex the territory of weaker inputs, resulting in single innervation of the muscle fiber [4, 8]. This transition from multiple to single innervation of muscle fibers does not occur until after the pattern of activity switches from synchronous stimulation of all inputs to low levels of asynchronous spiking by individual afferents [81].

2.2.2 Visual feature maps

The spiking activity of an RGC is generated from a web of interactions with nearby bipolar cells, amacrine cells, and horizontal interneurons that shape its characteristic response properties, which include retinal position, ocular origin, spatiotemporal precision, wavelength selectivity, and receptive field structure. For example, an ON-center RGC responds optimally to a circular patch of light in the center of its receptive field and is inhibited by illumination of the surrounding ring of photoreceptors, while an OFF-center RGC responds to the opposite pattern of stimulation. These properties lay the foundation for higher visual processing structures. In mammals, RGC axons are bundled into the optic nerve and routed to the LGN, where they segregate into property-specific layers according

to ocular origin and spatiotemporal precision.

ON- and OFF-center afferents from the same eye-specific layer in LGN project to primary visual cortex to form parallel ON- and OFF-subregions within cortical cell receptive fields that respond optimally to oriented light/dark edges presented at a particular retinal position in one eye. Cells tuned to the same orientation and ocular dominance stack vertically into columns perpendicular to the cortical surface. Columns are ordered tangentially as retinotopic maps in which orientation and eye preferences vary smoothly across the cortical surface except for jumps at pinwheel and fracture singularities. Cells in one column extend long range horizontal connections to cells in a different column with similar properties, coordinating selectivity between distant columns.

Patterned activity is in principle sufficient to organize cortical maps, since artificially rerouting retinal afferents to the auditory thalamus generates a two-dimensional retinotopic, ocularly segregated, and orientation selective map in an auditory cortical region that normally hosts a one-dimensional tonotopic map [101, 90, 96]. In practice, map formation tends to be assisted or instructed by intrinsic molecular cues and only subsequently revised by patterned activity. For example, ocular dominance map formation is activity-independent, developing normally even when driven by retinal activity unbalanced by monocular enucleation [22]. However, subsequent maintenance during the critical period after eye opening requires retinal activity [95].

2.2.2.1 Orientation map development

Orientation maps can be derived independently from a variety of compatible mechanisms [92, 31, 83], and perhaps for that reason are remarkably robust against experimental manipulation [102, 88]. Within these maps, the strength of orientation selectivity for in-

dividual cells is instructed by correlated activity. Artificially increasing RGC correlations by even 10% through chronic electrical stimulation of the optic nerve during development dramatically weakens orientation selectivity [102].

Retinal wave correlations disappear too early in development and lack the requisite spatial precision to instruct orientation selectivity [105]. Instead, corticothalamic feedback interacts with intrinsic LGN circuitry to reshape spontaneous retinal activity correlations between ON- and OFF-center RGCs from each eye. LGN cells in the same center-type sublamina of the same eye-specific layer have the highest correlations, while cells in opposite center-type sublaminae of the same eye-specific layer have weaker correlations. Cells in opposite center-type sublaminae of different eye-specific layers have the least, but still significant, correlations [103]. This observed pattern of LGN activity presented to cortex possesses the necessary correlations within and between celltypes to organize orientation-selective receptive fields [31]. Interfering with this correlation structure by selectively blocking ON-center retinal ganglion cell activity freezes the development of orientation selectivity in ferret visual cortex during the critical period [14].

The source of these instructive correlations may be spontaneous activity, visual experience, or both. White et al. examined orientation map maturation in ferrets raised in the dark, which eliminates all visual input, or with binocular lid suture, which activates the retina only with very low spatial and temporal frequencies. Dark-reared ferrets still develop weak orientation maps whose diminished selectivity can be attributed to a lack of intercolumn horizontal connections. Lid-sutured ferrets fail to develop orientation maps at all, implying that intracolumn selectivity can be disrupted by abnormal patterns of visually evoked activity. These results suggest that spontaneous activity alone can tune orientation selectivity within a column, but cannot coordinate long range horizontal connections between columns. Visual experience then sharpens tuning by providing intercolumn corre-

lations [104].

2.3 Models of neural map development

Computational models of neural development generate insight by boiling an opaque complex system down to a few transparent ingredients. Modelers strip away all system components that seem irrelevant to the property of interest to carve out a toy system that is simple enough to analyze or intuitive enough to explain. For example, growth cone dynamics are carefully detailed when describing axon guidance by diffusible gradients, and completely ignored when describing the dimension-reducing nature of the cortical surface. The justification for any simplified model is its ability to correctly postdict existing experimental data, and to generate novel predictions for future experiments.

2.3.1 Axon guidance models

When describing multiple axons navigating local guidance cue gradients, the standard strategy is to construct complicated dynamic equations for axon growth derived from the underlying reaction-diffusion processes, preserving mechanistic intuition at the cost of analytical tractability. At this level of complexity, modelers must content themselves with numerical iteration of these equations, which are tested by comparing the population statistics of the simulated axon trajectories with those of real axons for measurable quantities like neurite length and axon fasciculation [87, 50].

Goodhill and coworkers have been able to derive more analytical results from the simpler case of gradient detection by a single growth cone. They calculated the optimal shape of a guidance cue gradient for a growth cone that measures gradient steepness by sens-

ing either absolute concentration increments across its body or percentage increments, and found that in each case the maximum guidance range is about 1 cm. Conversely, they observed that the shape of the gradient *in situ* should predict the slope measurement technique actually employed by the growth cone [42, 43]. Goodhill also extracted the steepness constraints on the Slit gradient *in vivo* that are required for consistency between published experimental results [41].

Models of chemotropic axon guidance do not generalize well to the collective property of map formation, which is more fruitfully studied at a higher level by abstracting away the inner workings of the axons to focus on interactions between axons driven by correlated activity patterns.

2.3.2 Hebbian map models

A number of neural net models have been proposed that generate cortical feature maps by applying the Hebb rule to input activity correlations [97, 100]. We will briefly review several of these *Hebbian map models* in terms of a simple two-layer neural network that consists of a source layer, indexed by Greek letters α, β; and a target layer, indexed by Roman letters $\mathbf{x}, \mathbf{y}$. Source cell α is connected to target cell $\mathbf{x}$ by a synapse with weight $w_{\alpha\mathbf{x}}$. Source cell activity $s_\alpha(t)$ is called *presynaptic* and target cell activity $v_\mathbf{x}(t)$ is called *postsynaptic*.

The classical Hebbian learning rule [48] postulates that synapses are strengthened by correlations between their presynaptic and postsynaptic activity. Synaptic weight updates

take the form

$$\frac{d}{dt} w_{\alpha \mathbf{x}}(t) = \eta s_{\alpha}(t) \sum_{\mathbf{y}} I(\mathbf{x}, \mathbf{y}) v_{\mathbf{y}}(t) \tag{2.1}$$

where η is an update rate and $I(\mathbf{x}, \mathbf{y})$ is an interaction function that describes the effect on target cell $\mathbf{x}$ by activity from neighboring target cells $\mathbf{y}$. The range of lateral target cell connections described by $I(\mathbf{x}, \mathbf{y})$ can be used to divide Hebbian map models into two general categories [34].

Global Hebbian map models use a nonlinear $I(\mathbf{x}, \mathbf{y})$ to select for weight update the one target cell out of the entire population that receives the most activation from the current input pattern. Since this function involves a global optimization, each target cell requires information from every other target cell, implying an all-to-all connectivity within the target layer. *Local Hebbian map models* reduce the necessary lateral connectivity through the use of a softer selection criterion based on the degree of correlation between the activities of each target cell and its input afferents. This interaction function performs a more local optimization, which can be computed through a smaller number of mostly short-range connections, typically described by a linear kernel $I(\mathbf{x}, \mathbf{y}) = I(\mathbf{x} - \mathbf{y})$ centered on each target cell.

Global Hebbian map models are used when the optimality of the results is important and connections are cheap to implement, as in a software implementation of a pattern classifier. They have been extensively studied by the machine learning community. Local Hebbian map models are used when the biological plausibility of the classification procedure is important or connections are expensive to implement, as in a developmental neuromorphic silicon chip. They appear mainly in the computational neuroscience literature.

2.3.3 Local Hebbian map models

Since unconstrained Hebbian synapse strengthening causes synaptic weights to increase without bound, local Hebbian map models typically *normalize* the sum of all weights entering and/or exiting a cell to a constant value. These global normalization schemes induce competition between synapses, since any increase in one set of weights must be balanced by an equivalent decrease from the remaining set of weights. Another way to confine weight values is to restrict them to some allowed range $w_{\min} \leq w \leq w_{\max}$. Weights that reach either of the limits are said to *saturate*.

Hebbian models typically wave away the specifics of filopodial sprouting and retracting by initializing the weights with weak nonzero values for all connections. Subsequent application of the learning dynamics on this blank slate leads to a positive feedback cycle that either saturates weights at their maximum strength or prunes them completely. Sprouting exuberance can be spatially constrained by an *arbor function* that cuts weights off after a certain distance.

2.3.3.1 Correlation-based learning

Linsker [63, 62, 61] simulated weight vector evolution in a feedforward linear network using a Hebbian update rule based on an ensemble average of input activity pattern statistics.

$$\frac{\partial}{\partial t} w_\alpha = k_1 + \sum_\beta \left(Q_{\alpha\beta} + k_2 \right) w_\beta \tag{2.2}$$

Here, k_1 and k_2 are model parameters and $Q_{\alpha\beta}$ is the covariance between the activities of input cells α and β. The elements of the covariance matrix can be positive or negative,

so weights saturate at $\pm w_{\max}$. There is no explicit normalization, but the hard saturation limits restrict the choice of weight vectors to the interior of a hypercube in weight space. Linsker found that an isotropic Gaussian covariance function could lead to interesting center-surround or even bilobed receptive field structures, depending on parameter choice.

MacKay and Miller [65] recast Linsker's update rule as the matrix equation

$$\dot{\mathbf{w}} = (\mathbf{Q} + k_2\mathbf{I})\,\mathbf{w} + k_1\mathbf{n} \tag{2.3}$$

subject to the saturation constraints $-w_{\max} \leq w_\alpha \leq w_{\max}$. Here, $\mathbf{I}$ is the identity matrix whose elements $I_{\alpha\beta}$ are 1 for $\alpha = \beta$ and 0 for $\alpha \neq \beta$, and $\mathbf{n}$ is the vector with all elements $n_\alpha = 1$. $\mathbf{Q} + k_2\mathbf{I}$ is symmetric, so it has a complete set of orthonormal eigenvectors $\mathbf{e}^{(a)}$ with real eigenvalues λ_a. Weight vectors can therefore be written as a weighted sum of these eigenvectors, $\mathbf{w}(t) = \sum_a w_a(t)\mathbf{e}^{(a)}$, and the component of $\mathbf{w}$ in the direction of each eigenvector $\mathbf{e}^{(a)}$ grows or decays exponentially at a rate proportional to its eigenvalue λ_a, relative to the fixed point of the update equation.

$$w_a(t) - w_a^{\mathrm{FP}} = \left[w_a(0) - w_a^{\mathrm{FP}}\right] e^{\lambda_a t} \tag{2.4}$$

Weights that reach the saturation limits freeze their dynamics, so the final receptive field is determined by the first eigenvector to grow to those limits. The principal eigenvector is the eigenvector that corresponds to the largest positive eigenvalue and therefore grows at the fastest rate. Its weights are the first to saturate, locking the receptive field into an image of the winning eigenvector.

Miller subsequently applied this eigensystem analysis to a more biological linear up-

date model. Since spike frequencies cannot be negative, the Miller model relies only on positive activity correlations between and within separate source cell populations [72, 71]. Similarly, weights are restricted to non-negative values $0 \leq w^i_{\alpha\mathbf{x}} \leq w_{\max}$ to model a purely excitatory feed-forward projection. Total synaptic input to a given target cell is normalized to a constant to prevent all synapses within an arbor radius from converging onto the same target cell.

$$\frac{d}{dt}\sum_{\alpha} w^i_{\alpha\mathbf{x}}(t) = 0 \tag{2.5}$$

Labeling source cell populations with i, j, the weight update rule is

$$\frac{d}{dt}w^i_{\alpha\mathbf{x}}(t) = \eta A(\mathbf{x},\alpha)\sum_{\mathbf{y}} I(\mathbf{y},\mathbf{x})\sum_{j,\beta} C^{ij}(\alpha,\beta)w^j_{\beta\mathbf{y}}(t) \tag{2.6}$$

where η is a constant update rate, $A(\mathbf{x},\alpha)$ is a localized arbor function that enforces global retinotopy by restricting the spatial distribution in the target layer of synapses projected by each source cell α, and $I(\mathbf{y},\mathbf{x})$ is a Mexican hat interaction function that relates activity at target cell $\mathbf{x}$ to target cell $\mathbf{y}$. $C^{ij}(\alpha,\beta)$ is the correlation between the activities of source cell α in population i and source cell β in population j. For example, $C^{\{\text{LEFT,ON}\},\{\text{RIGHT,OFF}\}}(\alpha,\beta)$ might be the correlation between the ON-center RGC located at retinal position α in the left eye and the OFF-center RGC located at retinal position β in the right eye.

ON- and OFF-center inputs from each eye can assemble correlation modes that generate binocularly matched orientation maps. In this case, there are four source cell populations, one for each center-type/eye origin pairing. The sixteen corresponding correlation func-

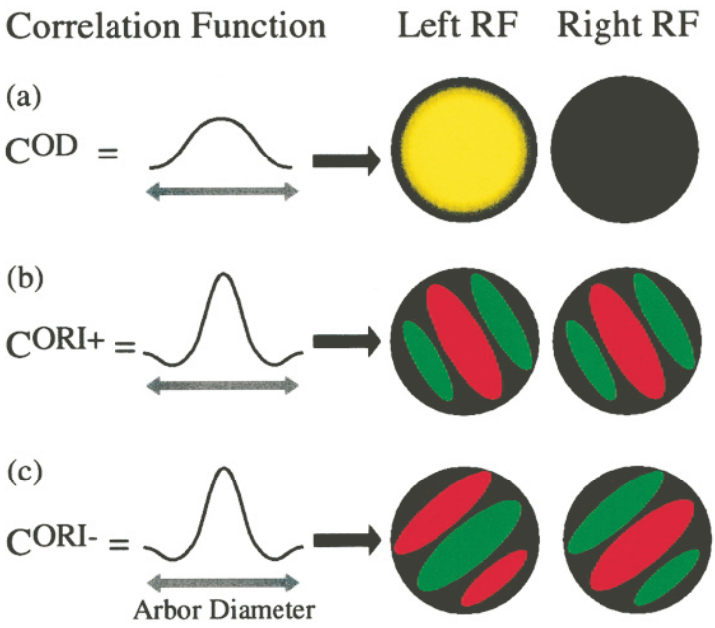


Figure 2.5: Correlation modes required for ocular dominance and orientation selectivity

(a) Monocular receptive fields segregate if C^{OD} does not oscillate within an arbor radius [72] (b) $C^{\mathrm{ORI}+}$ organizes oriented receptive fields that are in phase between eyes. (c) $C^{\mathrm{ORI}-}$ organizes oriented receptive fields that are antiphase between eyes. Binocularly matched orientation maps can form if $C^{\mathrm{ORI}+}$ or $C^{\mathrm{ORI}-}$ oscillate once within an arbor radius [31].

tions can be reduced to four correlation modes by exploiting symmetry:

$$
\begin{aligned}
C^{\mathrm{SUM}} &= C^{\mathrm{S_E,S_C}} + C^{\mathrm{S_E,O_C}} + C^{\mathrm{O_E,S_C}} + C^{\mathrm{O_E,O_C}} \\
C^{\mathrm{OD}} &= C^{\mathrm{S_E,S_C}} + C^{\mathrm{S_E,O_C}} - C^{\mathrm{O_E,S_C}} - C^{\mathrm{O_E,O_C}} \\
C^{\mathrm{ORI}+} &= C^{\mathrm{S_E,S_C}} - C^{\mathrm{S_E,O_C}} + C^{\mathrm{O_E,S_C}} - C^{\mathrm{O_E,O_C}} \\
C^{\mathrm{ORI}-} &= C^{\mathrm{S_E,S_C}} - C^{\mathrm{S_E,O_C}} - C^{\mathrm{O_E,S_C}} + C^{\mathrm{O_E,O_C}}
\end{aligned}
$$

Here, $C^{\mathrm{S_E,S_C}}$ is the correlation between source cells in the same eye with the same center-type, $C^{\mathrm{O_E,O_C}}$ is the correlation between source cells in opposite eyes with opposite center-types, and so on. The ocular dominance mode C^{OD} is just the difference in correlations within and between eyes. If intraocular correlations exceed interocular correlations,

monocular receptive fields form (Figure 2.5(a)). There are two orientation-selective modes: C^{ORI+} is the sum of the left-eye and right-eye ON/OFF correlation differences, and corresponds to ON/OFF subregions that are in phase between eyes (Figure 2.5(b)), and C^{ORI-} is the difference between the left-eye and right-eye ON/OFF correlation differences, corresponding to ON/OFF subregions that are antiphase between eyes (Figure 2.5(c)).

2.3.3.2 Competition for trophic sustenance

Local Hebbian models typically invoke some normalization scheme that is selected more for its stabilizing or competition-inducing properties than its attention to specific biological mechanisms. To dispel this abstraction, several modelers have studied the activity-dependent dynamics of neurotrophic factors (NTFs) as a candidate normalization process [100]. These trophic models tend to reproduce the predictions of more general correlation-based models, since both schemes employ similar Hebbian cooperation rules. Trophic models specify an NTF-dependent competition mechanism and therefore make the additional obvious prediction that flooding the system with an excess of NTF will prevent normal development.

Elliott and Shadbolt have extensively analyzed a simplified trophic model whose dynamics are described by

$$\frac{dw_{\alpha\mathbf{x}}}{dt} = -\eta w_{\alpha\mathbf{x}} + \eta \left[\frac{w_{\alpha\mathbf{x}} \left(a + a_\alpha\right) \rho_\alpha}{\sum_\beta w_{\beta\mathbf{x}} \left(a + a_\beta\right) \rho_\beta} \sum_{\mathbf{y}} \Delta_{\mathbf{xy}} \left(T_0 + T_1 \frac{\sum_\beta w_{\beta\mathbf{y}} a_\beta}{\sum_\beta w_{\beta\mathbf{y}}} \right) \right] \tag{2.7}$$

The basic idea is that the number of synapses $w_{\alpha\mathbf{x}}$ connecting afferent α to target $\mathbf{x}$ depends on the amount of NTF taken up by each synapse, under the assumptions that NTF release is gated by target cell activity and NTF uptake is gated by afferent cell activity. The first

term on the right represents activity-independent weight erosion, which causes synapses to wither unless supplied with the trophic support described by the second term, which models the proportion of the NTF supply available at $\mathbf{x}$ that is taken up by afferent α.

Target cell release of NTF has an activity-independent component and an activity-dependent component. Activity-dependent NTF release from target cell $\mathbf{y}$ is proportional to postsynaptic activity, which is approximated with the dendritic excitation $\sum_\beta w_{\beta\mathbf{y}} a_\beta$ and normalized against the total synaptic weight $\sum_\beta w_{\beta\mathbf{y}}$ innervating $\mathbf{y}$ to imply that each target cell can only synthesize enough NTF to sustain a limited synaptic area. The relative strengths of the activity-independent and activity-dependent release components is set by the parameter ratio T_0/T_1.

The sum over $\mathbf{y}$ represents the total NTF pool available at $\mathbf{x}$, which combines contributions from all possible release sites $\mathbf{y}$ to the current site $\mathbf{x}$. NTF diffusion from $\mathbf{y}$ to $\mathbf{x}$ is modeled by the function $\Delta_{\mathbf{xy}}$. The fraction in the brackets before the sum describes afferent competition for NTF at $\mathbf{x}$. Afferents compete for shares of the available NTF pool by expressing and activating synaptic NTF receptors. Afferent α binds NTF from $\mathbf{x}$ in direct proportion to the share of the total population of high-affinity receptors present at $\mathbf{x}$ that belong to α.

The number of high-affinity NTF receptors expressed by α at $\mathbf{x}$ is just the total synaptic area $w_{\alpha\mathbf{x}}$ multiplied by the binding affinity $(a+a_\alpha)$ and the receptor density ρ_α. The binding affinity has an activity-independent component, controlled by parameter a, and an activity-dependent component, controlled by the afferent activity a_α. NTF receptor expression ρ_α is activity-dependent, described by $\rho_\alpha = \bar{a}_\alpha / \sum_{\mathbf{x}} w_{\alpha\mathbf{x}}$, where $\bar{a}_\alpha$ is a sliding average of recent afferent α activity. ρ_α is normalized against the total weight strength maintained by source cell α to account for a limited pool of presynaptic resources.

To analyze their model mathematically, Elliott and Shadbolt examined the simple case of two axons converging onto the same neuron. There are three fixed points, two corresponding to single innervation from each neuron and one corresponding to multiple innervation. The stability of these fixed points depend on the relative balance of activity-dependent and activity-independent NTF release. If activity-dependent release dominates, the system tips to single innervation by the axon initialized with the most synapses. If activity-independent release dominates, both axons remain [29]. A unique prediction of this model is that axons that are about to be displaced from a target should immediately downregulate their receptor expression [28].

When driven by simulated retinal waves, a three layer version of this model develops retinogeniculocortical connectivity, self-organizing LGN retinotopy and cortical ocular dominance maps. The model is initialized with a global retinotopic bias to speed convergence and protect synapses from premature elimination. As expected, flooding the system with exogenous NTF prevents map formation [30].

2.3.4 Global Hebbian map models

Local Hebbian map models tend to get trapped in globally suboptimal states unless initialized with a considerable amount of prior information such as coarse global retinotopy. While this might be appropriate for biological models, which can assume such instruction is provided by independent genetic mechanisms, it cripples machine learning algorithms, which typically prefer to assume as little as possible. As a consequence, most engineering and mathematical effort focuses on global Hebbian map models, the most popular of which is the self-organizing map (SOM) algorithm introduced by Kohonen [59].

Kohonen observed that the intracortical connections between many cortical cells tend to be excitatory at short distances and inhibitory at long distances, and that this center-surround lateral connectivity might permit the most active target cell to recruit activity from nearby target cells while silencing more distant target cells. Applying the Hebb rule, the only target cells with enough activity to potentiate weights would be clustered around the target cell with the strongest response to the current pattern of source cell activity. Kohonen proposed the following heuristic algorithm:

1. At each time step, draw one pattern vector $\mathbf{v}$ from a prior distribution of source cell activity vectors, where the vector element v_α is the activity of source cell α during the time step.

2. Next, identify the target cell $\mathbf{y}$ that responds most strongly to the input $\mathbf{v}$. This is the neuron whose weights $\mathbf{w_y}$ have the most overlap with the participating source cells, so $\mathbf{y} = \min_{\mathbf{x}} \|\mathbf{v} - \mathbf{w_x}\|$.

3. Finally, update the weight vectors $\mathbf{w_x}$ of target cells $\mathbf{x}$ according to their distance from the winning cell $\mathbf{y}$:

$$\mathbf{w_x}(t+1) = \mathbf{w_x}(t) + \eta(t) h(\mathbf{x}, \mathbf{y}, t) \left[\mathbf{v} - \mathbf{w_x}\right] \tag{2.8}$$

   where the neighborhood function $h(\mathbf{x}, \mathbf{y}, t)$ is a decreasing function of separation $|\mathbf{x} - \mathbf{y}|$ and is usually taken to be a Gaussian. The update rate $\eta(t)$ and the width of $h(\mathbf{x}, \mathbf{y}, t)$ are typically initialized at large values and gradually reduced over time, encouraging rough global ordering early in the weight evolution and precise local optimization later, a procedure analogous to simulated annealing. The neighborhood function was originally inspired by cortical microcircuits, but can equivalently be interpreted as the spread of a target-derived diffusible factor such as nitric oxide.

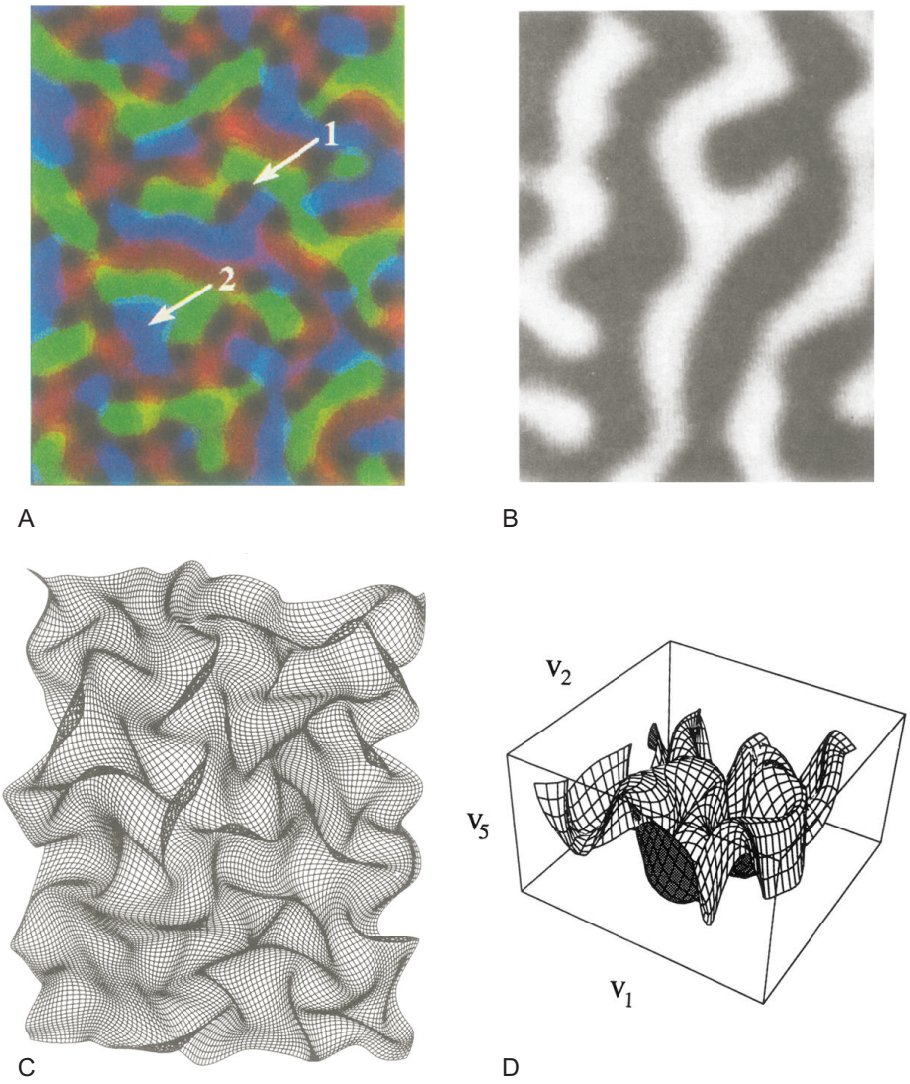


Figure 2.6: SOM as dimension-reducing map

The SOM algorithm compresses a high dimensional feature space into a two-dimensional cortical surface. (a) Orientation map. Arrows point to pinwheel (1) and fracture (2) singularities. (b) Ocular dominance map. (c) Locally distorted representation of retinal space. Receptive field centers of adjacent target cells are connected by lines. (d) The cortical sheet folds through five-dimensional feature space to try to cover competing feature maps while preserving local continuity. V1 and V2 are the retinotopic coordinates and V5 is the degree of monocularity. The orientation selectivity coordinates V3 and V4 are suppressed for visualization purposes [79].

More abstractly, the SOM model can be interpreted as a dimension-reducing algorithm that compresses a high-dimensional feature space into a two-dimensional cortical map. For example, a five-dimensional feature space might be defined by two retinotopic coordinates, the direction and magnitude of orientation selectivity, and the degree of monocular innervation. The cortical surface is represented as a two-dimensional sheet that folds and flattens in this five-dimensional feature space in an attempt to simultaneously satisfy conflicting constraints (Figure 2.6). The Hebbian correlation rule pulls the sheet to cover a representative sample of feature points, while the neighborhood function tries to keep the sheet locally smooth [79]. The tradeoff between these two constraints is controlled by the distribution of input pattern vectors in feature space, and can be analyzed with techniques borrowed from statistical physics such as energy functions, phase transitions and stochastic processes [33, 32, 79, 51, 44].

SOM algorithms derive much of their power from the winner-take-all computation, which tends to rescue the system from getting trapped in local optima. This winner-take-all function becomes increasingly unwieldy in larger networks, since each cell's activity must be compared with every other cell's activity, implying a more universal connectivity than is physiologically observed. Models that value biological plausibility over technical performance therefore tend to discard the winner-take-all scheme in favor of the more local computations that inspired it.

2.3.5 VLSI implementations

Most map formation chips implement some variant of the SOM algorithm. The first attempts were digital [59] and were restricted by low numerical precision to a limited range of weight values. In pursuit of compact circuits, subsequent work focused on individual

analog circuits for high precision operations like weight multiplications [35], distance measurements [13], neighborhood functions [99], and winner-take-all computations [13].

Despite early examples of fully analog SOM chips [66, 47] and mixed analog-digital vector quantization [35, 84] and clustering chips [89], more recent implementations tend to be fully digital, due to the more favorable tradeoff between numerical precision and silicon area. Current work focuses on accelerating the design cycle by shifting from full custom application-specific integrated circuit (ASIC) design [69] to more generic hardware such as digital signal processors [11], programmable logic [7], general purpose multiprocessor neurocomputers [82], and soft intellectual property (IP) cores [49].

The closest thing to a VLSI implementation of a local Hebbian map model was presented by Elliott and Kramer, who used output from a spiking silicon retina to drive self-organization of retinotopy and ocular dominance using the Elliott and Shadbolt model of NTF-dependent map formation [26, 27]. However, only the input layer was implemented in hardware, which was used to generate realistic spike trains whose variability could be tuned to test the robustness of the algorithm. Weight storage and update was implemented entirely in software running on a separate workstation. To date, there has been no fully hardware implementation of a local Hebbian map model.

Chapter 3

Neurotrope1 system

In this chapter we present Neurotrope1, a neuromorphic system based on a simple correlation-based learning rule in which cells that fire at the same time are wired to the same place. This rule is inspired by axon guidance during neural development.

3.1 Self-wiring axons

Our basic goal is to self-organize feed-forward excitatory connections between two layers of neurons. Neurons in the first layer, which we will call *source cells*, extend axons into the second layer of neurons, which we will call *target cells*. At one end of the axon, the source *cell body* generates spikes in response to excitation and transmits the spikes down the axon. At the other end, the *axon terminal* receives spikes from the axon and releases neurotransmitter, exciting nearby target cell bodies. The cell body remains fixed at some location in the source layer, while the axon terminal may move between locations in the target layer. The description of any axon can be reduced to the coordinates of its two

endpoints within their respective layers.

In a mature network, the axon terminal forms a *synapse* that anchors the axon to its designated target cell and enhances excitatory neurotransmission. In a developing network, the axon terminal forms a *growth cone* that guides axon motion within the target layer. Regardless of whether an axon terminal forms a synapse or a growth cone, we will refer to source cell activity as *presynaptic* and target cell activity as *postsynaptic*.

In the last chapter we saw that a growth cone is an amoeboid body located at the tip of a growing axon that continually extends and retracts fingers of membrane called *filopodia* to feel its way along its migration path, towing the axon behind it. The growth cone moves when it extends more filopodia in one direction than it retracts from that direction. Filopodia bind extracellular chemicals using membrane receptors that are activated by presynaptic spikes. Some chemicals stabilize filopodia when bound, reducing the filopodial retraction rate and causing net growth cone motion in the direction of increasing chemical concentration. Growth cone motion can thus be gated by presynaptic activity and guided by local extracellular cues.

The properties of these extracellular cues are obviously central to understanding how the brain is able to wire itself up. We are going to postulate the existence of a target-derived, activity-dependent, extracellular guidance chemical which we will call *neurotropin*. We will use neurotropin to construct the following automatic wiring program (Figure 3.1):

1. A layer of source cells projects axons to a layer of target cells. Elongating axons are tipped by growth cones that guide the axons to their targets.

2. Presynaptic spikes generated by a source cell trigger neurotransmitter release from its axon terminal, exciting neighboring target cells.

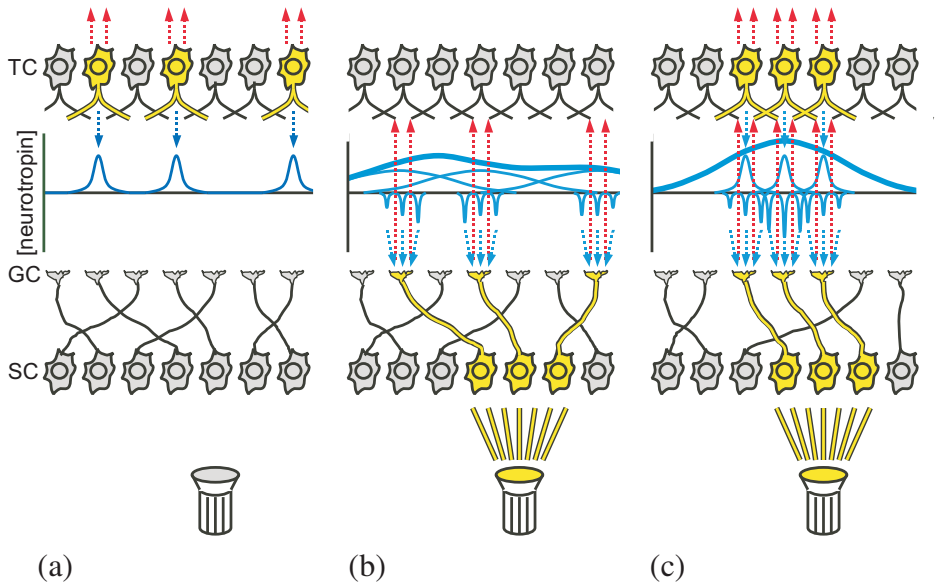


Figure 3.1: Neurotropic axon guidance

(a) Spiking (yellow) target cells (TC) release pulses of neurotropin into the surrounding medium. (b) Between spikes, neurotropin diffuses, establishing a spatial concentration profile. Active growth cones (GC) excite target cells (red) and measure the local neurotropin gradient. (c) Growth cones migrate up the detected gradient, tending to cluster correlated axon terminals.

3. Postsynaptic spikes generated by a target cell trigger neurotropin release from the target cell body.

4. Neurotropin diffuses spatially within the target layer until being consumed by glia or bound by active growth cones.

5. Presynaptic spikes generated by a source cell also cause its growth cone filopodia to bind extracellular neurotropin.

6. A growth cone moves up the local neurotropin gradient by comparing amounts of neurotropin bound by opposing filopodia. When there is no gradient, filopodia pull equally in all directions and there is no net motion. When there is a gradient, the net effect is that the growth cone climbs the local neurotropin gradient, looking for concentration peaks whose presence coincides with the growth cone's presynaptic

activation.

7. Neurotropin concentration peaks are centered on target cells that are being stimulated by active growth cones, so moving growth cones toward detected concentration peaks is a way to cluster coactive growth cones. Source cells that fire at the same time will end up wired to neighboring target cells ('cells that fire together, wire together').

8. Axons also take up physical space, meaning they can't maximize neurotropin release by all converging onto the same target cell. This density constraint requires a moving growth cone to displace other growth cones in its path. Axon targeting is therefore a balance between directed growth cone guidance and undirected displacement.

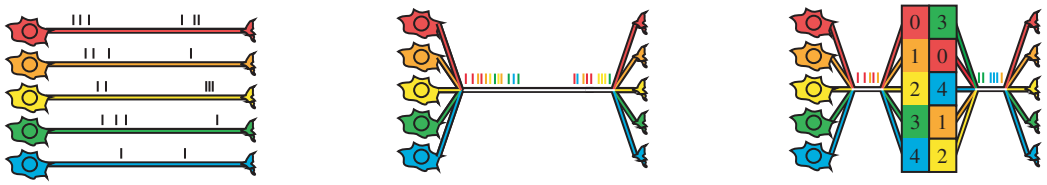
(An equivalent formulation of this algorithm on a longer time scale is to relabel growth cones as axon arbors and filopodia as axon branches. In this case, axon branches form synapses that are stabilized or destabilized by postsynaptically released survival signals called neurotrophins. Neurotrophins do not diffuse significantly from their release sites; instead, release sites are distributed spatially by a dendritic arbor.)

The Neurotrope1 system implements this algorithm using three major elements:

- *Soft wires:* Axons are implemented virtually as entries in memory that can be automatically altered to rewire connections.

- *Neurotropin chip:* Silicon growth cone circuits use a local target-derived signal to compute updates for the virtual axon table.

- *Correlated stimulus:* Presynaptic activity correlations drive self-organization of feedforward connections.



(a) (b) (c)

Figure 3.2: Address-event representation

(a) Biological networks implement a dedicated axon for every neuron. (b) Address-event representation requires all neurons to share the bandwidth of one common axon. Each spike must be tagged with the address of its neural source. (c) For non-trivial network connectivity, address-events must be filtered through a lookup table that translates spike source addresses into spike target addresses.

3.2 Soft wires

In biological neural networks, each neuron has its own dedicated axon that physically wires its cell body to every target. A signal consists of a voltage spike arriving from a particular wire at a particular time — an asynchronous one-hot code (Figure 3.2(a)).

When constructing large neuromorphic systems it remains infeasible to assemble a physical wire for every neural connection, due to space constraints. *Address-event representation* (AER) allows all cells on the same chip to share a common output axon (Figure 3.2(b)). As a consequence, each spike communicated on this shared axon must be tagged with its designated target address. However, neuromorphic chips typically label their output spikes with the address of the source cell, which is not necessarily the same as the address of the target cell on a different chip. When the source address space and the target address space differ, spikes must be filtered through a lookup table that decodes source addresses into target addresses. We call this lookup table the *forward map* (Figure 3.2(c)).

All the information we need to route a spike is stored in its neuron's entry in the forward map. We say the axon is implemented *virtually* in memory. Rewiring axon targets is just a

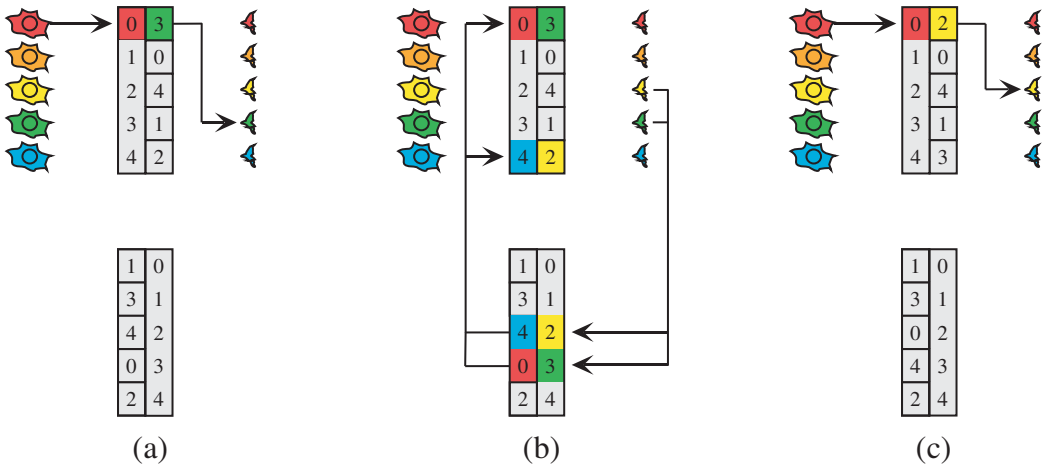


Figure 3.3: Address-event remapping

(a) A virtual axon encoded in the forward map (top) routes a spike from its cell body to its growth cone. (b) Axon migration originates in growth cones, which use the reverse map (bottom) to identify their innervating cell bodies. Axons move by modifying the appropriate forward and reverse map entries. (c) The updated forward map routes a spike from the cell body to its new growth cone location.

matter of changing entries in the forward map. We would like our system to update these entries automatically, based on some simple learning rule.

In the Neurotrope1 system, an update is computed by a growth cone circuit at the target end of the virtual axon, and is transmitted from the target chip to the lookup table using an address-event that encodes the desired growth cone motion relative to its current location. Since this address-event is generated in the target chip, it is tagged with a target address that must be filtered through a *reverse map* that decodes target addresses into source addresses. The resulting source address is used to index the relevant entry in the forward map, which is changed to the new target. The reverse map must also be updated to reflect the new connectivity (Figure 3.3).

This update protocol actually moves two virtual axons: the one whose growth cone requested the update, and the one already occupying the requested target. So one axon

moves closer to its target and one axon is bumped in a random direction. This displacement noise is equivalent to a hard density constraint and enforces competition between growth cones.

The forward and reverse maps are implemented with an off-chip SRAM. A computer can write to and read from the SRAM via a USB port controlled by a Ubicom ip2022 microcontroller. The microcontroller also performs all modifications to the lookup table. Target address lookup is handled separately for faster execution, using asynchronous logic implemented in a CPLD (complex programmable logic device).

3.3 Neurotropin chip

The neurotropin chip, Neurotrope1, interleaves a 48×20 array of growth cones with a 24×20 array of target neurons. The 11.5 mm^2 chip was fabricated through MOSIS using the TSMC 0.35 $\mu\mathrm{m}$ process.

The dominant feature of the Neurotrope1 chip is a monolithic diffusive p-type channel that models the extracellular medium. This channel is laid out as a continuous honeycomb lattice, with target cell and growth cone circuitry occupying the hexagonal cells of the lattice (Figure 3.4). Neurotropin is represented as charge diffusing in the lattice, and can be injected or drained at each node.

Neurotropin is released into the extracellular medium by the cell bodies of spiking target cells. Source cell spikes are routed through the lookup table to growth cone circuits that simultaneously excite nearby target cells and sample the neurotropin concentration in the surrounding extracellular medium. Each growth cone circuit extends three filopodia to nearest neighbor nodes (Figure 3.5).

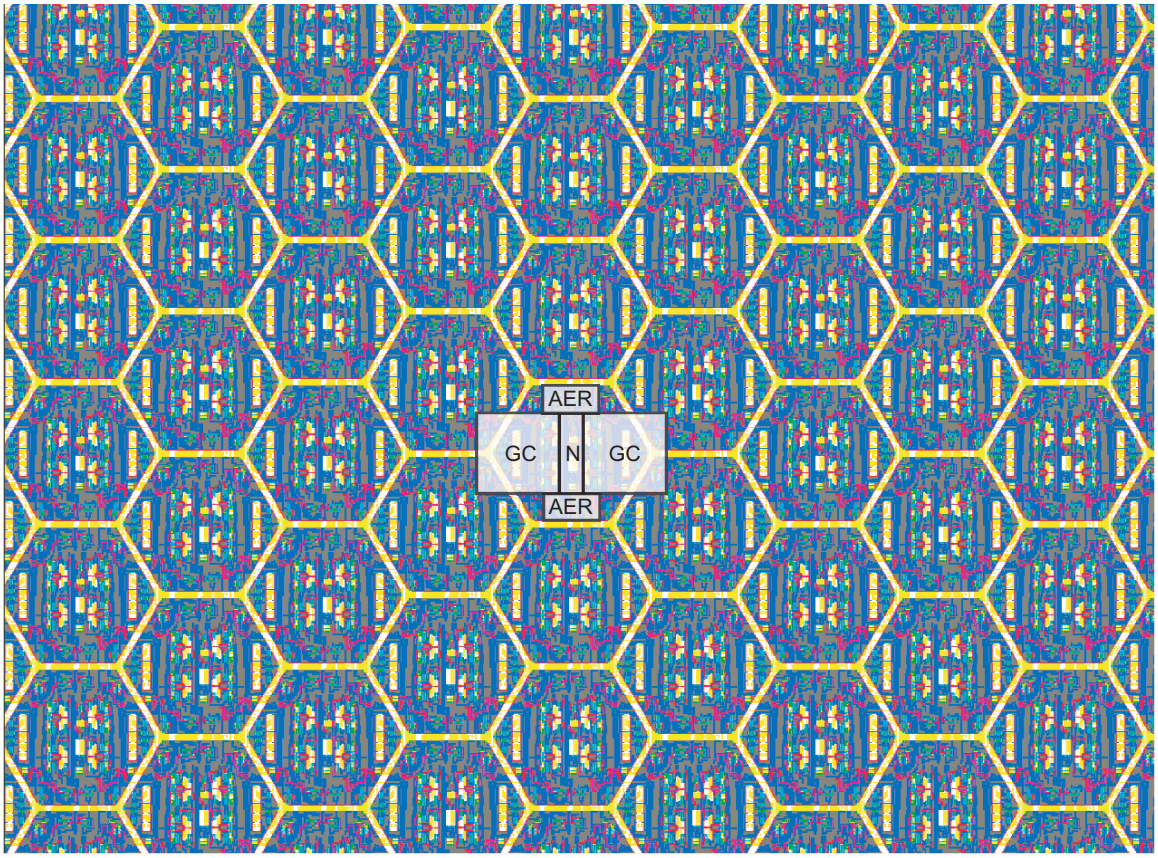


Figure 3.4: Neurotrope1 array layout

Detail of Neurotrope1 chip layout. Growth cone circuits (GC) occupy nodes of the monolithic honeycomb lattice, while target cell circuits (N) occupy lattice cells. The remaining space is filled by off-chip communication circuits (AER).

3.3.1 Neurotropin circuit

Figure 3.6 displays the circuits that add, diffuse, and remove charge in the extracellular medium modeled by the hexagonal charge-diffusing lattice M1. V_{spread} sets the maximum amount of charge M1 can hold. The total charge in M1 is determined by circuits that implement activity-dependent release and uptake. In addition, M11 and M12 provide a path for activity-independent release and uptake.

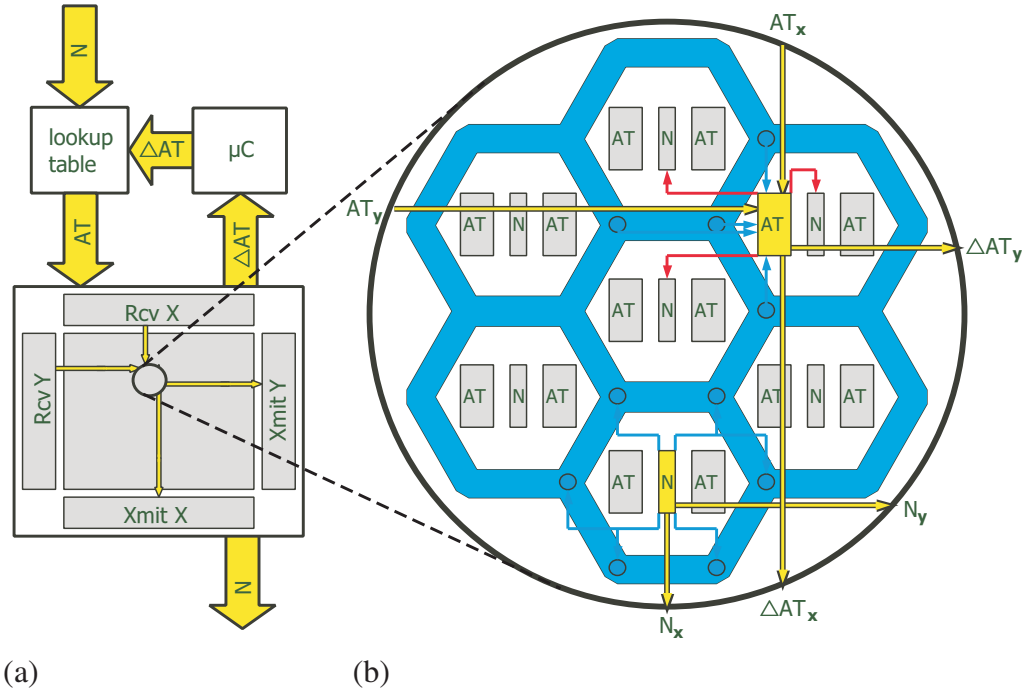


Figure 3.5: Block diagram

(a) The lookup table converts source cell addresses (N) into axon terminal addresses (AT) and passes them into the neurotropin chip. The neurotropin chip outputs target cell spikes (N) and axon terminal address updates (ΔAT) on the same shared port. Target cell spikes are routed to the next layer of cells, while address updates are sent to a microcontroller (μC) that writes them to the lookup table. (b) Upon receiving a presynaptic address-event, the neurotropin chip toggles row and column selects ($\mathrm{AT_{x,y}}$) to activate the indicated axon terminal, which excites nearby target cells (red) and samples neurotropin from the extracellular medium (blue). Upon completing a gradient measurement, the axon terminal requests an address update ($\Delta\mathrm{AT_{x,y}}$). Spiking target cells release neurotropin into the extracellular medium and transmit their spikes off-chip ($\mathrm{N_{x,y}}$).

Target cell activity triggers neurotropin release, as implemented by the circuit in the left box of Figure 3.6. Spikes from any of the three neighboring target cells pull C_{spost} to ground, opening M7 and discharging C_{fpost} through M4 and M5. As C_{fpost} falls, M6 opens, establishing a transient path from V_{dd} to M1 that injects charge into the hexagonal lattice. Upon termination of the target cell spike, C_{spost} and C_{fpost} are recharged by decay currents through M2 and M3. V_{ppost} and V_{fpostout} are chosen such that C_{spost} relaxes faster than C_{fpost}, permitting C_{fpost} to integrate several target cell spikes and facilitate charge injection

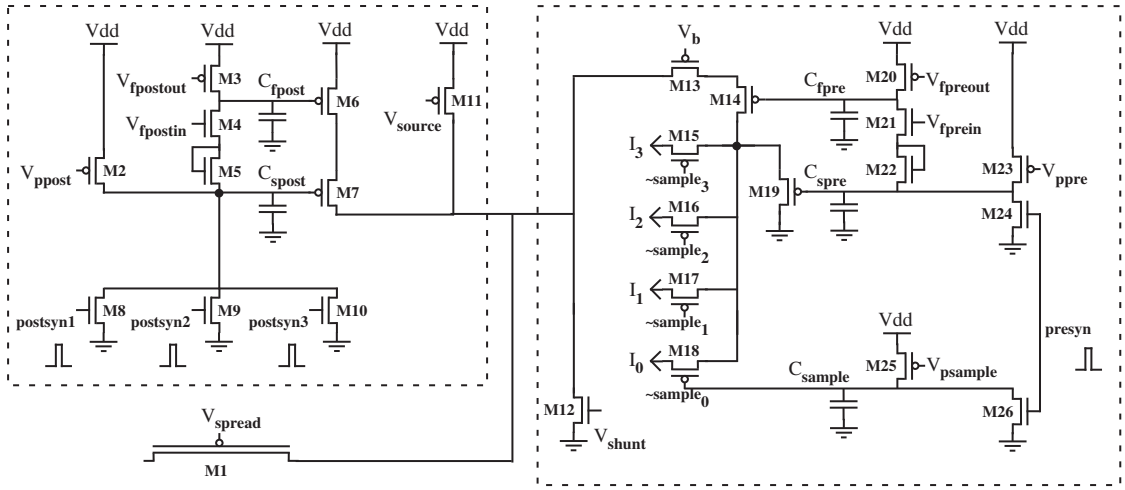


Figure 3.6: Neurotropin circuit

Neurotropin sample, diffuse, and uptake circuits. Target cell spikes inject charge (left box) into a monolithic pFET channel (M1), where it diffuses until being passively decayed (M12) or actively sampled by local growth cone circuits (right box).

if spikes arrive in a burst rather than singly. V_{fpostin} determines the contribution of a single spike to the facilitation capacitor C_{fpost}.

Source cell activity triggers neurotropin uptake, as implemented by the circuit in the right box of Figure 3.6. Charge is removed from the hexagonal lattice by a facilitation circuit similar to that used for target cell release. A source cell spike arriving at the growth cone pulls C_{spre} to ground through M24. C_{spre}, in turn, drains charge from C_{fpre} through M21 and M22. C_{fpre} removes charge from the hexagonal lattice through M14, up to a limit set by M13, which prevents the hexagonal lattice from being completely drained in order to avoid charge trapping. Current from M14 is divided between five possible sinks. Depending on source cell activation, up to four growth cones may sample a fraction of this current through M15-18; the remainder is shunted to ground through M19 in order to prevent a single source cell event from exerting undue influence on gradient measurements. The current sampled by the growth cone at its own site is gated by $\tilde{}sample_0$, which is

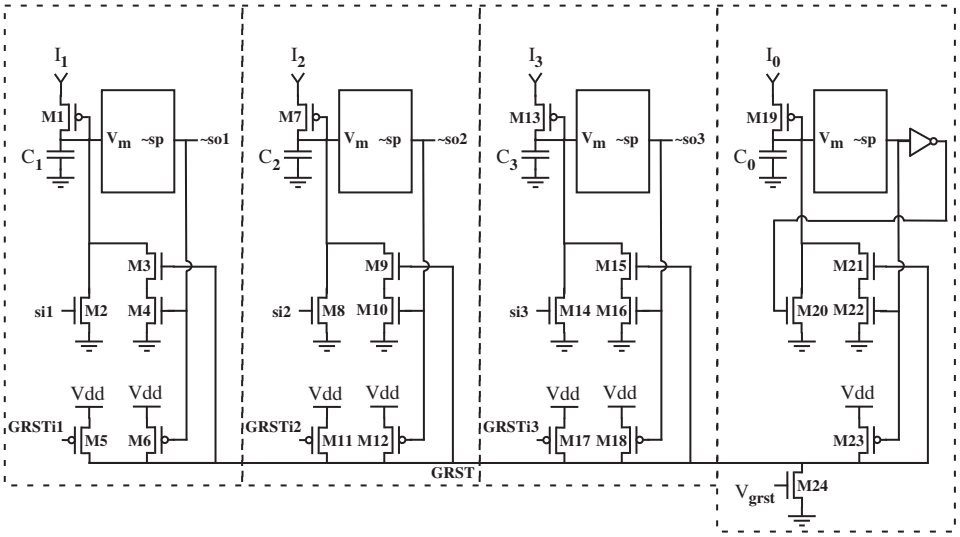


Figure 3.7: Gradient detection circuit

Each growth cone integrates neurotropin samples from its four adjacent nodes, including its own (right dotted box). The first integrator to cross threshold resets its competitors and transmits an address-event off-chip.

pulled low by a source cell spike through M26 and subsequently recovers through M25. Identical circuits in the other growth cones generate signals $\tilde{}sample_1$, $\tilde{}sample_2$, and $\tilde{}sample_3$. Sample currents I_0, I_1, I_2, and I_3 are routed to latency competition circuits in the four adjacent growth cones.

3.3.2 Gradient detection circuit

When a growth cone receives a spike, its filopodia sample charge from their nodes of the extracellular medium. These samples are integrated on separate capacitors, which race to cross a voltage threshold. The winner of this race will be the filopodium whose node tends to have the highest neurotropin concentration during the times the growth cone is presynaptically active. It resets the four filopodial integrators, including itself, and transmits its address-event off-chip.

A diagram of the gradient detection circuit is shown in Figure 3.7. Three of the four integrator circuits represent filopodia extended to the three nearest-neighbor nodes, Locations 1-3, and the fourth represents the growth cone body at the target node, Location 0. The growth cone moves to the node of the winning filopodium. If the growth cone body wins the race, the growth cone stays put.

In the circuit that samples neurotropin from Location 1 (left box of Figure 3.7), charge pulses I_1 arrive through diode M1 and accumulate on capacitor C_1 in an integrate-and-fire circuit described in [55]. Upon crossing threshold this circuit transmits a swap request $\tilde{}so1$, resets its competitors by using M6 to pull the shared reset line $GRST$ high, and disables M4 to prevent $GRST$ from using M3 to reset C_1. The swap request $\tilde{}so1$ remains low until acknowledged by $si1$, which discharges C_1 through M2. During the time that $\tilde{}so1$ is low, the other three capacitors are shunted to ground by $GRST$, preventing late arrivals from corrupting the declared gradient measurement before it has been transmitted off-chip. C_1 being reset restores $\tilde{}so1$ high, which releases $GRST$ to relax to ground through M24 with a decay time determined by V_{grst}.

C_1 is also reset if the neighboring axon terminal initiates a swap. $GRSTi1$ is pulled low if either the axon terminal at Location 1 decides to move to Location 0 or the axon terminal at Location 0 decides to move to Location 1. The accumulated neurotropin samples at both locations become obsolete after the exchange, and are therefore discarded when $GRST$ is pulled high through M5. Identical circuits integrate neurotropin from Locations 2 and 3 (center two boxes in Figure 3.7).

If C_0 (right box of Figure 3.7) wins the latency competition, the growth cone decides that its current location is optimal and therefore no action is required. In this case no off-chip communication occurs and C_0 immediately resets itself and its three rivals. Thus, the

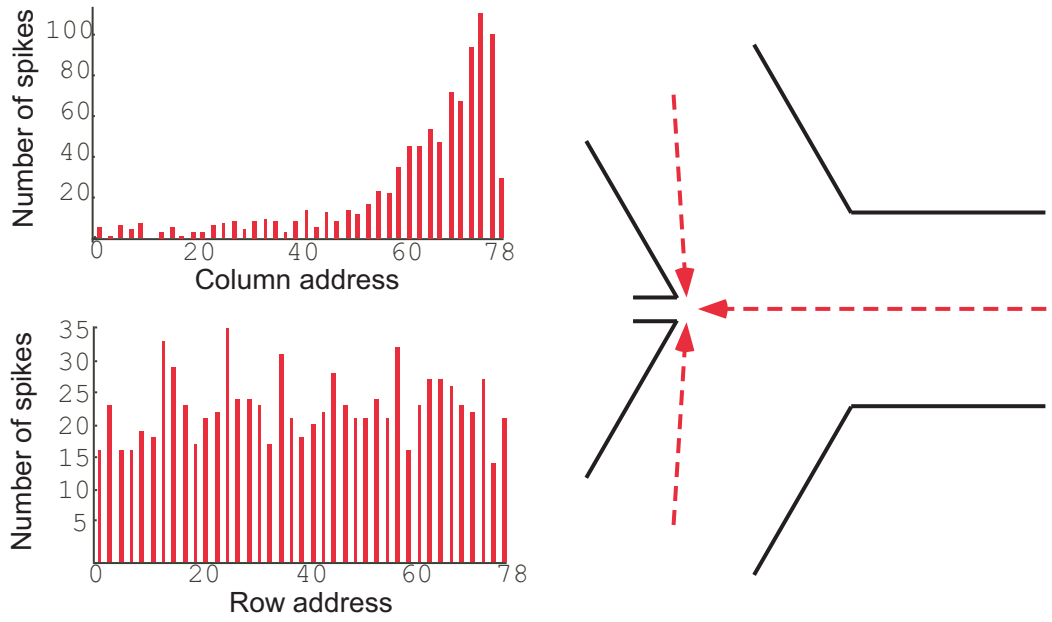


Figure 3.8: Chip anisotropies

(a) Target cell excitability increases with column address, but remains constant across rows. (b) Sampled charge travels a shorter distance from nodes adjacent to the tap than from the node opposite.

Location 0 circuit is identical to those of Locations 1-3 except that the inverted spike is fed directly back to the reset transistor M20 instead of to a communication circuit. Also, there is no $GRSTi0$ transistor since there is no swap partner.

3.3.3 Layout anisotropies

The chip exhibits two major anisotropies, one associated with spike generation and one associated with neurotropin diffusion. Each anisotropy significantly distorts growth cone motion and is due to poor layout.

Figure 3.8(a) displays a histogram of address-events binned by column and row address. When driven with the same excitation, circuits in the same row generate address-events at a rate that increases with column address. Within a column, circuits tend to fire at the

same rate. This horizontal excitability gradient is due to poor power distribution within the chip. The supply voltage V_{dd} is routed along a single metal bus that enters the array at column 0. Moving horizontally across the array, the accumulated load increases as each column draws current from the power bus, causing the supply voltage available to higher column addresses to sag. The neuron and growth cone circuits rely on a positive feedback circuit whose source is connected to V_{dd}, so lowering the effective V_{dd} reduces the threshold voltage required to trigger spike generation. The resulting position-dependent increase in activity establishes a permanent neurotropin gradient that tends to pull all growth cones to one side of the chip.

The workaround for this problem is to simply ignore the worst offenders by never stimulating circuits in the highest 16 column addresses. All future experiments will be performed on a reduced array of 24×20 growth cones.

The second major anisotropy arises from geometric properties of the hexagonal lattice (Figure 3.8(b)). The filopodial sampling circuit taps current from the hexagonal lattice using a transistor connected to one corner of the node. Charge sampled from the two nodes adjacent to the tap therefore travels a shorter path within the lattice than does charge sampled from the opposing node. The effect is to stretch the perceived node separation along the horizontal axis of the array so that growth cones are more likely to move vertically than horizontally.

The workaround for this problem is to program the microcontroller to accumulate successive swap requests from each growth cone and only update the lookup table after a threshold number of consecutive swap requests in the same direction. This threshold is set to be proportional to the relative frequency of each direction request observed over the entire array.

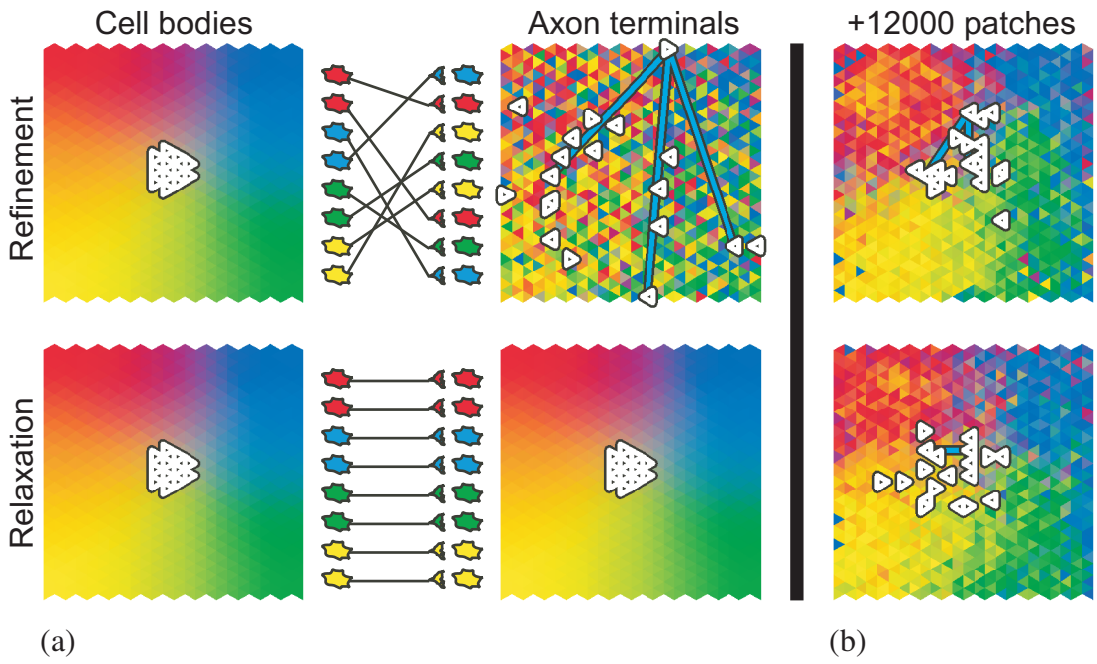


Figure 3.9: Topographic self-organization

(a) Contiguous patches of the input layer are activated (left), projecting a pattern of excitation to growth cone locations in the target layer (center). Lines connect growth cones corresponding to the four input cells at the center of the patch. Input cells and growth cones are colored according to the input cell location. (b) After presentation of many randomly centered input patches, the axon projection achieves a similar level of topography when initialized with a coarse projection (top) or a perfect projection (bottom).

3.4 Correlated stimulus

Since growth cone motion is directed by local target activity and since target cell activity is driven by source cell spikes, correlations between source cell spikes will determine the final connectivity. Different patterns of source activity will drive the system to self-organize different feed-forward axon projections. Ideally we would generate this correlated activity by presenting real world video to a silicon retina, using the retinal ganglion cells as the source layer. Unfortunately, the current version of the silicon retina is plagued by the same excitability gradient observed in Neurotrope1, due to similarly flawed power distribution.

We therefore simulate the source layer using a pattern generator board.

We programmed the pattern generator board to drive the chip with a sequence of randomly centered patches of source cell activity meant to simulate retinal waves (Figure 3.9(b)). Each patch consisted of 19 adjacent source cells: a randomly selected source cell and its nearest, next-nearest, and third-nearest neighbors on a hexagonal grid. Every patch participant generated a burst of 8192 spikes during each patch presentation, which were routed to the appropriate axon terminal circuit according to the connectivity map stored in the lookup table. About 100 patches were presented per minute.

We generated a coarsely topographic initial projection by starting with a perfectly topographic projection and executing a sequence of $(N/2)^2$ swaps between a randomly chosen axon terminal and one of its randomly chosen target layer neighbors, where N is the number of axon terminals used. We opted for a fanout of 1 and full synaptic occupancy, so 480 source cells projected axons to 480 synaptic sites. This perturbed connectivity map preserved a loose global bias, representing the formation of a coarse topographic projection from activity-independent cues. This new initial map was then allowed to evolve according to the swap requests generated by the chip. After approximately 12000 patches, a refined topographic projection emerged (Figure 3.9(b)).

To quantify this refinement, we defined the *topographic error* for a single source cell to be the average of the target layer distances between the axon terminals projected by the source cell and the three immediately adjacent source cells. A source cell in a perfectly topographic projection would therefore have unit error. Topographic error drops quickly at the beginning of the evolution as local clumps of correlated axon terminals crystallize. Further refinement requires the disassembly of locally topographic crystals that happened to nucleate in a globally inconvenient location. During this later phase, the error decreases

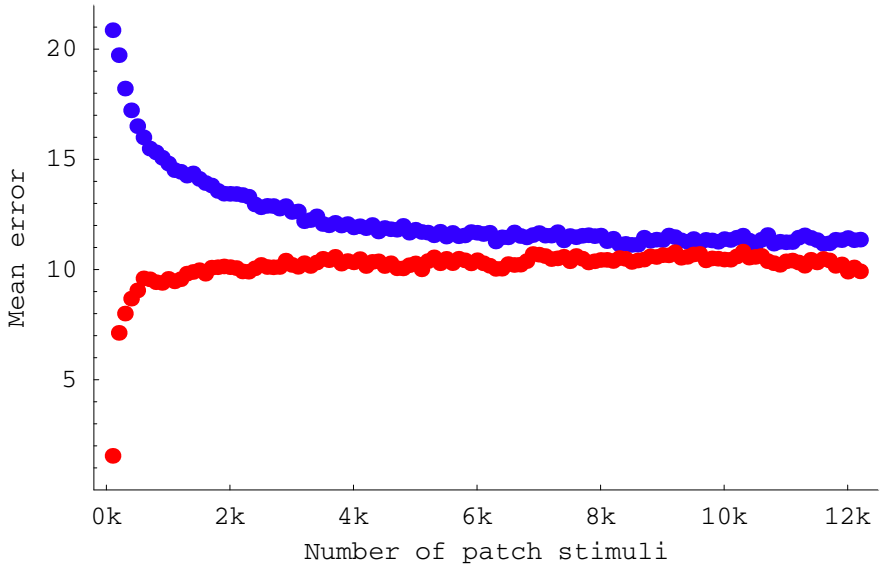


Figure 3.10: Topographic error evolution

The average topographic error converges to the same level from coarse initial topography (blue) and from perfect initial topography (red).

slowly toward an asymptote. To evaluate this limit we seeded the system with a perfect projection and let it relax to a sustainable decree of topography, which we found to have an error of about 10 units (Figure 3.10).

To characterize the system evolution more precisely, we will introduce some techniques of stochastic analysis in the next chapter.

Chapter 4

Stochastic Analysis

In this chapter we develop a model for growth cone motion in the Neurotrope1 system using a finite state Markov chain whose stochastic transition matrix is derived from neurotropic attraction. We will then illustrate the matrix construction procedure using a simple supervised attraction example. We will conclude by testing the model against actual data.

4.1 Definitions

A *random variable* X takes on different values x in different systems seeded with different initial conditions. The collection of all possible instantiations of the system is called the *ensemble*, and determines the statistical properties of the random variable. By counting the number of instances of a particular value x in the ensemble and dividing by the total number of instances in the ensemble, we can construct a relative frequency for x within the ensemble and call it the *probability* $P(x)$.

In a *stochastic process*, the random variable can assume different values x_i at different times t_i. Each instantiation of this sequence of values and their associated times is called the *sample path* traced by the process through the state space. We construct an ensemble of sample paths in order to calculate the probability $P(x_1, t_1; ...; x_n, t_n)$ that $X(t)$ takes on value x_1 at time t_1, x_2 at time t_2, and so on. For a given value x_1, we can separate out all the instances in the ensemble in which $X(t_1) = x_1$ and compute relative frequencies within that subensemble to generate the *conditional probability* $P(x_2, t_2; ...; x_n, t_n|x_1, t_1)$ for future values of $X(t)$ given a known initial value.

We typically sample the state of the process at regular time intervals τ, collecting a discrete sequence of random variables $X_0 = X(t), X_1 = X(t + \tau), X_2 = X(t + 2\tau)....$ The random variable X_n is called the *state* of the process at time step n. An *integer time process* is a stochastic process whose state changes only at integer multiples of the time step. When the range of the random variable is also discrete, each sample path is a sequence of transitions between discrete states X_n at discrete sample times $n\tau$.

A *stationary* function is independent of time. If the conditional probability $P(X_n = i|X_{n-1} = j)$ is stationary, we call it the *transition probability* $W_{ij}^{(\tau)}$, where $\tau = t_n - t_{n-1}$ is the fixed time interval separating consecutive observations. Our strategy will be to construct transition probabilities from a model and compare them with transition probabilities constructed from actual measurements.

4.1.1 Markov chains

A *Markov chain* is an integer time process in which the value of the current state X_n depends on previous states $X_{n-1}, X_{n-2}, ..., X_0$ only through the most recent state X_{n-1};

that is,

$$P(X_n = i|X_{n-1} = j, X_{n-2} = k, ...X_0 = m) = P(X_n = i|X_{n-1} = j) \tag{4.1}$$

A Markov chain can be expressed pictorially as a directed graph in which an arrow pointing from state j to state i represents a nonzero transition probability $W_{ij}^{(n)} = P(X_n = i|X_0 = j)$ that the process is observed in state i at time n given that it was previously observed in state j at time 0 (Figure 4.1(a)). If the graph contains a directed path of any length n leading from state j to state i ($W_{ij}^{(n)} > 0$), we say that state i is *accessible* from state j ($j \rightarrow i$).

A state i is *recurrent* if it is accessible from all the states it accesses; that is, if $i \rightarrow j$ implies $j \rightarrow i$. A process leaving a recurrent state will always return to that state given enough time. On the other hand, a process leaving a nonrecurrent state may exit to a state from which it cannot return. Since this permanent departure will occur eventually with probability 1, nonrecurrent states are *transient*.

States i and j *communicate* if i is accessible from j and vice versa ($i \leftrightarrow j$). Communication is transitive, so if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$. A *class* is a non-empty set of states in which every state communicates with every other state in the class. The set of all states can be partitioned into disjoint classes of communicating states. A class of states is recurrent if no member can access a state outside the class, and transient otherwise. A Markov chain with a finite number of states always has at least one recurrent class and can have any number of additional recurrent or transient classes.

The *period* of a recurrent class is defined as the greatest common divisor (gcd) of all n for which $P_{ii}^n > 0$. When the gcd is 1, the class is considered *aperiodic*. A periodic process

indefinitely perpetuates the influence of its initial state, which determines the phase of oscillation between states. A recurrent and aperiodic class is called *ergodic*, and eventually erases the memory of its initial state. In this thesis, we will use ergodic finite-state Markov chains to model growth cone population behavior.

4.1.2 Stochastic matrices

Under the Markov assumption, the transition probability that the process will move from state j to state i in 2 steps is

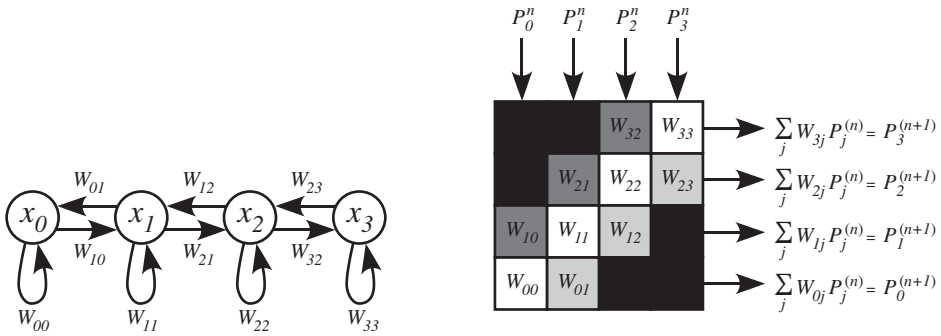
$$W_{ij}^{(2)} = \sum_k W_{ik} W_{kj} \tag{4.2}$$

We can express this relation more compactly by defining a *stochastic matrix* $\mathbf{W}$ whose elements W_{ij} are the transition probabilities. $\mathbf{W}$ is a square matrix with non-negative elements in which each column j corresponds to an initial state x_j and each row i corresponds to a final state x_i (Figure 4.1(b)). Since transition probabilities are conditioned on the initial state, the elements in each column sum to unity ($\sum_i W_{ij} = 1$). Equation 4.2 can now be written as a simple matrix multiplication, $\mathbf{W}^{(2)} = \mathbf{W}^2 = \mathbf{W} \cdot \mathbf{W}$. This is a special case of the *Chapman-Kolmogorov equation*,

$$W_{ij}^{(n+m)} \quad = \quad \sum_k W_{ik}^{(n)} W_{kj}^{(m)} \tag{4.3}$$

which we can use to iteratively construct transition probabilities for paths of arbitrary length through successive matrix multiplications.

The probability $P_i^{(n+m)}$ that the process will enter state i at time $(n+m)$ is the sum of all possible m-step transitions $W_{ij}^{(m)}$ into state i from every initial state j, each weighted by



(a) (b)

Figure 4.1: Markov chain

(a) Directed graph representation. Each arrow represents a possible transition from state x_j to state x_i, labeled with the transition probability W_{ij}. Arrows representing zero probability transitions are not drawn. (b) Stochastic matrix representation. Columns correspond to initial states and rows index final states. Matrix elements are tinted according to probability amplitude; darker elements have lower probability. Zero probability transitions are colored black.

the probability $P_j^{(n)}$ that the process actually occupied state j at time n:

$$P_i^{(n+m)} = \sum_j W_{ij}^{(m)} P_j^{(n)} \tag{4.4}$$

or in matrix notation,

$$\mathbf{P}^{(n+m)} = \mathbf{W}^m \mathbf{P}^{(n)} \tag{4.5}$$

where we have defined the state distribution vector $\mathbf{P}^{(n)}$ to be the column vector whose ith element is the probability $P_i^{(n)}$ that the process occupies state i at time n. Equation 4.5 predicts any future state distribution $\mathbf{P}^{(n)} = \mathbf{W}^n \mathbf{P}^{(0)}$ given only the transition matrix $\mathbf{W}$ and the initial distribution $\mathbf{P}^{(0)}$.

For a memoryless process like an ergodic Markov chain, we expect the system to forget its initial state given sufficient time, so that repeated iteration of $\mathbf{W}$ on an initial state dis-

tribution $\mathbf{P}^{(0)}$ will eventually converge to a final state distribution $\mathbf{P}^{(\infty)}$ that is independent of time.

$$\mathbf{P}^{(\infty)} = \lim_{n\to\infty} \mathbf{P}^{(n)} = \lim_{n\to\infty} \mathbf{W}^n\mathbf{P}^{(0)} = \lim_{n\to\infty} \mathbf{W}\mathbf{P}^{(n-1)} = \mathbf{W}\mathbf{P}^{(\infty)} \tag{4.6}$$

We say $\mathbf{P}^{(\infty)}$ is the *stationary distribution* with respect to $\mathbf{W}$, since it is unaffected by subsequent multiplication by $\mathbf{W}$. When multiplication of a matrix $\mathbf{W}$ by a column vector $\mathbf{P}$ yields a scalar multiple of $\mathbf{P}$ — that is, when $\mathbf{W}\mathbf{P} = \lambda\mathbf{P}$ — we say that $\mathbf{P}$ is an *eigenvector* of $\mathbf{W}$ with *eigenvalue* λ. The stationary probability vector $\mathbf{P}^{(\infty)}$ is the eigenvector of $\mathbf{W}$ with eigenvalue 1. Eigenvectors and their eigenvalues are intrinsic to each matrix, and can be computed using standard matrix techniques.

The eigenvectors of a matrix represent an alternate set of generally nonorthogonal coordinate axes that are specifically aligned to match the matrix operation. We can rewrite an input probability vector $\mathbf{P}$ in the new coordinates as the sum of its eigenvector components:

$$\mathbf{P} = \sum_{\lambda} a_\lambda \mathbf{\Lambda}_\lambda \tag{4.7}$$

where $\mathbf{\Lambda}_\lambda$ is the eigenvector of $\mathbf{W}$ with eigenvalue λ. Every matrix multiplication scales individual eigenvector components by their eigenvalues, so eigenvectors corresponding to $|\lambda| < 1$ die off exponentially with n.

$$\mathbf{W}^n\mathbf{P} = \sum_{\lambda} \lambda^n a_\lambda \mathbf{\Lambda}_\lambda \tag{4.8}$$

According to the Perron-Frobenius theorem, all of the eigenvalues of a stochastic matrix have magnitudes less than or equal to 1, so repeated multiplication by $\mathbf{W}$ carves away their eigenvector components exponentially with n until only the stationary probability vector

$\mathbf{P}^{(\infty)} = \mathbf{\Lambda}_1$ remains.

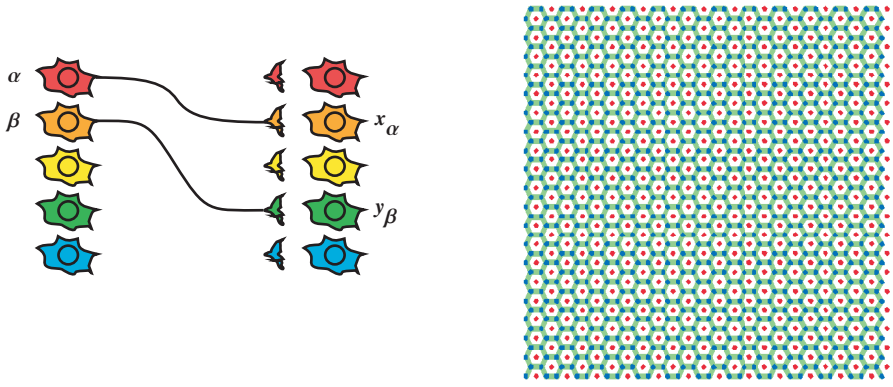
4.1.3 Continuous-time Markov chains

State transitions of integer-time processes are defined to occur at deterministic intervals, an assumption chosen mainly to facilitate sampling. This assumption becomes less plausible when dealing with heterogenous populations of noisy components. For example, different growth cones move at different speeds in identical gradients, due to device mismatch, so the integer-time transition probabilities for a fast growth cone are higher than for a slow growth cone in the same integer time interval. Since there is no way to predict which growth cones are fast and which are slow, the intervals between transitions are themselves random variables.

In a *continuous-time Markov chain*, the intervals between state transitions are distributed exponentially with state-dependent transition rates λ_i. That is, the probability that the process transitions from state i to state j over time τ is

$$P(X_{t+\tau} = j | X_t = i) = (1 - \exp[-\lambda_i \tau]) P_{ij} \tag{4.9}$$

where $(1 - \exp[-\lambda_i \tau])$ is the probability that a transition occurs within time τ, and P_{ij} is the probability that the transition out of i was directed to j. We can recover an integer-time Markov chain representation by sampling the underlying continuous-time Markov chain at regular time intervals τ. We will derive the transition rates λ_i from a model of neurotropic attraction based on a set of assumptions described in the next section.



(a) (b)

Figure 4.2: Neurotrope1 array

(a) Source cells are indexed with lower case Greek letters and target cells are indexed with lower case Roman letters. A target cell index may be subscripted with the source cell index of the innervating axon terminal. (b) Axon terminal and target cell arrays interleave in the Neurotrope1 chip. Axon terminals occupy nodes on a honeycomb lattice (blue). Target cells occupy nodes on a triangular lattice (red). Neurotropin diffuses along the edges of the honeycomb lattice (green).

4.2 Model transition matrix

In this section we present a Markov chain model of neurotropic growth cone attraction in the Neurotrope1 system. We first state some assumptions that allow us to derive an expression for the neurotropin sampled by an active growth cone, and then outline a procedure to construct a transition matrix that models growth cone motion under the influence of this sampled neurotropin.

We index locations in each layer with their two-dimensional coordinates, using lower case Greek letters $\alpha, \beta \in \mathbb{Z}^2$ to denote locations in the source layer and lower case Roman letters $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^2$ to denote locations in the target layer (Figure 4.2(a)). Axon terminals in the Neurotrope1 system occupy discrete target layer sites that are laid out in a honeycomb lattice, while target cells are arranged in a triangular lattice interleaved between the axon

terminal nodes (Figure 4.2(b)). For simplicity we will approximate the target layer as being a single honeycomb lattice in which each node hosts one axon terminal and one target cell. A subscripted target index $\mathbf{x}_\alpha$ denotes the target layer site occupied by the axon terminal projected by source cell α.

4.2.1 Sampled neurotropin

We use four assumptions to derive a general expression for the neurotropin sampled by an active growth cone.

1. **Target cell activity is proportional to source cell activity**

   Each source cell α fires at rate $a_\alpha(t)$ and projects an axon into the target layer that elaborates an arbor of $A(\mathbf{x} - \mathbf{x}_\alpha)$ excitatory synapses onto a neighborhood of target cells centered on $\mathbf{x}_\alpha$. Each target cell $\mathbf{x}$ fires at a rate $a_\mathbf{x}(t)$ that is linearly proportional to the sum of its excitation:

   $$a_\mathbf{x}(t) \propto \sum_\alpha A(\mathbf{x} - \mathbf{x}_\alpha) a_\alpha(t) \tag{4.10}$$

   The arbor function $A(\mathbf{x} - \mathbf{x}_\alpha)$ is a nonincreasing function of distance $\|\mathbf{x} - \mathbf{x}_\alpha\|$ from the arbor center.

2. **Neurotropin release is proportional to target cell activity**

   Postsynaptic spikes trigger active target cell bodies to release neurotropin, whose spatiotemporal spread establishes the instantaneous extracellular neurotropin concentration profile $N(\mathbf{x}, t)$. We assume that neurotropic contributions from separate target cells sum linearly, so that the extracellular neurotropin concentration present

at target site x at time t is proportional to a weighted sum of local target cell activity:

$$N(\mathbf{x}, t) \propto \sum_{\mathbf{y}} G(\mathbf{x} - \mathbf{y}) a_{\mathbf{y}}(t) \tag{4.11}$$

The neurotropin spreading kernel $G(\mathbf{x} - \mathbf{y})$ describes the fraction of neurotropin released from target site $\mathbf{y}$ that spreads to target site $\mathbf{x}$, and is a nonincreasing function of distance $\|\mathbf{x} - \mathbf{y}\|$.

3. **Neurotropin uptake is proportional to source cell activity**

   A growth cone's neurotropin uptake is gated by its presynaptic activity. The neurotropin sampled by growth cone α at time t is

$$N_\alpha(t) \propto a_\alpha(t) N(\mathbf{x}_\alpha, t) \tag{4.12}$$

4. **Presynaptic rate correlations are independent of time**

   To remove the explicit time dependence from $N_\alpha(t)$, we assume that presynaptic rate correlations remain unchanging over time. Source cells fire spikes in fixed duration bursts whose time-averaged overlap defines their mutual coactivity $C_{\alpha\beta} = \langle a_\alpha(t) a_\beta(t) \rangle_t$, which we take to be fixed for each pair of source cells (α, β). Combining equations, the neurotropin sampled by growth cone α from its location $\mathbf{x}_\alpha$ in the target layer is

$$N_\alpha(\forall_\beta \mathbf{x}_\beta) \propto \sum_\beta C_{\alpha\beta} \sum_{\mathbf{y}} G(\mathbf{x}_\alpha - \mathbf{y}) A(\mathbf{y} - \mathbf{x}_\beta) \tag{4.13}$$

   where $\forall_\beta \mathbf{x}_\beta$ is the set of all axon terminal locations $\mathbf{x}_\beta$.

The behavior of this expression is relatively insensitive to the exact shapes of $A(\mathbf{x} -$

$\mathbf{y})$ and $G(\mathbf{x}-\mathbf{y})$ so long as both are monotonically nonincreasing functions of distance $\|\mathbf{x}-\mathbf{y}\|$. Both functions distribute the influence of the same target cell activity over a wider spatial neighborhood, and can be merged into a single spreading kernel without loss of generality. For convenience, we assume that each axon terminates onto a single target site, so that $A(\mathbf{x}-\mathbf{x}_\alpha)=\delta(\mathbf{x}-\mathbf{x}_\alpha)$. We typically take the neurotropin spreading kernel to be a peak-normalized Gaussian $G(\mathbf{x}-\mathbf{y})=\exp\left[-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma_{\mathrm{NT}}^2}\right]$, a choice inspired by the transistor circuit analysis outlined in Appendix B[1].

In this thesis, we will always sample growth cone positions at regular intervals τ, collecting the nth sample at time $t=n\tau$. We say that $\mathbf{x}_\alpha^{(n)}$ is the position of growth cone α observed at the nth sample, during which

$$N_\alpha(\forall_\beta \mathbf{x}_\beta^{(n)}) \propto \sum_\beta C_{\alpha\beta} G(\mathbf{x}_\alpha^{(n)}-\mathbf{x}_\beta^{(n)}) \tag{4.14}$$

4.2.2 Transition matrix construction

We are interested in the transition probability $W_{\mathbf{y}\mathbf{x}_\alpha}$ that a growth cone initially located at target site $\mathbf{x}_\alpha$ will have jumped to an adjacent site $\mathbf{y}$ after one time step. To derive $W_{\mathbf{y}\mathbf{x}_\alpha}$, we will make the following assumptions:

1. **Growth cone jumps arrive according to a Markov process whose rate is linear in the sampled neurotropin concentration $N_\alpha(\forall_\beta \mathbf{x}_\beta)$.**

   Growth cones move by rerouting their address-events to different target site circuits. Due to device mismatch, every filopodial neurotropin integrator in every target site circuit counts up to a different threshold of sampled neurotropin before triggering a

[1] Note that it is the height of the Gaussian peak that is normalized, not the area under the peak.

growth cone jump. As a consequence, the time it takes for a growth cone to exit a target site varies unpredictably from site to site. In identical gradients, growth cones speed through target site circuits with low jump thresholds and stall at circuits with high jump thresholds. To model this variance in jump timing, we assume that growth cone jumps arrive according to an exponential distribution with a position-dependent rate.

This jump rate should increase with the sampled neurotropin concentration $N_\alpha(\forall_\beta \mathbf{x}_\beta)$, since filopodia are able to bind larger neurotropin samples and therefore integrate to threshold faster by using fewer, larger steps. We assume this increase is linear, so that the probability that an active growth cone α requests a jump from target site $\mathbf{x}_\alpha$ within a time step τ is

$$P_{\text{jump}}(\forall_\beta \mathbf{x}_\beta) = (1 - \exp\left[-\lambda_{\text{NT}} N_\alpha(\forall_\beta \mathbf{x}_\beta)\tau\right]) \tag{4.15}$$

λ_{NT} is a free parameter relating neurotropin concentration to time.

2. **Growth cone jumps are explicitly directed with a probability that depends on the magnitude of the local neurotropin gradient $\|\nabla_{\mathbf{x}_\alpha} N_\alpha(\forall_\beta \mathbf{x}_\beta)\|$.**

   Since a growth cone requests an update every time one of its neurotropin accumulators overflows, extracellular neurotropin alone is sufficient to trigger an unending stream of jump requests, even without a detectable gradient. The presence of a gradient biases these automatically generated jumps toward a preferred direction. The stronger the gradient, the stronger the bias.

   For simplicity, we assume that direction biasing is all or nothing. Directed jumps act as signal, and point unerringly up the gradient. Undirected jumps act as noise, and their orientations are uniformly distributed over all available directions. Direction

biasing is a race between two independent processes; if the gradient picks a direction before the concentration triggers the actual jump, the jump is directed; otherwise it is undirected. We treat this switch from noise to signal as a random event that occurs with a probability that depends on the gradient magnitude. The probability that a given jump is directed is

$$P_{\text{dir}}(\forall_\beta \mathbf{x}_\beta) = 1 - \exp\left[-\lambda_{\text{dir}} \|\nabla_{\mathbf{x}_\alpha} N_\alpha(\forall_\beta \mathbf{x}_\beta)\|\right] \tag{4.16}$$

The proportionality constant λ_{dir} is a free parameter.

3. **A growth cone can jump at most once per time step.**

   We pick a time step τ that is short enough to allow a growth cone to jump at most once per time step from its initial site $\mathbf{x}_\alpha$, always to some adjacent target site $\mathbf{y}$, so $W_{\mathbf{y}\mathbf{x}_\alpha} = 0$ if $\|\mathbf{x}_\alpha - \mathbf{y}\| > 1$. The growth cone remains at $\mathbf{x}_\alpha$ if it fails to jump during the time step, so $W_{\mathbf{x}_\alpha \mathbf{x}_\alpha} = 1 - P_{\text{jump}}(\forall_\beta \mathbf{x}_\beta)$. The remaining probability $P_{\text{jump}}(\forall_\beta \mathbf{x}_\beta)$ is distributed among the adjacent sites according to jump type. Jumps are directed with probability $P_{\text{dir}}(\forall_\beta \mathbf{x}_\beta)$, and flow to the adjacent site $\mathbf{y}^*$ that is most aligned with the sampled gradient direction, where

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} [(\mathbf{y} - \mathbf{x}_\alpha) \cdot \nabla_{\mathbf{x}_\alpha} N_\alpha(\forall_\beta \mathbf{x}_\beta)] \tag{4.17}$$

   Jumps are undirected with probability $1 - P_{\text{dir}}(\forall_\beta \mathbf{x}_\beta)$, and distribute evenly among

all adjacent sites. Summarizing,

$$W_{\mathbf{y}\mathbf{x}_\alpha} = \begin{cases} 1 - P_{\text{jump}}(\forall_\beta \mathbf{x}_\beta), & \mathbf{y} = \mathbf{x}_\alpha \\ P_{\text{jump}}(\forall_\beta \mathbf{x}_\beta) \left\{ P_{\text{diff}}(\mathbf{y}, \mathbf{x}_\alpha)(1 - P_{\text{dir}}(\forall_\beta \mathbf{x}_\beta)) \right. & \\ \qquad \left. + P_{\text{dir}}(\forall_\beta \mathbf{x}_\beta) \right\}, & \mathbf{y} = \mathbf{y}^* \\ P_{\text{jump}}(\forall_\beta \mathbf{x}_\beta) \left\{ P_{\text{diff}}(\mathbf{y}, \mathbf{x}_\alpha)(1 - P_{\text{dir}}(\forall_\beta \mathbf{x}_\beta)) \right\}, & \|\mathbf{y} - \mathbf{x}_\alpha\| = 1, \\ & \mathbf{y} \neq \mathbf{y}^* \\ 0, & \|\mathbf{y} - \mathbf{x}_\alpha\| > 1 \end{cases}$$

where $P_{\text{diff}}(\mathbf{y}, \mathbf{x}_\alpha)$ is the probability that an undirected jump originating at $\mathbf{x}_\alpha$ terminates at $\mathbf{y}$. For now, $\mathbf{x}_\alpha$ and $\mathbf{y}$ refer to target site coordinates, so $P_{\text{diff}}(\mathbf{y}, \mathbf{x}_\alpha) = 1/Y(\mathbf{x}_\alpha)$, where $Y(\mathbf{x}_\alpha)$ is the number of target sites $\mathbf{y}$ adjacent to the current site $\mathbf{x}_\alpha$. For a honeycomb lattice, $Y(\mathbf{x}_\alpha) = 3$ everywhere except the array boundaries. $P_{\text{diff}}(\mathbf{y}, \mathbf{x}_\alpha)$ is less trivial when the state vectors $\mathbf{x}_\alpha$ and $\mathbf{y}$ encode more complicated order parameters.

4.3 Population parameters

Typically we are less interested in tracing individual growth cone trajectories than in recruiting some aggregate population behavior to perform a useful computation whose progress can be characterized by a single population parameter. It would be easier to work with a transition matrix $\mathbf{W}$ that is assembled directly from this population parameter instead of cobbling the parameter together after the fact from a $\mathbf{W}$ assembled from individual growth cone positions $\mathbf{x}_\alpha$. In this section, we introduce two parameters to define and control population behavior: an *uptake parameter* that describes growth cone motivation, and an *order parameter* that describes the problem to be solved. We can program the growth

cone population to execute a particular computation by shaping the form of the uptake parameter to match the form of the order parameter for the desired operation.

4.3.1 Uptake parameter

The first goal of every growth cone α is to maximize its equilibrium neurotropin uptake $N_\alpha^{(\infty)}$. For a given set of presynaptic coactivities $C_{\alpha\beta}$ and target layer distances $r_{\alpha\beta}^{(n)} = \|\mathbf{x}_\alpha^{(n)} - \mathbf{x}_\beta^{(n)}\|$ separating growth cones α and β at sample n, the average amount of neurotropin growth cone α can expect to receive from all coactive growth cones β during sample n is

$$N_\alpha^{(n)} = \sum_\beta C_{\alpha\beta} G(r_{\alpha\beta}^{(n)}) \tag{4.18}$$

Since $G(r)$ is a monotonically nonincreasing function of target layer distance r, a perfectly guided growth cone should move to minimize the distances separating it from coactive neurotropin sources. We can formalize this statement by defining the *uptake parameter* U_α for a given growth cone α to be the average distance separating growth cone α from another growth cone β, weighted by the coactivity between the two growth cones:

$$U_\alpha^{(n)} = \frac{\sum_\beta C_{\alpha\beta} r_{\alpha\beta}^{(n)}}{\sum_\beta C_{\alpha\beta}} \tag{4.19}$$

We average the uptake parameters of the individual growth cones to obtain the population uptake parameter $U^{(n)}$ for the full axon projection:

$$U^{(n)} = \frac{\sum_{\alpha,\beta} C_{\alpha\beta} r_{\alpha\beta}^{(n)}}{\sum_{\alpha,\beta} C_{\alpha\beta}} \tag{4.20}$$

The density constraint restricts the number of neighbors allowed at a given separation r, forcing growth cones to prioritize distances for minimization according to coactivity. The shape of the final configuration is implicitly encoded in the set of input coactivities $C_{\alpha\beta}$, which we can manipulate directly to specify a desired solution.

4.3.2 Order parameter

In order to harness collective growth cone motion to solve a specific problem, we need to select coactivity patterns that align the goals pursued by the growth cone population with those implied by the task. To facilitate this choice of presynaptic stimuli, we devise an *order parameter* whose value decreases to a minimum as the system approaches the intended endstate of the desired computation. The form of the order parameter is obviously specific to each problem and generally not unique. In this thesis we focus on the problem of *retinotopy*, the degree to which a population of axon terminals reprise the topological relations between their projecting source cells. We can parameterize retinotopy in two different ways.

4.3.2.1 Relative retinotopy

In a retinotopic mapping from source layer to target layer, adjacent source cells project axon terminals to adjacent target sites. We can describe the retinotopy of an axon terminal projected by a single source cell α in relative terms by collecting the pairwise distances $r^{(n)}_{\alpha(\alpha+\mathbf{f}_i)}$ separating it from the axon terminals projected by adjacent source cells $\alpha + \mathbf{f}_i$ in the nth sample, where $\mathbf{f}_i$ is the unit vector oriented along the ith edge exiting node α in the lattice of source cells (Figure 4.3(a)). Perfect retinotopy is attained when each of these pairwise separations is minimized, subject to the density constraint. We average the instantaneous pairwise separations $r^{(n)}_{\alpha(\alpha+\mathbf{f}_i)}$ to obtain the retinotopic order parameter $\Phi^{(n)}_{\alpha}$

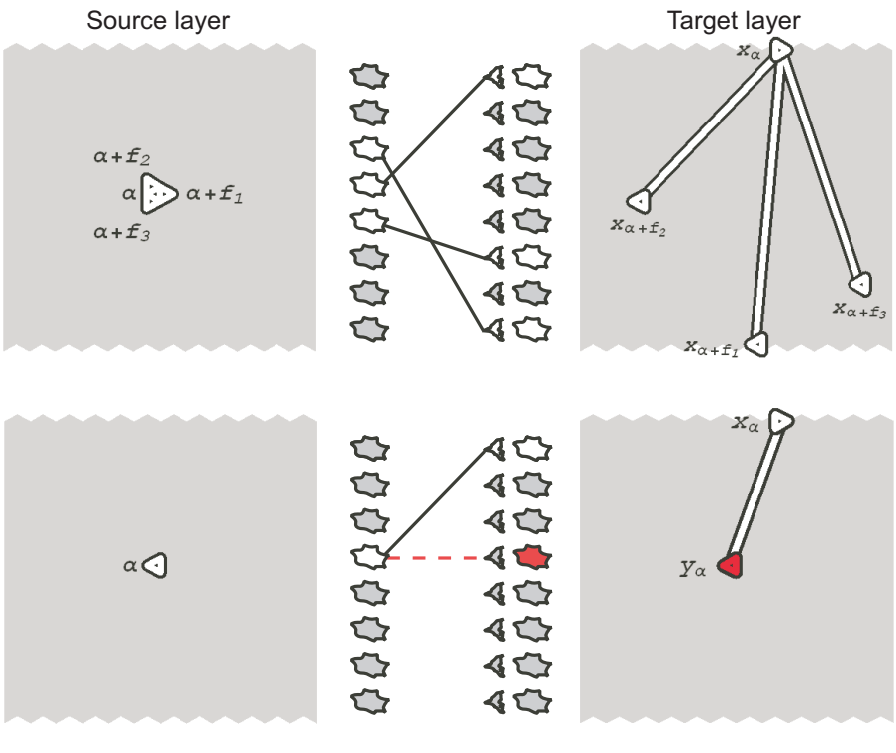


Figure 4.3: Retinotopic order parameters

Retinotopic order parameters Φ_α for the axon terminal projected by source cell α. Left: source layer; right: target layer. (a) Relative retinotopy. Φ_α is the average of the target layer distances $r_{\alpha(\alpha+\mathbf{f}_i)} = \|\mathbf{x}_\alpha - \mathbf{x}_{\alpha+\mathbf{f}_i}\|$ that separate the axon terminal projected by source cell α from the axon terminals projected by adjacent cell bodies $(\alpha + \mathbf{f}_i)$ in the source layer. (b) Absolute retinotopy. Φ_α is the distance $r_\alpha = \|\mathbf{x}_\alpha - \mathbf{y}_\alpha\|$ separating the growth cone (white) from a fixed reference site (red) in the target layer.

for axon terminal α during sample n:

$$\Phi_\alpha^{(n)} = \frac{1}{Z_\alpha} \sum_i r_{\alpha(\alpha+\mathbf{f}_i)}^{(n)} \tag{4.21}$$

where Z_α is a normalization constant. We can track the retinotopic refinement of the entire axon projection by averaging all of the pairwise separations into a single population order parameter $\Phi^{(n)}$:

$$\Phi^{(n)} = \frac{1}{Z} \sum_{\alpha,i} r_{\alpha(\alpha+\mathbf{f}_i)}^{(n)} \tag{4.22}$$

where $Z = \sum_\alpha Z_\alpha$ is a normalization constant.

The density constraint restricts the range of $r_{\alpha(\alpha+\mathbf{f}_i)}$ to nonzero values, so $\Phi^{(n)}$ is minimized when all of the $r_{\alpha(\alpha+\mathbf{f}_i)} = 1$, indicating that the axon terminals projected by adjacent cells in the source layer remain adjacent in the target layer. Thus, the value of the order parameter is minimized by the state of perfect retinotopy and increases monotonically with retinotopic deviation, exactly as desired.

We will return to this order parameter when we study unsupervised self-organization in Chapter 7.

4.3.2.2 Absolute retinotopy

Another way to characterize retinotopy is to select a perfectly retinotopic axon ordering[2] and compare each axon terminal's position $\mathbf{y}_\alpha$ within that desired endstate with its current position $\mathbf{x}_\alpha^{(n)}$. Each axon terminal is associated with a unique absolute position $\mathbf{y}_\alpha$ in the target layer, and its retinotopy is measured by its distance $r_\alpha^{(n)} = \|\mathbf{x}_\alpha^{(n)} - \mathbf{y}_\alpha\|$ off this mark (Figure 4.3(b)). The order parameter for axon terminal α at sample n is simply

$$\Phi_\alpha^{(n)} = r_\alpha^{(n)} \tag{4.23}$$

$\Phi_\alpha^{(n)}$ is defined only in terms of the axon terminal's own position relative to a fixed target site, without reference to the positions of the other axon terminals, so individual axon terminals are free to minimize their own order parameters independently.

[2]Map retinotopy is invariant under rotation, translation, and mirror reversal, but each of these symmetries can be broken by imposing sufficiently irregular layer boundaries or by fixing the desired position of one or more axon terminals within the target layer. A perfectly retinotopic solution is unique once all of these symmetries have been broken.

We will use this order parameter to characterize supervised attraction in the next chapter.

4.4 Empirical transition matrix

In this section, we introduce a procedure to measure an empirical transition matrix directly from data collected from the Neurotrope1 system. We construct this experimental $\mathbf{W}$ by reading connection maps from the lookup table at regular sampling intervals τ and observing the change in each axon terminal order parameter $\Phi_\alpha^{(n)}$ between consecutive samples. The procedure is as follows:

1. **Construct the marginal probability $P^{(n)}(r)$.**

   Pair order parameter values from consecutive maps. Count the number $\#(*, r)$ of observed pairs $\{\Phi_\alpha^{(n+1)}, \Phi_\alpha^{(n)}\}$ with initial order parameter $\Phi_\alpha^{(n)} = r$. This number divided by the total number $\#(*, *)$ of pairs is the probability that a randomly selected growth cone possessed order parameter r at time $t = n\tau$.

$$P^{(n)}(r) = \frac{\#(*, r)}{\#(*, *)} \tag{4.24}$$

2. **Construct the conditional probability $P^{(n)}(s|r)$.**

   Count the number $\#(s, r)$ of pairs with initial order parameter $\Phi_\alpha^{(n)} = r$ and final order parameter $\Phi_\alpha^{(n+1)} = s$. This number divided by $\#(*, r)$ is the conditional probability that a randomly selected growth cone jumped from order parameter r at time t to order parameter s at time $t + \tau$.

$$P^{(n)}(s|r) = \frac{\#(s, r)}{\#(*, r)} \tag{4.25}$$

3. **Construct the stationary transition probability W_{sr}.**

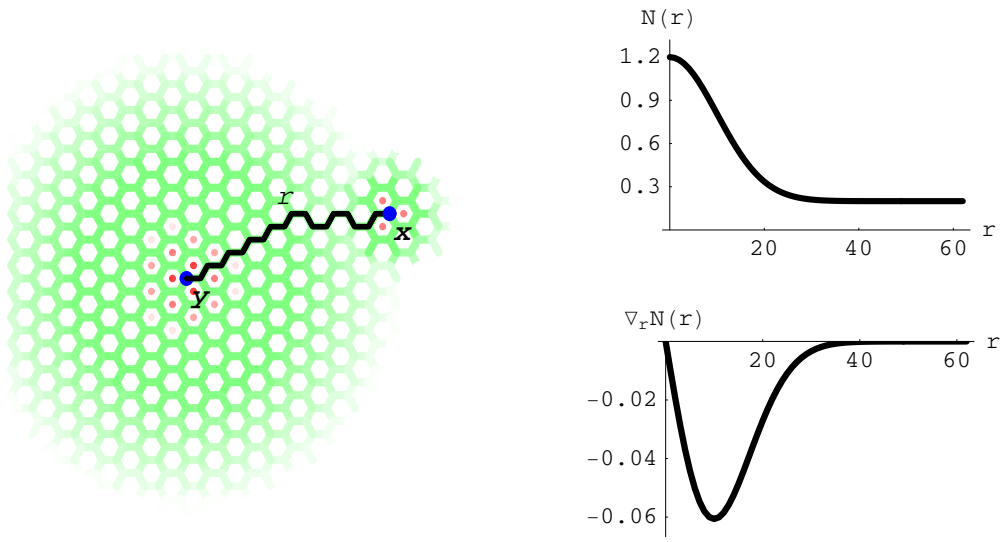
   If the conditional probability distribution remains the same for consecutive observations over the course of the experiment, then we can assume that the conditional probability is stationary and therefore the system is Markov. Under this assumption, we can treat the time series of system states as an ensemble. We average $P^{(n)}(s|r)$ over all n to find the stationary transition probability W_{sr}.

$$W_{sr} = \frac{1}{Z} \sum_{n=0}^{Z-1} P^{(n)}(s|r) \tag{4.26}$$

   Z is the total number of samples used to construct $\mathbf{W}$.

4.5 Example: Single growth cone-synapse pair

To demonstrate these transition matrix techniques, we will analyze the simplest non-trivial neurotropic attraction scenario, that of a single mobile growth cone coactive with a single anchored synapse (Figure 4.4). The synapse is permanently fixed at target site $\mathbf{y}$ and stimulates neurotropin release from the surrounding target cells, creating a concentration peak centered on $\mathbf{y}$. The growth cone also induces neurotropin release from target cells near its own location $\mathbf{x}$, but is free to shift its location over time. Without competing attractors, a perfect gradient-climbing growth cone should move to the synapse position and stay there, so the uptake parameter is simply the number of edges r currently separating the growth cone from the synapse. We would like to characterize how well the synapse can attract the growth cone, so the order parameter is also r. Since the synapse attracts the growth cone to a predetermined position, we say that the synapse instructs growth cone migration and that this is a simple case of *supervised learning*.



(a) (b)

Figure 4.4: Pair attraction

An anchored synapse attracts a mobile growth cone. (a) A growth cone located at $\mathbf{x}$ is separated by r edges from a synapse located at $\mathbf{y}$. Both axon terminals (blue) excite postsynaptic activity in neighboring target cells (red), releasing neurotropin to spread along the lattice edges (green). (b) Top: neurotropin concentration $N(r)$ sampled by the growth cone; bottom: neurotropin gradient $\nabla_r N(r)$ sampled by the growth cone.

4.5.1 Model transition matrix

The transition matrix for this order parameter is a 63×63 element square matrix whose columns correspond to an initial separation r and whose rows correspond to the subsequent separation s. Since the growth cone can only occupy the limited number of discrete target sites implemented on the Neurotrope1 chip, r and s assume integer values ranging from 0 up to a maximum of 62, which is the number of edges separating diagonally opposite corners of the rectangular array. Boundary conditions for the finite array introduce a dependence on the synapse location $\mathbf{y}$, which we will eliminate when we generalize to population behavior in the next chapter.

4.5.1.1 Construct W(y)

We choose a time step short enough to permit the growth cone to jump at most once between consecutive observations. The resulting transition matrix $\mathbf{W}(\mathbf{y})$ has nonzero elements only on or immediately adjacent to its diagonal. Elements $W_{(r\pm1)r}(\mathbf{y})$ immediately below or above the diagonal are the probabilities that a growth cone initially at r jumps one edge toward or away from its synapse. An element $W_{rr}(\mathbf{y})$ on the diagonal is the probability $1 - P_{\text{jump}}(r)$ that a growth cone at r has not jumped to a new location after one time step.

$P_{\text{jump}}(r)$ can be broken down into the probabilities for directed and undirected jumps, which must be apportioned between the adjacent transition probabilities $W_{(r\pm1)r}(\mathbf{y})$. From the perspective of the growth cone, the sampled neurotropin gradient points directly toward its synapse, which is conveniently located at the zero of the separation coordinate r. Directed jumps automatically orient toward lower r, so we can assign all directed jump probability $P_{\text{dir}}(r)$ to $W_{(r-1)r}(\mathbf{y})$. Undirected jumps are divided between adjacent nodes according to the probability $P_{\text{diff}}(r, r \pm 1|\mathbf{y})$ that the adjacent node is located at separation $(r \pm 1)$ from a synapse located at $\mathbf{y}$.

Our recipe for the elements of $\mathbf{W}(\mathbf{y})$ can be summarized as

$$W_{sr}(\mathbf{y}) = \begin{cases} 0, & s < r-1 \\ P_{\text{jump}}(r)\left\{P_{\text{diff}}(r, s|\mathbf{y})(1 - P_{\text{dir}}(r)) + P_{\text{dir}}(r)\right\}, & s = r-1 \\ 1 - P_{\text{jump}}(r), & s = r \\ P_{\text{jump}}(r)\left\{P_{\text{diff}}(r, s|\mathbf{y})(1 - P_{\text{dir}}(r))\right\}, & s = r+1 \\ 0, & s > r+1 \end{cases}$$

4.5.1.2 Compute $P_{\text{jump}}(r)$ and $P_{\text{dir}}(r)$

$P_{\text{jump}}(r)$ and $P_{\text{dir}}(r)$ both depend on the neurotropin sampled by the growth cone when it is located r edges away from the synapse. Neurotropin emanates isotropically from a release site, so the neurotropin concentration at any target site depends on the radial distances separating it from the release sites at each of the two axon terminals. The growth cone samples neurotropin from the center of its own release site, so the neurotropin observed by the growth cone depends solely on its distance from the synapse release site:

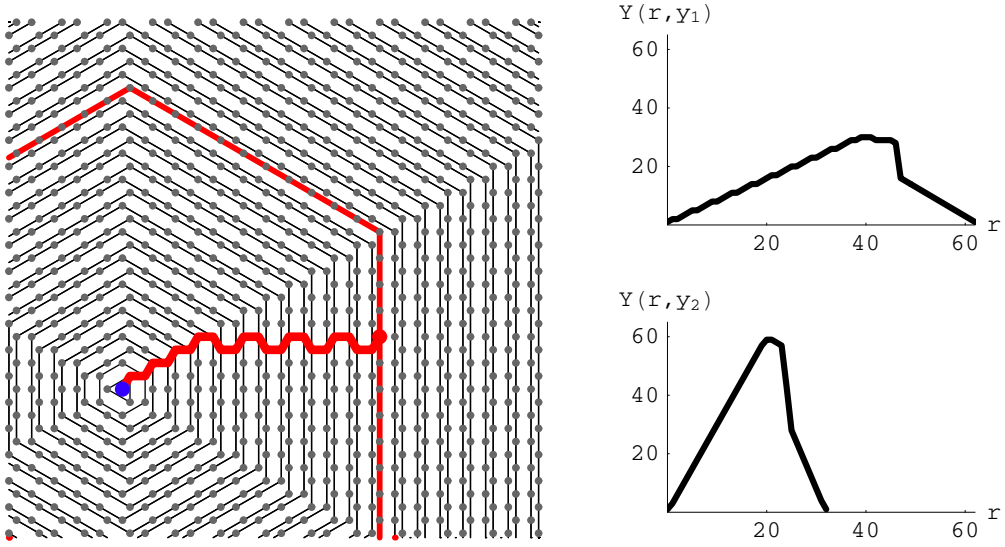
$$N(r) \propto \exp\left[-\frac{r_\alpha^2}{2\sigma_{\text{NT}}^2}\right] + \rho_{\text{GC}} \tag{4.27}$$

The first term in this expression is the peak-normalized Gaussian spreading kernel that represents the portion of synapse-induced neurotropin release that percolates r edges to the growth cone from the synapse position. The width σ_{NT} of the Gaussian is a free parameter. The ρ_{GC} term represents the neurotropin release stimulated by the growth cone from its own location, and is the ratio of the postsynaptic spike rates induced by the growth cone and the synapse. ρ_{GC} is a free parameter, and is presumed to be less than one since neurotransmission is more efficient at a synapse.

Given this expression for the sampled neurotropin concentration, the probability $P_{\text{jump}}(r)$ that the growth cone jumps away from its initial position at separation r during a unit time step is

$$P_{\text{jump}}(r) \;=\; 1 - \exp\left[-\lambda_{\text{NT}} N(r)\right] \tag{4.28}$$

where λ_{NT} is a free parameter. The probability $P_{\text{dir}}(r)$ that a particular growth cone jump



(a) (b)

Figure 4.5: $Y(r, \mathbf{y})$

(a) The target site $\mathbf{y}$ (blue dot) occupied by a synapse acts as a static attractor for a coactive growth cone. All other target sites $\mathbf{x}$ (red dot) can be grouped in concentric rings according to the number of lattice edges $r = |\mathbf{x} - \mathbf{y}|$ separating them from the attractor (red lines). (b) Number $Y(r, \mathbf{y})$ of target sites located r edges away from a reference target site $\mathbf{y}$ located in a corner of the array (top) or the center of the array (bottom).

initiated from a node at separation r is directed is

$$P_{\text{dir}}(r) \quad = \quad 1 - \exp\left[-\lambda_{\text{dir}} \|\nabla_r N(r)\|\right] \tag{4.29}$$

where λ_{dir} is a free parameter.

4.5.1.3 Compute $P_{\text{diff}}(r, s|\mathbf{y})$

$P_{\text{diff}}(r, s|\mathbf{y})$ is the probability that an undirected growth cone moves by chance from a node at separation r to a node at separation s, given that the synapse is located at $\mathbf{y}$. For short time steps, growth cones can make at most one jump, so $P_{\text{diff}}(r, s|\mathbf{y}) = 0$ for $|r - s| > 1$. Moreover, in a honeycomb lattice there are no edges connecting nodes at the

same r, so jumps are directed either inward or outward and $P_{\text{diff}}(r, r|\mathbf{y}) = 0$.

An undirected jump is equally likely to follow any of the available edges out of its node, so the probability $P_{\text{diff}}(r, r \pm 1|\mathbf{y})$ that an unguided growth cone jumps by chance from r to $r \pm 1$ is the fraction of edges exiting its current node at r that lead to adjacent nodes at $r \pm 1$, which we will approximate as

$$P_{\text{diff}}(r, r \pm 1|\mathbf{y}) \approx \frac{Y(r \pm 1, \mathbf{y})}{Y(r-1, \mathbf{y}) + Y(r+1, \mathbf{y})} \tag{4.30}$$

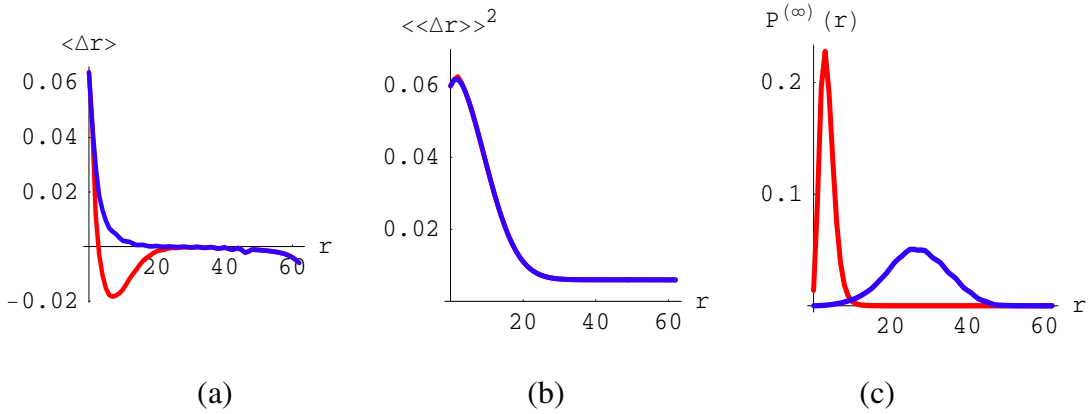
$Y(r, \mathbf{y})$ is the number of lattice nodes located r edges away from target site $\mathbf{y}$ (Figure 4.5(a)). Each target site defines its own set of concentric rings of equidistant nodes, whose number varies with the position of $\mathbf{y}$ relative to the boundaries. A synapse located in a corner of the array can be separated from its growth cone by up to twice as many edges as one located in the center of the array (Figure 4.5(b)).

4.5.2 Jump moments of W

To obtain some insight into the dynamic properties of our transition matrix, we introduce indicators called *jump moments*. The kth jump moment $\mu_k(r)$ of a transition matrix $\mathbf{W}$ is defined to be

$$\mu_k(r) = \sum_s (s - r)^k \mathbf{W}_{sr} \tag{4.31}$$

The first jump moment $\mu_1(r) = \langle \Delta r \rangle$ is the expected jump Δr in the value of r over one time step. That is, if a growth cone is initially separated from the synapse by a distance r, at the next time step we expect to find it at a site located $r + \langle \Delta r \rangle$ away from the synapse. In the absence of noise, $\langle \Delta r \rangle$ describes deterministic growth cone motion. Sample $\langle \Delta r \rangle$



(a) (b) (c)

Figure 4.6: Jump moments

Sample transition matrix jump moments and predicted stationary distribution for $P_{\text{dir}}(r) = 0$ (blue) or $P_{\text{dir}}(r) > 0$ (red). (a) $\langle \Delta r \rangle$ is the expected change in r over one time step. (b) $\langle \langle \Delta r \rangle \rangle^2$ is the variance in Δr. (c) The stationary distribution $P^{(\infty)}(r)$ is the eigenvector of $\mathbf{W}$ with eigenvalue 1.

functions are plotted in Figure 4.6(a). The red line is the first jump moment of a transition matrix constructed from $P_{\text{dir}}(r) > 0$, corresponding to neurotropic attraction, and the blue line is constructed from $P_{\text{dir}}(r) = 0$, corresponding to purely random motion. Deviations from the blue line reveal the presence of directed jump biasing.

Notice that $\langle \Delta r \rangle$ is generally nonzero even when $P_{\text{dir}}(r) = 0$. The inherent geometry of our order parameter combine with the boundary conditions of the array to bias the growth cone toward intermediate r even in the absence of explicit direction. $\langle \Delta r \rangle$ is positive at low r, since there are more sites located at intermediate r than at low r. Each hexagonal lattice node has three neighbors, so in an infinite hexagonal array there are $3r$ sites located a distance $r > 0$ from an attractor and 1 site at $r = 0$. If a growth cone is equally likely to jump to any of its three neighbors, the probability that a growth cone will jump outward to an attractor separation $r + 1$ instead of inward to $r - 1$ is proportional to the number of neighbors at the higher separation. This probability is 1 at $r = 0$ and falls asymptotically toward $1/2$ as r increases. At high r, there is no net bias to jump inward or outward, so

$\langle \Delta r \rangle \approx 0$. In a finite array, the number of sites at high r falls to zero as r exceeds the array bounds, so there are more sites located at intermediate r than at high r and $\langle \Delta r \rangle$ becomes negative at high r. The zero crossing for $\langle \Delta r \rangle$ in a finite array is located near the separation r with the largest number of sites.

The variance $\langle\langle \Delta r \rangle\rangle^2 = \mu_2(r) - \mu_1^2(r)$ is plotted in Figure 4.6(b). $\langle\langle \Delta r \rangle\rangle^2$ reflects the propensity for a growth cone to make any jump, directed or not, and therefore tracks the sampled neurotropin concentration $N(r)$. Higher moments describe deviations from the Gaussian distribution. The third moment encodes skewness, the asymmetry of the distribution peak, while the fourth moment encodes kurtosis, the thickness of the distribution tails. We will not concern ourselves with higher moments, since their information is subsumed by the full transition matrix.

We use the full transition matrix to predict the stationary distribution $P^{(\infty)}(r)$ by computing the principal eigenvector of $\mathbf{W}$, which is plotted in Figure 4.6(c). For random motion, $P^{(\infty)}(r)$ is proportional to $Y(r, \mathbf{y})$. Neurotropic attraction shifts this probability toward $r = 0$.

4.5.3 Empirical transition matrix

To test our model transition matrix, we compare its predictions with actual chip data. Figure 4.7(a) shows the empirical transition matrix assembled from connection maps downloaded from the Neurotrope1 system over the course of a refinement experiment consisting of 225000 randomly selected stimuli presented according to the population supervised attraction protocol that will be introduced in the next chapter. During this time, no growth cones were observed at certain large separations, so we have no estimate of W_{sr} for those

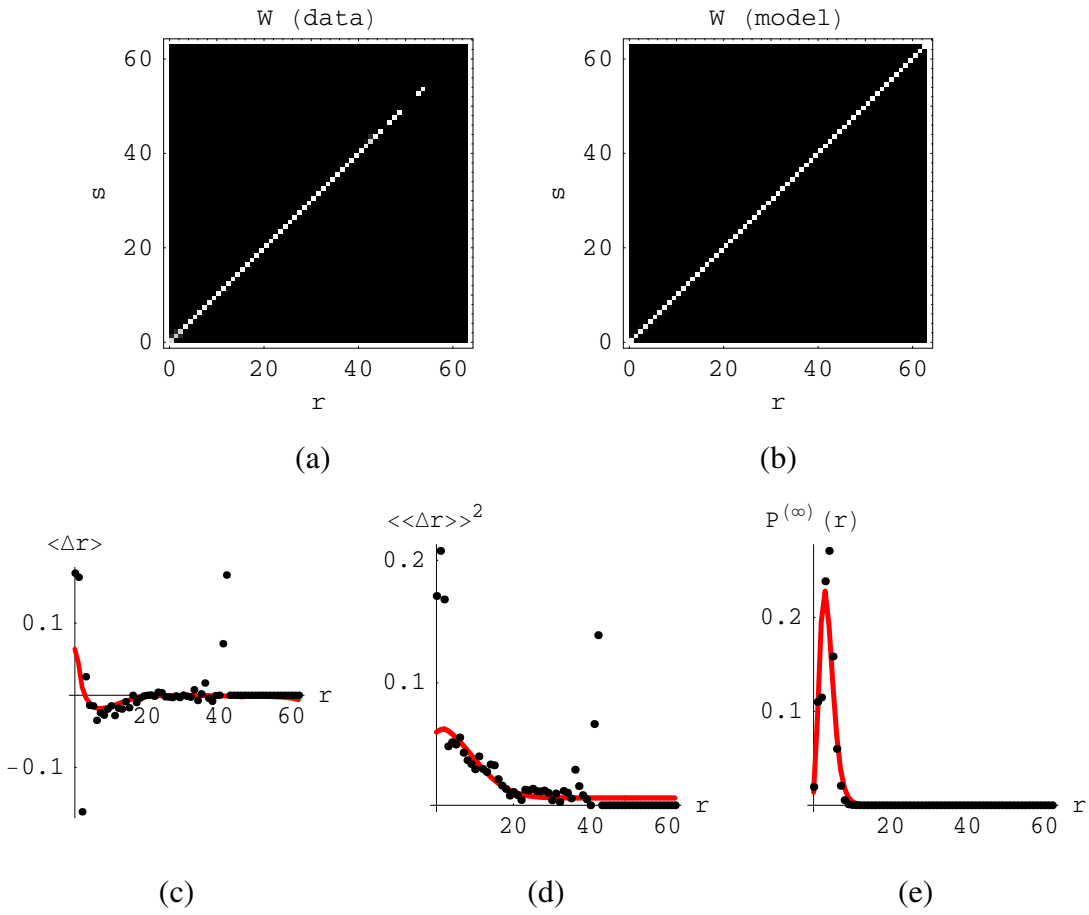


Figure 4.7: Model fit to data

(a) Empirical transition matrix. (b) Model transition matrix. (c) Expected separation change $\langle \Delta r \rangle$ for empirical (black dots) and model (red line) transition matrices. (d) Separation change variance $\langle \langle \Delta r \rangle \rangle^2$. (e) Stationary separation distribution $P^{(\infty)}(r)$, computed from transition matrix principal eigenvector.

values of r[3]. Extremely large growth cone-synapse separations are seldom observed during short experiments, since there are few growth cone configurations that permit these separations. However, such rare events exert little influence on system dynamics, so once we have accumulated enough data to construct a reasonable estimate of $\mathbf{W}$, we can use its principal eigenvector to predict the stationary distribution $P^{(\infty)}$ without having to wait for

[3]A value of r that is not observed during an experiment constitutes an inaccessible state, for which the entry transition probability W_{rs} is 0 for all s. We typically assign the exit transition probability W_{sr} to be 0 for all s, dropping the state out of the transition matrix entirely, although an equally valid heuristic would be to assign $W_{rr} = 1$ and $W_{(s \neq r)r} = 0$.

the system to completely converge.

4.5.4 Model fitting

We fit the empirical transition matrix (Figure 4.7(a)) with the model transition matrix (Figure 4.7(b)) by varying the free parameters by hand until the jump moments of the model roughly match the data, as shown in Figure 4.7(c) and (d). The stationary distributions predicted by the principal eigenvectors of the two transition matrices are overlaid in Figure 4.7(e), and show reasonably good agreement.

We can validate some of the assumptions of our model by comparing its predicted dynamics with those of the data. Figures 4.8 and 4.9 show the jump probability $P^{(n)}_{\text{jump}}(r)$ and one step transition probabilities $W^{(n)}_{(r\pm1)r}$ as functions of time step n and initial separation r from a single static release site. The top row of each figure is compiled from chip data collected using the supervised attraction protocol that will be introduced in the next chapter, and the bottom row is the model whose free parameters have been tweaked by hand to fit the data. In Figure 4.8, the jump probabilities asymptote toward 1 exponentially with time, as befits a Markov process. Figure 4.9(a) shows that the jump rates fall off with distance from the release site, suggesting a Gaussian concentration peak centered on the release site. Figure 4.9(b) shows the probability that the growth cone jumps toward its synapse. This probability is high at intermediate distances, but falls sharply at short and long distances, which is consistent with a function of the gradient of a Gaussian.

In the next chapter we will apply these techniques to a supervised learning problem.

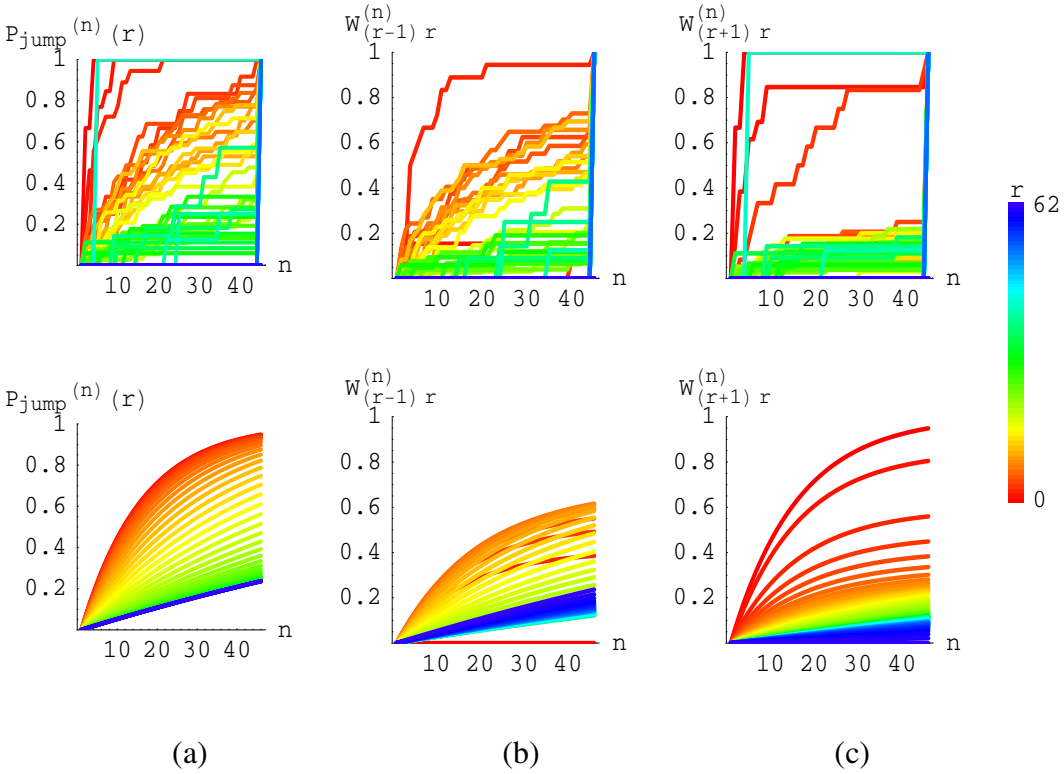


Figure 4.8: Time dependence of P_{jump} and $\mathbf{W}$

Top: data. Bottom: model. Color denotes separation r. (a) $P^{(r)}_{\text{jump}}(n)$. The probability of a jump asymptotes to 1 exponentially with n. (b) $W^{(n)}_{(r-1)r}$ is the probability that a growth cone has jumped one step toward its synapse after n time steps. (c) $W^{(n)}_{(r+1)r}$ is the probability that a growth cone has jumped one step away from its synapse after n time steps.

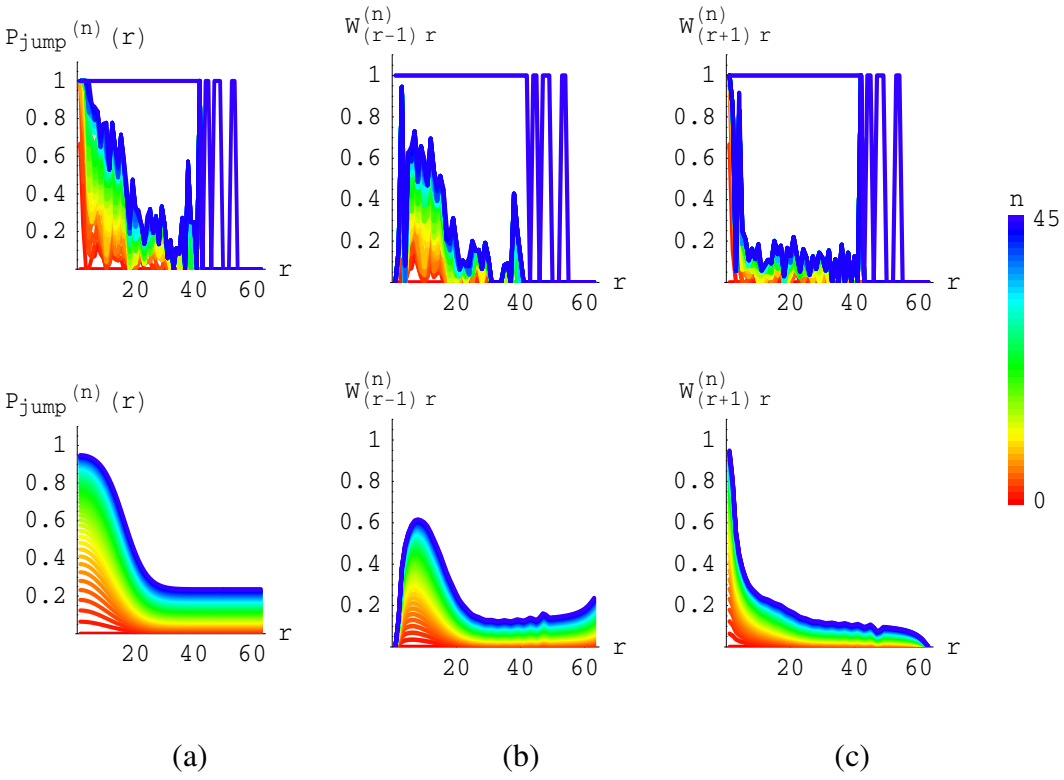


Figure 4.9: Separation dependence of P_{jump} and $\mathbf{W}$

Top: data. Bottom: model. Color denotes initial separation r. (a) $P^{(n)}_{\text{jump}}(r)$. (b) $W^{(n)}_{(r-1)r}$. (c) $W^{(n)}_{(r+1)r}$.

Chapter 5

Supervised Pair Attraction

In this chapter we apply the transition matrix techniques developed in the previous chapter to analyze supervised pair attraction in the Neurotrope1 system, with and without the density constraint.

5.1 Stimulus protocols

Our eventual goal is to study the general case of unsupervised behavior of a population of migrating growth cones that interact via neurotropic attraction and volume repulsion. This is a complicated problem, which we will build up to in two stages.

5.1.0.1 Unsupervised self-organization

In the general case of unsupervised self-organization introduced in Chapter 2, a cluster of growth cones is activated simultaneously and its coactive members converge spatially within the target layer. A growth cone participates in multiple clusters, each tugging it in a different direction until it achieves a consensus-satisfying position. To prevent all growth cones from collapsing onto the same target cell, a *density constraint* states that only one growth cone can occupy a given target site at a time. A growth cone that wishes to move must first push aside any other growth cone obstructing its path.

There are thus two types of growth cone interactions: mutual attraction between coactive growth cones, and unilateral displacement of an inactive growth cone by an active growth cone. Attraction between coactive growth cones fosters cooperation within clusters, whose members band together to defend territory in the target layer; displacement of inactive growth cones fosters competition between clusters, whose members expel nonmembers from their territory. The migration of an individual growth cone is guided by its membership in different sets of overlapping clusters. This cluster membership roster is the program that instructs the development and final configuration of the growth cone population, using only an initial configuration as input.

5.1.0.2 Supervised pair attraction

Our first simplification is to select a set of simple, nonoverlapping coactivity clusters that will be easy to analyze. The simplest nontrivial cluster consists of a single pair of axon terminals that are coactive with one another and no other. When perfectly guided by a rule that translates correlated activity into spatial proximity, the two axon terminal

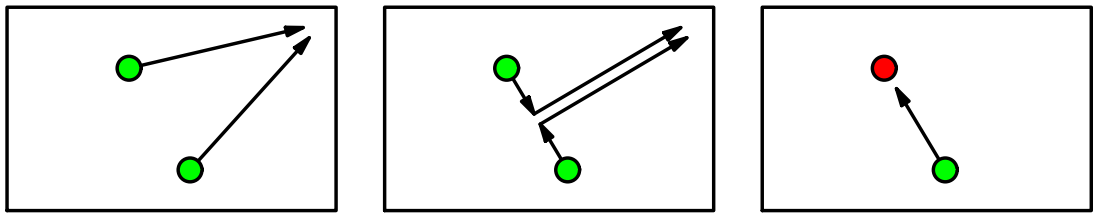


Figure 5.1: Paired attraction mechanisms

Paired attraction in the presence of a weak global attractor. (a) Tidal attraction. Growth cones arrive at adjacent locations by moving independently to a weak global attractor. (b) Unsupervised local attraction. Growth cones mutually attract, then move as a pair to a weak global attractor. (c) Supervised local attraction. A fixed synapse (red) strongly attracts a mobile growth cone (green), overriding the weak global attractor.

positions should exactly coincide. The target layer distance r separating these two positions is therefore a good measure of the mutual attraction between the axon terminals.

The simplest paired stimulus protocol would be to activate two growth cones simultaneously and observe their attraction, as quantified by r. However, in an imperfect system like a neuromorphic chip, systematic anisotropies can establish activity-independent gradients that tug mobile growth cones toward a corner of the array. Under these conditions the corner acts as a global attractor, creating tidal effects that can be difficult to distinguish from attraction between growth cones (Figure 5.1(a,b)). We can override a weak global gradient with a strong local attractor if we fix the position of one axon terminal and let the other roam freely, as if one axon has already formed stable synapses that anchor it in place, while the other is tipped by a motile growth cone. This type of paired attraction constitutes a *supervised learning* task in which the fixed synapse instructs the migration of the mobile growth cone (Figure 5.1(c)).

5.1.0.3 Supervised pair attraction without bumps

Our second simplification is to remove the density constraint, in order to separate and

characterize distinct growth cone motion mechanisms. There are two mechanisms to move a growth cone within the target layer, one internal and one external.

Internally generated moves are called *jumps*, and are initiated by the growth cone itself once it determines the direction of the local neurotropin gradient. A growth cone binds neurotropin from its location only when it is presynaptically active, and is blind to neurotropin that appears while it is silent. Upon accumulating a threshold of bound neurotropin, the growth cone jumps from its current node on the discrete lattice of target layer sites to an adjacent lattice node selected by its calculation of the local neurotropin gradient. The growth cone sometimes miscalculates the gradient, so some jumps cause undirected motion.

Externally generated moves are called *bumps*, and are initiated by one of the growth cone's neighbors in order to satisfy the constant density constraint. A growth cone is bumped from its current node whenever another growth cone jumps in from an adjacent node. A bumped growth cone is unilaterally displaced to the adjacent node without regard to its own activity or the relative fitness of its former position, so all bumps represent undirected motion.

Jumps are fundamental to supervised pair attraction, since they mediate the neurotropic interaction that draws the growth cone to its synapse. Bumps are not, since they only mediate the density constraint that disambiguates overlapping clusters. Since the activity of growth cone-synapse pairs do not overlap, each growth cone is uniquely attracted to a distinct synapse, rendering the density constraint superfluous and permitting us to dispense with bumps entirely in this special case[1]. This second simplification allows us to isolate

[1]Each update request generated by the Neurotrope1 chip identifies a pair of adjacent growth cones whose positions are to be swapped. To eliminate bumps, we simply program the system to only update the lookup table entries corresponding to the growth cone that requested the update, leaving its swap partner untouched. Multiple growth cones can occupy the same target site, but their individual jump requests can be disam-

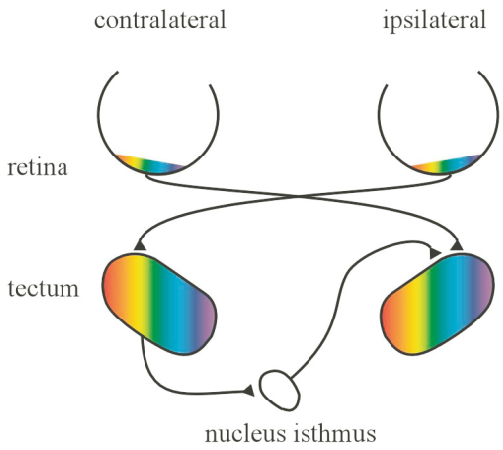


Figure 5.2: Binocular map formation in *Xenopus* optic tectum

Retinotectal axons form a retinotopic projection onto contralateral tectum, which then relays axons via nucleus isthmus to ipsilateral tectum, where the isthmotectal axon terminals align retinotopically with the stable contralateral retinotectal projection.

the jump mechanism.

Our strategy will be to first characterize jumps in this second special case of *supervised attraction without bumps*, and then generalize to the first special case of *supervised attraction with bumps*.

5.1.1 Experimental protocol

Examples of supervised learning in the brain include classical conditioning in the cerebellum, calibration of the vestibule-ocular reflex, and spatial receptive field formation in the inferior colliculus [58]. The supervised attraction protocol described in this chapter is most analogous to binocular map formation in *Xenopus* frog optic tectum (Figure 5.2) [45]. Neighboring retinal ganglion cells project axons to neighboring locations in contralateral tectum, using activity-independent cues to form a stable retinotopic map. Cells in this map

biguated because only one growth cone is active at a time under the supervised pair attraction protocol.

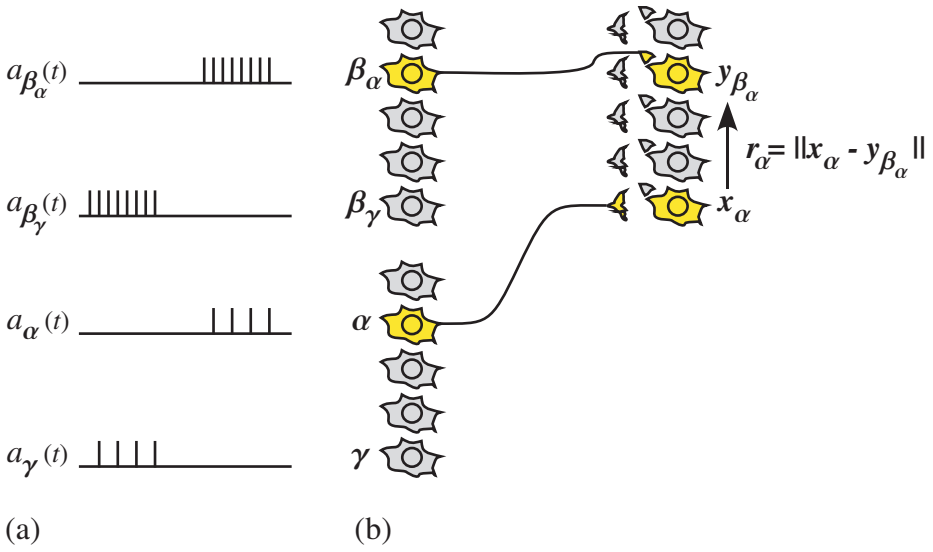


Figure 5.3: Supervised attraction protocol

Two separate source cell populations innervate the target layer, one projecting synapses and the other projecting growth cones. Each growth cone α, γ is coactive with a unique synapse $\beta_\alpha, \beta_\gamma$. (a) Presynaptic spikes arrive in bursts whose average overlap defines coactivity. A growth cone α is coactive only with its own synapse β_α. (b) A synapse β_α remains at a fixed target site $\mathbf{y}_{\beta_\alpha}$, neurotropically attracting its growth cone from location $\mathbf{x}_\alpha$. The order parameter for this protocol is the target layer distance $r_\alpha = \|\mathbf{x}_\alpha - \mathbf{y}_{\beta_\alpha}\|$ separating the growth cone from its synapse, averaged over all growth cone-synapse pairs.

subsequently project to the nucleus isthmus, and thence to ipsilateral tectum, where the retinotopic map projected from the contralateral eye provides an activity template for the isthmotectal projection to mimic, resulting in a retinotopic map of the ipsilateral eye that aligns exactly with the contralateral map. Isthmotectal axons are attracted to their contralateral counterparts.

We will characterize growth cone motion in the Neurotrope1 system by simulating supervised retinotopic refinement of the isthmotectal projection. In these experiments, two populations of source cells project axons into the target layer (Figure 5.3). Axons projected by the first population are tipped by mobile growth cones, corresponding to the dynamic isthmotectal projection. Axons projected by the second population form stable synapses, corresponding to the static contralateral retinotectal projection. Each growth cone α is

coactive with a unique synapse β_α, which acts as a fixed attractor that reels the growth cone toward its retinotopically correct target.

5.1.1.1 Stimulus

All of the experiments described in this chapter use the following paired stimulus protocol to instruct topographic development:

1. Select a growth cone-synapse pair at random. Every pair is selected with equal probability, so that on average each of the Z pairs is equally active.

2. Stimulate the growth cone-synapse pair with a burst of spikes at some constant frequency. The burst length should be chosen to elicit at most one jump from the growth cone.

3. Route each spike to either the synapse or the growth cone. The fraction of spikes routed to each axon terminal determines its relative neurotransmission efficacy.

4. Repeat.

Under this stimulus protocol, each growth cone-synapse pair is activated separately and equally, so the coactivity between growth cone α and axon terminal β is

$$C_{\alpha\beta} \propto \delta(\beta - \beta_\alpha) + \rho_{\mathrm{GC}}\delta(\beta - \alpha) \tag{5.1}$$

where ρ_{GC} measures the neurotransmission efficacy of a growth cone relative to a synapse. The resulting uptake parameter $U_\alpha^{(n)}$ for an individual growth cone is just the target layer distance $r_\alpha^{(n)} = \|\mathbf{x}_\alpha^{(n)} - \mathbf{y}_{\beta_\alpha}\|$ separating the growth cone α from its synapse β_α, which is conveniently identical to the order parameter for absolute retinotopy $\Phi_\alpha^{(n)} = r_\alpha^{(n)}$. The

population uptake and order parameters are equal to the average separation $r^{(n)} = \langle r_\alpha^{(n)} \rangle$ over all Z growth cone-synapse pairs.

$$\Phi^{(n)} = U^{(n)} = \frac{1}{Z} \sum_\alpha \| \mathbf{x}_\alpha^{(n)} - \mathbf{y}_{\beta_\alpha} \| \tag{5.2}$$

5.1.1.2 Update

Growth cones are free to move anywhere in the target array, while synapses are anchored to fixed locations. The Neurotrope1 system implements this update rule by storing synapse and growth cone target addresses in separate blocks of the lookup table. Lookup table update requests computed by the Neurotrope1 chip are applied exclusively to the growth cone address block. Synapse addresses ignore update requests, remaining constant over the course of the experiment to instruct the migration of the growth cone population. Each synapse target address is initialized with the address of the corresponding source cell to establish a perfectly topographic mapping from source cell to synapse.

5.1.1.3 Sampling

The contents of the growth cone address block are read out at fixed sampling intervals τ. Each sample consists of the Z growth cone positions indexed in target layer coordinates, which are immediately translated into relative growth cone-synapse separations r. The population distribution $P^{(n)}(r)$ of separations r within the nth sample is constructed by counting the number of appearances of each value of r and normalizing to Z.

We select the sampling interval τ to include approximately Z presynaptic burst stimuli, which is enough time to allow every growth cone on average one opportunity to jump. The actual number of opportunities for a given growth cone to jump between samples is

stochastic, since bursts are targeted randomly.

5.1.1.4 Initialization

Growth cone addresses are initialized with different distributions $P^{(0)}(r)$ for different experiments. In a topographic refinement experiment, the growth cone address block is initialized with a coarsely topographic mapping from source layer to target layer, and converges over time to a more faithful representation of the source topography. In a topographic relaxation experiment, the growth cone address block is initialized with a perfect image of the source layer, and is scrambled by noise to a more sustainable degree of topography. We will only examine refinement experiments in this chapter.

5.2 Supervised pair attraction without bumps

In this section we characterize the growth cone jump mechanism in isolation by examining supervised pair attraction without bumps. We construct a model transition matrix and fit it to an empirical transition matrix compiled from chip data.

5.2.1 Model transition matrix

For supervised pair attraction, only two neurotropin release sites are active at a time, one hosted by each axon terminal in a growth cone-synapse pair. Without bumps, axon terminals in different growth cone-synapse pairs do not interact, so each growth cone is free to move toward its coactive synapse without being displaced by other growth cones or distracted by other synapses. The model transition matrix $\mathbf{W}$ that describes this population of independently operating growth cone-synapse pairs is almost identical to the transition

matrix $\mathbf{W}(\mathbf{y})$ that described a single growth cone-synapse pair in the previous chapter. The only difference is that the order parameter for the single pair task is tied to the location $\mathbf{y}$ of a specific synapse, while the order parameter for the population task is averaged over all synapse sites.

We can eliminate the synapse-specificity of $\mathbf{W}(\mathbf{y})$ by modifying the diffusion probability $P_{\text{diff}}(r, r \pm 1|\mathbf{y})$. Randomly oriented jumps move the growth cone to an adjacent site at a new separation of $r \pm 1$ with a probability $P_{\text{diff}}(r, r \pm 1)$ that depends on the lattice topology and boundary conditions. At the population level, these geometric considerations are approximated by

$$P_{\text{diff}}(r, r \pm 1) \approx \frac{Y(r \pm 1)}{Y(r-1) + Y(r+1)} \tag{5.3}$$

where $Y(r)$ is the number of ways to select synapse sites and arrange growth cones within the array such that the radial distance separating the synapse from the growth cone is r.

All of the other components of the single pair transition matrix remain the same. The sampled neurotropin concentration is

$$N(r) \propto \exp\left[-\frac{r^2}{2\sigma_{\text{NT}}^2}\right] + \rho_{\text{GC}} \tag{5.4}$$

where the peak-normalized Gaussian spreading kernel width σ_{NT} and the relative growth cone-synapse activity ratio ρ_{GC} are free parameters. The probability for a growth cone to initiate a jump between consecutive samples is

$$P_{\text{jump}}(r) = 1 - \exp\left[-\lambda_{\text{NT}} N(r) \tau\right] \tag{5.5}$$

where the jump rate λ_{NT} and the sampling interval τ are free parameters. Finally, the

probability for a given growth cone jump to be directed is

$$P_{\text{dir}}(r) \quad = \quad 1 - \exp\left[-\lambda_{\text{dir}} \|\nabla_r N(r)\|\right] \tag{5.6}$$

where the attraction bias λ_{dir} is a free parameter.

From these components we assemble the one step transition probabilities W_{sr} for the sampling interval τ.

$$W_{sr} = \begin{cases} 0, & s < r-1 \\ P_{\text{jump}}(r)\left\{P_{\text{diff}}(r,s)\{1 - P_{\text{dir}}(r)\} + P_{\text{dir}}(r)\right\}, & s = r-1 \\ 1 - P_{\text{jump}}(r), & s = r \\ P_{\text{jump}}(r)\left\{P_{\text{diff}}(r,s)\{1 - P_{\text{dir}}(r)\}\right\}, & s = r+1 \\ 0, & s > r+1 \end{cases} \tag{5.7}$$

5.2.2 Parameter fits

The model as stated has no fewer than five free parameters: the sampling interval τ, the relative growth cone-synapse activity ratio ρ_{GC}, the neurotropin spreading range σ_{NT}, the jump rate λ_{NT}, and the attraction bias λ_{dir}. The sampling interval τ is specified by our sampling program, and is selected to provide every growth cone on average one opportunity to jump between consecutive samples. τ is long enough for Z presynaptic burst stimuli, each of which contains enough spikes to elicit at most one jump from the stimulated growth cone.

Our strategy is to construct a model $\mathbf{W}$ by extracting values for the remaining four parameters from features of the expected separation change $\langle \Delta r \rangle$ and the separation change

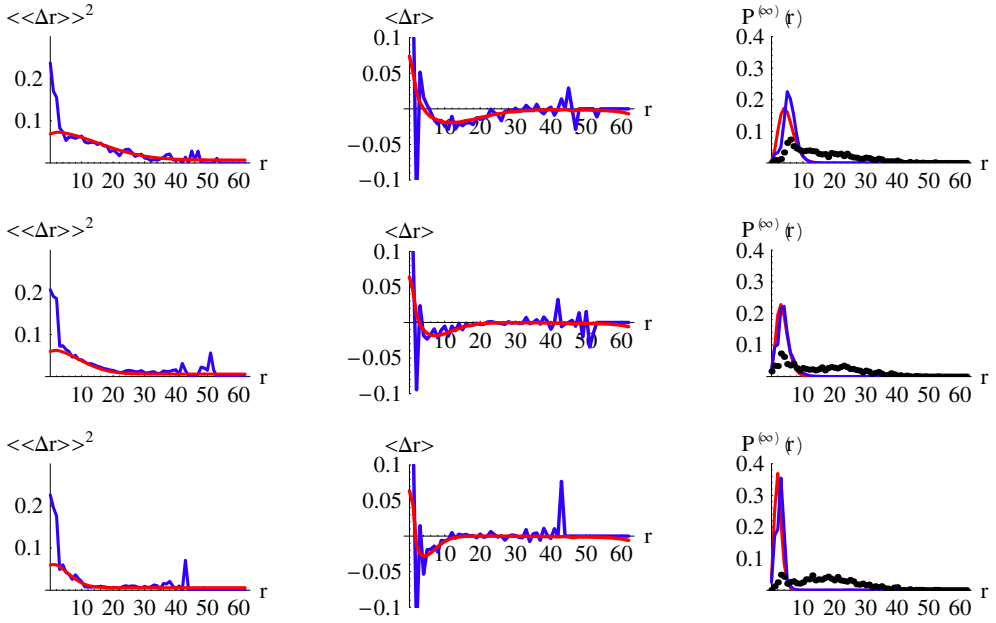


Figure 5.4: Moment fits and predicted $P^{(\infty)}(r)$ without bumps

Moments and predicted stationary distribution for empirical (blue) and model (red) transition matrices $\mathbf{W}$. Top: $V_{\text{shunt}} = 0.20V$; middle: $V_{\text{shunt}} = 0.30V$; bottom: $V_{\text{shunt}} = 0.40V$. (a) Separation change variance $\langle\langle\Delta r\rangle\rangle^2$. (b) Expected separation change $\langle\Delta r\rangle$. (c) Predicted stationary distribution $P^{(\infty)}(r)$ (solid lines) and measured distribution $P^{(100)}(r)$ (black dots).

variance $\langle\langle\Delta r\rangle\rangle^2$ for a single sampling interval τ, and to compare the $P^{(\infty)}(r)$ predicted by the model $\mathbf{W}$ with that predicted by an empirical $\mathbf{W}$ compiled from chip data. The parameter values we extract from the special case of supervised attraction without bumps will carry over to more general cases.

5.2.2.1 Separation change variance $\langle\langle\Delta r\rangle\rangle^2$

The separation change variance $\langle\langle\Delta r\rangle\rangle^2$ tracks all jumps, directed or not, initiated by growth cones from an initial separation r, and therefore acts as a proxy for the jump probability $P_{\text{jump}}(r)$, which in turn reflects the neurotropin concentration $N(r)$ observed by the growth cone (Figure 5.4(a)). Accordingly, we will use features of $\langle\langle\Delta r\rangle\rangle^2$ to fit the three

free parameters associated with $P_{\text{jump}}(r)$: the relative growth cone-synapse activity ratio ρ_{GC}, the spreading range σ_{NT}, and the jump rate λ_{NT}.

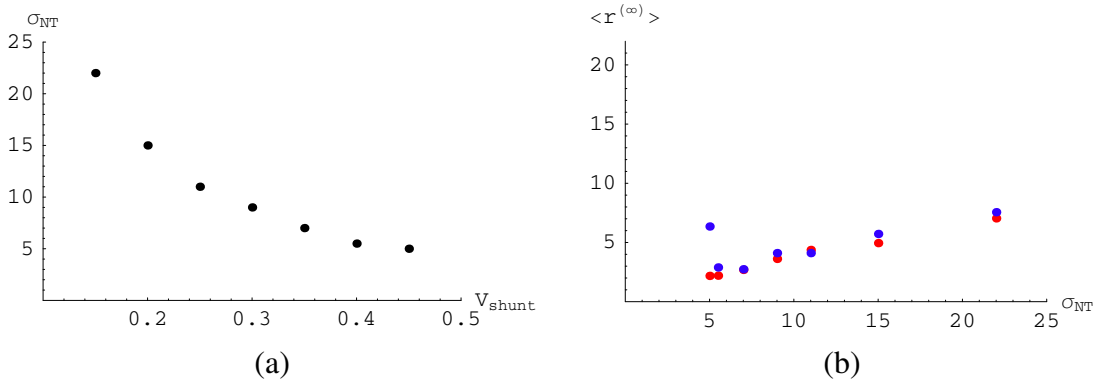
$\langle\langle\Delta r\rangle\rangle^2$ is flat at large r, where the only neurotropin available to the growth cone is the weak constant release it triggers from its own location. This small constant directly reflects the growth cone's presynaptic activity, which determines ρ_{GC}. At low r, $\langle\langle\Delta r\rangle\rangle^2$ consists of a sharp spike at $r = 0$ riding atop a broader, shallower peak. The sharp spike is an implementation-specific artifact that occurs at short distances where the growth cone and synapse are forced to share transistor circuits. The broad peak tracks the spread of neurotropin released from the synapse position. The width of the broad peak determines the spreading range σ_{NT}, and the height determines the jump rate λ_{NT}.

5.2.2.2 Expected separation change $\langle\Delta r\rangle$

The sign and magnitude of the expected separation change $\langle\Delta r\rangle$ reflects the balance between directed and undirected growth cone jumps (Figure 5.4(b)). Negative $\langle\Delta r\rangle$ pull the growth cone closer to its synapse and positive $\langle\Delta r\rangle$ push it away. The geometry of the radial coordinate r generates weak positive and negative $\langle\Delta r\rangle$ for low and high r, respectively. In addition, there is a broad negative trough centered on $r = \sigma_{\text{NT}}$ that follows the gradient of the neurotropin spreading kernel. The net attractive bias is reflected by the depth of this negative trough, a feature that determines the attraction bias λ_{dir}. Larger bias rates increase the share of directed jumps, pushing $\langle\Delta r\rangle$ more negative.

5.2.2.3 Stationary distribution $P^{(\infty)}(r)$

We test our model $\mathbf{W}$ by comparing the $P^{(\infty)}(r)$ predicted by its principal eigenvector with that predicted by the measured $\mathbf{W}$. Figure 5.4(c) shows the empirical and model $\mathbf{W}$



(a) (b)

Figure 5.5: Spreading range σ_{NT}

(a) σ_{NT} dependence on V_{shunt}. (b) Stationary $\langle r \rangle$ predicted by empirical $\mathbf{W}$ (blue) and model $\mathbf{W}$ (red) for different σ_{NT}.

eigenvector predictions for $P^{(\infty)}(r)$ (blue and red lines), as well as the recorded distribution $P^{(100)}(r)$ of the 100th sample (black dots). The two eigenvector predictions do not fit the recorded $P^{(100)}(r)$ because growth cones move slowly without bumps and the population distribution does not approach equilibrium within 100 samples. However, the model $\mathbf{W}$ prediction does roughly line up with the empirical $\mathbf{W}$ prediction, which is the aim of the parameter extraction.

5.2.2.4 Parameter values

Applying these heuristics to data collected from the chip, we extract the following parameter values: $\lambda_{\mathrm{NT}} = 0.06$ jumps/τ, $\lambda_{\mathrm{dir}} = 10$, $\rho_{\mathrm{GC}} = 0.1$. We manipulate the spreading range σ_{NT} directly with the gate bias V_{shunt}, which controls the strength of the constant decay current at each lattice node. Increasing V_{shunt} reduces the range of neurotropin spread by shunting more current closer to the release site. The relation between σ_{NT} and V_{shunt} is plotted in Figure 5.5(a).

Reducing σ_{NT} improves growth cone localization. Figure 5.5(b) shows the stationary

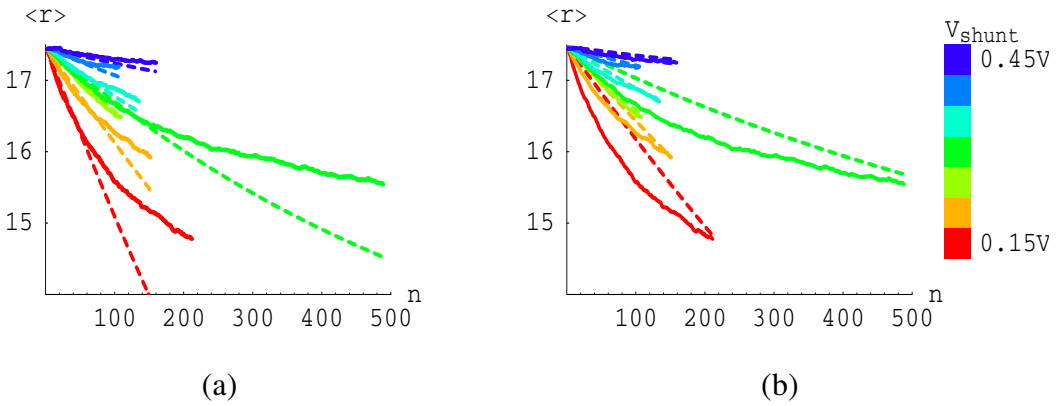


Figure 5.6: Time evolution of $\langle r \rangle$ without bumps

Solid lines: Average growth cone separation $\langle r \rangle$ within the nth sampled distribution. Dashed lines: Average growth cone separation $\langle r \rangle$ after n iterations of the empirical $\mathbf{W}$ on the initial distribution. (a) Empirical $\mathbf{W}$ constructed from first 50 samples only. (b) Empirical $\mathbf{W}$ constructed from all samples.

value of the average separation $\langle r^{(n)} \rangle = \sum_r r P^{(n)}(r)$, as computed from the eigenvector predictions of $P^{(\infty)}(r)$ for the empirical $\mathbf{W}$ (blue) and the model $\mathbf{W}$ (red). The model $\mathbf{W}$ predictions line up reasonably well with the empirical $\mathbf{W}$ predictions for large σ_{NT}, but transition probabilities are harder to estimate when growth cone jump events are infrequent, so at low σ_{NT} the model $\mathbf{W}$ prediction deviates from the empirical $\mathbf{W}$ prediction. Nevertheless, the general trend is that smaller σ_{NT} yield lower predicted $\langle r^{(\infty)} \rangle$ values.

5.2.3 Time evolution

The price of better localization is slower convergence. Figure 5.6 shows the time evolution of $\langle r \rangle$ from its initial value for different choices of V_{shunt}, as measured directly from samples recorded during a refinement experiment and as predicted by iterating an empirical $\mathbf{W}$ on the initial distribution $P^{(0)}(r)$. $\langle r \rangle$ decreases toward its steady-state value more slowly at higher values of V_{shunt}, which are associated with lower predicted $\langle r^{(\infty)} \rangle$ values. Even at low values of V_{shunt}, the recorded $\langle r \rangle$ does not approach its steady-state value

within the sampled timeframe. The sampling interval in these experiments is about 5 minutes, so it is impractical to simply collect samples until the system converges. Instead, we must rely on the eigenvector prediction of the empirical $\mathbf{W}$ to estimate the final $\langle r \rangle$.

The validity of the eigenvector prediction relies on the assumption that $\mathbf{W}$ is stationary. This assumption is necessary both to construct $\mathbf{W}$ from the time series average of distribution samples $P^{(n)}(r)$ and to predict long term behavior from short sample runs. We can evaluate this assumption by comparing $\mathbf{W}$ constructed over different lengths of time. We divide our time series of collected $P^{(n)}(r)$ samples into nonoverlapping training and test subsets. The first 50 samples constitute the training set, which we use exclusively to estimate $\mathbf{W}$ using our standard procedure. We then construct a new time series of predicted $P^{(n)}(r)$ by iterating $\mathbf{W}$ on $P^{(0)}(r)$ and compare this predicted time series with the actual time series.

Figure 5.6(a) shows the time evolution of $P^{(n)}(r)$, as actually recorded (solid lines) and as predicted by the $\mathbf{W}$ assembled from the first 50 samples (dashed lines). The time course predictions track the data in the training set fairly well, but do not generalize well to the test set data, consistently undershooting the remaining samples. The estimated $\mathbf{W}$ tends to overfit the endpoints of its training set, as illustrated in Figure 5.6(b), in which $\mathbf{W}$ is constructed from the entire sample set. However, the final $\langle r \rangle$ predictions for both $\mathbf{W}$ are very similar (data not shown). Thus, although the measured $\mathbf{W}$ depends on our selection of training set samples and is therefore not strictly stationary, it is sufficiently stationary to produce adequate predictions.

5.3 Supervised attraction with bumps

Having characterized internally generated growth cone jumps, we now reinstate externally generated growth cone bumps.

5.3.1 Model transition matrix

To model the effect of the density constraint on growth cone motion, we assume that a nonjumping growth cone gets randomly bumped with uniform probability, regardless of its initial position relative to its synapse or the boundaries. We use a four step procedure to construct a model transition matrix $\mathbf{W}$ for supervised attraction with bumps.

1. **Construct jump matrix $\mathbf{J}^{(T)}$.**

   Growth cone jumps during a time step $T \ll \tau$ are described by the same one step transition matrix we used for supervised attraction without bumps in (5.7).

2. **Construct bump matrix $\mathrm{B}^{(T)}$.**

   A growth cone may also be bumped to an adjacent site via displacement by migrating neighbors. Every bump corresponds to a jump initiated by a mobile neighbor, so the probability that a growth cone is displaced depends on the rate at which its neighbors jump, which in turn depends on their instantaneous separations from their own targets. In principle we could compute a joint probability distribution for the separations of the growth cone and its neighbors, but this would require too much data to be practical. Instead, we will use a constant mean field approximation P_{bump} for the average jump probability within the growth cone population. Any value assumed by $P_{\mathrm{jump}}(r)$ represents a viable choice for P_{bump}; since we are interested in growth cone behavior near their targets, we select $P_{\mathrm{bump}} = P_{\mathrm{jump}}(0)$, which also corresponds to

the highest possible volatility[2].

Assuming all its neighbors are equally active, a growth cone is equally likely to get bumped from any of the incoming lattice edges. Since this is equivalent to uniformly distributed noise, bumps displace the growth cone to adjacent states $r \pm 1$ according to $P_{\text{diff}}(r, r \pm 1)$. We can summarize this bump mechanism in a one-step transition matrix $\mathbf{B}^{(T)}$ whose nonzero elements are

$$\begin{aligned} B^{(T)}_{(r\pm1)r} &= P_{\text{jump}}(0) P_{\text{diff}}(r, r \pm 1) \\ B^{(T)}_{rr} &= 1 - P_{\text{jump}}(0) \end{aligned} \tag{5.8}$$

All other elements $B^{(T)}_{sr}$ in this bump matrix are zero.

3. **Construct motion matrix $\mathbf{W}^{(T)}$.**

   Growth cones motion during one time step is a combination of the jumps generated during that time step and their associated bumps. We merge these two mechanisms into a single transition matrix $\mathbf{W}^{(T)}$ by simply multiplying the jump matrix $\mathbf{J}^{(T)}$ by the bump matrix $\mathbf{B}^{(T)}$.

$$\mathbf{W}^{(T)} = \mathbf{B}^{(T)} \mathbf{J}^{(T)} \tag{5.9}$$

4. **Iterate $\mathbf{W}^{(T)}$.**

   We previously selected the sampling interval τ to permit each of the Z growth cones in the population to initiate at most one jump between consecutive samples. In prin-

[2] A more principled estimate would be the population average jump probability $\sum_r P_{\text{jump}}(r) P^{(n)}(r)$, but this depends on the instantaneous separation distribution $P^{(n)}(r)$ and is therefore nonstationary. Our use of a constant parameter P_{bump} to model an inherently nonstationary property is justified solely by its surprisingly good fit to the data.

ciple, any or all of these jumps could bump the same growth cone, so a growth cone can potentially move several steps between samples: up to one step due to its own jump, and up to $Z - 1$ steps due to bumps from its neighbors. To describe the resulting multistep path, we chop the sampling interval τ into Z identical slices and construct a transition matrix $\mathbf{W}^{(T)}$ for this smaller time step $T = \tau/Z$. We then iterate $\mathbf{W}^{(T)}$ Z times to obtain the transition matrix $\mathbf{W}$ for the full sampling interval τ.

$$\mathbf{W} = \left[\mathbf{W}^{(\frac{T}{Z})}\right]^{Z} \tag{5.10}$$

5.3.2 Parameter fits

Adding bumps does not introduce any new free parameters but does require a new value for the jump rate λ_{NT}, which is selected to match features of the separation change variance $\langle\langle \Delta r \rangle\rangle^2$ and the expected separation change $\langle \Delta r \rangle$. The remaining free parameters retain their previously fitted values.

5.3.2.1 Separation change variance $\langle\langle \Delta r \rangle\rangle^2$

The separation change variance $\langle\langle \Delta r \rangle\rangle^2$ consists of a small peak at σ_{NT} that is generated by the growth cone's own jumps, riding on top of a much larger baseline that is generated by bump activity (Figure 5.7(a)). We choose λ_{NT} to roughly match this constant baseline.

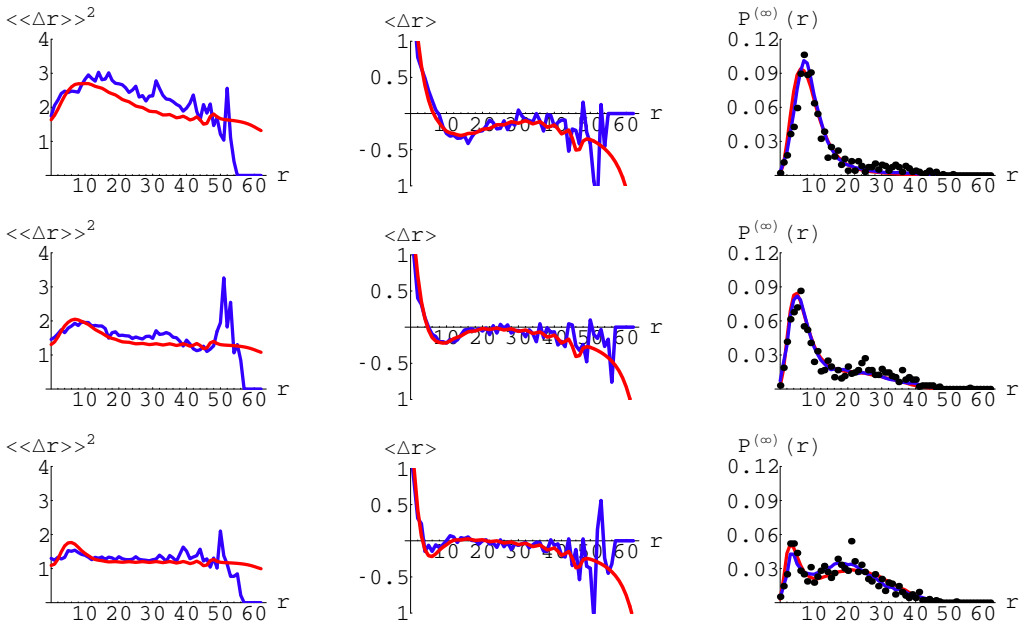


Figure 5.7: Moment fits and predicted $P^{(\infty)}(r)$ with bumps

Moments and predicted stationary distribution for empirical (blue) and model (red) transition matrices $\mathbf{W}$. Top: $V_{\text{shunt}} = 0.20V$; middle: $V_{\text{shunt}} = 0.30V$; bottom: $V_{\text{shunt}} = 0.40V$. (a) Separation change variance $\langle\langle \Delta r \rangle\rangle^2$. (b) Expected separation change $\langle \Delta r \rangle$. (c) Predicted stationary distribution $P^{(\infty)}(r)$ (solid lines) and measured distribution $P^{(100)}(r)$ (black dots).

5.3.2.2 Expected separation change $\langle \Delta r \rangle$

The expected separation change $\langle \Delta r \rangle$ is also magnified by the addition of bumps (Figure 5.7(b)). The positive peak at $r = 0$ and the negative dip at high r are both caused by random growth cone motion, as constrained by the geometry of the radial coordinate and the array boundaries. Directed growth cone jumps reprise the negative trough centered on σ_{NT}, whose depth is used to fine tune λ_{NT}.

5.3.2.3 Stationary distribution $P^{(\infty)}(r)$

The $P^{(\infty)}(r)$ predictions of the empirical and model $\mathbf{W}$ match each other fairly well

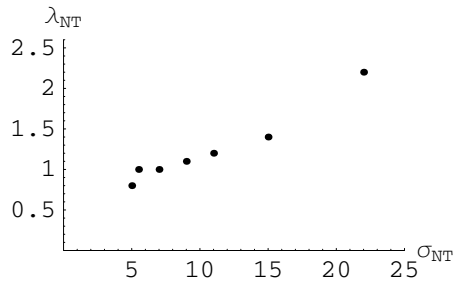


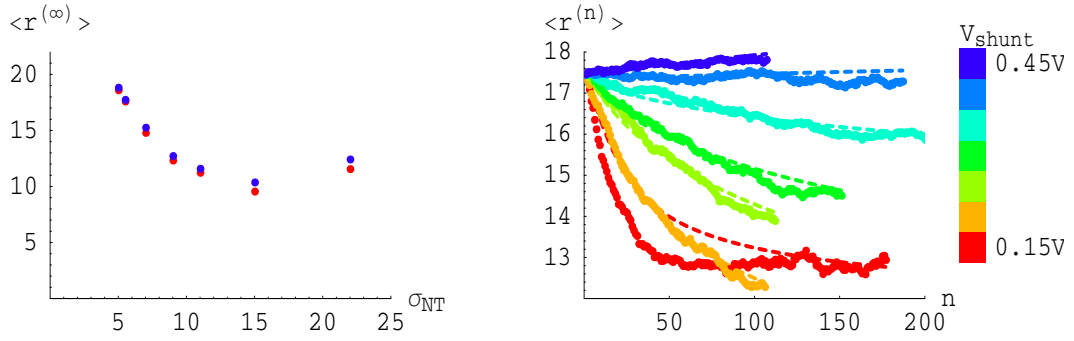
Figure 5.8: λ_{NT} dependence on σ_{NT} with bumps

(Figure 5.7(c)). They also agree with the recorded distribution $P^{(100)}(r)$ of the 100th sample, since bumps significantly enhance growth cone volatility, causing the population distribution to converge toward $P^{(\infty)}(r)$ much more quickly.

The shape of $P^{(\infty)}(r)$ depends on σ_{NT}. At low σ_{NT} $P^{(\infty)}(r)$ is bimodal, with a broad peak at intermediate r that mirrors the distribution expected if growth cones diffused randomly within the target layer, and a sharper peak at low r caused by neurotropic attraction of growth cones to their synapses. Increasing σ_{NT} suppresses the diffusion peak and broadens the attraction peak.

5.3.2.4 Parameter values

Introducing bumps dramatically magnifies the base jump rate λ_{NT} beyond its fitted no-bump value. This increase occurs because a growth cone that would normally be trapped for several sampling intervals at a high jump threshold target site can be bumped prematurely by another growth cone to a lower threshold target site that is easier to escape. This positive feedback endows the fitted values of λ_{NT} with the σ_{NT} dependence plotted in Figure 5.8, which it inherits from the mean field bump probability P_{bump}. Because the neurotropin spreading kernel $G(r)$ is normalized to its peak and not its area, broadening the peak width σ_{NT} increases the total amount of neurotropin released by each postsynaptic spike and



(a) (b)

Figure 5.9: Time evolution of $\langle r \rangle$ with bumps

(a) Equilibrium $\langle r \rangle$ as a function of σ_{NT}, as predicted by empirical $\mathbf{W}$ (blue) and model $\mathbf{W}$ (red). (b) $\langle r \rangle$ evolution recorded from data samples (solid lines) and predicted by empirical $\mathbf{W}$ (dashed lines).

elevates the aggregate neurotropin concentration. Since jumps are generated as a function of the available neurotropin and each jump generates a reciprocal bump, P_{bump} increases with σ_{NT}, pulling λ_{NT} with it.

5.3.3 Time evolution

Bumps elevate growth cone volatility, accelerating convergence at the expense of localization. Figure 5.9(a) plots $\mathbf{W}$ predictions for the final average separation $\langle r^{(\infty)} \rangle$ against σ_{NT}. The predicted $\langle r^{(\infty)} \rangle$ values are all considerably higher than their no-bump counterparts from Figure 5.5(b). The $\langle r^{(\infty)} \rangle$ values plotted in Figure 5.9(a) are the projected endpoints of the time evolutions displayed in Figure 5.9(b), as recorded from chip data (solid lines) and as retrodicted by the measured $\mathbf{W}$ constructed from all available samples (dashed lines) for different values of V_{shunt}. Each $\langle r \rangle$ evolution is faster than its no-bump counterpart from Figure 5.6(b); at $V_{\mathrm{shunt}} = 0.15V$, $\langle r \rangle$ achieves its equilibrium value within 100 samples. The price of faster convergence is poorer performance; at $V_{\mathrm{shunt}} = 0.45V$,

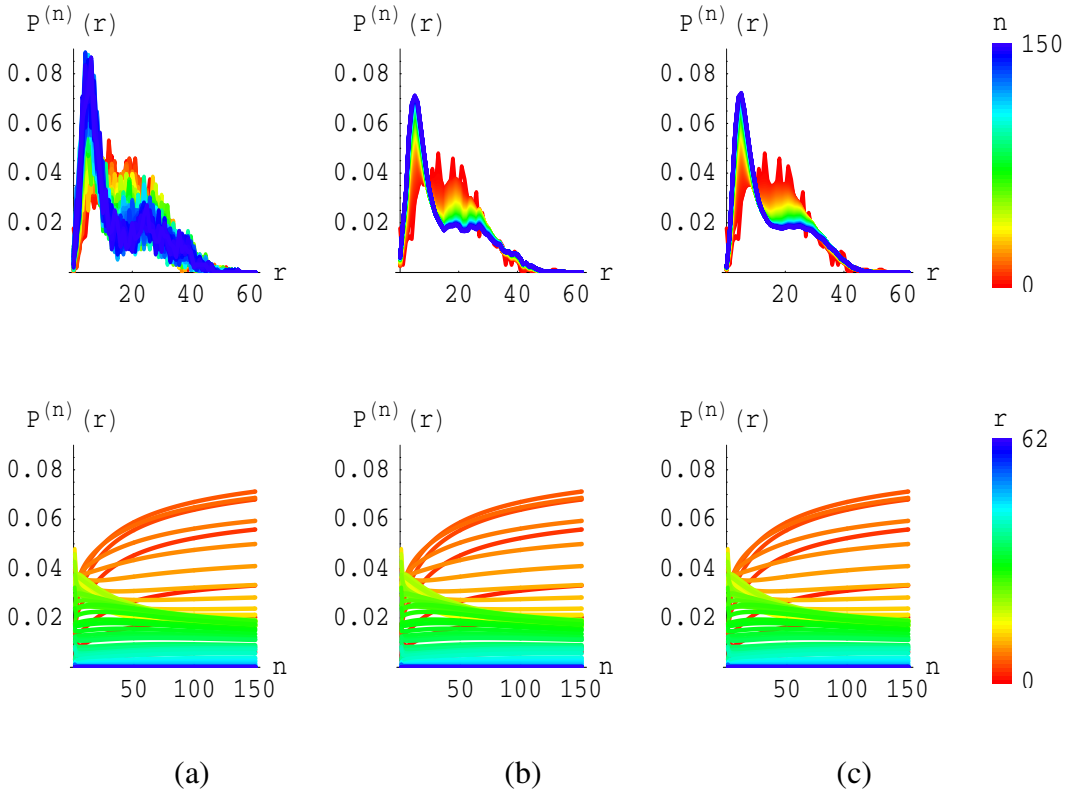


Figure 5.10: Time evolution $P^{(n)}(r)$ with bumps

Time evolution of separation distribution $P^{(n)}(r)$ with bumps at $V_{\text{shunt}} = 0.30V$. Top: $P^{(n)}(r)$ as a function of initial separation r; bottom: $P^{(n)}(r)$ as a function of sample number n. (a) Measured $P^{(n)}(r)$. (b) $P^{(n)}(r)$ predicted from n iterations of empirical $\mathbf{W}$ on measured $P^{(0)}(r)$. (c) $P^{(n)}(r)$ predicted from n iterations of model $\mathbf{W}$ on measured $P^{(0)}(r)$.

$\langle r \rangle$ actually increases from its initial value.

Convergence dynamics within the full population distribution $P^{(n)}(r)$ are shown in Figure 5.10, which plots the evolution of the population distribution $P^{(n)}(r)$ from its initial distribution $P^{(0)}(r)$, as recorded directly from chip data (Figure 5.10(a)) and as predicted by iterating the empirical $\mathbf{W}$ (Figure 5.10(b)) or the model $\mathbf{W}$ (Figure 5.10(c)) on $P^{(0)}(r)$. Over time, smaller values of r become more probable at the expense of intermediate r as the growth cone moves toward its synapse. Neurotropin does not spread beyond intermediate

r, so there is little change in $P^{(n)}(r)$ at large r.

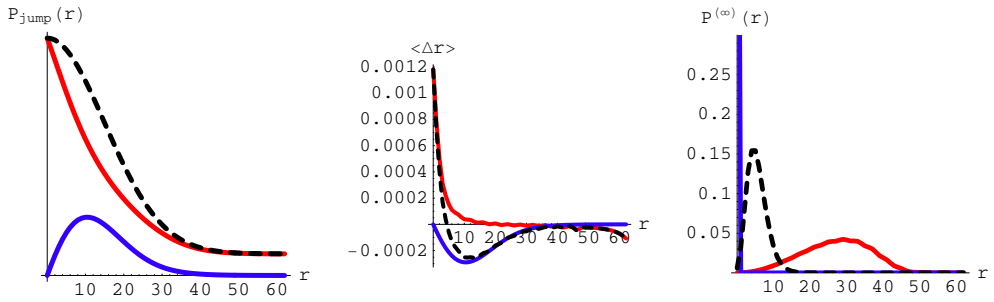
Reassuringly, the three $P^{(n)}(r)$ time series shown in the panels of Figure 5.10 display qualitatively similar features, affirming the validity of the transition matrix approach. The power of this analysis is that it takes the $63N$ data points contained in the N sample time series of $P^{(n)}(r)$, reduces them to the 63×63 elements of the empirical $\mathbf{W}$, and then reduces them further to the 5 free parameters that are required to specify the model $\mathbf{W}$. The fact that these 5 numbers suffice to reproduce the major features of $P^{(n)}(r)$ is a testament to the descriptive capacity of our model.

5.4 Growth cone motion mechanisms

In this section, we take a closer look at how different growth cone motion mechanisms constrain the performance predicted by our model. There are three sources of growth cone motion: directed jumps, undirected jumps, and bumps. Directed and undirected jumps are initiated internally, and bumps are initiated externally. Undirected jumps and bumps both generate random motion; only directed jumps mediate neurotropic guidance. We explain the origin of an optimal spreading range in terms of the behavior of each growth cone motion mechanism.

5.4.1 Directed and undirected jumps

Our model states that without bumps, growth cones move purely on their own initiative, generating jumps at a rate that depends on the local neurotropin concentration $N(r)$. A fraction of these concentration-fueled jumps are directed, meaning that their orientation is deliberately selected to point up the local neurotropin gradient $\nabla_r N(r)$, and the rest are



(a) (b) (c)

Figure 5.11: Directed and undirected motion without bumps

Black dashed: all jumps; blue: directed jumps only; red: undirected jumps only. (a) Probability of a growth cone jump during one time step. (b) Expected separation change $\langle \Delta r \rangle$ during one time step. (c) $P^{(\infty)}(r)$ predicted by model **W**.

undirected, meaning that their orientation is randomly selected. The share of jumps that are directed depends on the local gradient magnitude $\|\nabla_r N(r)\|$.

Figure 5.11(a) decomposes the aggregate jump probability $P_{\text{jump}}(r)$ into its constituent probabilities for directed and undirected jumps. A growth cone initiates a jump within one time step with a probability $P_{\text{jump}}(r)$ (black dashed line) that tracks the neurotropin concentration $N(r)$, which is a Gaussian function of distance r. The number of jumps peaks at $r = 0$ and decays with increasing r. The fraction of jumps that are preferentially directed toward the synapse (blue) follows the gradient magnitude $\|\nabla_r N(r)\|$, which begins at zero at $r = 0$, rises to a peak at an intermediate $r \approx \sigma_{\text{NT}}$, and falls to zero again at large r. The remaining jumps are undirected (red).

Directed and undirected jumps exert opposing influences on net growth cone motion. Figure 5.11(b) plots the expected change in growth cone-synapse separation $\langle \Delta r \rangle$ over a single time step, overlaid with the $\langle \Delta r \rangle$ separately predicted by directed and undirected jumps. Directed jumps all point inward (blue), so their $\langle \Delta r \rangle$ contribution exactly tracks their probability.

An undirected jump is equally likely to carry a growth cone to any of its adjacent target sites, whose distribution of separations relative to the attractor depends on the geometry of the array. If there are $Y(r)$ sites in the array located r edges from a given synapse, the probability that a random motion displaces its growth cone by chance from an initial separation at r to an adjacent site with separation $r \pm 1$ is approximately $\frac{Y(r\pm1)}{Y(r-1)+Y(r+1)}$. Target sites can be sorted into concentric rings of constant separation from any given reference site, and each ring can contain up to $3r$ sites, depending on whether or not it intersects the array boundaries. At low r, $Y(r+1) \gg Y(r-1)$, so undirected jumping is structurally biased to push the growth cone away from its synapse. At intermediate r, $Y(r+1) \approx Y(r-1)$, so random jumps have little net effect. At large r, the array boundaries cut off the number of available sites, so $Y(r+1) \ll Y(r-1)$, creating a net inward bias whose effect is largely negated by the low jump rate at high r. This r dependence is reflected in the $\langle \Delta r \rangle$ for undirected jumps (red), which is strongly positive at low r, mostly flat at intermediate r, and slightly negative at high r.

The aggregate $\langle \Delta r \rangle$ sums contributions from both jump types (black dashed line). The separation dependence of $\langle \Delta r \rangle$ divides the r domain into three bands of values. At low r, random motion dominates, creating a strongly positive $\langle \Delta r \rangle$ that repels the growth cone from its synapse. As r rises, the share of directed jumps increases, creating an intermediate band of r values at which directed jumps dominate, pushing $\langle \Delta r \rangle$ negative. A growth cone occupying this *directed jump band* is strongly attracted to its synapse. Further increases in r reduce both the directed jump share and the aggregate jump rate, so random motion again dominates but jumps are very rare and have little net effect except at very large r where the array boundaries bias random motion to point inward.

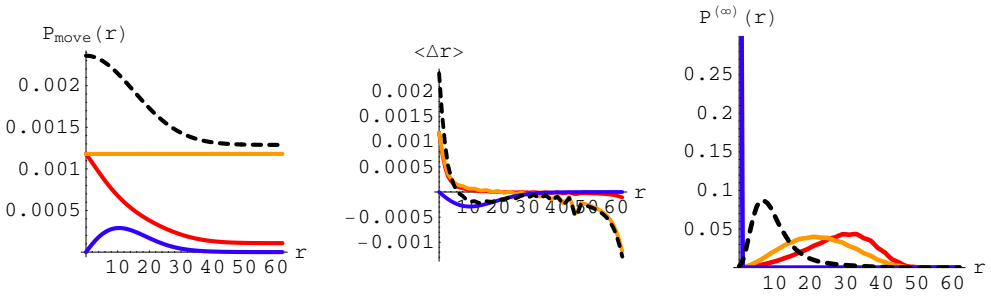
The system has a stable fixed point at the zero crossing of $\langle \Delta r \rangle$, where random outward motion balances directed inward motion. A growth cone that is closer to its synapse is

randomly displaced by undirected jumps, and a growth cone that is further from its synapse is deliberately reeled in by directed jumps. The location of this fixed point anchors the predicted peak of the stationary separation distribution $P^{(\infty)}(r)$, as shown by the black dashed line in Figure 5.11(c). With only directed jumps (blue), the growth cone would move unerringly to its synapse and remain at that site, so all of the probability in $P^{(\infty)}(r)$ would be located at $r = 0$. With only undirected jumps (red), the growth cone would tend to occupy all possible target sites with equal frequency and $P^{(\infty)}(r)$ would be approximately equal to the fraction of sites in the array that are located a distance r from the synapse. The actual $P^{(\infty)}(r)$ is a compromise between these two extremes.

5.4.2 Bumps

Neurotropic attraction is only effective when a growth cone enters spreading range of its synapse. Without bumps, a growth cone must generate its own motion, relying on neurotropin received from its synapse to fuel its jumps. If the growth cone is out of range, it never sees its attractor and rarely moves, stagnant and hopelessly lost.

Adding bumps introduces an external source of random motion that can rescue distant growth cones from their neurotropic poverty. Bumps continually buffet a distant growth cone, increasing its opportunities to explore new sites and eventually stumble into its direction band. Once in range, the growth cone jumps straight for its target and tries to stay there. However, bumps are omnipresent even when the growth cone is inactive and a silent growth cone has no defense against external bumps, so the growth cone can be kicked out of range as easily as it is kicked in. Its only recourse is to resume activity and generate enough directed jumps to recover the ground lost to bumps during periods of inactivity. Thus, like undirected jumps, bumps are helpful at long r and harmful at short r, but unlike



(a) (b) (c)

Figure 5.12: Directed and undirected motion with bumps

Black dashed: all moves; blue: directed jumps only; red: undirected jumps only; orange: bumps only. (a) Probability of a growth cone motion during one time step. (b) Expected separation change $\langle \Delta r \rangle$ during one time step. (c) $P^{(\infty)}(r)$ predicted by model $\mathbf{W}$.

undirected jumps, bumps are uniformly distributed over all r.

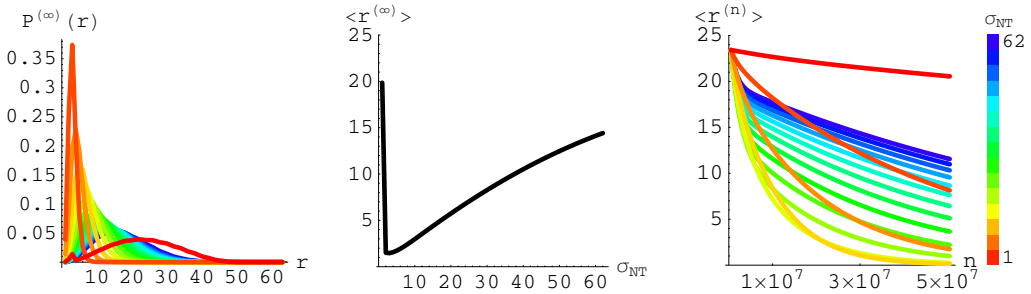
Figure 5.12(a) shows the contributions of each growth cone motion mechanism to the probability $P_{\text{move}}(r)$ that the growth cone moves from its initial position within one time step. As before, the undirected jump probability (red) peaks at $r = 0$ and trails off with increasing r, while the directed jump probability (blue) peaks near $r = \sigma_{\text{NT}}$ and falls to zero at low and high r. A motionless growth cone is bumped by a jumping neighbor with a constant probability (orange).

The omnipresent bump probability exerts a strong influence on the expected separation change $\langle \Delta r \rangle$ (Figure 5.12(b)). Bumps are randomly oriented and behave like undirected jumps, deflecting the growth cone from its synapse at low r, and from the array boundaries at high r. Adding bumps reinforces the effect of undirected jumps, degrading localization performance by pushing the zero crossing of $\langle \Delta r \rangle$ to higher r.

At high r, where jumps are rare and random, bumps accelerate convergence. A growth cone at high r receives scant neurotropic sustenance or guidance from its distant synapse,

and therefore generates undirected jumps at a glacially slow rate. Eventually, one of these jumps propels it by chance into its directed jump band, which reels the growth cone in toward its synapse. The convergence speed of the growth cone population is limited by the time required for distant growth cones to encounter their basins of attraction. Without bumps, $\langle \Delta r \rangle$ (black dashed) would track the negligible undirected jump contribution (red) at high r. Adding bumps increases the rate of random motion, shaking the growth cone loose from its current position and forcing it to guess a new one. This bump noise is detrimental at low r, where the initial position is fairly good and will probably be degraded by random motion. Bump noise is beneficial at high r, where growth cones are stagnant from lack of neurotropin and it is relatively easy to improve on the current position, which is known to be very wrong. In each case, adding bumps forces growth cones to blindly probe their environment at a much faster pace (orange).

The stationary separation distributions $P^{(\infty)}(r)$ predicted by each growth cone motion mechanism are plotted in Figure 5.12(c). Undirected jumps (red) and bumps (orange) are both noise, and push $P^{(\infty)}(r)$ toward the site distribution expected from purely random motion. The $P^{(\infty)}(r)$ peak for undirected jumps favors slightly larger separations than the $P^{(\infty)}(r)$ peak for bumps because undirected jumps are strongly repulsive at low r and only weakly attractive at high r due to the falloff in jump rate, while bumps are present equally at all r. Directed jumps (blue) pull all growth cones to their synapses, creating a delta function peak at $r = 0$. The final $P^{(\infty)}(r)$ (black dashed) peak is intermediate between these noise and signal peaks, and its height is substantially lower than the $P^{(\infty)}(r)$ peak without bumps (Figure 5.11(c)).



(a) (b) (c)

Figure 5.13: σ_{NT} dependence without bumps

(a) Predicted stationary separation distribution $P^{(\infty)}(r)$. (b) Predicted average stationary separation $\langle r^{(\infty)} \rangle$. (c) Predicted evolution of average separation $\langle r^{(n)} \rangle$.

5.4.3 Neurotropin spreading range σ_{NT}

A growth cone moves differently depending on its position relative to its synapse. Jumps are abundant but undirected at low r, plentiful and mostly directed at intermediate $r \approx \sigma_{\mathrm{NT}}$, and rare and undirected at large r. The distances delimiting these three bands of r values determine the convergence properties of the growth cone population, and vary with the neurotropin spreading range σ_{NT}. The addition of bumps speeds convergence but degrades equilibrium performance.

5.4.3.1 Jumps only

Growth cone localization performance, as measured by the predicted stationary separation $\langle r^{(\infty)} \rangle$, is optimized by a small non-zero value of σ_{NT} (Figure 5.13(b)). When σ_{NT} is too small, neurotropin cannot spread far enough from its release site to communicate the activity of a synapse to nearby growth cones, so neurotropic attraction fails. Increasing σ_{NT} broadens the Gaussian $N(r)$ peak, extending the range at which a synapse can elicit jumps from the growth cone, but also pushing both edges of the directed jump band to

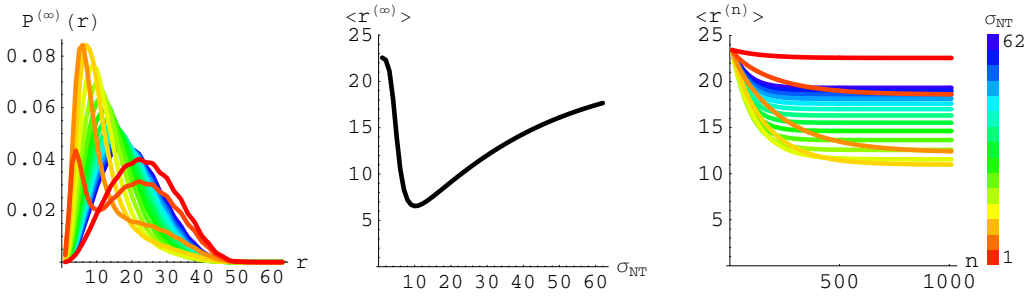
higher r. At low r, this effect is detrimental, since it increases the range at which undirected jumps dominate and pushes the system fixed point to higher r. Increasing σ_{NT} too far thus degrades localization performance.

However, at long distances there are no directed jumps to pull the growth cone toward the synapse, so the growth cone must rely on the infrequent undirected jumps to randomly jostle it into the directed jump band. Increasing σ_{NT} extends the outer edge of the directed jump band to higher r, enabling the synapse to attract more distant growth cones. Higher σ_{NT} also elevate the neurotropin concentration at long distances, generating more undirected jumps that can randomly propel the distant growth cone into its directed jump band. Since the time required for the population to reach steady state is limited primarily by the slow jump rate of its most distant growth cones, increasing σ_{NT} accelerates convergence (Figure 5.13(c)).

5.4.3.2 Jumps and bumps

Although a growth cone can initiate jumps only during its occasional bursts of activity, it may be bumped at any time during the long intervening stretches of inactivity. The correctness of the growth cone's position is therefore continually undermined by random motion, and the growth cone can only repair this damage during its brief periods of activity. A growth cone that generates insufficient directed jumps while it is active will be overwhelmed by random bumps while it is inactive.

At short σ_{NT}, jumps are rare and directed jumps are rarer, so bumps dominate and $P^{(\infty)}(r)$ reverts to the site distribution expected from purely random motion (Figure 5.14(a)). Jump frequency increases with σ_{NT}, creating a bimodal $P^{(\infty)}(r)$, with one peak caused by jumps and the other peak caused by bumps. As the number of directed jumps rises with



(a) (b) (c)

Figure 5.14: σ_{NT} dependence with bumps

(a) Predicted stationary separation distribution $P^{(\infty)}(r)$. (b) Predicted average stationary separation $\langle r^{(\infty)} \rangle$. (c) Predicted evolution of average separation $\langle r^{(n)} \rangle$.

σ_{NT}, the jump peak of $P^{(\infty)}(r)$ suppresses the bump peak. Further increases in σ_{NT} shift and broaden the unimodal $P^{(\infty)}(r)$ peak.

At high σ_{NT}, the population converges very quickly to achieve a fairly poor performance, as measured by $\langle r^{(\infty)} \rangle$. At low σ_{NT}, the population converges very slowly and achieves an even worse performance. The optimal σ_{NT} combines fairly quick convergence with fairly good performance (Figure 5.14(c)). Performance is substantially worse than without bumps, but convergence is much faster in all cases.

5.4.3.3 Optimal σ_{NT}

If we consider directed jumps to be signal and random motion to be noise, reducing σ_{NT} lowers signal strength until it hits the irreducible noise level established by the undirected jump and bump mechanisms. However, reducing σ_{NT} also increases the precision encoded by the signal. This opposition between signal precision and signal strength implies the existence of an optimal σ_{NT} that minimizes the achievable $\langle r^{(\infty)} \rangle$ (Figure 5.14(b)). We will elaborate on the origins of this optimal σ_{NT} as well as the constraints that fundamentally

limit $\langle r^{(\infty)} \rangle$ in the next chapter.

Chapter 6

Optimal Neurotropin Spreading Range

In this chapter, we examine the constraints that generically limit neurotropic growth cone attraction. We first investigate the role of random motion on equilibrium performance, both experimentally and theoretically. We then examine the system parameters that collectively specify population behavior and the resource considerations that constrain their values. We conclude by distilling these insights into a simple attraction basin model that explains the origin of the optimal spreading range. We will find that for a given set of parameter values, the performance of the specified system is dictated by its ability to trap freely diffusing growth cones within their basins of attraction, and subsequently guide captive growth cones to their targets within those basins. Growth cone capture and growth cone guidance have opposite dependences on the basin radius, implying the existence of an optimal signal range that balances these two complementary roles.

6.1 Empirical performance

In this section, we examine the long term behavior of the growth cone population under supervised pair attraction with bumps. Up to this point we have examined the short term dynamics of the growth cone population and the long term behavior they predict. To facilitate parameter extraction on this short time scale, we tuned our system for very conservative behavior, sacrificing convergence speed for equilibrium performance, but under these settings the system converged too slowly to observe empirically, forcing us to extrapolate from the principal eigenvector of the transition matrix $\mathbf{W}$. In this section, we pick more aggressive system settings in order to observe convergence empirically within a reasonable time. Specifically, we reduce the number of jump requests required to trigger a lookup table update, increasing the jump rate but decreasing the directed jump share and increasing the bump rate. We also increase the presynaptic burst length to allow growth cones to jump multiple steps within a single sampling interval, increasing jump rate but increasing bump rate. The net effect is to accelerate convergence but degrade performance.

Figure 6.1 displays population convergence under these new settings for three different spreading ranges σ_{NT}. Figure 6.1(a) shows convergence at a long σ_{NT}, as controlled by V_{shunt}; Figure 6.1(b) shows convergence at the optimal σ_{NT}[1]; and Figure 6.1(c) shows convergence at a short σ_{NT}. The top row of color maps plots the target layer positions of the growth cones, which are colored according to the source layer coordinates of their projecting cell bodies. The bottom row shows the evolution of the average separation $\langle r^{(n)} \rangle$ within the full growth cone population (black solid line). At each sample n, growth cones are sorted into quartiles according to individual separations $r^{(n)}$, and the average separations within each quartile are also plotted (black dots). For reference, population and

[1] Increasing random motion also shifts the optimal value of V_{shunt}, in this case from $0.2V$ to $0.3V$.

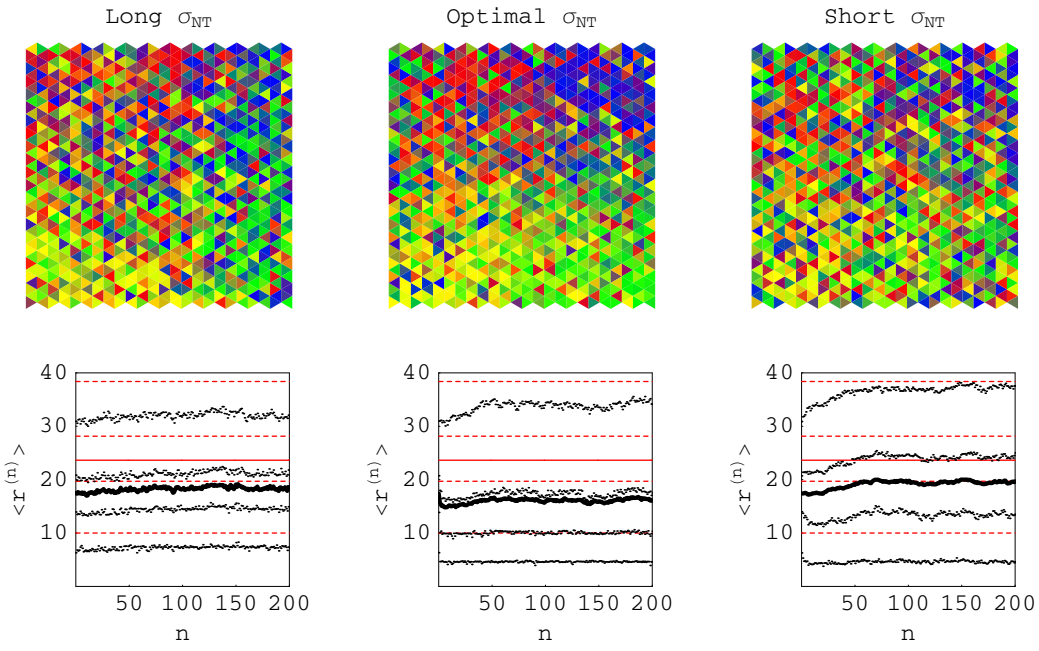


Figure 6.1: Population and quartile $\langle r^{(n)} \rangle$ evolution

Coarse initial topography refines or degrades, depending on V_{shunt}. Top: Topographic color map 200 samples after initialization. Bottom: Evolution of growth cone-synapse separation $\langle r^{(n)} \rangle$, averaged over the full population (black line) or quartile populations (black dots). Red lines denote $\langle r \rangle$ for a completely random separation distribution, averaged over the full population (solid) and quartile populations (dashed). (a) $V_{\text{shunt}} = 0.20V$. (b) $V_{\text{shunt}} = 0.30V$. (c) $V_{\text{shunt}} = 0.40V$.

quartile $\langle r^{(n)} \rangle$ are plotted in red for the stationary distribution expected from purely random motion. If there were no neurotropic attraction, the measured population and quartile $\langle r^{(n)} \rangle$ would converge to these chance levels. Instead, the population $\langle r^{(n)} \rangle$ improves from its initial value at the optimal σ_{NT}, but deteriorates at low and high σ_{NT}. However, even low and high σ_{NT} are able to maintain some non-trivial level of topography, since their population $\langle r^{(n)} \rangle$ stabilizes at a value well below the chance $\langle r^{(n)} \rangle$.

Comparing the long σ_{NT} (Figure 6.1(a)) with the optimal σ_{NT} (Figure 6.1(b)), we observe that while the lowest quartile performs worse for long σ_{NT}, the highest quartile per-

forms better. This is consistent with the idea that long σ_{NT} are able to rope distant growth cones in from further away, improving the performance of the highest quartile, but less able to hold growth cones at their synapses, degrading the performance of the lowest quartile. Similarly, comparing the optimal σ_{NT} (Figure 6.1(b)) with the short σ_{NT} (Figure 6.1(c)), we observe that the lowest quartile performs equally well in both cases, but the highest quartile of the short σ_{NT} unravels almost to chance performance. Again, this is consistent with the idea that short σ_{NT} are able to hold growth cones at their synapses, but unable to influence distant growth cones, which wander freely.

6.1.1 Equilibrium convergence

The notion that long σ_{NT} attract distant growth cones and short σ_{NT} confine close growth cones immediately suggests a simulated annealing scheme in which we initialize the system with a large σ_{NT} in order to catch distant growth cones, and gradually reduce σ_{NT} in order to tie growth cones closer and closer to their synapses. Unfortunately, while such a scheme might speed convergence, the irreducible and omnipresent undirected jumps and bumps prevent it from improving equilibrium performance.

We can demonstrate this failure by comparing population evolution at short σ_{NT} from two different initial conditions: coarse topography and perfect topography. Perfect initial topography starts every growth cone at zero separation from its synapse, automatically placing it within its own attraction basin. If the failing of the short σ_{NT} is that initially distant growth cones never wander into range of a strong short range synaptic attraction and remain lost at high r, trapping the system in a local energetic minimum, performance should improve if all growth cones occupy their global optima to begin with.

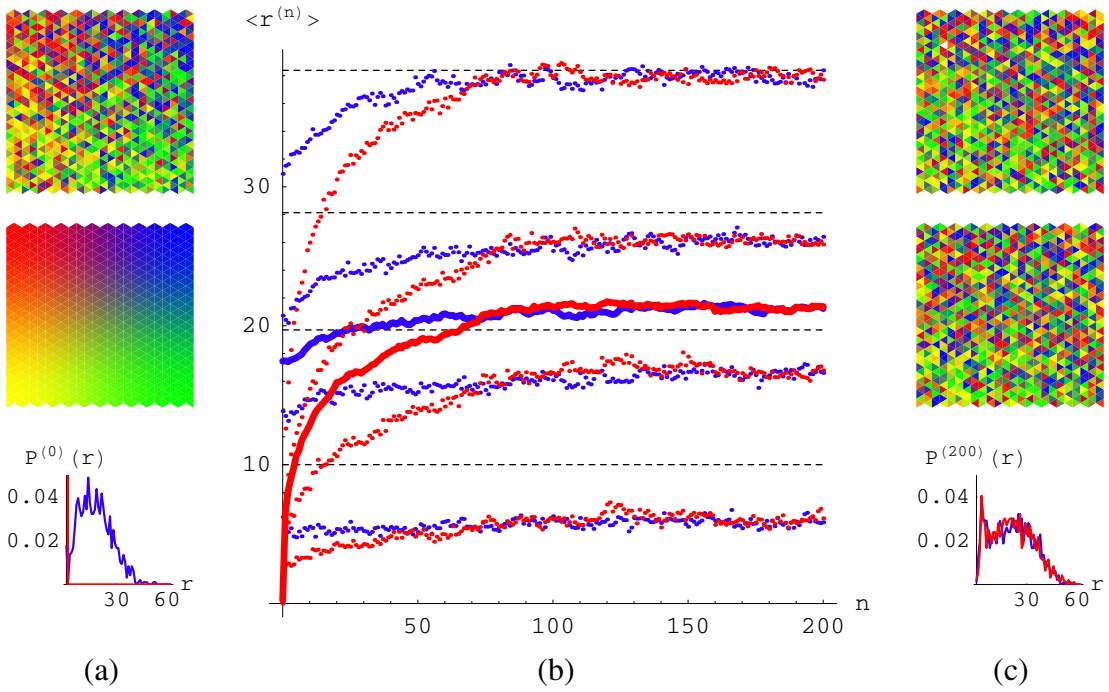


Figure 6.2: Evolution from coarse and perfect topography

Population evolution from coarse (blue) or perfect (red) initial topography for $V_{\text{shunt}} = 0.45V$. (a) Population state at initialization. Top, middle: Color maps of coarse (top) and perfect (middle) initial topographic projections. Bottom: Initial separation distribution $P^{(0)}(r)$. For perfect initial topography, $P^{(0)}(0) = 1$ off chart. (b) Mean separation $\langle r^{(n)} \rangle$ (solid line) and quartile (dots) evolutions from initially coarse or perfect topography. Black dashed lines denote quartile breaks for a completely random separation distribution. (c) Population state at the 200th sample. Top, middle: Color maps at the 200th sample for coarse (top) and perfect (middle) initial topography. Bottom: Measured separation distribution $P^{(200)}(r)$.

Figure 6.2(a) shows the color maps and separation distributions $P^{(0)}(r)$ for coarse (blue) and perfect (red) initial topography. Population and quartile $\langle r^{(n)} \rangle$ evolutions from both initializations are overlaid in Figure 6.2(b), and the color maps and measured separation distributions $P^{(200)}(r)$ at the 200th sample are shown in Figure 6.2(c). Despite initialization at zero separation, perfect topography relaxes to the same degraded separations as coarse topography for all quartiles, indicating that short σ_{NT} does not suffice to bind growth cones to their synapses. More generally, localization performance is intrinsic

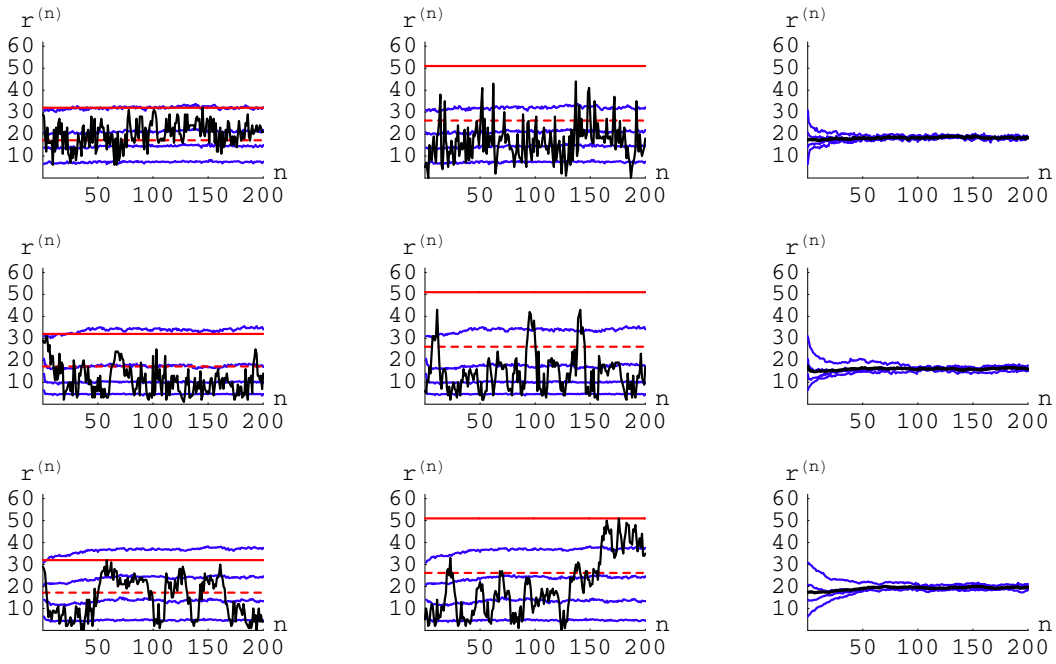


Figure 6.3: $r^{(n)}$ trajectories

Top row: $V_{\text{shunt}} = 0.20V$; middle row: $V_{\text{shunt}} = 0.30V$; bottom row: $V_{\text{shunt}} = 0.40V$. (a) Instantaneous separation $r^{(n)}$ (black) for synapse $(24, 20)$, located at the array center. Red lines indicate maximum possible separation (solid) and average random separation (dashed) from this synaptic site. Blue lines plot mean separations within instantaneous population quartiles. (b) Instantaneous separation $r^{(n)}$ for synapse $(24, 40)$, located at the array edge. (c) Separation evolution for initial quartile groups (blue) and full population (black).

to the specific value of σ_{NT}, and independent of the initial distribution.

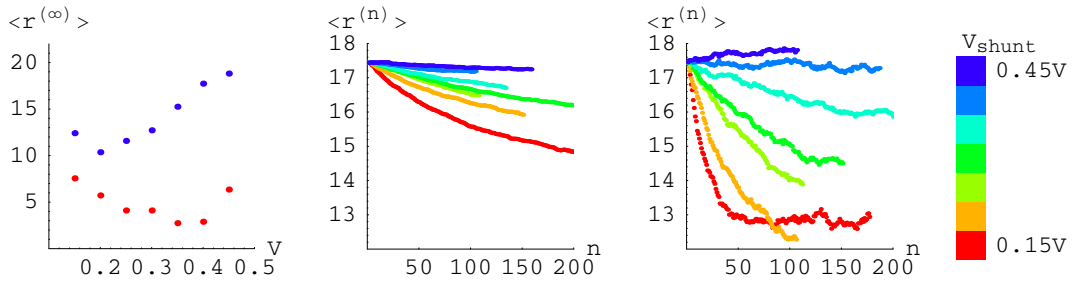
Figure 6.3(a) plots the instantaneous separation $r^{(n)}$ at different σ_{NT} for a synapse located at the center of the array, equidistant from the boundaries (black). The maximum possible error is achieved when the growth cone is located at the target site most distant from the synapse and is plotted for reference (red solid line). For this centrally located synapse, the most distant target site is located in one of the four array corners, and the maximum error is roughly half the length of the array diagonal. The separation expected

by chance for a growth cone moving randomly in the array is also determined by synapse location. A randomly moving growth cone would be expected to visit every target site in the array with equal frequency, so the chance error is the average distance between the synapse and all possible target sites (red dashed line).

At long σ_{NT} (top), the synapse error fluctuates with high frequency around the chance error, consistent with fast undirected jump generation. At the optimal σ_{NT} (middle), the synapse error spends most of its time below chance, consistent with directed attraction. At low σ_{NT} (bottom), the synapse error tends to plateau at all levels between its maximum and minimum, moving between plateaus in sharp, infrequent hops. The low amplitude jitter on these plateaus is consistent with jump generation, and the sharp hop between plateaus is consistent with a long string of consecutive bumps that propel the growth cone a great distance before it has time to react. A synapse located on the edge of the array exhibits similar behavior for a correspondingly higher maximum and chance error, as shown in Figure 6.3(b). Instantaneous quartile breaks are plotted in blue. Quartile membership is highly fluid, as individual synapse separations freely cross quartile breaks over the course of the evolution.

In Figure 6.3(c) we classify each growth cone according to its quartile in the initial map and track the $\langle r^{(n)} \rangle$ for each group over the course of the evolution. If growth cone attraction were strictly monotonic at some short range, we would expect the lowest initial quartile group error to remain distinct, since growth cones within that group are more likely to have been captured by their synapses. Instead, all four groups converge to the same global error as growth cones shuffle between instantaneous quartiles and equalize the mix of separations within each initial quartile group.

The high turnover of quartile membership indicates that growth cones do not remain



(a) (b) (c)

Figure 6.4: $\langle r^{(n)} \rangle$ convergence with and without bumps

Convergence of average separation $\langle r \rangle$ at different values of V_{shunt} with and without bumps. (a) Equilibrium mean separation $\langle r^{(\infty)} \rangle$ predicted by empirical $\mathbf{W}$ with bumps (red) and without bumps (blue), as a function of V_{shunt}. (b) Time evolution of mean separation $\langle r^{(n)} \rangle$ recorded from chip data without bumps. (c) Time evolution of $\langle r^{(n)} \rangle$ recorded from chip data with bumps.

tethered to their synapses once attracted. This loss of confinement is associated with intrinsic bump noise, and prevents simulated annealing schemes from improving performance.

6.1.2 Bumps

To improve performance we need to reduce the random motion injected by bumps and undirected jumps. One way to do this is to eliminate bumps entirely, as we did in the previous chapter. Reverting to the original system settings, Figure 6.4(a) plots the empirical $\mathbf{W}$ prediction for the equilibrium separation $\langle r^{(\infty)} \rangle$, our proxy for performance, against the decay transistor gate bias V_{shunt}, our proxy for spreading range σ_{NT}. $\langle r^{(\infty)} \rangle$ is minimized by some intermediate V_{shunt}, which corresponds to an *optimal spreading range*. Adding bumps shifts this optimum to lower V_{shunt}, and elevates $\langle r^{(\infty)} \rangle$ at all values of V_{shunt}, quantitatively degrading the sustainable level of retinotopy. At large V_{shunt}, $\langle r^{(\infty)} \rangle$ actually exceeds the initial value $\langle r^{(0)} \rangle$, showing that bumps can unravel even the coarse initial retinotopy.

Bumps also affect convergence speed. Without bumps, growth cones can only move by

generating their own jumps, an excruciatingly slow process at large values of V_{shunt} (Figure 6.4(b)). Adding bumps considerably hastens convergence to equilibrium (Figure 6.4(c)).

6.1.3 Gradient measurement averaging

Another way to reduce random motion is to increase the directed jump share. Every jump automatically generates a complementary bump that disrupts the guidance of other growth cones. A growth cone jump is only worthwhile if the loss incurred by the bumped growth cone is exceeded by the gain accrued by the jumping growth cone. By this measure, each directed jump is guaranteed to at least break even, since it reduces the separation of its own growth cone by exactly one edge and increases the separation of the bumped growth cone by at most one edge. On the other hand, an undirected jump can potentially damage itself as well as its neighbor. Better performance therefore requires an higher proportion of directed jumps, which we can obtain by averaging multiple jump requests to increase the precision of the gradient measurement.

The Neurotrope1 chip generates jump requests by transmitting a signal every time a growth cone circuit accumulates a threshold of sampled neurotropin. Since a presynaptically active growth cone automatically binds neurotropin even in the absence of a detectable gradient, many of the requested jump directions will be random, selected by stochastic concentration fluctuations at the filopodia. These undirected concentration-driven jump requests constitute noise in the gradient measurement. The obvious way to reduce this noise is to average over many gradient measurements by requiring the microcontroller to accumulate multiple jump requests from the same growth cone before actually updating the relevant lookup table entries. This program tweak improves measurement precision at the cost of update speed and the additional memory required to store the jump request accu-

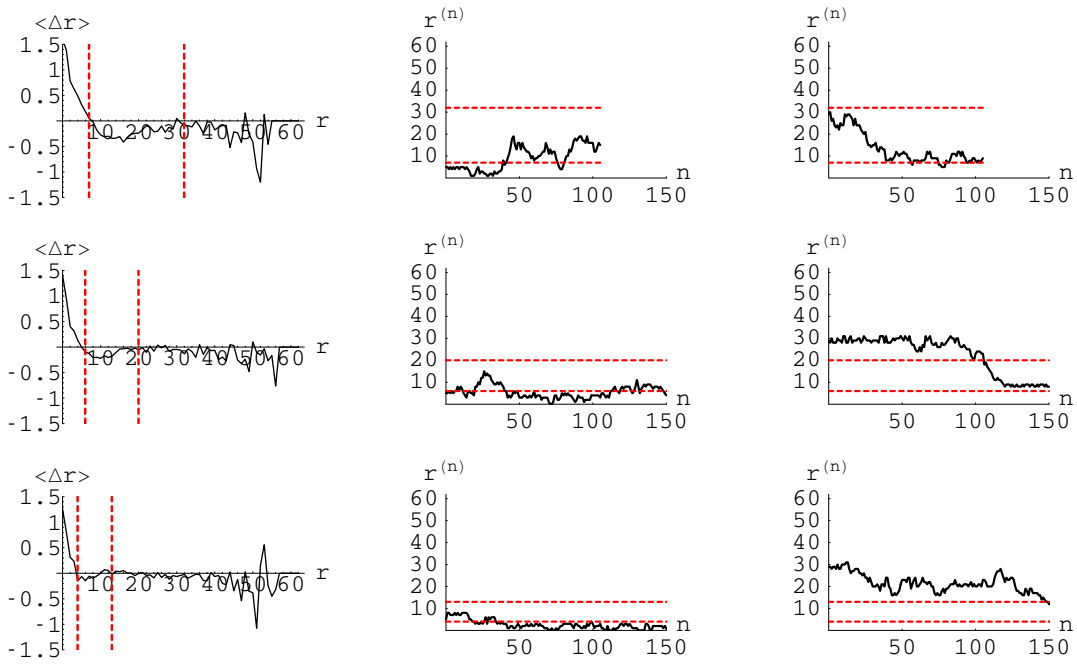


Figure 6.5: $r^{(n)}$ trajectories with averaged gradient measurements

Top: $V_{\text{shunt}} = 0.20V$; middle: $V_{\text{shunt}} = 0.30V$; bottom: $V_{\text{shunt}} = 0.40V$. (a) Expected separation change $\langle \Delta r \rangle$ predicted by empirical $\mathbf{W}$. Vertical dashed lines delimit the region dominated by directed jumps. (b) Measured $r^{(n)}$ trajectory for a growth cone initially located at low r. Horizontal dashed lines delimit the directed jump region. (c) Measured $r^{(n)}$ trajectory for a growth cone initially located at high r.

mulators. From a resource perspective, the memory requirement quickly becomes onerous, since each bit of extra measurement precision requires an additional 3 bits of memory to be allocated to every growth cone in the population, one bit per filopodium.

Increasing measurement precision causes a drastic decline in the total number of jumps, since it takes much longer for a growth cone to accumulate the multiple thresholds of neurotropin required to initiate a jump. Since the lost jumps encoded a much noisier gradient measurement, most of them would have been undirected and their elimination increases the share of directed jumps in the surviving pool of jumps. Random motion is doubly diluted

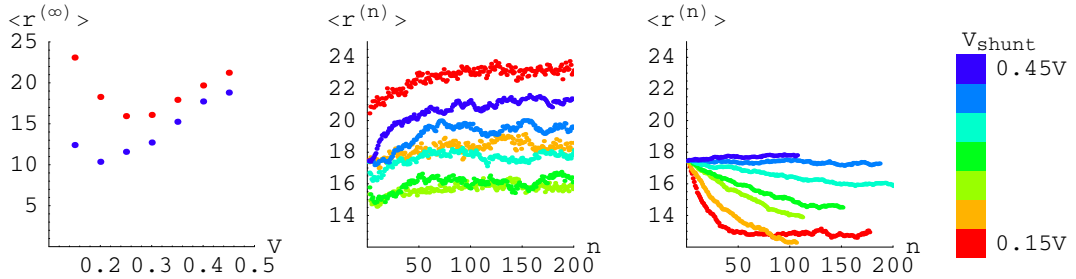
because the bumps associated with the lost jumps are also eliminated.

Averaging gradient measurements constrains growth cones to move more deliberately, as shown in Figure 6.5[2]. Figure 6.5(a) plots the expected separation change $\langle \Delta r \rangle$ for three different values of V_{shunt}. Vertical red dashed lines demarcate the directed jump band, whose lower bound is located at the zero crossing of $\langle \Delta r \rangle$ to the left of the attractive trough and whose upper bound is located at the highest value of $\langle \Delta r \rangle$ to the right of the trough. Growth cones to the left of the directed jump band bounce rapidly around the system fixed point at the zero crossing of $\langle \Delta r \rangle$. Growth cones within the directed jump band jump toward their synapses. Growth cones to the right of the directed jump band diffuse randomly until they fall into the directed jump band. The directed jump band cuts off at low r because the diminishing gradient magnitude $|\nabla_r N(r)|$ reduces the share of jumps that are directed, and at high r because the diminishing concentration $N(r)$ reduces both the directed jump share and the total number of jumps. Increasing σ_{NT} broadens the directed jump band and shifts it toward higher r.

Sample $r^{(n)}$ trajectories are plotted in Figure 6.5(b) and (c), in which horizontal dashed lines denote the directed jump band. Note the complete absence of the plateau-hopping behavior that characterized the short σ_{NT} trajectories in Figure 6.3. The reduction in jump rate eliminates the long bump chains that can push a growth cone away from its synapse before it can react. The high frequency jitter associated with undirected jump generation at long σ_{NT} has also vanished. Under these conditions, a growth cone's motion is dominated not by bump noise, but by its position relative to its directed jump band.

A growth cone starting from an initial separation $r^{(0)}$ located below the directed jump

[2]Without averaging, we accumulate 7 jump requests in each vertical direction or 2 in the horizontal direction before executing the requested update, in order to smooth out the effects of the layout anisotropy described in Chapter 3. With averaging, we raise the execution threshold to 150 jump requests in a vertical direction and 3 in the horizontal direction.



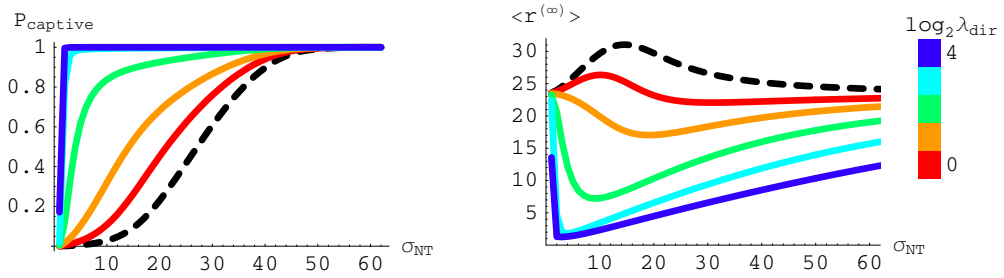
(a) (b) (c)

Figure 6.6: $\langle r^{(n)} \rangle$ convergence with and without gradient measurement averaging

Convergence of average separation $\langle r \rangle$ at different values of V_{shunt} with and without averaging. (a) Stationary mean separation $\langle r^{(\infty)} \rangle$ predicted by empirical $\mathbf{W}$ with averaging (blue) and without averaging (red), as a function of V_{shunt}. (b) Time evolution of mean separation $\langle r^{(n)} \rangle$, as recorded from chip data without averaging. (c) Time evolution of $\langle r^{(n)} \rangle$, as recorded from chip data with averaging.

band is randomly shuffled inward and outward by undirected jumps and bumps (Figure 6.5(b)). Random outward motion is bounded by the lower edge of the directed jump band, which herds wayward growth cones back toward their synapses. Raising σ_{NT} moves this fence further out, allowing the growth cone to roam further from its synapse and thereby degrading localization performance. A growth cone starting from an $r^{(0)}$ located above the directed jump band is randomly jostled by bumps that eventually nudge it by chance into its direction band to slide toward its synapse (Figure 6.5(c)). Raising σ_{NT} extends the grasp of the directed jump band, increasing the chances for the growth cone to topple into range and thereby speeding convergence.

Generalizing to the full growth cone population, averaging improves equilibrium performance, slows convergence, and lengthens the optimal σ_{NT} (Figure 6.6). Averaging effectively increases the sensitivity with which a growth cone can measure a gradient. More sensitive growth cones can perceive an attractor from further away, lengthening the optimal σ_{NT}. In the next section we will systematically vary the sensitivity λ_{dir} in the model



(a) (b)

Figure 6.7: Model $\mathbf{W}$ dependence on gradient strength λ_{dir} and range σ_{NT}

Influence of gradient strength λ_{dir} and range σ_{NT} on equilibrium growth cone guidance, as predicted by the model $\mathbf{W}$. (a) Equilibrium probability $P_{\text{captive}} = P(r^{(\infty)} \leq \sigma_{\text{NT}})$ that a growth cone is found within its attraction basin, as a function of spreading range σ_{NT} and parameterized by λ_{dir}. Black dashed line indicates $\lambda_{\text{dir}} = 0$. (b) Average equilibrium separation $\langle r^{(\infty)} \rangle$.

transition matrix $\mathbf{W}$ and observe its effect on predicted equilibrium performance.

6.2 Model transition matrix

In the previous section, we observed empirically that the predicted equilibrium performance $\langle r^{(\infty)} \rangle$ depends on the neurotropin spreading range σ_{NT}, and that this dependence can be shifted by manipulating various parameters. In this section, we observe a similar σ_{NT} dependence in our model transition matrix $\mathbf{W}$. We vary the growth cone sensitivity λ_{dir}, the shape of the neurotropin spreading kernel $N(r)$, and the concentration dependence of the jump probability $P_{\text{jump}}(r)$ and observe the consequences predicted by the model $\mathbf{W}$. We will find that the optimal σ_{NT} shifts with λ_{dir}, but that the predicted equilibrium performance is surprisingly insensitive to perturbations in $N(r)$ and $P_{\text{jump}}(r)$.

6.2.1 Growth cone sensitivity

In the model transition matrix $\mathbf{W}$ from the previous chapter, the probability that any

given jump is explicitly directed is a function of $\lambda_{\mathrm{dir}}\|\nabla_r N(r)\|$, where the attraction bias λ_{dir} is a free parameter that scales the measured gradient magnitude. Larger values of λ_{dir} increase the growth cone's response to smaller values of $\|\nabla_r N(r)\|$, so we can think of λ_{dir} as the growth cone's sensitivity to the local neurotropin gradient.

Increasing λ_{dir} monotonically improves growth cone guidance, as shown in Figure 6.7. Figure 6.7(a) plots the equilibrium probability P_{captive} that a growth cone is found within a circular attraction basin of radius σ_{NT} centered on its target. Random motion allows growth cones to shuttle in and out of this basin with some probability that depends on λ_{dir}. Weaker attraction biases λ_{dir} loosen a basin's hold on captured growth cones, lowering P_{captive} for all basin sizes σ_{NT}. Outside the attraction basin, growth cones move randomly, if at all. Since a diffusing growth cone is less likely to encounter smaller attraction basins, P_{captive} decreases with σ_{NT}.

The equilibrium separation $\langle r^{(\infty)}\rangle$ predicted by $\mathbf{W}$ depends on an attraction basin's ability to capture diffusing growth cones and on its ability to guide captive growth cones. Inside the attraction basin, growth cone localization depends on the local gradient magnitude, which depends inversely on σ_{NT}. σ_{NT} parameterizes the distance required for the neurotropin concentration to fall a set level to a detection threshold from its peak at a release site, so increasing σ_{NT} weakens the gradient of a monotonically decreasing function. Growth cones have a harder time finding smaller basins, but are better guided once found.

The optimal σ_{NT} balances growth cone capture with growth cone guidance, as shown in Figure 6.7(b). $\langle r^{(\infty)}\rangle$ rises at low σ_{NT}, consistent with the idea that diffusing growth cones have trouble finding small basins, and at high σ_{NT}, consistent with the idea that large spreading ranges yield weaker gradients. The minimum value of $\langle r^{(\infty)}\rangle$ for a given attraction bias λ_{dir} is located at some intermediate σ_{NT}. Increasing λ_{dir} pushes this optimal

σ_{NT} closer to 0, as if raising the potential barrier that blocks a growth cone from escaping from a basin, and lowers the $\langle r^{(\infty)} \rangle$ achievable at high σ_{NT}, as if increasing the salience of weaker gradients.

Interestingly, when $\lambda_{\mathrm{dir}} = 0$, corresponding to no growth cone attraction, $\langle r^{(\infty)} \rangle$ actually rises to a maximum at an intermediate σ_{NT} (black dashed line in Figure 6.7(b)). This is a consequence of the concentration dependence of the jump mechanism.

6.2.2 Concentration dependence

In our conception of neurotropic axon guidance, a growth cone moves by preferentially sprouting new filopodia in one direction and retracting old filopodia from the opposite direction. Filopodial sprouting is promoted by the presence of neurotropin, biasing a growth cone to climb the local neurotropin gradient by sprouting more filopodia in the direction of increasing neurotropin. Since growth cone size is fixed, each growth cone must construct its filopodia from a conserved pool of cytoskeletal resources. New filopodia cannibalize existing filopodia for structural components, so any increase in the sprouting rate is exactly matched by an increase in the retraction rate and the overall effect of neurotropin is to accelerate filopodial turnover. Since filopodial longevity depends inversely on the local neurotropin concentration, more neurotropin makes a growth cone more mobile.

Concentration-driven mobility obeys the risk-averse philosophy that a growth cone should never move unless directed. Every jump is presumed to explicitly increase the fitness of the jumping growth cone, so a growth cone should remain immobile in the absence of neurotropin in order to avoid making active missteps. Under this philosophy, a growth cone only generates a jump when it has some idea where to go and otherwise remains silent, an appealing property for a bandwidth-limited implementation like the Neurotrope1 system

in which growth cone jump requests must compete for access to a shared output link.

The contrary philosophy – a growth cone should always move unless directed to remain still – is represented by concentration-damped mobility. One criticism of concentration-driven growth cone mobility is that excessive neurotropin uptake renders a growth cone hyperactive, causing it to fidget compulsively atop the neurotropin peaks where it should be the most content. A stabler growth cone motion mechanism would reverse the concentration dependence of filopodial longevity, so that filopodia turn over at a high rate in the absence of neurotropin and are stabilized by its presence. This concentration-damped mobility would allow a growth cone to spend more time in the neurotropin-rich region surrounding its attractor but pass swiftly through neurotropin-starved regions that are self-evidently the wrong places to be.

From an implementation perspective, concentration-damped mobility offers a more elegant allocation of system bandwidth. Concentration-driven mobility permits the system bandwidth to be monopolized by the growth cones least in need of attention, those that have already found their targets. Once a growth cone has found its target, further concentration-driven jump requests cannot improve the growth cone's position, but merely affirm its continuing correctness. Nevertheless, correctly placed growth cones continue to generate jump requests at a high rate, so that as more and more growth cones settle into their targets, a greater share of the system bandwidth must be spent to service their self-affirming jump requests, crowding out more critical jump requests from less voluble growth cones marooned in low neurotropin regions. By contrast, concentration-damped mobility preferentially assigns bandwidth to the growth cones most in need of attention, those that are furthest from their targets. A wildly misplaced growth cone generates many random jump requests in an effort to escape its neurotropin-starved location, but this jump rate tapers off as the growth cone ascends a neurotropin peak, automatically annealing its bandwidth consumption and

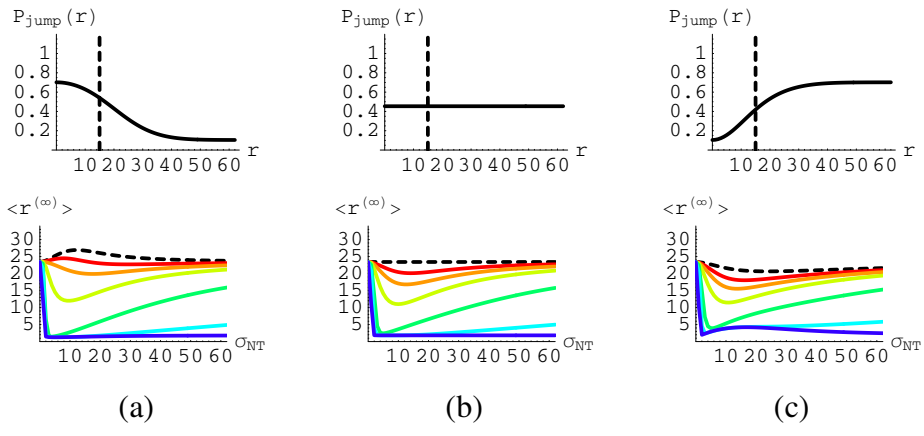


Figure 6.8: Concentration-dependent growth cone mobility

Equilibrium performance predicted by model **W** using different growth cone mobility functions $P_{\text{jump}}(r)$. Top: Jump probability $P_{\text{jump}}(r)$. Dashed line indicates σ_{NT}. Bottom: Average equilibrium separation $\langle r^{(\infty)} \rangle$. Colors denote different values of direction bias λ_{dir}. Black dashed line corresponds to $\lambda_{\text{dir}} = 0$; blue line corresponds to $\lambda_{\text{dir}} = 10^6$. (a) Concentration-driven mobility. (b) Constant mobility. (c) Concentration-damped mobility.

freeing the system to attend to direr placement errors.

6.2.2.1 Model predictions

We can compare the predicted equilibrium performances of concentration-driven and concentration-damped mobilities by modifying our expression for the jump probability $P_{\text{jump}}(r)$ in the model **W**. For concentration-driven mobility,

$$P_{\text{jump}}(r) \quad = \quad 1 - \exp\left[-\lambda_{\text{NT}} N(r)\right] \tag{6.1}$$

while for concentration-damped mobility,

$$P_{\text{jump}}(r) \quad = \quad 1 - \exp\left[-\lambda_{\text{NT}} (N_0 - N(r))\right] \tag{6.2}$$

where N_0 is the peak neurotropin concentration. The average equilibrium separation $\langle r^{(\infty)} \rangle$

predicted by the model $\mathbf{W}$ for each mobility function is plotted in Figure 6.8 for different growth cone guidance strengths, as parameterized by the jump direction bias λ_{dir}.

The model predicts that a concentration-damped mobility will perform better when growth cone guidance is weak, but a concentration-driven mobility will perform better when growth cone guidance is strong. When λ_{dir} is small, a growth cone diffuses almost randomly, spending more time in regions where it has lower mobility. A concentration-damped mobility allows the growth cone to spend more time in the high neurotropin regions surrounding its target, while a concentration-driven mobility strands the growth cone in low neurotropin regions far away from its target. When λ_{dir} is large, a growth cone moves straight for its target at a speed that depends on its mobility. A concentration-driven mobility allows the growth cone to capitalize on increasingly strong guidance signals by accelerating the growth cone up the local gradient, while a concentration-damped mobility brakes the growth cone as it nears its target, reducing the rate at which it can jump toward its target but also reducing the rate at which it can repair damage incurred through displacement by other migrating growth cones.

In short, the quality of the growth cone guidance mechanism determines which concentration dependence is best. Concentration-driven mobility rewards the successes of a strong guidance mechanism, while concentration-damped mobility punishes the failures of a weak guidance mechanism.

6.2.3 Gradient shape

One obvious criticism of a Gaussian neurotropin spreading kernel is that its gradient peaks at $r = \sigma_{\mathrm{NT}}$ but then falls to zero at a neurotropin release source, precisely the point at which the growth cone generates the most concentration-driven jump requests. In the

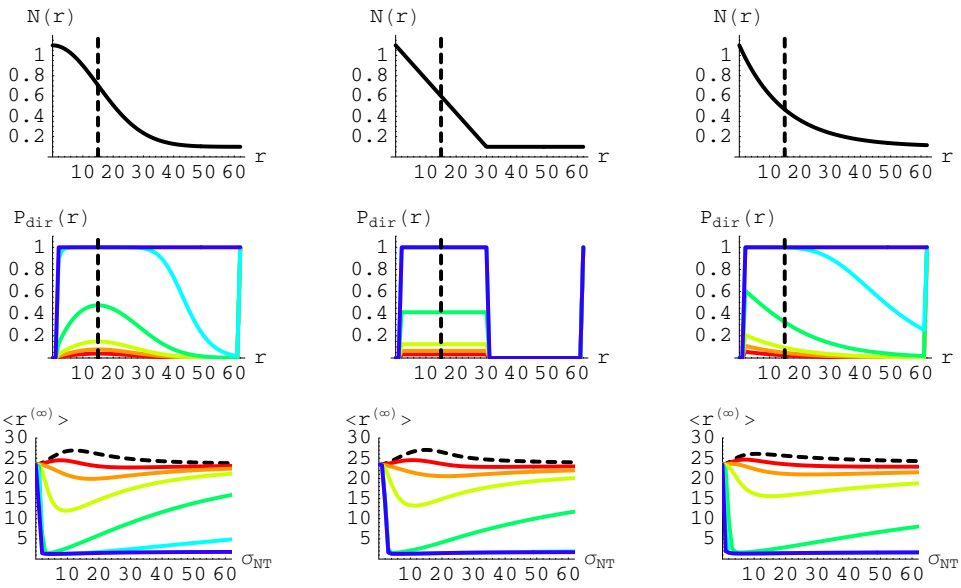


Figure 6.9: Neurotropin spreading kernel shape

Top: Neurotropin spreading kernel $N(r)$. Dashed line denotes σ_{NT}. Middle: Directed jump probability $P_{\mathrm{dir}}(r)$. Dashed line denotes σ_{NT}. Bottom: Average equilibrium separation $\langle r^{(\infty)} \rangle$. Colors indicate different values of λ_{dir}. Dashed line indicates $\lambda_{\mathrm{dir}} = 0$. (a) Gaussian spreading kernel. (b) Triangular spreading kernel. (c) Exponential spreading kernel.

absence of a discernable guidance signal, all of these jumps are randomly oriented, causing a correctly positioned growth cone to jitter energetically around its mark. This hyperactive fidgeting eventually topples the growth cone from its perch, to be swept away by bumps and other noise sources.

We might expect performance to improve if we replaced the Gaussian neurotropin spreading kernel (Figure 6.9(a)) with a function whose gradient is high at $r = 0$, such as a triangular kernel (Figure 6.9(b))

$$N(r) = \begin{cases} 1 - \frac{r}{2\sigma_{\mathrm{NT}}} & \text{if } r \leq 2\sigma_{\mathrm{NT}} \\ 0 & \text{if } r > 2\sigma_{\mathrm{NT}} \end{cases} \tag{6.3}$$

or an exponential kernel (Figure 6.9(c))

$$N(r) = \exp\left[-\frac{r}{\sigma_{\mathrm{NT}}}\right] \tag{6.4}$$

Plugging each expression into the model $\mathbf{W}$, we find that the predicted qualitative dependence of $\langle r^{(\infty)} \rangle$ on the neurotropin spreading range σ_{NT} and jump direction bias λ_{dir} is surprisingly insensitive to the shape of the kernel. Gaussian, triangular, and exponential spreading kernels all yield $\langle r^{(\infty)} \rangle$ curves that rise at low and high values of σ_{NT}. The optimal value of σ_{NT} that minimizes $\langle r^{(\infty)} \rangle$ can be reduced in each case by increasing the direction bias λ_{dir}.

We can evaluate the relative performance of the different kernel shapes by comparing the minimum values of $\langle r^{(\infty)} \rangle$ achieved by each kernel for identical λ_{dir}. By this measure, the Gaussian kernel performs slightly better when λ_{dir} is small, and the exponential kernel performs slightly better when λ_{dir} is large. The difference lies in the way each kernel shape plays to the neurotropin gradient's complementary roles of growth cone capture and growth cone guidance.

When λ_{dir} is small, the only feature a growth cone can discern is the presence or absence of neurotropin, so the neurotropin gradient behaves like a discrete potential barrier located near its peak. The gradient of a Gaussian kernel peaks at $r = \sigma_{\mathrm{NT}}$, making its attraction basin appear larger than that of an exponential kernel, whose gradient peaks at $r = 0$. Since larger basins are easier to find, the Gaussian kernel performs better in this weak attraction limit.

When λ_{dir} is large, the growth cone can discriminate a wide range of concentration values, so the neurotropin gradient acts as a continuous guidance signal that directs captive

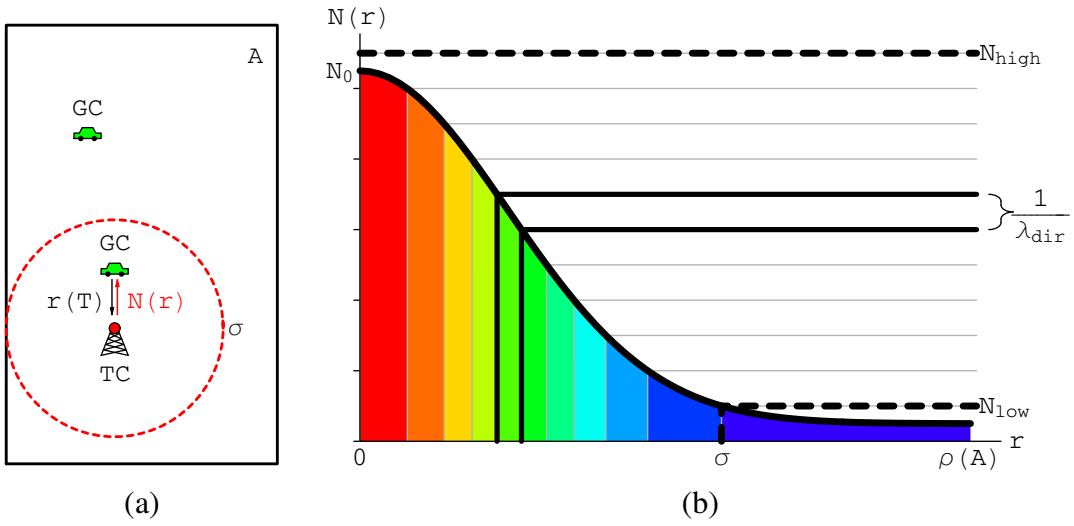


Figure 6.10: Neurotropic attraction system

(a) An active target cell (red) broadcasts a homing signal that is received by all coactive growth cones (green) within range σ (dashed circle). Growth cones move within a bounded target layer of area A, and their positions r relative to the target cell depends on the neurotropin signal $N(r)$ and the level of thermal noise T. (b) System parameter definitions. Vertical axis: neurotropin concentration $N(r)$ bound by growth cone; horizontal axis: distance r separating source cell from growth cone. Dashed lines bound sensitivity domain of growth cone receptor population. Colors denote discriminable concentration bands. $1/\lambda_{\text{dir}}$ is the minimal discriminable concentration increment and $\rho(A)$ is the maximum radial distance in a target layer of size A.

growth cones within their basins. The gradient of the exponential kernel increases continuously to a peak at $r = 0$, where the gradient of a Gaussian kernel falls to 0. In this strong attraction limit, a growth cone is more concerned with its guidance within the attraction basin than its ability to remain there, so the exponential kernel performs better.

6.3 Resource constraints

In this section, we outline the resource constraints that limit the values of parameters like λ_{dir} that govern neurotropic growth cone attraction in any physical implementation. We cast neurotropic attraction as a simple system in which an active target cell acts as a fixed guidance beacon that broadcasts a neurotropic homing signal to all coactive growth

cones within range (Figure 6.10(a)). Growth cones act as mobile receivers that shift their positions toward the direction of increasing signal strength. System performance is governed by a growth cone's ability to decode the position information encoded in the local neurotropin signal and then update its position accordingly. The system and its ultimate performance are completely specified by a handful of parameters: transmitter strength N_0, receiver sensitivity λ_{dir}, target layer size A, signaling range σ, and a temperature-like parameter T.

6.3.1 Position measurement

For the special case of a single active target cell, the neurotropin level $N(r)$ at any point in the target layer encodes the distance r separating that point from the transmitting target cell (Figure 6.10(b)). $N(r)$ attenuates with distance from the transmitter, decreasing monotonically from a peak value of N_0 at the active target cell site. The position information embedded in $N(r)$ can be decoded by a coactive growth cone, which is sensitive to a restricted domain of $N(r)$ values bounded by a detection threshold N_{low} and a saturation limit N_{high}. Within this domain, the growth cone measures the local neurotropin signal with some finite precision that chunks the continuous dynamic range of $N(r)$ values into a discrete set of discriminable increments whose size depends on the receiver sensitivity λ_{dir}[3]. Each increment encodes membership in a disjoint band of equiprobable separations r whose neurotropin levels are indistinguishable (colors in Figure 6.10(b)). A growth cone records a nonzero neurotropin gradient when its filopodia occupy different equiprobable separation bands. The perceived gradient strength is proportional to the number of incre-

[3]For simplicity, we model the fraction of bound receptors with a piecewise linear function of concentration whose slope is $1/(N_{\text{high}} - N_{\text{low}})$ for $N_{\text{low}} \leq N(r) \leq N_{\text{high}}$ and 0 elsewhere. We divide each unit of neurotropin concentration into a constant number λ_{dir} of discriminable gradations, which we define as the receiver sensitivity. N_{low} is taken to be the minimum discriminable concentration increment $1/\lambda_{\text{dir}}$ and N_{high} is taken to be equal to N_0.

ments spanned by the growth cone body.

The distribution of signal increments over the domain of separations r is controlled by the shape of the neurotropic attenuation function $N(r)$, as parameterized by the signaling range σ, which is defined as the distance at which $N(r)$ drops below the growth cone's detection threshold N_{low}[4]. The optimal function shape spreads the signal increments to cover the entire target layer. Extending neurotropic signaling beyond the finite dimensions $\rho(A)$ of the array is a waste of dynamic range, since low signal values are expended on unreachably long pair separations. Excessively short signaling ranges are similarly profligate, spending the entire dynamic range to encode a small number of short distances with unnecessary precision.

6.3.2 Position update

Mobile growth cones continually update their positions as instructed by their decoded neurotropin signals. In principle, a level-based neurotropin signal contains enough information to localize a growth cone's position to the width of an equiprobable separation band, as set by λ_{dir}. In practice, the growth cone's ability to hold this position is compromised by bumps, which we model as thermal fluctuations whose average strength depends on the temperature T.

We define thermal fluctuations to include any growth cone motion that is not explicitly and correctly directed toward a neurotropic attractor. Fluctuations may arise from implementation-specific issues like device mismatch and layout errors, or generically from stochastic spike arrival timing and stimulus presentation order. We lump all of these ther-

[4]Note that σ refers to the radius of the attraction basin, while σ_{NT} parameterizes the width of the neurotropin spreading kernel $G(r)$. Typically, $\sigma \approx 2\sigma_{\text{NT}}$, but the exact relation depends on the form of $G(r)$.

mal noise mechanisms into a single parameter T that is analogous to the temperature of a confined population of gas molecules. In the absence of guidance, T sets the rate at which a growth cone diffuses through the target layer. Lowering T reduces the average fluctuation size, slowing random growth cone motion.

The presence of a gradient biases the growth cone to drift in the direction of increasing neurotropin. The stronger the gradient, the stronger the bias. Since the measured gradient magnitude $\lambda_{\text{dir}}\|\nabla_r N(r)\|$ is proportional to $\lambda_{\text{dir}} N_0$, increasing either the transmitter strength N_0 or the receiver sensitivity λ_{dir} strengthens the attraction bias at a fixed temperature T.

6.3.3 Resource constraints

Each choice of parameter values $\{N_0, \lambda_{\text{dir}}, T, A, \sigma\}$ implicitly specifies an achievable system performance and the associated resource schedule. We can purchase better system performance by committing additional resources to secure more favorable parameter values.

6.3.3.1 Transmitter strength N_0

At the transmitting end of the signal, simply raising the initial concentration peak N_0 would expand the dynamic range of the target cell broadcast. However, increasing raw broadcast power requires the target cell to pump out more neurotropin molecules, which must be procured from a finite manufacturing capacity. In a silicon chip, N_0 is capped by the peak current that can be sourced by the power supply during a postsynaptic spike; in a biological neuron, N_0 would ultimately be limited by the protein synthesis rate.

6.3.3.2 Receiver sensitivity λ_{dir}

At the receiving end of the signal, increasing the precision λ_{dir} of the growth cone gradient sensor would stretch the existing dynamic range to encode more states at the cost of the time and space required to collect and store the additional information. In a biological growth cone, λ_{dir} is limited by the membrane receptor density that can be sustained by a finite protein synthesis capacity, and by the signal-specific amplification cascades that can be encoded in a finite genome; in a silicon chip, λ_{dir} is limited by the die area available for the integrating capacitors in the filopodial transistor circuits, or alternately by the memory and logic available to store and average multiple samples.

6.3.3.3 Temperature T

Thermal motion allows a growth cone to explore the target layer for its correct position but hinders its ability to consolidate that position once found. For this reason, learning systems typically anneal the level of thermal motion over the course of development to some minimum achievable value. In a biological system, this minimum value is determined by the resource flow available to stabilize and maintain synaptic contacts. In a silicon chip, the minimum error rate is limited by the time and labor available to design and implement reliable circuits, and the die space available to accommodate the larger devices or compensation circuits required for better matching.

6.3.3.4 Target layer size A

Growth cones located out of range of their coactive target cells wander aimlessly within the target layer until sliding by chance into their basins of attraction. A wandering growth cone may remain lost for a very long time if the target layer area A is much larger than the

basin area, so reducing A automatically improves performance by restricting the options of a diffusing growth cone. However, unlike the transmitter strength N_0 and the receiver sensitivity λ_{dir}, the size of the target layer is chosen to optimize the performance of the adult circuit, not the developmental mechanism. A is typically taken to be as large as possible, constrained only by metabolism or die area.

6.3.3.5 Signaling range σ

For fixed N_0, the nonincreasing nature of the attenuation function imposes a tradeoff between signal range and signal strength. Uniformly stretching $N(r)$ to cover a larger range σ weakens the gradient at each point by spreading the same concentration increment over a longer distance. This tradeoff implies the existence of an optimal signal range σ^* that minimizes the expected equilibrium separation $\langle r^{(\infty)} \rangle$ within the growth cone population for a given allocation of resources. In the next section, we will describe σ^* in terms of the system parameters N_0, λ_{dir}, A, and T.

6.4 Attraction basin model

In this section we derive the optimal signal range σ^* for an attraction basin model of neurotropic growth cone guidance. We consider the simple case of a single mobile growth cone moving in the presence of a fixed coactive release site. The neurotropin concentration $N(r)$ peaks at the release site and falls off with distance r, crossing below the growth cone's detection threshold at a radius $r = \sigma$ that defines a circular basin of attraction centered on the release site. Outside of this basin, the growth cone is free to diffuse randomly at some base rate, but once captured by the basin, a portion of the growth cone's motion is biased toward the attractor by the local neurotropin gradient. Successful growth cone attraction

depends on a basin's ability to capture free growth cones and guide captive growth cones.

Our strategy in this section is to first construct a toy model to describe growth cone capture by an attraction basin, and then augment that model to describe growth cone guidance within the basin. We will then compare the features of this toy model with those of the model transition matrix $\mathbf{W}$ from the previous chapter. We will find that small values of σ reduce the footprint of the attraction basin within the target layer, degrading its ability to capture free growth cones. Large values of σ weaken the gradient within the basin, degrading its ability to guide captive growth cones. The optimal value σ^* emerges from the confluence of these two mechanisms.

6.4.1 Growth cone capture

The steady-state probability P_{captive} that the growth cone is found within the attraction basin depends on the probability P_{capture} for a free growth cone to be captured by the basin and the probability P_{escape} for a captive growth cone to escape from the basin. At equilibrium, the probability currents into and out of the basin are equal:

$$P_{\text{captive}}P_{\text{escape}} = (1 - P_{\text{captive}})P_{\text{capture}} \tag{6.5}$$

Rearranging terms,

$$P_{\text{captive}} \;=\; \frac{P_{\text{capture}}}{P_{\text{escape}} + P_{\text{capture}}} \tag{6.6}$$

Outside the basin, a free growth cone diffuses randomly, and can be found anywhere in the array with equal probability. The equilibrium probability P_{capture} for a free growth

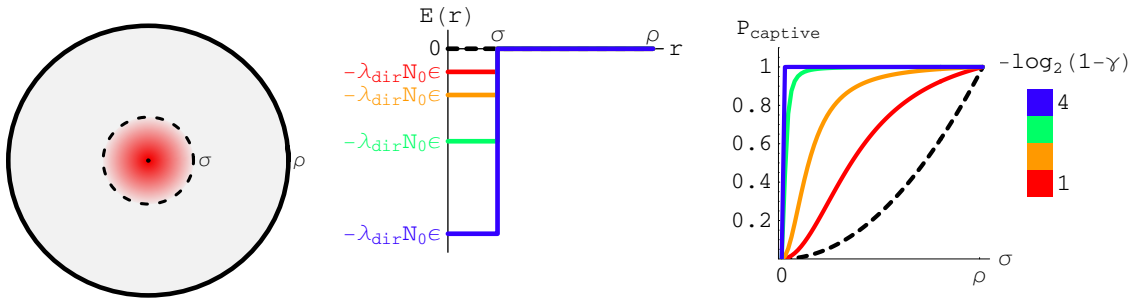


Figure 6.11: Potential well model

(a) Neurotropin (red) released from a single active target cell defines a circular attraction basin of radius σ (dashed line) at the center of a circular target layer of radius ρ (solid line). (b) Potential energy profile $E(r)$ for attraction basins modeled as square potential wells of depth $\lambda_{\text{dir}} N_0 \epsilon$. (c) Equilibrium probability P_{captive} for a growth cone to be found inside the attraction basin for the well depths depicted in (b). Black dashed line plots the case of no attraction ($\lambda_{\text{dir}} N_0 = 0$).

cone to stumble into its attraction basin is

$$P_{\text{capture}} = \frac{B}{A} \tag{6.7}$$

where B is the area of the attraction basin and A is the area of the target layer. For the special case of a circular attraction basin of radius σ located at the center of a circular target layer of radius ρ (Figure 6.11(a)), $B = \pi\sigma^2$ and $A = \pi\rho^2$.

Inside the basin, a captive growth cone attempts to escape but collides with the potential barrier surrounding the basin perimeter. The growth cone successfully escapes the basin with probability $(1 - \gamma)$, while with probability γ the growth cone lacks the energy to hurdle the potential barrier and rebounds toward the center of the basin instead. A sufficiently energetic growth cone can move anywhere in the array, leaving the basin area with probability $(A - B)/A$, so the equilibrium probability P_{escape} for a captive growth cone to leak

out of its basin is

$$P_{\text{escape}} = \frac{A - B}{A}(1 - \gamma) \tag{6.8}$$

We can relate γ to the peak neurotropin concentration N_0 and growth cone receptor sensitivity λ_{dir} if we think of the attraction basin as a potential well with uniform depth $\lambda_{\text{dir}} N_0 \epsilon$ (Figure 6.11(b)), where ϵ is a unit conversion ratio[5]. To escape the well, a captive growth cone must first acquire enough thermal energy to vault over the potential barrier. The probability that a given thermal fluctuation is large enough to propel the growth cone out of the well is $\exp[-\lambda_{\text{dir}} N_0 \epsilon / T]$, where the temperature T parameterizes the average fluctuation strength. Hence,

$$\gamma = 1 - \exp\left[-\frac{\lambda_{\text{dir}} N_0 \epsilon}{T}\right] \tag{6.9}$$

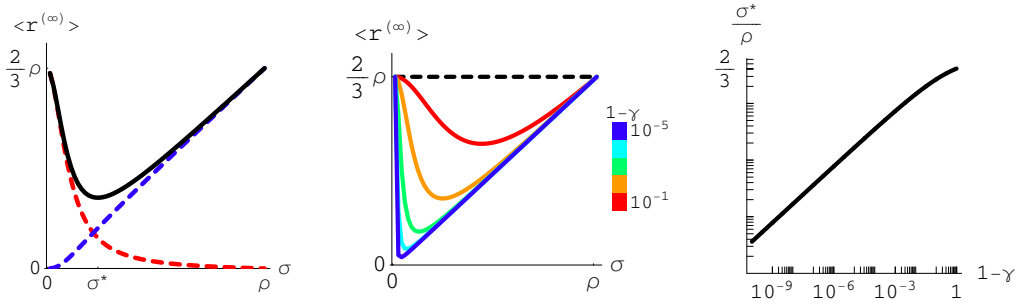
Increases in transmitter strength N_0 or receiver sensitivity λ_{dir} strengthen the attraction bias γ, while increases in thermal noise T degrade it.

Substituting (6.7) and (6.8) into (6.6),

$$P_{\text{captive}} = \frac{B}{(1 - \gamma)A + \gamma B} \tag{6.10}$$

P_{captive} behaves according to our intuition about the problem (Figure 6.11(c)). A freely diffusing growth cone is more likely to stumble into a basin that covers a larger fraction of the available area, so $P_{\text{captive}} \to 1$ as $B \to A$, and conversely, $P_{\text{captive}} \to 0$ as $B \to 0$. Once caught, the time a growth cone remains in captivity depends on the attraction bias γ.

[5]This is equivalent to modeling the neurotropin spreading kernel $N(r)$ with a step function whose gradient magnitude is N_0 at $r = \sigma$ and 0 elsewhere.



(a) (b) (c)

Figure 6.12: Optimal basin radius σ^* in square potential well model

(a) Average equilibrium separation $\langle r^{(\infty)} \rangle$ as a function of basin radius σ. Red dashed line indicates contribution $(1 - P_{\text{captive}})\langle r \rangle_{\text{free}}$ from free growth cones; blue dashed line indicates contribution $P_{\text{captive}}\langle r \rangle_{\text{captive}}$ from captive growth cones. (b) $\langle r^{(\infty)} \rangle$ for different attraction biases γ. (c) Optimal basin radius σ^* as a function of $1 - \gamma$.

With perfect attraction, the captive growth cone never escapes its basin, so $P_{\text{captive}} \to 1$ as $\gamma \to 1$. Without attraction, the basin is indistinguishable from the rest of the target layer, so P_{captive} reverts to the uniform probability $B/A \propto \sigma^2$ when $\gamma = 0$.

6.4.1.1 Optimal basin radius σ^*

Our use of a step function for $N(r)$ implies that a captive growth cone is located with equal probability at any point inside the basin, and that a free growth cone is located with equal probability at any point outside the basin. Given these two permissible states, the expected distance $\langle r^{(\infty)} \rangle$ separating a growth cone from the basin center at equilibrium is

$$\langle r^{(\infty)} \rangle = P_{\text{captive}}\langle r \rangle_{\text{captive}} + (1 - P_{\text{captive}})\langle r \rangle_{\text{free}} \tag{6.11}$$

where the equilibrium distance $\langle r \rangle_{\text{captive}}$ separating a captive growth cone from the basin center is $\int_0^{2\pi} \int_0^{\sigma} r^2 d\theta dr/(\pi\sigma^2) = \frac{2}{3}\sigma$, and the equilibrium distance $\langle r \rangle_{\text{free}}$ separating a free

growth cone from the basin center is $\int_0^{2\pi}\int_\sigma^\rho r^2 d\theta dr/(\pi\rho^2 - \pi\sigma^2) = \frac{2}{3}(\rho^3 - \sigma^3)/(\rho^2 - \sigma^2)$.

$\langle r^{(\infty)}\rangle$ and its constituent components are plotted against σ in Figure 6.12(a). $\langle r\rangle_{\text{free}}$ dominates at low σ, which allow more growth cones to escape their basins, while $\langle r\rangle_{\text{captive}}$ dominates at high σ, which allow captive growth cones to roam within larger basins. Increasing γ traps more growth cones in their potential wells, reducing $\langle r\rangle_{\text{free}}$ but leaving $\langle r\rangle_{\text{captive}}$ untouched (Figure 6.12(b)).

Combining terms,

$$\langle r^{(\infty)}\rangle \;=\; \frac{2}{3}\frac{\sigma^3\gamma + \rho^3(1-\gamma)}{\sigma^2\gamma + \rho^2(1-\gamma)} \tag{6.12}$$

There are two real values of σ that extremize this expression. The first real extremum is located at $\sigma = 0$ and is a maximum if $\gamma > 0$ and a minimum if $\gamma = 0$. Since the ideal of $\gamma = 0$ is unrealizable in a physical implementation, we are more interested in the other real extremum located at

$$\sigma^* = \frac{(1-\gamma)^{1/3}}{\sqrt{\gamma}}\left[(1+\sqrt{\gamma})^{1/3} - (1-\sqrt{\gamma})^{1/3}\right]\rho \tag{6.13}$$

σ^* ranges from 0 to $\frac{2}{3}\rho$, and is the *optimal basin radius* that minimizes $\langle r^{(\infty)}\rangle$ for a given choice of release strength N_0, receptor sensitivity λ_{dir}, target layer radius ρ, and noise level T (Figure 6.12(c)).

6.4.2 Growth cone guidance

Comparing the flat attraction basin model predictions for $\langle r^{(\infty)}\rangle$ (Figure 6.12(b)) with those of the model **W** (Figure 6.7(b)), we observe that as λ_{dir} increases, the flat attraction basin model successfully describes the improvement in $\langle r^{(\infty)}\rangle$ at low σ, where performance

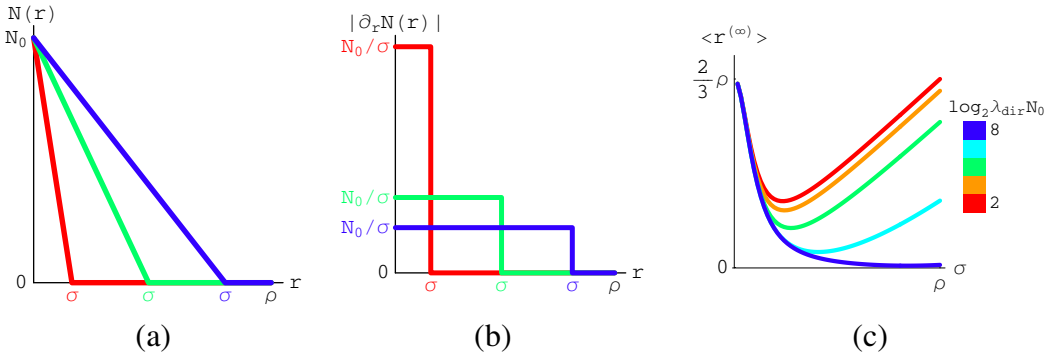


Figure 6.13: Guidance within attraction basin

(a) Linear spreading kernel $N(r)$ for different attraction basin radii σ and fixed N_0. (b) Gradient magnitude $|\nabla_r N(r)|$ for the σ in (a). Increasing σ decreases $|\nabla_r N(r)|$ for fixed N_0. (c) Average equilibrium separation $\langle r^{(\infty)} \rangle$ for different values of $\lambda_{\text{dir}} N_0$ and fixed $\lambda_{\text{dir}} N_0 \epsilon / T$.

is limited by a growth cone's ability to remain in the basin, but fails to capture the improvement at high σ, where performance is limited by a growth cone's lack of guidance within the basin. In practice, growth cones are not entirely unguided within their attraction basins, but are biased toward their attractors by the local gradient. The strength of this guidance signal is limited by the dynamic range $\lambda_{\text{dir}} N_0$ of bound neurotropin and by the basin radius σ.

For simplicity, we will describe the neurotropin concentration $N(r)$ with a triangular spreading kernel that falls linearly from a peak of N_0 at $r = 0$ to zero at $r \geq \sigma$, defining a circular attraction basin of radius σ (Figure 6.13(a)). Higher values of σ weaken the gradient magnitude within the basin by spreading the drop from N_0 over a greater distance (Figure 6.13(b), an inverse dependence that holds for any monotonically decreasing function of r.

To compute the average equilibrium separation $\langle r \rangle_{\text{captive}}$ within an attraction basin, we classify captive growth cones as either guided or unguided. A growth cone is guided with probability P_{dir} and unguided with probability $1 - P_{\text{dir}}$, where $P_{\text{dir}} = 1 - \exp(-\lambda_{\text{dir}} |\nabla_r N(r)|)$

is a function of the local gradient magnitude $|\nabla_r N(r)|$. For a linear gradient, $|\nabla_r N(r)| = N_0/\sigma$. Guided captive growth cones are always found at $r = 0$, while unguided captive growth cones may be found anywhere inside the basin with equal probability, for a combined equilibrium separation of

$$\langle r \rangle_{\text{captive}} = (1 - P_{\text{dir}})\frac{2}{3}\sigma. \tag{6.14}$$

We can improve growth cone guidance, as measured by $\langle r^{(\infty)} \rangle$, by increasing the guidance probability P_{dir} (Figure 6.13(c)). There are two ways to do this. First, increasing receptor sensitivity λ_{dir} allows the growth cone to detect and amplify shallower gradients. Second, raising the peak concentration N_0 increases the salience of the guidance signal by steepening the gradient. Reducing the basin radius σ would also steepen the gradient.

The fundamental conflict is that the capture probability P_{captive} is a function of σ but the guidance probability P_{dir} is a function of $1/\sigma$. These opposing dependences imply the existence of an optimal basin radius σ^* that balances an attraction basin's ability to capture free growth cones with its ability to guide captive growth cones.

In summary, the attraction basin model establishes the existence of an optimal signaling range σ^* and relates it to a set of parameters that collectively prescribe the convergence properties of the system. Equilibrium performance at low σ can be predicted by treating the basin as a flat potential well, but predictions for high σ require consideration of the gradient within the basin. Consistent with this model, we observed that equilibrium performance can be improved empirically by eliminating bumps to reduce the temperature T, or by averaging multiple gradient measurements to increase the growth cone sensitivity λ_{dir}. Reducing T enhances growth cone capture, improving performance at low σ_{NT} (Figure

6.4), while increasing λ_{dir} enhances growth cone guidance, improving performance at high σ_{NT} (Figure 6.6). Now that we understand the constraints governing supervised pair attraction, we are finally ready to tackle unsupervised self-organization, which we will cover in the next chapter.

Chapter 7

Unsupervised Self-Organization

In this chapter, we examine unsupervised growth cone motion under the influence of multiple mobile neurotropin release sites of varying strength, when driven by presynaptic activity correlations generated by an ideal stimulus generator and by a silicon retina.

7.1 Unsupervised patch attraction

In unsupervised self-organization, there are no static attractors to direct active growth cone motion. Instead, active growth cones direct each other, exerting mutually attractive forces whose average strength depends on the degree to which their presynaptic activities coincide. Growth cone motion is driven by two opposing mechanisms: cooperative attraction of coactive growth cones, and competitive displacement of inactive growth cones.

7.1.1 Order parameter

Without fixed instructors to mark absolute positions within the target layer, the retinotopic order parameter must be assembled solely from relations between axon terminals. As defined in Chapter 4, the relative retinotopic order parameter for the axon terminal projected by source cell α is the average of the instantaneous pairwise separations $r^{(n)}_{\alpha(\alpha+\mathbf{f}_i)}$ separating it from the axon terminals projected by its source cell's immediate neighbors $\alpha + \mathbf{f}_i$:

$$\Phi^{(n)}_{\alpha} = \frac{1}{N_\alpha} \sum_i r^{(n)}_{\alpha(\alpha+\mathbf{f}_i)} \tag{7.1}$$

where N_α is a normalization constant. We average the individual axon terminal order parameters to obtain the population order parameter for the full axon projection:

$$\Phi^{(n)} = \frac{1}{N} \sum_{\alpha,i} r^{(n)}_{\alpha(\alpha+\mathbf{f}_i)} \tag{7.2}$$

where $N = \sum_\alpha N_\alpha$ is a normalization constant.

7.1.2 Uptake parameter

To match the relative retinotopic order parameter that defines the problem with the uptake parameter that programs growth cone attraction, we must choose a presynaptic stimulus protocol that generates the appropriate coactivities $C_{\alpha\beta}$.

7.1.2.1 Pair stimulus

Since our retinotopic order parameter is composed of pair relations between growth

cones, one obvious strategy is to address each relation individually by stimulating each growth cone pair in turn. We randomly select and simultaneously activate a source cell α and one of its immediately adjacent neighbors $\alpha + \mathbf{f}_i$, each of which projects a mobile growth cone into the target layer. The two coactive growth cones are mutually attractive and move to minimize their target layer separation $r_{\alpha(\alpha+\mathbf{f}_i)}$. Retinotopic self-organisation is induced through sequential activation of every growth cone pair $\{\alpha, \alpha + \mathbf{f}_i\}$ whose separation is contained in $\Phi^{(n)}$. The resulting set of coactivities $C_{\alpha\beta} = \delta[\|\alpha - \beta\| - 1]$ assembles an uptake parameter $U_\alpha^{(n)}$ with the same form as the retinotopic order parameter $\Phi_\alpha^{(n)}$. (This choice is not unique, since any monotonically decreasing function of source layer separation $\|\alpha - \beta\|$ would achieve similar results.)

$$\begin{aligned} U_\alpha^{(n)} &= \sum_\beta \delta[\|\alpha - \beta\| - 1] r_{\alpha\beta}^{(n)} \\ &= \sum_i r_{\alpha(\alpha+\mathbf{f}_i)}^{(n)} \end{aligned} \tag{7.3}$$

The advantage of this approach is that we can build up any arbitrary coactivity pattern out of individual pair stimuli. The disadvantage is that it takes a very long time to present all of the necessary pair stimuli with sufficient repetition to achieve convergence.

7.1.2.2 Patch stimulus

We can present stimuli more quickly by activating several pairs in parallel, a strategy that has some biological plausibility. During mammalian development, spontaneous bursts of coordinated activity called *retinal waves* sweep across contiguous patches of the retina, correlating the spiking of neighboring retinal ganglion cells (RGCs) [105]. Consecutive retinal waves do not overlap, but eventually tile the entire retina, establishing sufficient statistical structure to encode the relative retinotopic distance separating each pair of RGC

cell bodies in the coincidence of their spontaneous spike activities.

Since we are only interested in the static property of retinotopy, we represent each retinal wave as a spatially contiguous patch of coactive source cells, ignoring its temporal dynamics. Each patch is centered on a randomly chosen source cell and uniformly activates all neighboring source cells within a fixed radius p. Every patch is selected with equal probability, so the coactivity $C_{\alpha\beta}$ between two source cells α and β is proportional to the number of patches shared by both source cells. Consequently, the uptake parameter includes some extra terms not present in the order parameter. For example, the uptake parameter for a patch stimulus of radius 1 is

$$U_{\alpha}^{(n)} \;=\; \sum_{i} r_{\alpha(\alpha+\mathbf{f}_i)}^{(n)} + \frac{1}{4}\sum_{j} r_{\alpha(\alpha+\mathbf{g}_j)}^{(n)} \tag{7.4}$$

where $\alpha + \mathbf{f}_i$ indexes the three source cells located one edge away from source cell α and $\alpha + \mathbf{g}_j$ indexes the six source cells located two edges away. For small patches, these extra terms are closely linked to the order parameter terms, and minimizing one tends to minimize all. In larger patches, the extra terms can sometimes distract the growth cone from the primary goal of minimizing the order parameter.

7.1.3 Unsupervised retinotopic evolution

Under patch stimulation, an unsupervised population of mobile growth cones converges to the same level of sustainable retinotopy from very different initial configurations. Figure 7.1 shows the evolution of the growth cone population under stimulation by patches of radius 1. We seed the system with either a completely random initial projection (Figure 7.1(a), blue line in Figure 7.1(b)) or a perfectly retinotopic initial projection (Figure 7.1(c), red line in Figure 7.1(b)) and release the system to refine or relax its retinotopy.

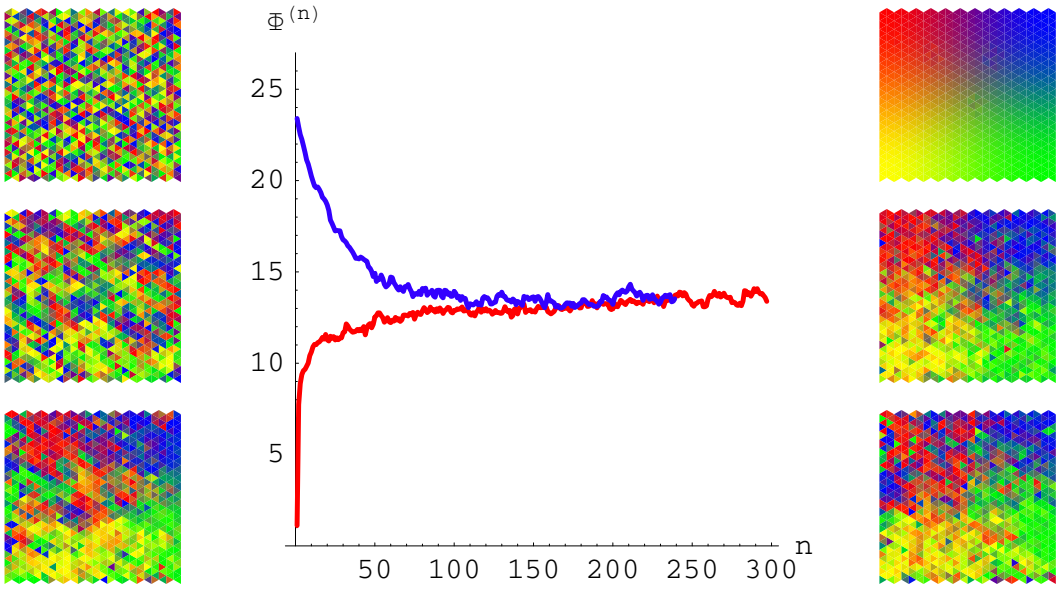


Figure 7.1: Unsupervised retinotopic evolution

Population refinement and relaxation to a common sustainable retinotopic level. (a) Source cell color maps generated from no initial retinotopy. Source cells are colored by the target layer coordinates of their growth cones. Top: $n = 0$; middle: $n = 20$; bottom: $n = 200$. (b) Evolution of retinotopic order parameter $\Phi^{(n)}$ from no initial retinotopy (blue) and perfect initial retinotopy (red). (c) Source cell maps generated from perfect initial retinotopy. Top: $n = 0$; middle: $n = 20$; bottom: $n = 200$.

Starting from random initial locations (top color map in Figure 7.1(a)), growth cones quickly crystallize into small clusters that are locally topographic but globally misplaced (middle color map). Over time these locally topographic crystals melt, migrate, and merge to form larger, better positioned topographic clusters until only a single globally retinotopic map remains (bottom color map). A perfectly retinotopic initial map relaxes to a similar state of disorder, as illustrated by the color maps in Figure 7.1(c).

The evolution of the order parameter is plotted in Figure 7.1(b). $\Phi^{(n)}$ converges to the same stable value $\Phi^{(\infty)}$ from both initial configurations, indicating that the achievable retinotopic endstate is a property of the system settings, not the initial state. In other words, the retinotopy of the final state cannot be improved merely by initializing growth cones

closer to their intended targets.

7.1.4 Model transition matrix

For small patch radii, we can directly apply our model transition matrix for supervised pair attraction to describe unsupervised patch attraction after only a minor adjustment to the form of the sampled neurotropin concentration $N(r)$. We treat the patch stimulus as if its participating growth cones had been activated in a sequence of nonoverlapping pair stimuli between the patch center and each of the other patch participants in turn, ignoring the cross terms in the uptake parameter between pairs of noncentral patch participants. The neurotropin concentration sampled by each growth cone in a pair stimulus is a function of the radial distance r separating the growth cone from its coactive partner. There are only two sources of neurotropin release at a time, one stimulated by each growth cone. A growth cone detects the neurotropin released from its own site $C_0G(0)$ and a portion $C_1G(r)$ of the neurotropin that spreads from the other growth cone site. Since each growth cone is equally active, $C_0 = C_1$ and the ratio $\rho_{\mathrm{GC}} = C_0/C_1$ is identically 1 and is no longer a free parameter.

Of the remaining free parameters, the sampling interval τ and the attraction bias λ_{dir} retain the values previously fitted in the supervised attraction experiments, while the jump rate λ_{NT} and the signalling range σ_{NT} are fit to features of the expected change $\langle \Delta\Phi \rangle$ in the order parameter between samples and the variance $\langle\langle \Delta\Phi \rangle\rangle^2$ of that change. Each growth cone is activated by multiple patches during each time step, so λ_{NT} is higher than for supervised pair attraction. σ_{NT} captures the effective growth cone signalling range, which

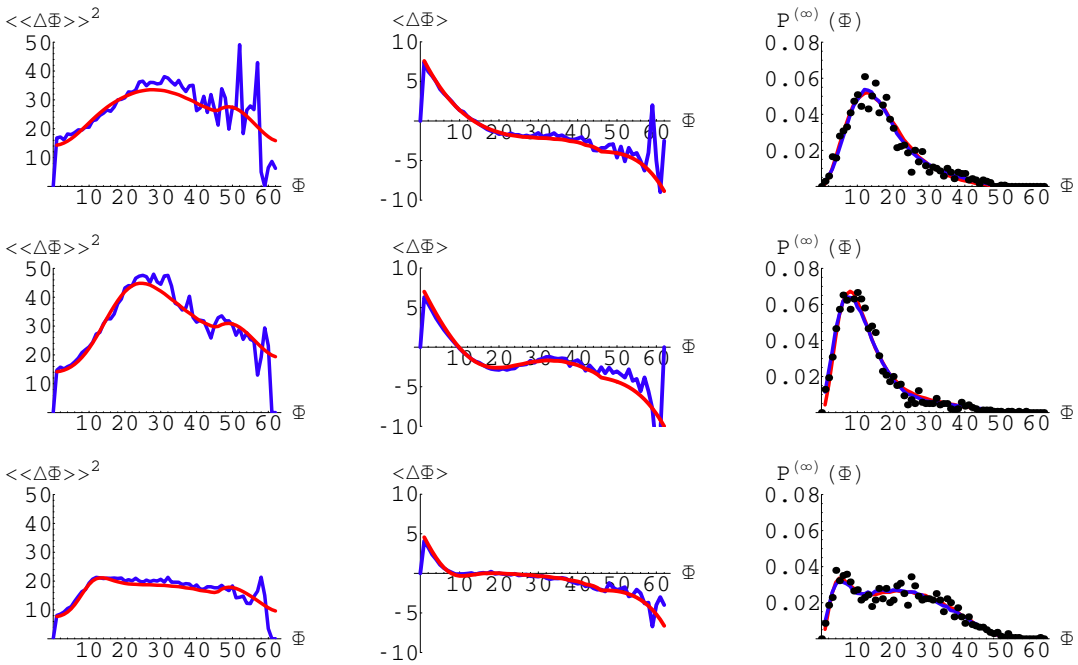


Figure 7.2: Moment fits and predicted $P^{(\infty)}(r)$ for patch radius=1

Moments and predicted stationary distribution for empirical (blue) and model (red) transition matrices $\mathbf{W}$. Top: $V_{\text{shunt}} = 0.20V$; middle: $V_{\text{shunt}} = 0.30V$; bottom: $V_{\text{shunt}} = 0.40V$. (a) Order parameter change variance $\langle\langle\Delta\Phi\rangle\rangle^2$. (b) Expected order parameter change $\langle\Delta\Phi\rangle$. (c) Predicted stationary distribution $P^{(\infty)}(\Phi)$ (solid lines) and measured distribution $P^{(100)}(\Phi)$ (black dots).

is a combination of the neurotropin spreading range and the patch radius.

7.1.4.1 Change variance $\langle\langle\Delta\Phi\rangle\rangle^2$

The change variance $\langle\langle\Delta\Phi\rangle\rangle^2$ again consists of a small peak at σ_{NT} due to the growth cone's own jumps, riding on top of a much larger baseline due to bump activity (Figure 7.2(a)). The height of the baseline sets λ_{NT}. $\langle\langle\Delta\Phi\rangle\rangle^2$ takes on much larger values for patch stimuli because each growth cone is activated more often between samples and therefore has more opportunities to move.

7.1.4.2 Expected change $\langle \Delta \Phi \rangle$

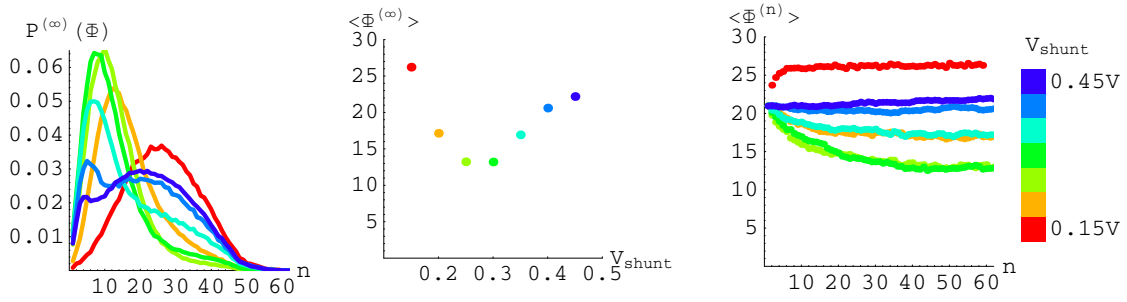
The expected order parameter change $\langle \Delta \Phi \rangle$ under patch stimulation reprises the major features of the expected separation change $\langle \Delta r \rangle$ under supervised pair stimulation (Figure 7.2(b)). The amplitudes of these features are magnified by the increased stimulation between samples. The position of the attractive trough reveals the effective signalling range σ_{NT}.

7.1.4.3 Stationary distribution $P^{(\infty)}(\Phi)$

The stationary distributions $P^{(\infty)}(\Phi)$ predicted by the principal eigenvectors of the empirical and model $\mathbf{W}$ both fit the recorded distribution $P^{(100)}(\Phi)$ of the 100th sample, as shown in Figure 7.2(c). $P^{(n)}(\Phi)$ converges more quickly than in the supervised case because more pairs are activated between samples. As before, performance peaks at an intermediate value of V_{shunt} that corresponds to an optimal neurotropin spreading range. We will discuss this result in the next section after exploring the contribution of patch size to the effective signaling radius.

7.2 Effective signaling range

In theory, uptake parameter minimization heavily penalizes large separations, implying that coactive growth cones can robustly communicate their attraction across arbitrarily long distances. In practice, this global information flow becomes prohibitively expensive in a large physical system, so growth cones approximate uptake parameter minimization with the more local mechanism of neurotropic gradient ascent. Growth cones operating under these theoretical and practical techniques navigate similar paths near a coactive attractor,



(a) (b) (c)

Figure 7.3: Optimal spreading range

Time evolution of $\langle\Phi\rangle$ under stimulation by patches of radius 1. (a) $P^{(\infty)}(\Phi)$ predicted by empirical $\mathbf{W}$. (b) $\langle\Phi^{(\infty)}\rangle$ as a function of V_{shunt}. (c) $\langle\Phi^{(n)}\rangle$ evolution recorded from data samples (solid lines) and predicted by empirical $\mathbf{W}$ (dashed lines).

but diverge in their assessment of more distant targets.

Neurotropin-mediated growth cone signaling is limited by the neurotropin spreading range, which establishes a perception horizon beyond which a growth cone cannot detect distant release sites. Since growth cones cannot pursue invisible targets, the cost of large separations is effectively zero under neurotropic gradient ascent. Growth cones that are separated from their coactive attractors by long target layer distances must rely on random motion to jostle them into signaling range. Increasing the patch size recruits more coactive attractors, increasing the likelihood that one will appear on the horizon to guide an active growth cone.

In this section, we investigate the influence on retinotopic convergence of the effective growth cone signaling range, which consists of the neurotropin spreading range and the radius of the patch stimulus.

7.2.1 Neurotropin spreading range σ_{NT}

Varying the decay transistor gate bias V_{shunt} to control the neurotropin spreading range σ_{NT}, we find that retinotopic convergence is optimized at an intermediate value of σ_{NT} (Figure 7.3). As in the supervised case, smaller values of σ_{NT} yield stronger, slower attraction at shorter distances, while larger values of σ_{NT} yield weaker, faster attraction at longer distances. Performance is best at an intermediate value of σ_{NT}.

7.2.2 Patch radius

Another way to extend the effective attraction range of a growth cone is to increase the size of the presynaptic patch stimulus. To isolate the effect of patch radius on topographic convergence, we reduce the neurotropin spreading range to a small fraction of the array dimensions and stimulate the system with patches of increasing size. Larger patches are stimulated with shorter bursts to maintain a constant absolute presynaptic activity level.

Each patch of radius p consists of a source cell and all of its neighbors up to p edges away, for a total of $1 + \sum_{r=1}^{p} 3r = 1 + 3p(p+1)/2$ coactive cells if the patch is not cut off by the array boundaries (Figure 7.4(a)). Conversely, each source cell participates in up to $1 + 3p(p+1)/2$ distinct patches. The coactivity between a pair of growth cones is proportional to the number of patches shared by both growth cones. Neighboring source cells participate in overlapping patches, so $C_{\alpha\beta}$ decreases monotonically with source layer separation $\|\alpha - \beta\|$, dropping to zero at twice the patch radius (Figure 7.4(b)).

Source cell coactivity leads to target layer proximity. Figure 7.4(c) plots the stationary average distance $\langle r_{\alpha\beta} \rangle$ separating growth cones α and β in the target layer as a function of the source layer separation $|\alpha - \beta|$ of their projecting cell bodies. Larger patches correlate

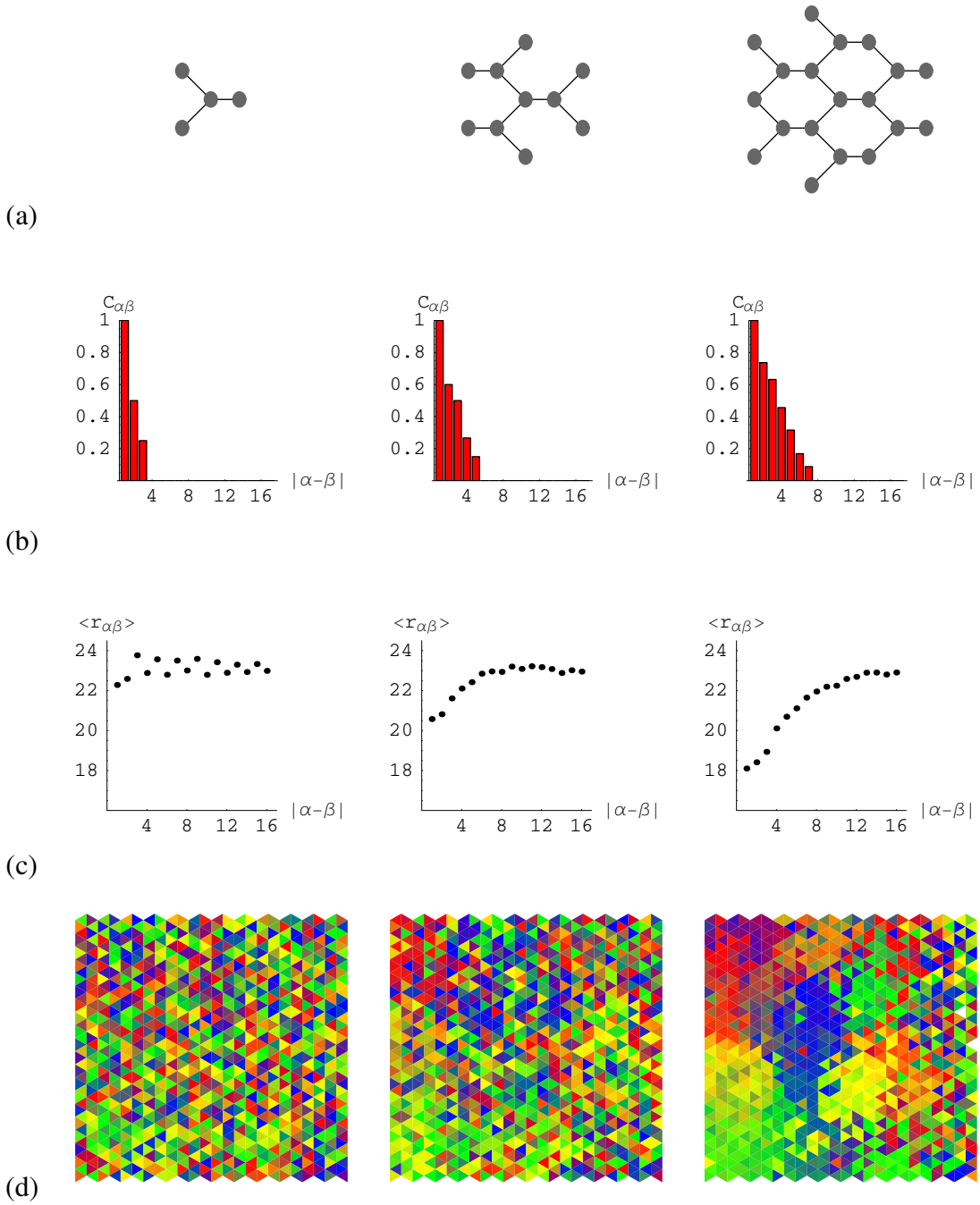


Figure 7.4: Patch radius

Coactivity dependence on patch radius. Data collected at $V_{\text{shunt}} = 0.45V$. Left: patch radius=1; middle: patch radius=2; right: patch radius=3. (a) Sample patches of coactive source cells. (b) Coactivity $C_{\alpha\beta}$ as a function of source layer separation $|\alpha - \beta|$. (c) Measured average target layer distance $\langle r^{(240)} \rangle_{\alpha\beta}$ separating growth cones projected by source cells α and β, as a function of $|\alpha - \beta|$. (d) Growth cone positions in the target layer during the 240th sample, colored by source cell coordinates.

the activities of more widely spaced source cells, clustering their growth cones in the target layer. Large patches can compensate for short neurotropin spreading ranges, as vividly illustrated in Figure 7.4(d). We select a spreading range that is too short to refine retinotopy using a patch stimulus of radius 1. Increasing patch size rescues performance, aggregating growth cones into larger, more refined topographic clusters. The effective neurotropic signaling range is a combination of patch radius and neurotropin spreading range.

7.2.2.1 Neurotropic connectivity

We can envision the network of growth cone pair relations as an undirected graph in which two coactive growth cones are neurotropically connected if they are within spreading range of each other (Figure 7.5(a)). Connected growth cones mutually attract, while unconnected growth cones rely on random motion to eventually nudge them into range of an attractor. A pair of coactive growth cones whose attraction basins do not directly overlap may still be connected indirectly if both are within view of a third coactive growth cone.

Increasing the patch size increases the number of coactive intermediaries available to bridge long pair separations. For example, in the middle panel of Figure 7.5(a), the attraction basins of three of the radius 1 patch growth cones overlap (black, red, green), but the fourth growth cone (blue) is disconnected from the first three, being too distant to communicate directly. However, the fourth growth cone overlaps with other members of the radius 2 patch (grey), which also overlap with the original three growth cones, so the four radius 1 patch growth cones are connected indirectly and can actively attract each other, minimizing the three pairwise separations that collectively comprise the order parameter. By contrast, the fourth growth cone (black) remains disconnected in the similar configuration for a radius 1 patch shown in the left panel of Figure 7.5(a).

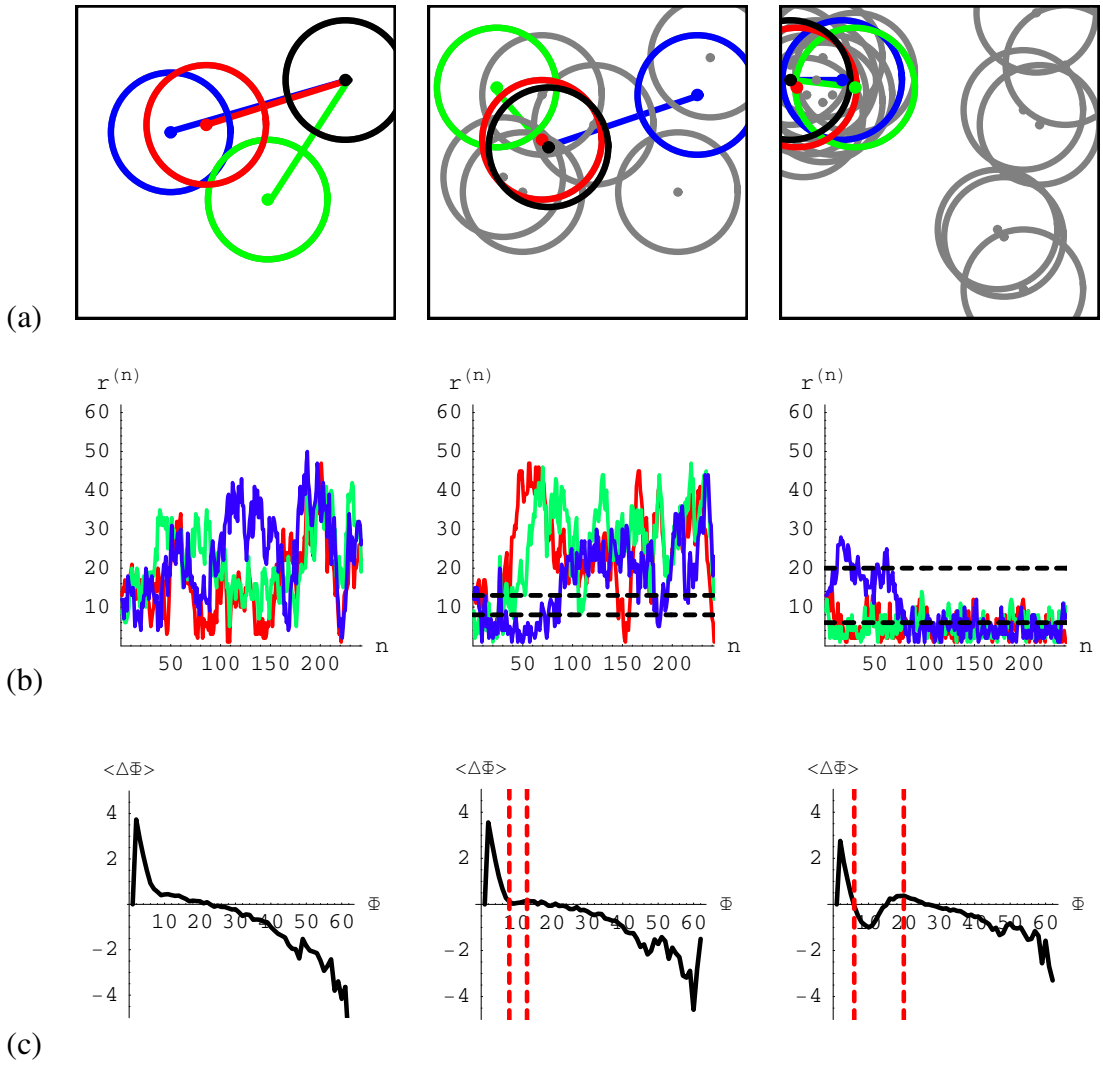


Figure 7.5: Patch trajectory

Trajectory dependence on patch radius. Data collected at $V_{\text{shunt}} = 0.45V$. Left: patch radius=1; middle: patch radius=2; right: patch radius=3. (a) Positions (dots) and detection horizons (circles) for growth cones participating in a patch at $n = 240$. Each growth cone is located at the center (dot) of an attraction basin of radius σ_{NT} (circle). Black dot marks position of growth cone projected by center of patch. Red, green, and blue lines identify pair separations $r_{\alpha\beta}$ that compose order parameter Φ_α. (b) Evolution of pair separations $r_{\alpha\beta}$ that compose order parameter Φ_α. Dashed lines bound directed jump band. (c) Expected order parameter change $\langle \Delta\Phi \rangle$ as a function of initial value. Dashed lines bound directed jump band.

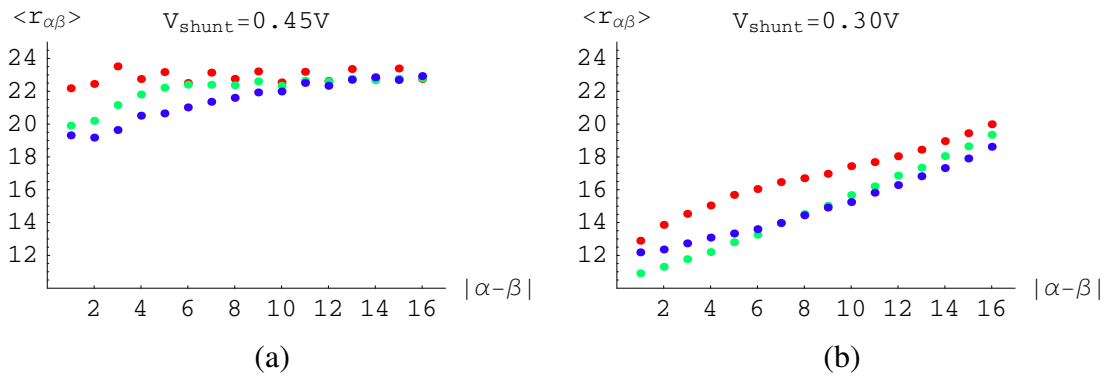


Figure 7.6: Signaling range

Average target layer distance $\langle r_{\alpha\beta} \rangle$ separating growth cones α and β, as a function of the source layer distance $|\alpha - \beta|$ separating their projecting cell bodies. Red: patch radius = 1; green: patch radius = 2; blue: patch radius = 3. (a) Short σ_{NT}. (b) Intermediate σ_{NT}.

The probability that a growth cone falls within spreading range of a neurotropic attractor is proportional to the density of attractors in the target layer. The Neurotrope1 array has a fixed volume, so any decrease in the number of neurotropic attractors within this bounded arena also decreases attractor density. Lowering attractor density reduces the number of pair relations available to tie a growth cone to a cluster of coactive neighbors, making it easier for a growth cone to slip away. Growth cones in a radius 1 patch easily escape each other's orbit, wandering in and out of range many times over the course of the evolution (Figure 7.5(b), left panel).

Larger patches offer more opportunities for a growth cone to connect to at least one of its fellow patch members, which all cooperate to hold each other together within the patch, preserving their coherence (Figure 7.5(b), right panel) and widening the band of pair separations at which directed jumps dominate growth cone motion (Figure 7.5(c)).

7.2.2.2 Attractors and distractors

For short spreading ranges, each neurotropic attraction basin covers a comparatively

small fraction of the array area, making it difficult for a stray growth cone to rediscover a lost attractor. Increasing patch radius compensates for the short spreading range by adding more attractors to broaden the coactive coverage of the array, rescuing performance (Figure 7.6(a)). At longer spreading ranges, an initial increase in patch radius has a similarly salutary effect, but a second increase partially reverses this trend. Figure 7.6(b) plots the average target layer distance $r_{\alpha\beta}$ separating two growth cones as a function of the source layer distance $\|\alpha - \beta\|$ separating their projecting cell bodies for a longer σ_{NT}. Larger patches coactivate growth cones projected by more widely spaced source cells. When the patch gets too large, these additional attractive forces actually distract growth cones of more proximate cell bodies from their primary attraction to the center growth cone.

For example, the patch of radius 1 centered on source cell α recruits all four of the nearest-neighbor source cells affiliated with the retinotopic order parameter Φ_α. Each patch elicits two types of interactions between participating growth cones. First, the center source cell α is coactive with each of its three nearest neighbors $\alpha+\mathbf{f}_i$, so their growth cones move to minimize the three distances $r_{\alpha(\alpha+\mathbf{f}_i)}$ that comprise the order parameter Φ_α for the center source cell. Changes in these center-neighbor separations constitute radial motion relative to the center growth cone. Second, the three neighboring source cells $\alpha+\mathbf{f}_i$ are also coactive with one another, so their growth cones attempt to minimize three separations $r_{(\alpha+\mathbf{f}_i)(\alpha+\mathbf{f}_j)}$ that do not contribute to the order parameter. Changes in these neighbor-neighbor separations mix radial and angular motion relative to the center growth cone.

From the perspective of the order parameter, the supplemental linkage provided by these extraneous neighbor-neighbor interactions is beneficial when a proximate neighbor attracts a distant neighbor closer to the center growth cone, and detrimental when the distant neighbor draws the proximate neighbor away from the center growth cone. However, as long as more distant neighbors are less coactive, the center growth cone should be able

to dominate the attention of more proximate neighbors. Problems arise when there are too many simultaneously active neurotropin sources to distinguish between their relative coactivities. Since a growth cone pursues all equally coactive neurotropin sources with equal intensity, saturating the measurable coactivity prevents the growth cone from localizing positions within a patch.

In summary, as in the supervised case, the general principle is that signals in the real world are physically bounded to a finite dynamic range of distinguishable signal values. Since a growth cone's guidance depends on its ability to discriminate level differences in neurotropin concentration, the patch size and spreading range must be chosen to prevent the combined guidance signal from saturating at one of its bounds.

7.3 Silicon retina

Up to this point we have driven the Neurotrope1 system with near-ideal presynaptic coactivity patterns supplied by a programmable stimulus generator. However, the real test of any neuromorphic algorithm is its performance when burdened with the imperfections of real world systems. We therefore conclude this chapter by replacing the ideal stimulus generator with a more physical signal source in the form of a silicon retina.

7.3.1 Multichip system

We use the silicon retina described in [107, 108] to translate patterned illumination into appropriately correlated spike trains that are fed into the Neurotrope1 chip as presynaptic input. Activity generated by two different populations of retinal ganglion cells (RGCs) drives growth cones in the combined system to self-organize retinotopy, a task that is hin-

dered by three implementation-specific chip pathologies.

7.3.1.1 Retinal ganglion cells

The silicon retina transduces incident photons into electrical spike activity generated by four different RGC types. ON-center RGCs are excited by a spot of light in the center of their receptive fields and inhibited by light in the surrounding annulus, while OFF-center RGCs are excited and inhibited by the opposite polarity stimulus. ON- and OFF-center RGCs are further divided into sustained or transient cell types depending on their temporal response to a change in stimulus. We will use the sustained ON- and OFF-center RGCs as our population of source cells, which we will stimulate by projecting patterns of light onto the silicon retina.

ON- and OFF-center RGCs respond to opposite polarity stimuli, so the activity of an RGC of one cell type is anticorrelated with the activity of another RGC of the opposite cell type at the same location. This statistical difference segregates growth cones from each cell type into subpopulations that can be separately instructed with appropriate manipulation of the visual stimulus.

7.3.1.2 Presynaptic activity gradient

The silicon retina exhibits several of the same pathologies as the Neurotrope1 chip, including a systematic excitability gradient that causes RGCs on the left side of the silicon retina to spike more frequently in response to a given stimulus. Enhanced presynaptic activity increases growth cone mobility for two reasons. First, growth cones accumulate extant neurotropin more quickly, since neurotropin binding is gated by presynaptic activity. Second, there is more extant neurotropin, since neurotropin release is gated by postsynaptic

activity, which is triggered by presynaptic activity. Consequently, more active growth cones can systematically displace less active growth cones from mutually preferred territory in the target layer.

7.3.1.3 Postsynaptic excitability gradient

The Neurotrope1 chip exhibits a similar gradient of postsynaptic excitability that increases from left to right across the target cell array. A given presynaptic stimulus elicits more neurotropin release from more excitable target cells, so neurotropin-hungry growth cones are globally attracted to the highly excitable region on the right side of the array. The density constraint forces growth cones to fight over this neurotropin-rich territory, pulling other growth cones back when active and being pulled back when inactive. This churning is purely disruptive if all source cells are equally active, but any imbalance in source cell activity allows the postsynaptic excitability gradient to sort growth cones by relative presynaptic activity. More active growth cones remain closer to the right side of the array, so a growth cone's horizontal position reflects its relative level of presynaptic activity within the population.

7.3.1.4 Quiescent RGC activity

Device mismatch in the silicon retina renders some RGC circuits more excitable than others. At the bias settings required to sustain a reasonable mean firing rate within the RGC population, some RGCs never fire under any provocation. Their silent growth cones do nothing to improve their own position but at least do not interfere with the guidance of more active growth cones. At the other extreme, some RGCs are so excitable that they fire spikes even in the absence of a stimulus. Their perpetually active growth cones eventually clump together in the most excitable corner of the Neurotrope1 array, creating a global

attractor that actively disrupts the guidance of more transiently active growth cones.

Device mismatch in the Neurotrope1 chip also causes random variability in growth cone mobility, but this effect is nonsystematic and therefore less pernicious.

7.3.2 Retinotopic self-organization

To induce retinotopic self-organization within the population of growth cones projected by the same cell-type, we illuminate the retina with an sequence of stimulus images. Each image consists of a randomly centered spot of light or dark presented against a uniform gray background, which stimulates a contiguous patch of ON- or OFF-center RGCs. RGCs are coactive with neighboring RGCs of the same cell type, so in the absence of other considerations we expect their growth cones to form disjoint retinotopic projections.

7.3.2.1 Gaussian bump stimulation

We begin by attempting to induce retinotopic self-organization within a subpopulation of growth cones projected by a single cell type. We initialize the system with a coarsely retinotopic projection and present randomly centered dark spot stimuli to excite patches of OFF-center RGCs. Light spot stimuli are not presented, so ON-center RGCs remain inactive.

Figure 7.7(a) and (b) plot the target layer locations of RGC growth cones for each cell type at three different samples, colored according to the coordinates of their projecting cell bodies in the silicon retina. ON-center RGC growth cones are plotted in Figure 7.7(a) and OFF-center RGC growth cones in Figure 7.7(b).

Seeded with a coarsely topographic initial projection that evenly distributes growth

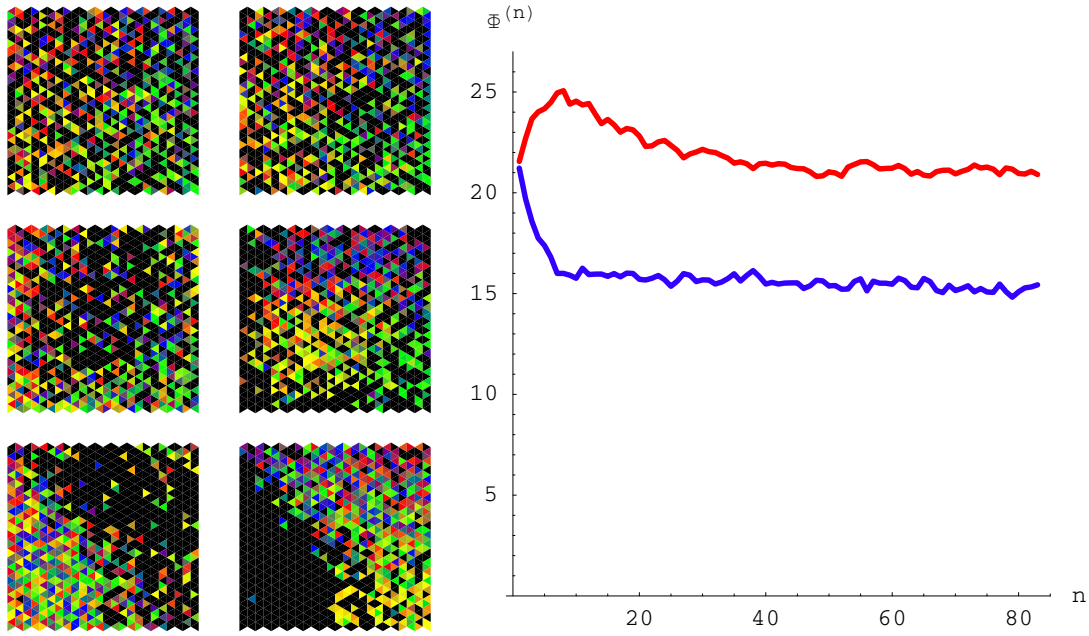


Figure 7.7: OFF-center RGC stimulation

ON- and OFF-center RGC growth cone evolution under Gaussian bump stimulation of OFF-center RGCs. (a) Target layer color maps of ON-center RGC growth cones, colored by the source layer coordinates of their cell bodies. Top: $n = 0$; middle: $n = 7$; bottom: $n = 80$. (b) Target layer color maps of OFF-center RGC growth cones. (c) Order parameter evolution for ON-center (red) and OFF-center (blue) RGC growth cone subpopulations.

cones of both cell types across the whole target layer (top color maps), the active OFF-center RGC growth cones immediately clump together in the center of the array, where they refine their internal topography at the expense of the inactive ON-center RGC growth cones, which are exiled to the edges of the array (middle color maps). Once gathered, the unitary mass of OFF-center RGC growth cones is free to drift as a whole toward the more excitable right half of the array, displacing the inactive ON-center RGC growth cones from that edge to join their compatriots on the less excitable left side. Since all of the ON-center RGC growth cones are being herded into the same place, their internal topography gradually

improves as the active OFF-center RGC growth cones continue to expel straggling ON-center RGC growth cones from the right side of the array (bottom color maps). A handful of ON-center RGCs are robustly excited by the gray background and their perpetually active growth cones congregate in the upper right corner of the array, creating a global attractor for nearby OFF-center RGC growth cones.

Figure 7.7(c) plots the evolution of the order parameters $\Phi_{\mathrm{ON}}^{(n)}$ and $\Phi_{\mathrm{OFF}}^{(n)}$ for retinotopy within each subpopulation, which are defined as the average target layer distance separating growth cones projected by adjacent RGCs of the same cell type. The initial segregation by cell type is reflected in the sharp drop in $\Phi_{\mathrm{OFF}}^{(n)}$ and rise in $\Phi_{\mathrm{ON}}^{(n)}$. As the body of OFF-center RGC growth cones shifts to the right, $\Phi_{\mathrm{ON}}^{(n)}$ returns to its initial level.

Strictly speaking, any imbalance between ON- and OFF-center RGC activity could have induced the segregation by cell type that is responsible for the sharp initial improvement in OFF-center retinotopy and gradual refinement of ON-center retinotopy. However, the subsequent maintenance and slow refinement of this retinotopy requires the topographic correlation structure of the dark spot stimulus, since the ON-center RGC growth cones are segregated to the same degree but display much coarser retinotopy. The difference between $\Phi_{\mathrm{ON}}^{(n)}$ and $\Phi_{\mathrm{OFF}}^{(n)}$ at high n is a measure of the active topographic refinement within the OFF-center subpopulation.

7.3.2.2 Center-surround stimulation

We conclude by attempting to self-organize retinotopy for both cell types simultaneously. We initialize the system with a random projection that spreads growth cones of both cell types over the entire target layer (Figure 7.8(a)). We then apply a difference of Gaussians to generate a center-surround stimulus that oscillates radially from light to

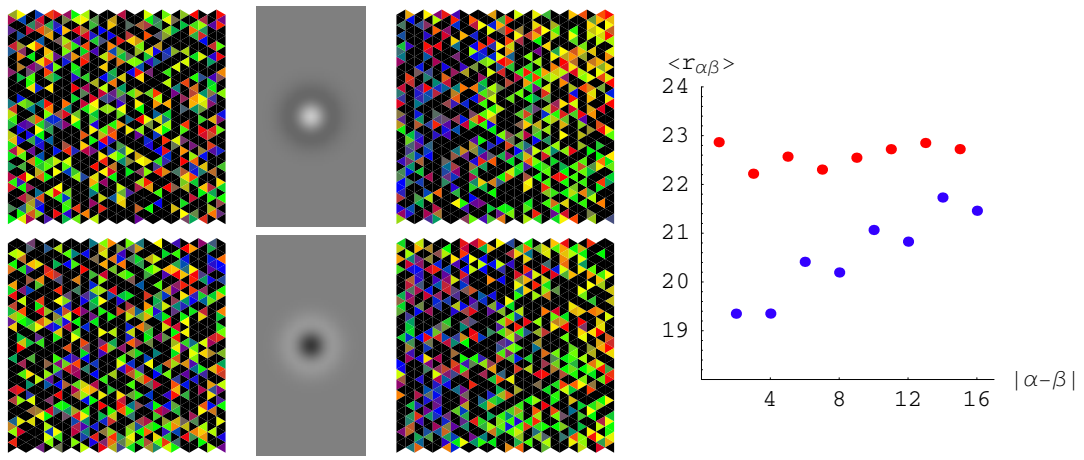


Figure 7.8: Coordinated retinotopic development of ON- and OFF-center RGC growth cones

Co-evolution of ON- and OFF-center growth cone subpopulations under oscillating spatial phase stimulation. (a) ON-center RGC growth cones (top) and OFF-center RGC growth cones (bottom) uniformly cover the entire target layer in the random initial projection. (b) The visual stimulus consists of an oscillating phase kernel presented against a gray background and centered over randomly selected regions of the silicon retina. Top: ON-center, OFF-surround; bottom: OFF-center, ON-surround. (c) Target layer color maps of ON- and OFF-center RGC growth cones at the 420th sample. (d) Average target layer distance $\langle r_{\alpha\beta} \rangle$ separating growth cones of the same cell type (blue) or the opposite cell type (red), as a function of source layer separation $|\alpha - \beta|$ of the projecting cell bodies.

dark or vice versa, coactivating neighboring ON- and OFF-center RGCs (Figure 7.8(b)). The location and polarity of each stimulus are independently drawn from a uniform distribution. The light-dark oscillation in the center-surround stimulus injects just enough coactivity between ON- and OFF-center RGCs to prevent the two cell types from segregating into disjoint retinotopic maps. Instead, both subpopulations develop and maintain coarse retinotopic maps that cover the entire target layer and are oriented in register with one another (Figure 7.8(c)).

The ability of the Neurotrope1 system to automatically extract this level of retinotopy from even the severely nonideal silicon retina stimuli is a powerful demonstration of the

robustness of the neurotropic guidance algorithm in physical implementations.

Chapter 8

Conclusion

We conclude this thesis with a discussion of our results and some suggestions for future research directions.

8.1 Discussion

Engineers have long admired the computational elegance and efficiency with which the brain effortlessly adapts to novel situations, and have made many attempts to endow conventional digital computers with a similar self-organizing capacity. Prominent examples include Grossberg's adaptive resonance theory (ART) [12] and Kohonen's self-organizing map (SOM) [59], both of which achieved popularity by distilling neurophysiological insights into a simple set of mathematical equations that require no biological knowledge to analyze and implement. However, by shrouding the messy biological details under a veil of mathematical notation, these models obscure the powerful computational role played by cellular structures such as motile growth cones.

In this thesis, we presented Neurotrope1, the first silicon model of self-organization based on growth cone migration. The Neurotrope1 system is a neuromorphic implementation of a two layer neural network with feedforward excitatory connections that automatically update their own weights based on a Hebbian learning rule. The core of the system is the Neurotrope1 chip, a full custom VLSI design that integrates axon terminals, target cells, and extracellular medium on the same silicon die. Axons connecting the two layers are implemented virtually, as soft wires stored in a set of off-chip lookup tables that can dynamically reroute address-events on the fly. The novel infrastructure developed to implement these soft wires is broadly applicable to any address-event-based system and has already been adapted for use with other neuromorphic devices [70, 2].

Our faithful adherence to cellular level details yields an algorithm that is better suited to physical implementation. All previous SOM chips have implemented a winner-take-all computation to enforce competition, requiring a global information flow that does not scale well to larger systems [35, 69, 66, 47, 99, 89, 13, 84]. By contrast, growth cones in the Neurotrope1 system employ a local information flow that scales effortlessly to larger populations. Growth cone motion is a purely cellular operation, drawing information only from the growth cone's own target site and the ring of immediately adjacent sites.

We characterized growth cone guidance in the Neurotrope1 system using a transition matrix analysis of supervised pair attraction. For this simplest nontrivial case of neurotropic attraction, neurotropic guidance fidelity is measured by a single active target cell's ability to attract and hold a single coactive growth cone. The neurotropin concentration surrounding an active target cell falls from a peak at the release site to a detection threshold at some radius σ that defines the edge of the target cell's basin of attraction. Growth cones inside this basin are guided toward the central attractor with a fidelity that depends on the local gradient magnitude. Growth cones outside the basin diffuse freely until stumbling into

the basin by chance. A captured growth cone can also be ejected from its basin by its own mistaken jumps and by random bumps from other growth cones. This non-negligible escape probability defeats any attempt to improve equilibrium performance by annealing σ.

Growth cone capture and growth cone guidance have opposite dependences on basin size. Diffusing growth cones are captured at a rate that depends on the fraction of the total target layer area covered by the basin, so growth cone capture is strongest at large σ. Larger basins have shallower internal gradients, since the concentration drop from peak to detection threshold is spread over a longer distance σ, so growth cone guidance is strongest at small σ. The optimal basin radius is the value of σ that best balances a basin's ability to guide captive growth cones with its ability to recapture escapees. This insight draws empirical support from our observation of an intermediate neurotropin spreading range that optimizes equilibrium performance in the Neurotrope1 system.

Generalizing to multiple coactive target cells, we found that the effective σ combines the spatial extent of presynaptic coactivity in the source layer with the spatial extent of neurotropin spread in the target layer. Larger presynaptic patches can compensate for shorter neurotropic spreading ranges. However, the combined guidance signal is still mediated by the neurotropin concentration, which saturates at high and low values and can only be measured with a finite precision. Consequently, the effective σ for multiple coactive target cells obeys the same optimality criterion that obtained in the single active target cell case. Since signals in any physical system are limited to a finite dynamic range, this optimality criterion is present in any hardware implementation of this algorithm, silicon or carbon.

We concluded by demonstrating the first fully hardware implementation of retinotopic self-organization. We interfaced the Neurotrope1 system directly with the silicon retina

described in [107, 108], which is the most elaborate neuromorphic model of the mammalian retina to date, mimicking thirteen different cell types in all five retinal layers and computing four distinct spike-encoded output streams whose biologically realistic activity correlations are ideally suited to program growth cone migration in the Neurotrope1 system. By illuminating the silicon retina with appropriately patterned stimuli, we induced separate populations of ON- and OFF-center retinal ganglion cells to organize their axon terminals into coordinated retinotopic maps. This is the first time every stage of the process from patterned illumination through neural map formation has been implemented entirely in hardware. The only comparable system was described in [26], which processed silicon retina data offline using a software model of neurotrophic guidance running on a workstation. Our system computes results in real-time at low power, the two prerequisites for autonomous mobile applications.

Perhaps the most surprising result of the Neurotrope1 system is that it even generates results to begin with. The sheer complexity of the system is daunting: a 48×20 array of growth cones, each containing four distinct filopodia; a 24×20 array of spiking target cells; a monolithic chip-wide transistor channel laid out in a honeycomb lattice that spans the entire array. Remember that, in a physical implementation, each added component introduces another opportunity for design error or device mismatch. For example, the measured activity of two filopodia in the same silicon growth cone can differ by over two orders of magnitude. Coupling this system with another neuromorphic device only multiplies these nonidealities. The combined silicon retina-Neurotrope1 system was rife with systematic signal corruption that would have been fatal for a more brittle learning heuristic.

Given all of these issues, the Neurotrope1 system's demonstrated ability to robustly refine retinotopy powerfully validates the neurotropic attraction algorithm for use in systems assembled from unreliable components. Such fault-tolerant algorithms have applications in

emerging fields like nanotechnology, which typically lack the process maturity to fabricate well-matched devices.

Biological algorithms operate on similarly unreliable components, so it is no surprise that several of the techniques that improve performance in the Neurotrope1 system have been proposed independently as mechanisms for biological growth cone navigation, including sample averaging to increase measurement precision [43], addition of coactive signals [41], and concentration-damped mobility [85]. The fundamental strategy is the same: expand the finite dynamic range of the guidance signal to cover a bounded physical domain with sufficient precision, satisfying an optimality criterion that is present in any physical instantiation of this system, be it the Neurotrope1 chip or the human brain.

8.2 Insights into biology

We can apply lessons learned from the Neurotrope1 system to describe an effective signalling range for biological morphogenesis, and to explain the role played by synaptogenesis during neural map development.

During the morphogenetic phase of neural map development, sparse, widely overlapping axon arbors shrink into dense, tightly packed termination zones. We can construct an effective signalling range by analogizing the neurotropin spreading range to the size of the dendritic arbors in the target tissue and the presynaptic patch size to the number of branches initially projected by each axon. All else being equal, we predict that smaller target cell dendritic arbors would require the innervating axons to elaborate more initial branches, leading to a higher proportion of pruned branches. The optimal axonal and dendritic arbor sizes would depend on the size of the target tissue.

The nature of the optimal signalling range offers some insight into the role played by synaptogenesis during biological neural map development. The fundamental constraint that limits a growth cone's performance is its ability to find and hold its optimal target. Any source of random motion, be it internal miscalculation or external competition, eventually allows a growth cone to leak away from this target. The existence of an optimal signalling range is a direct consequence of this nonzero escape probability, which forces a growth cone to balance its ability to remain at its target once found with its ability to return to its target once lost.

A biological growth cone directly attacks the escape probability by annealing its own mobility. As neurites wire themselves into their desired circuits, morphogenesis segues smoothly into a synaptogenetic phase in which neurites stabilize their connections by forming more permanent synapses. Neurites use morphogenesis to explore their surroundings for better targets and synaptogenesis to consolidate satisfactory connections. By separating the mechanisms that mediate exploration and consolidation, biological growth cones avoid the tradeoff that leads to an optimal signalling range when the two functions are mediated by the same neurotropic signal, as in the Neurotrope1 system.

8.3 Future work

Our experience with Neurotrope1 suggests several possible upgrades for incorporation into a successor system.

1. **Implementation issues**

   Flaws in the Neurotrope1 chip substantially degrade system performance. Most notably, poor power distribution creates an excitability gradient that renders almost half

the chip unusable. These errors would be easy to fix in future iterations.

2. **Neurotropin spreading kernel**

   The Gaussian shape of the neurotropin spreading kernel is suboptimal for gradient descent because the gradient drops to zero at a release site, precisely where the signal should be the strongest. Changing the kernel shape to an exponential, for example, could improve growth cone stationkeeping by marking the intended target with the peak of the guidance signal instead of its trough. The circumstances under which a change in kernel shape would improve performance require further study.

3. **Concentration-dependent mobility**

   Concentration-driven growth cones generate the most jump requests from the top of neurotropin peaks, destabilizing correctly placed growth cones and wasting bandwidth on content-free communications. Since growth cones should be less restless in regions that supply them with more neurotropin, concentration-damped mobility might be more effective. Growth cones would decelerate as they approach their targets, allowing them to consolidate their tropic gains and shift smoothly into a synaptogenetic phase. In addition, correctly placed growth cones would automatically anneal their bandwidth consumption, allowing the system to focus on improperly placed growth cones. The computational aspects of concentration-damped mobility can actually be simulated with the current system[1], but the bandwidth advantages

[1]For example, to improve gradient measurement precision, the current system accumulates multiple filopodial jump requests before executing a change in growth cone position. To store these requests, the system includes enough memory to associate each filopodium with a digital count that increments every time the filopodium requests a jump. The jump request of the first filopodial incrementor to overflow is serviced and the rest of the jump requests for that growth cone are dropped. To switch from concentration-driven to concentration-damped jump generation, each growth cone could be supplemented with a single bidirectional count in memory that encodes neurotropin concentration. Each presynaptic address-event routed to a growth cone would increment its counter, while each filopodial jump request would decrement it by some constant. If its counter overflowed, the growth cone would service its most frequently requested jump, as directed by its three filopodial incrementors. If its counter underflowed, the growth cone would clear all outstanding jump requests. As a result, the growth cone would execute a stream of jumps at a rate that is based on its

would require a new chip.

4. **Synaptogenesis**

   The best way to systematically improve performance would be to deepen the potential well of the attractor basin, preventing growth cones from escaping or being displaced through random motion. Potential wells can be deepened by improving a growth cone's gradient sensitivity or by enhancing its resistance to external displacement. In the limit of infinitely deep wells, we could even dispense with gradient-directed growth cone guidance entirely, since growth cones could be permanently localized to arbitrarily small attraction basins, as if they had formed stable synapses onto their attractors.

5. **Multiple arbor branches**

   One-to-one connectivity is fairly unforgiving, allowing each source cell only a single guess for its axonal target. A more distributed representation might yield better performance. For example, each source cell could project several growth cone-tipped axon branches into the target layer, each of which could move independently. Acceptable performance could be achieved if most of the branches clustered at the correct location. Given a mechanism to identify misplaced branches, such as concentration-damped mobility, hopelessly lost branches could be pruned and correctly placed branches could be reinforced.

6. **Multichip systems**

   Spike-encoded data is communicated as address-events, making them compatible with a family of neuromorphic devices [70, 107, 17, 2] that can be used to drive the system with correlated input activity from a variety of modalities. In particular,

presynaptic activity and attenuated by its neurotropic uptake, as measured by the frequency of filopodial jump requests.

the ability of the silicon retina to process arbitrary visual stimuli into growth cone-compatible spike trains in real-time opens the door to more ambitious experiments like using natural video correlations to drive growth cones to organize more complicated visual feature maps.

In larger systems, target cells will most likely be located on their own neuromorphic chips, not on the neurotropin chip itself. A future iteration might include only the growth cones, the extracellular medium, and the postsynaptic neurotropin release sites, receiving both presynaptic and postsynaptic spikes from off-chip. This generic axon chip would be a temporary plug-in module that could be inserted between neuromorphic chips to developmentally wire their connections, and removed upon completing the mature circuit.

8.4 Conclusion

The ability of the Neurotrope1 system to self-organize coarse retinotopy from even the highly nonideal input activity generated by the silicon retina is a powerful demonstration of the robustness of the neurotropic guidance algorithm against the vicissitudes of physical implementation. As such, this technique represents a practical addition to the neuromorphic toolbox.

Appendix A

Neurotrope1 system description

This appendix describes the Neurotrope1 system at the circuit board level. The Neurotrope1 system is divided between two circuit boards: the *router board*, which translates source-indexed address-events into target-indexed address-events using a lookup table called an *axon map*; and the *update board*, which computes axon map updates using the Neurotrope1 chip. The board level signal flow is diagrammed in Figure A.1.

A.1 Router board

The router board receives and transmits address-events through six independent asynchronous *address-event ports* (of which three are shown in Figure A.1). Source-indexed address-events are translated into target-indexed address-events using the axon map stored in an asynchronous *random access memory* (RAM) that is controlled by a *complex programmable logic device* (CPLD). Individual elements of the RAM can be written or read by a *microcontroller* (uC) acting through the CPLD. The microcontroller also handles com-

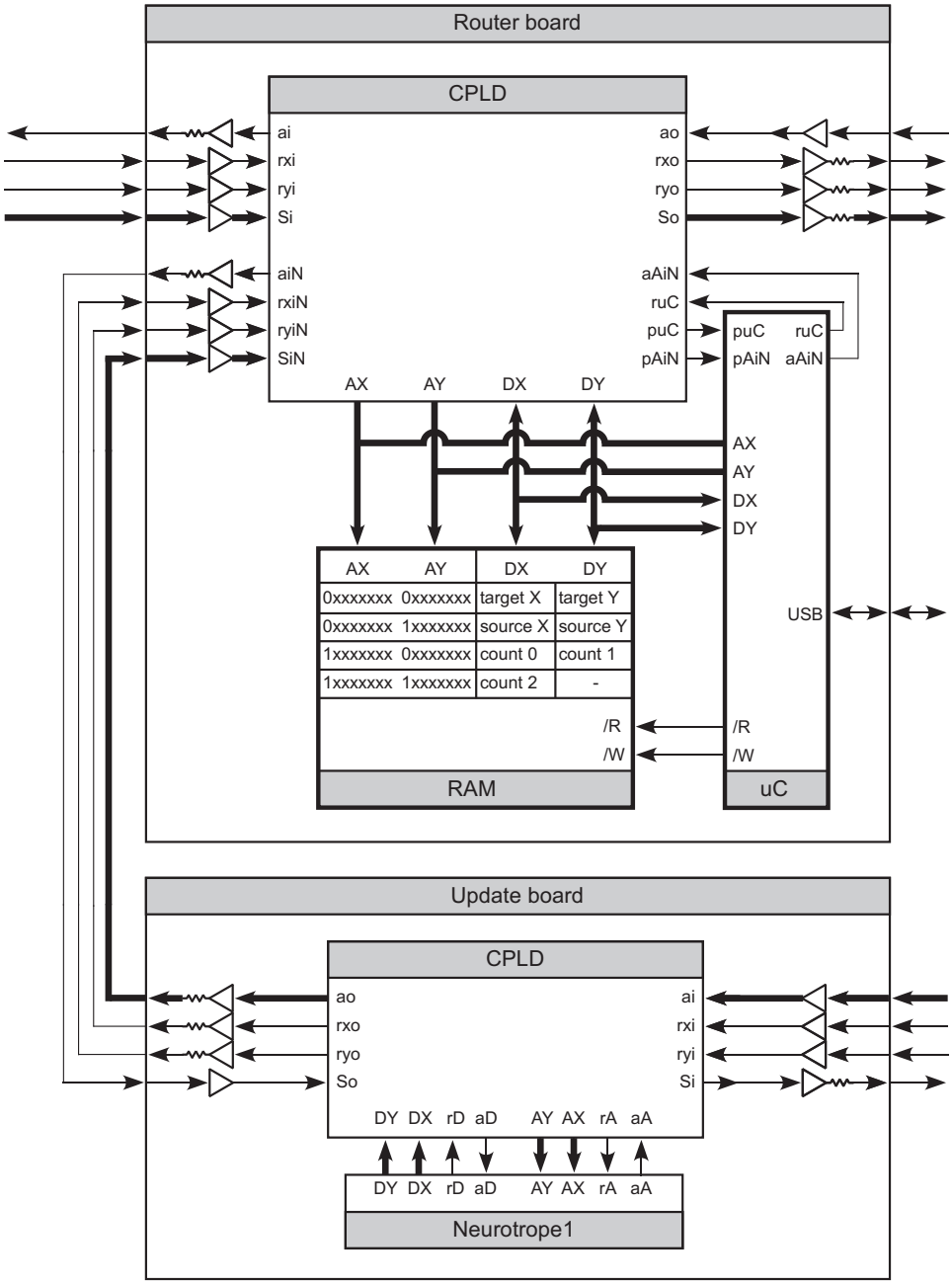


Figure A.1: System signal flow

The router board uses an axon map to translate source-indexed address-events into target-indexed address-events and forwards them to their destination. The update board transmits axon map updates to the router board based on target-derived feedback.

munication with an external computer through a *universal serial bus* (USB).

The router board performs three operations: the *route cycle*, which translates source-indexed address-events into target-indexed address-events; the *read/write cycle*, which reads or writes axon map entries based on USB communications with the external computer; and the *update cycle*, which modifies axon map entries based on feedback signals from the update board.

A.1.1 Address-event ports

The router board implements six independent address-event ports, three of which receive address-events and three of which transmit address-events. Each address-event port is composed of eight data lines and three control lines, according to the word-serial address-event protocol described in [6]. Control lines are tied to their inactive logic levels through $3.3k\Omega$ resistors. Every port output is driven with an LVT573 buffer whose output is series terminated by a 50Ω resistor in order to attenuate cable reflections. Similarly, every port input is routed through an LVT573 buffer to regenerate the signal. The LVT logic family operates at 3.3V but can tolerate logic inputs ranging from 2.0V to 5.5V, so the input buffers also insulate more sensitive and expensive board components from potentially dangerous line noise, particularly when interfacing with external components that utilize higher supply voltages.

The Neurotrope1 system requires three of the six address-event ports for normal operation. The *source port* receives address-events and passes them to the CPLD, which translates the source address into its target address using the axon map stored in the asynchronous RAM. The CPLD repackages this address as an address-event and routes it to the *target port* for transmission to the next stage in the system. The *update port* receives

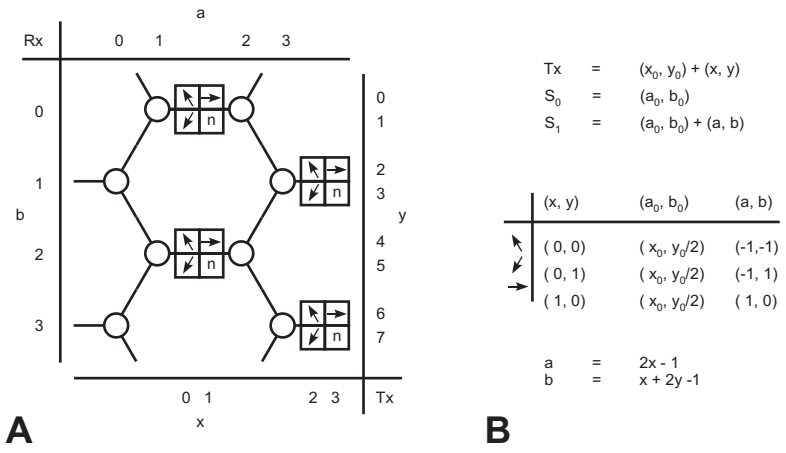


Figure A.2: Growth cone swaps

(a) Input and output address spaces. Receiver addresses (Rx) index growth cones by row and column. Transmitter addresses (Tx) index event-generators by row and column and use the lowest row and column bits to specify the event type. Half of the allotted receiver and transmitter addresses are unoccupied. (b) Transmitter address-events Tx can be decomposed into a base address (x_0, y_0) and an offset (x, y), as can the addresses of the two growth cones S_0 and S_1 to be swapped. The growth cone base address (a_0, b_0) and offset (a, b) are computed from (x_0, y_0) and (x, y).

address-events from the update board that encode feedback-derived modifications to the axon map. In addition, an optional *monitor port* can be configured to echo address-events to an external monitor controller or logic analyzer for visualization or data collection.

A.1.2 Microcontroller

A Ubicom ip2022 microcontroller handles axon map updates and USB communication with an external computer. Address-events arriving at the update port are passed through the CPLD to the microcontroller, which interprets the instructions encoded in the address-events and executes the requisite updates to the axon map stored in the RAM.

A.1.2.1 Axon map updates

The Neurotrope1 chip transmits address-events of four different types (Figure A.2(a)). The first type is the standard source-indexed neuron spike. The other three types are *swap requests*, which are instructions to update the axon map by swapping adjacent growth cones along one of three directions: diagonally up, diagonally down, or horizontally.

Each address-event can be written as the sum of a base address (x_0, y_0) that encodes row and column position in the array, and an offset (x, y) that encodes address-event type. Swap request address-events uniquely encode the identity of the two growth cones S_0 and S_1 to be exchanged. The S_1 address is obtained by adding an offset (a, b) to the S_0 address (a_0, b_0). The growth cone base address (a_0, b_0) and offset (a, b) are computed from the swap request base address (x_0, y_0) and offset (x, y), as outlined in Figure A.2(b).

The microcontroller maintains a count for every growth cone of the number of swap requests received in each direction. Each swap request modifies the swap counters for both of its growth cone participants by incrementing the swap counter corresponding to its own direction and clearing the swap counters for the two opposing directions. The microcontroller only executes an exchange when one of the swap counters exceeds a software-defined threshold, which requires an unbroken chain of consecutive swap requests in the same direction. Each direction has a different swap threshold, which is selected to balance the layout anisotropy described in Chapter 3.

A.1.2.2 USB communication

The microcontroller also executes commands received via USB from the external computer. During a write command, the microcontroller receives a two byte address and two

bytes of data and writes the data to the memory cell in the RAM indexed by the address. During a read command, the microcontroller receives an address from the computer, reads the data in the indexed memory cell, and transmits the data back to the computer.

The microcontroller code implementing the USB protocol is adapted from the standard software library provided by Ubicom, Inc. This code generates control and data signals for the Phillips Semiconductor PDIUSBP11A USB controller, which drives the USB bus. On the computer end, a graphical user interface program is written in C++ around the USBIO driver provided by Thesycon Gmbh.

A.1.3 RAM

The axon map is stored in an ISSI IS61LV25616-10T single-chip asynchronous RAM which can store up to 256K $\times$ 16 bits of data, of which only 64K is utilized by the router board, which grounds the highest two address bits. The remaining sixteen address bits are grouped into two eight bit address busses, AX for X addresses and AY for Y addresses. Data stored in the memory cell indexed by the address presented on AX and AY is input or output on sixteen data bits, which are similarly grouped into data busses DX and DY.

The RAM address space is further divided into four pages, indexed by the highest X and Y address bits. The first page stores the *forward map*, which is a list of target addresses indexed by source address. The second page stores the *reverse map*, which is a list of source addresses indexed by target address. The third and fourth pages store *data maps*, which are lists of data fields indexed by source address. The data fields on the third page contain the eight bit diagonal swap counters for each growth cone and the data fields on the fourth page contain the eight bit horizontal swap counter. Only the forward map is required for address-event translation; the contents of the other three pages are used for map updates.

The asynchronous RAM is normally operated in read mode, meaning that the data busses DX and DY are continually driven with the contents of the memory cell indexed by the address busses AX and AY. In write mode, the RAM places the contents of DX and DY into the memory cell indexed by AX and AY. In each case, after the initial address presentation, a setup time under 10 ns is required before the operation takes effect. Read mode and write mode are toggled by separate active-low enables, $/R$ and $/W$, that are controlled by the microcontroller.

The source port, update port, and microcontroller all compete for write access to the address busses, and this competition is mediated by an arbiter module in the CPLD.

A.1.4 CPLD

The CPLD controls the address-event ports and the RAM using four internal asynchronous logic modules: an *s2p module*, which converts word-serial address-events into word-parallel address-events; an *arbiter module*, which controls write access to the address bus of the RAM; a *delay module*, which establishes the necessary setup time for the RAM; and a *p2s module*, which converts word-parallel address-events into word-serial address-events (Figure A.3). All internal busses use dual rail data and all external interface busses use bundled data. A brief description of each module follows.

A.1.4.1 s2p module

Under the serial address-event protocol, an upward transition on the input Y request ryi indicates the presence of a Y address on the eight bit bundled data input bus Si. The s2p module latches this Y address on the dual rail output bus AYi and acknowledges receipt by raising the input acknowledge ai. The address-event source then drives Si with an

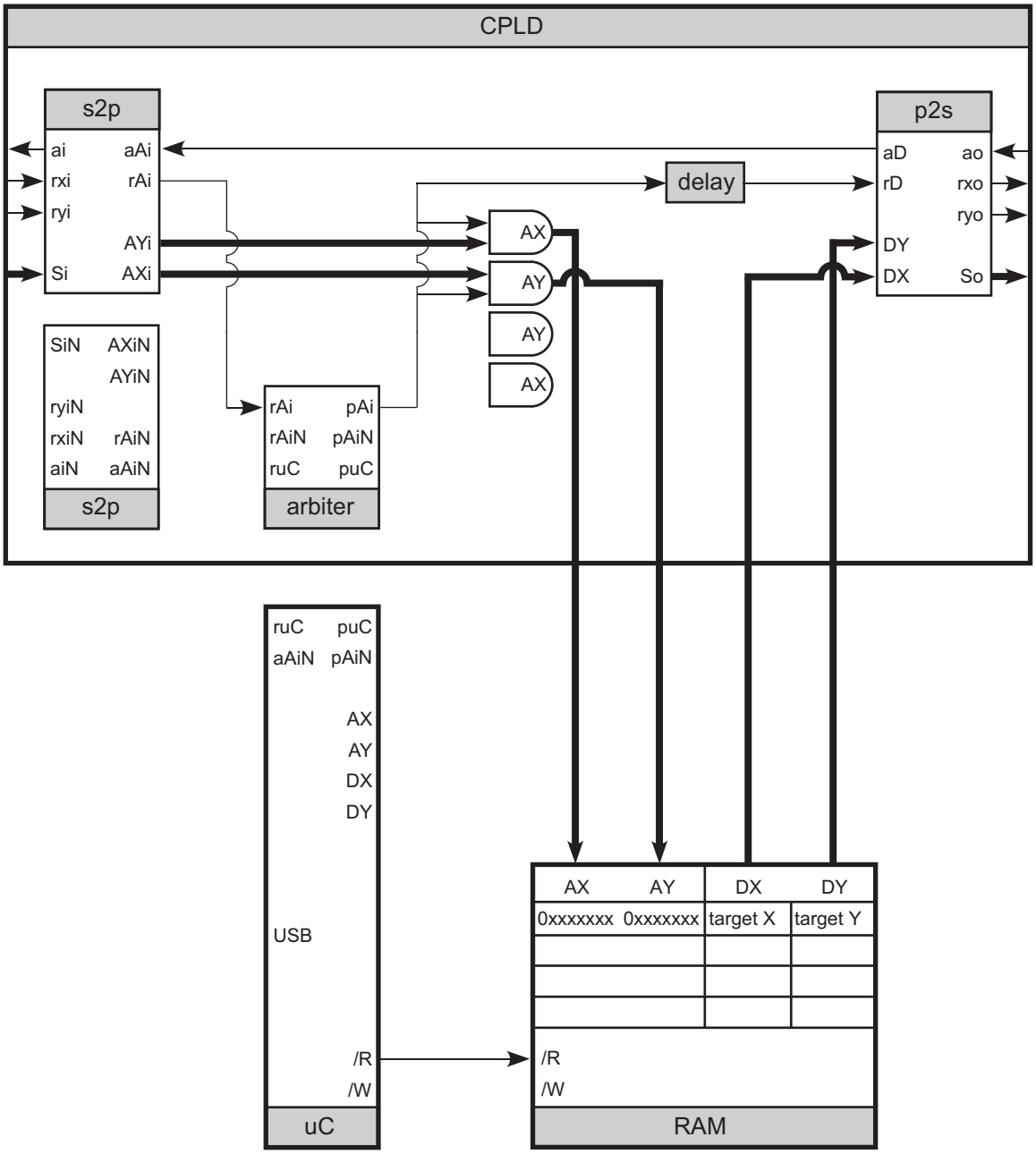


Figure A.3: Route cycle

Source-indexed address-events arriving at the source port are translated through the axon map stored in the RAM into target-indexed address-events that are retransmitted through the target port.

X address and lowers the active-low input X request rxi. The s2p module latches the X address on the dual rail output bus AXi and lowers ai. The full X-Y address is now present on the output busses, so the s2p module raises the output request rAi, signalling the presence of data to the next stage in the system. The s2p module then waits for the output acknowledge aAi to go high, clears the dual rail data from AX, lowers rAi, waits for aAi to go low, and finally waits for rxi to go high before raising ai.

At this point, the address-event source can either lower rxi, beginning another spike in the same row, or lower ryi, ending the serial address-event burst. If rxi goes low, the s2p module again latches the X address presented on Si into AX and raises rAi, beginning another parallel address-event. If ryi goes low, the s2p module clears AY and lowers ai, terminating the burst.

There are actually two s2p modules in the router CPLD, one associated with the source port and one associated with the update port. The modules are internally identical and are distinguished externally by appending the letter N to the signal names of the update port s2p module.

A.1.4.2 arbiter module

The arbiter module controls write access to the RAM address bus. It can receive a bus access request rAi from the s2p module of the source port, a request $rAiN$ from the s2p module of the update port, and a request ruC from the microcontroller. It returns one of three permission signals at a time; pAi grants bus access to the source port, $pAiN$ grants bus access to the update port, and puC grants bus access to the microcontroller. For simultaneous requests, permission is granted first to the microcontroller, then the update port, then the source port. Regardless of origin, a new request cannot preempt an existing

permission signal, which persists until its own request is cleared.

A.1.4.3 delay module

The delay module receives an input signal and echoes it to its output after some delay, which is chosen to exceed the setup time of the RAM. The CPLD implements a constant delay by routing the signal off-chip, through an LVT573 buffer, and back into the CPLD, a process which takes on the order of 10 ns. This delay can be extended by running the signal through the buffer more than once.

A.1.4.4 p2s module

The p2s module converts a dual rail parallel address-event into a bundled data serial address-event. Upon receiving an input request rD, the p2s module drives its bundled data output bus So with the address on the dual rail input bus DY and raises the output Y request ryo. When the output acknowledge ao goes high in response, the p2s module drives So with the address on the dual rail input bus DX and lowers the output X request rxo. When ao goes low, the p2s module clears So, raises rxo, waits for ao to go high, raises the input acknowledge aD, waits for rD to go low, lowers ryo, waits for ao to go low, and finally lowers aD, terminating the address-event.

A.1.5 Route cycle

The route cycle translates a source-indexed address-event into a target-indexed address-event, as diagrammed in Figure A.3.

An address-event arriving at the source port is first converted from serial address-event

format into parallel address-event format by the s2p module, which then raises rAi to requests permission from the arbiter module to access the RAM address busses AX and AY. Once the arbiter module grants permission by raising pAi, the parallel address is presented to the asynchronous RAM which outputs the target address stored in the addressed memory cell onto its data bus after some setup time, which is implemented by running the permission signal pAi through the delay module and thence to the p2s module, which bundles the target address as a serial address-event and transmits it off-board through the target port. Upon transmission, the p2s module raises aD to acknowledge the s2p module, which lowers rAi to return bus control to the arbiter module. The arbiter module lowers pAi, causing the p2s module to lower aD, terminating the cycle.

A.1.6 Read/write cycle

The external computer reads and writes to the RAM by sending commands via USB to the microcontroller, which implements these read/write commands as diagrammed in Figure A.4.

The microcontroller requests write access to the RAM address busses AX and AY from the arbiter module of the CPLD by raising ruC. When the arbiter module grants access by raising puC, the microcontroller drives AX and AY with the address supplied by the external computer.

If the computer sent a write command, the microcontroller now raises the RAM read enable $/R$ and clears the RAM write enable $/W$, which are both active low. It then writes the data supplied by the computer to the RAM data busses DX and DY. After waiting out the setup time, the microcontroller raises $/W$, clears DX and DY, clears AX and AY, and lowers $/R$. It then lowers ruC, returning bus control to the arbiter module, which lowers

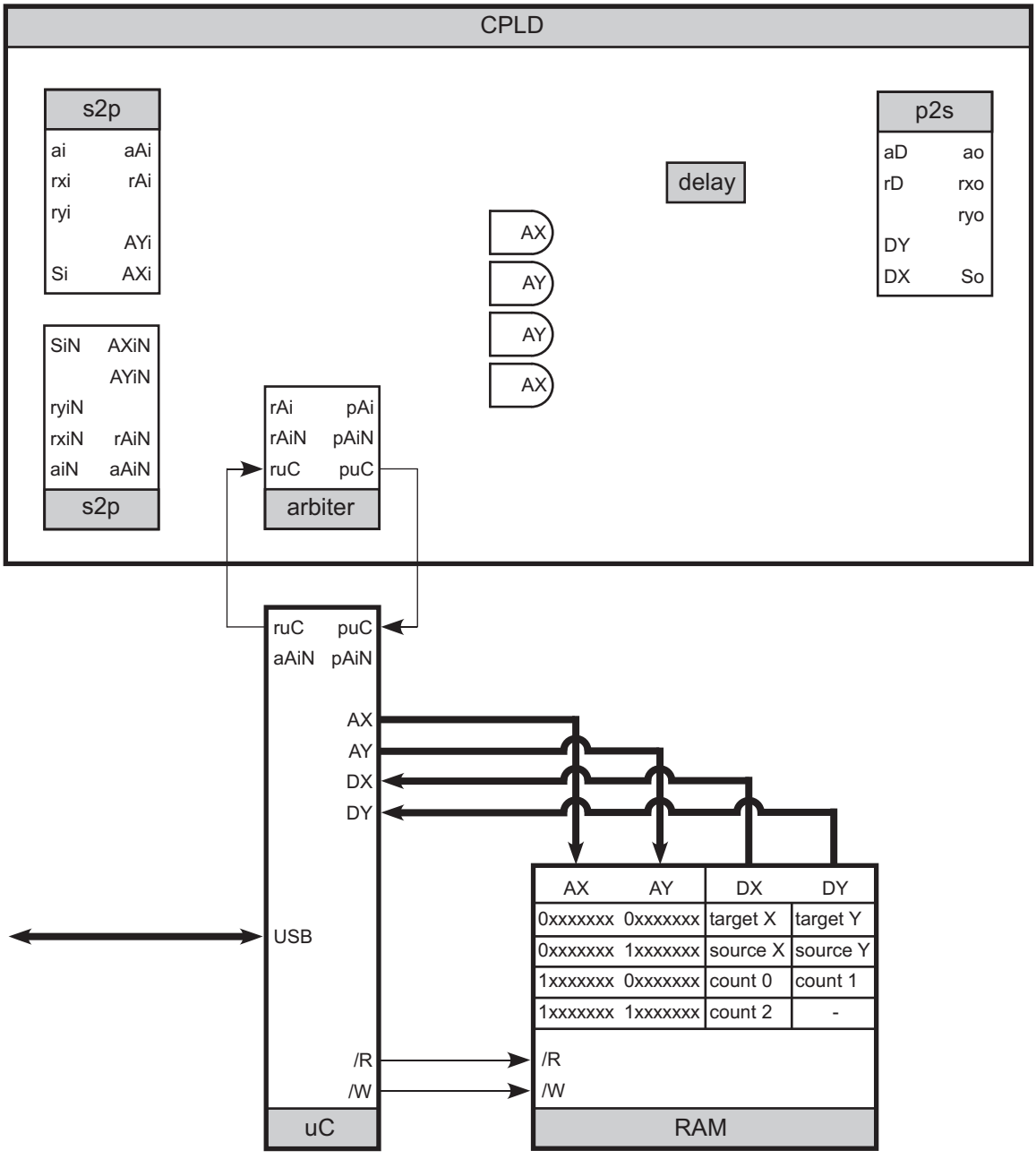


Figure A.4: Read/write cycle

The external computer sends commands via USB to the microcontroller to read or write to the RAM.

puC.

If the computer sent a read command, the microcontroller simply waits out the setup time and then reads the contents of DX and DY into an internal buffer for transmission via USB to the computer. It then lowers ruC, causing the arbiter module to lower puC.

A.1.7 Update cycle

The update cycle uses the microcontroller to interpret and execute the instruction encoded in an address-event arriving at the update port, as diagrammed in Figure A.5.

The s2p module associated with the update port converts an arriving serial address-event into parallel format and requests RAM address bus access from the arbiter module by raising $rAiN$. The arbiter module grants bus access by raising $pAiN$, which is forwarded to the microcontroller to request update handling. The microcontroller responds by reading the contents of AX and AY, which are now being driven by the parallel address-event from the update port. If the address corresponds to a swap request, the microcontroller requests bus control by raising ruC. It then acknowledges receipt of the instruction by raising $aAiN$, causing the s2p module to lower $rAiN$ and release AX and AY. Since the arbiter gives priority to the microcontroller, it lowers $pAiN$ and immediately raises puC, granting the microcontroller bus control.

The microcontroller then executes a sequence of RAM reads and writes to read and update the affected swap counters and, if necessary, to exchange entries in the forward and reverse maps. Upon completion, it returns bus control to the arbiter module by lowering ruC. The arbiter module lowers puC and waits for the next request.

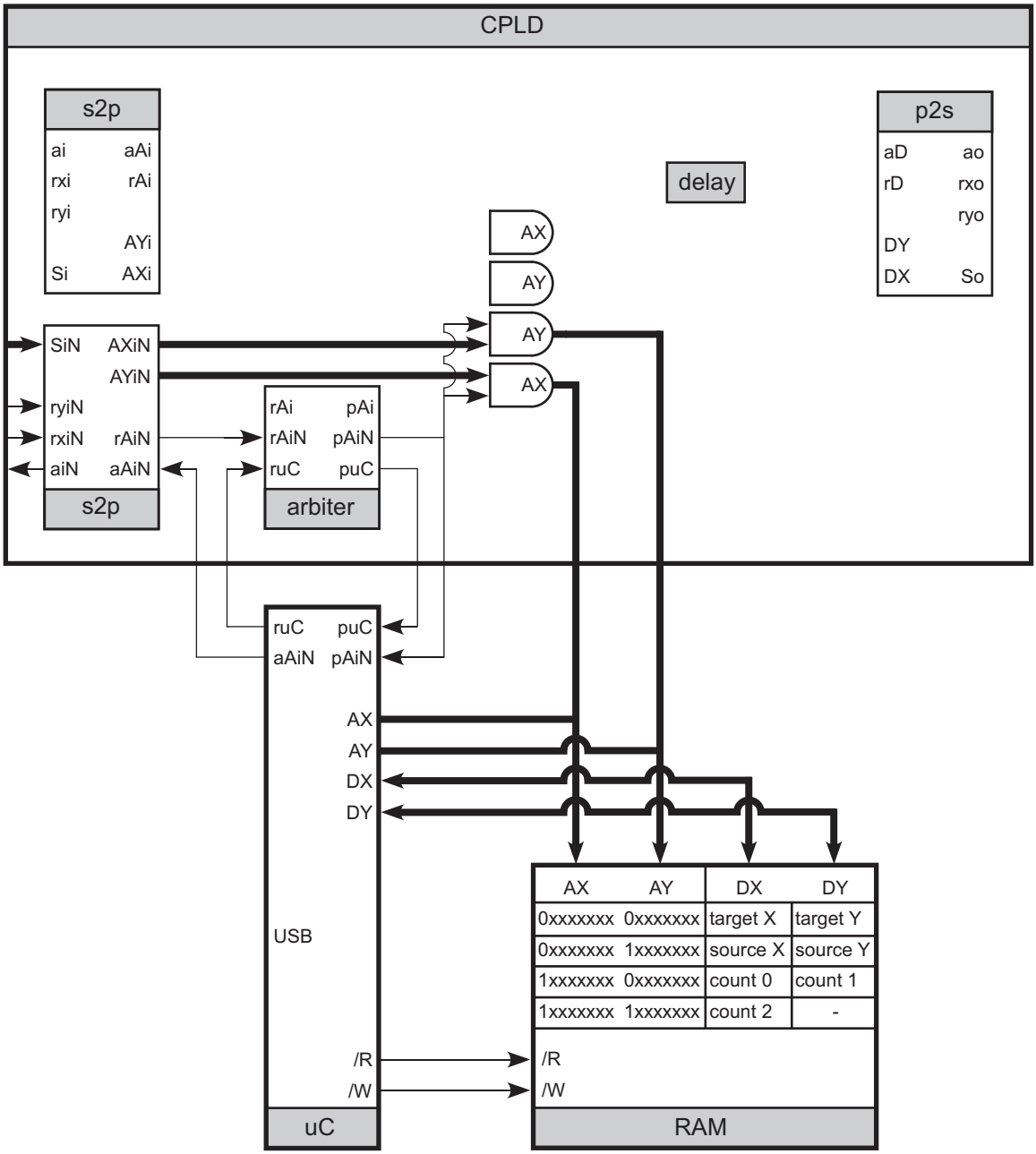


Figure A.5: Update cycle

Address-events arriving at the update port are interpreted by the microcontroller.

A.2 Update board

The update board contains the Neurotrope1 chip and its associated logic and analog components. The Neurotrope1 chip communicates using a parallel address-event protocol that has been superseded by the current serial address-event protocol and therefore all address-events must be filtered through a CPLD that converts serial address-events into parallel address-events and vice-versa.

A.2.1 Address-event ports

The update board includes one input serial address-event port and one output serial address-event port. All port lines are buffered through an LVT573 and output lines are series terminated with 50Ω resistors.

A.2.2 Neurotrope1

The Neurotrope1 chip receives source neuron spikes from the CPLD and returns target neuron spikes and growth cone updates. Layout errors in the receiver circuit inverted the sense of the input signals and disabled the normal handshake signals. As a result, addresses presented to the Neurotrope1 chip must be active low, and the address-event handshake is simulated by tying the receiver request to the receiver reset and removing it without acknowledgment after some delay, implemented by routing the receiver request to the input of an LVT573 buffer and using the buffer output as the receiver acknowledge. The transmitter is implemented normally.

All analog biases are set off-chip by $10k\Omega$ potentiometers. Current for source biases is supplied by LM6464 amplifiers. Board power is supplied through a DC jack and controlled

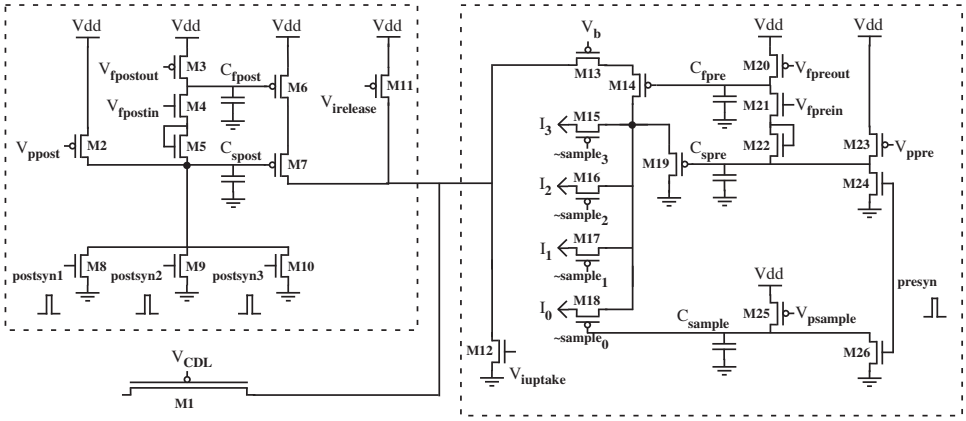


Figure A.6: Neurotropin circuit

Neurotropin sample, diffuse, and uptake circuits. Target cell spikes inject charge (left box) into a monolithic pFET channel (M1), where it diffuses until being passively decayed (M12) or actively sampled by local growth cone circuits (right box).

by five separate LM1086 3.3V voltage regulators, one for the analog circuits in the pixel array, one for the digital circuits in the pixel array, one for the address-event receiver, one for the address-event transmitter, and one for the output pads. The CPLD and buffers draw power from the transmitter voltage regulator, while the amplifiers draw power from the analog circuit voltage regulator.

A.2.2.1 Sampling requires fanout

A growth cone participates in four charge sampling circuits, one for its own lattice node $\mathbf{x}$, and one for each of the three adjacent lattice nodes $\mathbf{x} + \mathbf{f}$. In each charge sampling circuit, reproduced in Figure A.6, a pulse at node $\mathbf{x}$ pulls the voltage on the facilitation capacitor C_{fpre} low, diverting charge from the diffusive lattice M1 through M13 and M14 into a node-specific charge pool that is shared by the four adjacent sample circuits. Sample currents $I_{\mathbf{f}}$ are drawn from this charge pool, gated by pulses $\sim sample_{\mathbf{f}}$ that are generated at nodes $\mathbf{x} + \mathbf{f}$. Unused charge remaining in the pool is drained to ground through M19.

Since the charge pool is filled by a pulse at node $\mathbf{x}$ but the sample current is gated by a pulse at node $\mathbf{x} + \mathbf{f}$, charge transfer from $\mathbf{x}$ to $\mathbf{x} + \mathbf{f}$ requires two address-events. First, an address-event must arrive at the filopodial node $\mathbf{x}$ in order to fill the charge pool at $\mathbf{x}$ by pulling C_{fpre} low. Second, an address-event must arrive at the growth cone node in order to generate the specific filopodial sample pulse $\sim sample_{\mathbf{f}}$ that diverts charge from the pool onto the filopodial integration capacitor.

Growth cone neurotropin sampling therefore requires four address-events, one at the growth cone and one at each of its filopodia. Every growth cone-targeted address-event arriving at the update board must be fanned out to all four filopodial locations. This fanout is implemented by the CPLD.

A.2.3 CPLD

The update board CPLD uses seven modules to process address-events entering and exiting the Neurotrope1 chip, as diagrammed in Figure A.7. The s2p module and the p2s module are the same as those used in the router board CPLD. The *filter module*, *sequence module*, *addendX module*, *addendY module*, and *adder module* are described below.

A.2.3.1 filter module

Growth cones in the Neurotrope1 chip occupy nodes on a honeycomb lattice, which can be decomposed into dual triangular lattices (Figure A.8(a)). Nodes are assigned to each lattice according to the relative positions of their nearest neighbors. Type A nodes have one neighbor to the right and two to the left, while type B nodes have one neighbor to the left and two to the right. Node types are readily distinguished in the rectangular input address space because type A nodes all have even column addresses and type B nodes all

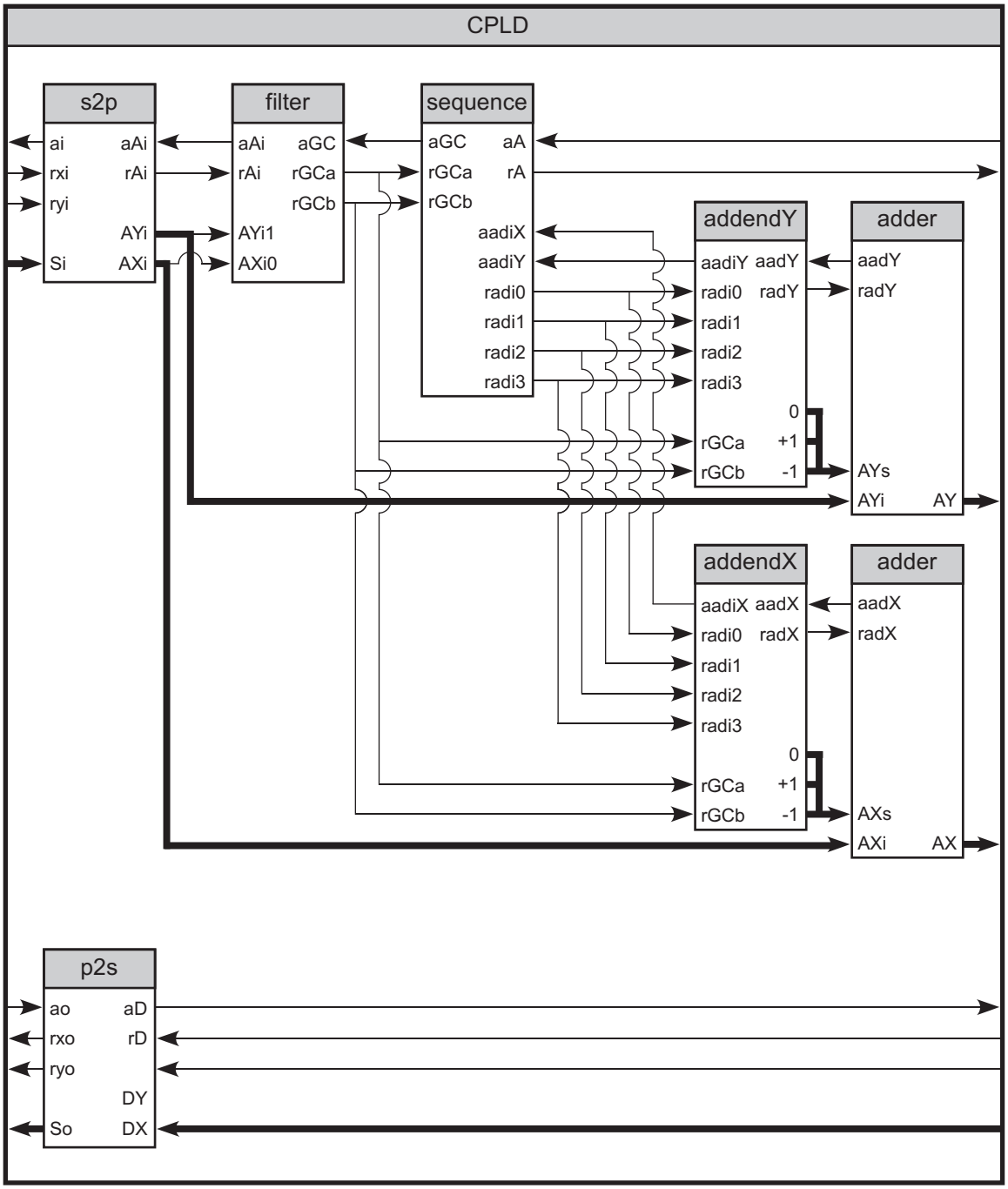


Figure A.7: Update board CPLD signal flow

The update board CPLD fans out growth cone address to four adjacent filopodial locations by adding an offset to the original address.

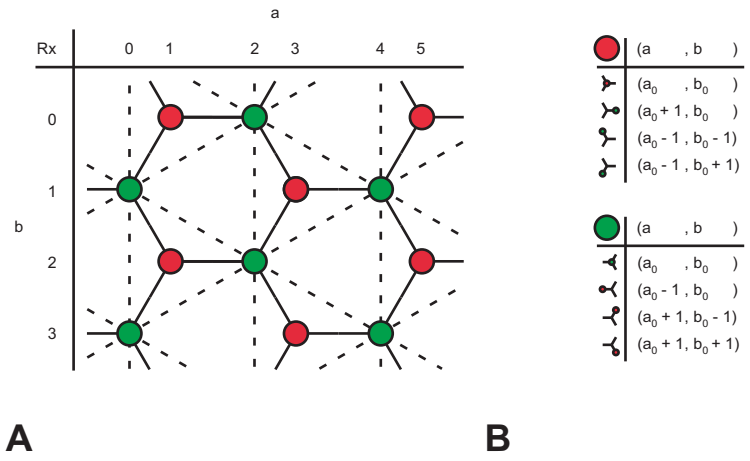


Figure A.8: Honeycomb lattice

(a) A honeycomb lattice can be decomposed into dual triangular lattices. Type A nodes (red) occupy one triangular lattice and type B nodes (green) occupy the other. (b) Fanout for type A nodes (top) and type B nodes (bottom).

have odd column addresses.

The filter module examines the lower address bits of a parallel address-event to determine the node type of the encoded growth cone address. When the input request rAi goes high, the filter module raises one of two output requests, $rGCa$ for type A addresses or $rGCb$ for type B addresses. The filter module then waits for the output acknowledge aGC to go high before removing the output request and acknowledging the input by raising the input acknowledge aAi, which is lowered after rAi goes low.

A.2.3.2 sequence module

The sequence module is a state machine that successively executes four different parallel address-event handshake cycles, one for each node in the filopodial fanout of the growth cone-targeted address-event.

When the sequence module receives an input request of either node type on $rGCa$ or $rGCb$, it raises the first fanout request $radi0$, which is sent to the addendX and addendY modules. The sequence module then waits for the fanout acknowledges $aadiX$ and $aadiY$ from both addend modules, indicating that the filopodial X and Y addresses have been prepared. The sequence module then raises the output request rA, sending the filopodial address into the Neurotrope1 chip.

When the output acknowledge aA goes high, the sequence module lowers $radi0$ and waits for $aadiX$ and $aadiY$ to go low before lowering rA. When aA goes low in response, the sequence module begins the next fanout event by raising the second fanout request $radi1$.

After completing the fourth fanout address-event, the sequence module raises the input acknowledge aGC and waits for the input request to go low before lowering aGC, ending the filopodial fanout sequence.

A.2.3.3 addendX/Y module

Filopodial addresses are computed by adding offset constants to the growth cone X and Y addresses. The specific constants depend on the node type of the growth cone and the identity of the filopodium (Figure A.8(b)). The addendX module accepts filopodial requests $radi0$, $radi1$, $radi2$, and $radi3$, and growth cone node types $rGCa$ and $rGCb$, and calculates the carry in bit and eight bit addend for a binary addition operating on the growth cone X address. The addendY module does the same for the growth cone Y address.

For type A growth cones ($rGCa$ high), the addendX module outputs an addend of 0x00 and carry in bit of 0 in response to the first filopodial request $radi0$, 0x01 and 0 for $radi1$, 0xfe and 1 for $radi2$, and 0xfe and 1 for $radi3$. For type B growth cones ($rGCb$ high), the

addendX module outputs 0x00 and 0 for $radi0$, 0xfe and 1 for $radi1$, 0x01 and 0 for $radi2$, and 0x01 and 0 for $radi3$.

The addendY module outputs identical values for type A and B growth cones: 0x00 and 0 for $radi0$, 0x00 and 0 for $radi1$, 0xfe and 1 for $radi2$, and 0x01 and 0 for $radi3$.

Upon completing an addend and carry in bit computation, the addendX module raises its output request $radX$, telling the adder module to begin its operation. When the addition is complete, the adder module raises the acknowledge $aadX$, which the addendX module passes back to the sequence module by raising $aadiX$. When the sequence module removes the filopodial request, the addendX module clears its output and lowers $radX$. It waits for $aadX$ to go low before lowering $aadiX$, completing the filopodial handshake cycle. The addendY passes its output to its adder module in the same way.

A.2.3.4 adder module

The adder module receives two addends and a carry in bit and outputs the results of a binary addition. There are two identical adder modules in the update board CPLD, one operating on the X address and one operating on the Y address.

The X address adder module executes its addition when it receives an input request from the addendX module on $radX$, indicating the presence of data on the addend and carry in bits AXs. The other addend AXi is the original growth cone X address output by the s2p module. The adder module outputs the sum AX and raises the acknowledge $aadX$. When $radX$ goes low, the adder module clears AX and lowers $aadX$.

A.2.4 Fanout cycle

In the fanout cycle, an arriving address-event is fanned out to its encoded target and the three immediately adjacent sites. The CPLD receives a growth cone address (a, b) and transmits a train of filopodial addresses (a, b), $(a \pm 1, b)$, $(a \mp 1, b - 1)$, $(a \mp 1, b + 1)$ to the Neurotrope1 chip.

An address-event arriving at the CPLD is first converted from serial to parallel format by the s2p module, which passes the parallel address to the filter and adder modules. The filter module examines the growth cone address to determine its node type and then raises one of two request lines, $rGCa$ for type A nodes or $rGCb$ for type B nodes. The sequence module accepts these requests and generates four consecutive filopodial address-events. It begins each filopodial address-event by raising one of four filopodial requests, $radi0$, $radi1$, $radi2$, or $radi3$, which collectively act as a one-hot encoding of filopodial identity. The addendX and addendY modules compute the offset constants AXs and AYs to be added to the growth cone X and Y addresses AXi and AYi, based on the filopodial identity encoded in $radi0$, $radi1$, $radi2$, and $radi3$, and the growth cone node type encoded in $rGCa$ and $rGCb$.

The addendX and addendY modules send their computed offsets to their respective adder modules, which add the offsets to the growth cone X and Y addresses and present the resulting filopodial X and Y addresses AX and AY to the Neurotrope1 chip. To begin the output address-event handshake cycle, the adder modules acknowledge the addendX and addendY modules, which pass the acknowledges back to the sequence module. When the sequence module has received both acknowledges $aadiX$ and $aadiY$, indicating that both AX and AY are ready, it raises the output request rA, telling the Neurotrope1 chip to latch the address.

When the Neurotrope1 chip raises the output acknowledge aA, the sequence module removes the filopodial request, causing the addend and adder modules to clear their outputs. When both addend acknowledges $aadiX$ and $aadiY$ have cleared, the sequence module either raises the next filopodial request or, if the last request $radi3$ has already been raised, acknowledges the filter module on aGC. The filter module passes the acknowledge back to the s2p module, which clears its output and acknowledges its input, completing the fanout cycle.

A.2.5 Output cycle

In the output cycle, a parallel address-event from the Neurotrope1 chip is converted to serial address-event format by the p2s module and transmitted to the output port.

Appendix B

Neurotropin spreading circuit

In this appendix we describe the transistor circuit that implements the neurotropin spreading kernel in the Neurotrope1 chip.

The Neurotrope1 chip models extracellular neurotropin spread as current flow in a discrete transistor lattice. A target cell circuit injects current into the network by tying a lattice node to the supply voltage. A fraction of this input current is shunted to ground by an activity-independent current sink, and the remaining current either charges the node capacitance or flows to adjacent nodes. Each lattice node can shunt only a small fixed current, so neurotropin spreads laterally from its release site to cover as many nodes as are required to sink the injected current. The size of the fixed shunt current is controlled by the gate voltage V_{shunt} shared by the leak transistors at every node. The higher V_{shunt}, the larger the shunt current and the shorter the range to which neurotropin can spread.

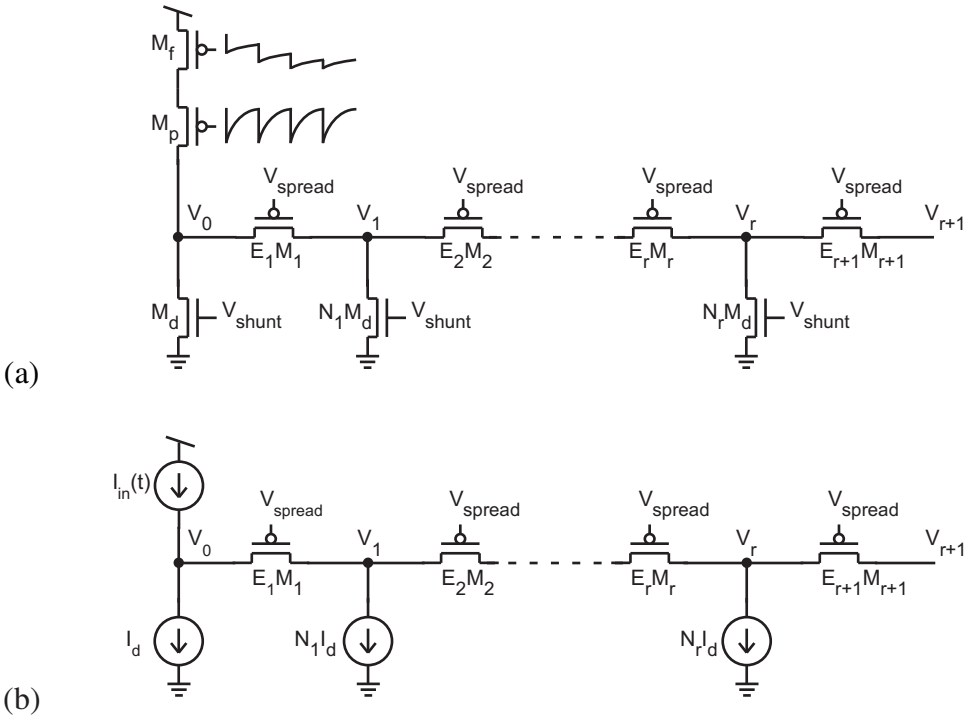


Figure B.1: Neurotropin spreading circuit

(a) Transistor circuit. (b) Equivalent current source representation.

B.0.5.1 Pulse range

The transistor circuit for the extracellular lattice is shown in Figure B.1(a). Lattice nodes are sorted into disjoint rings according to distance, with ring r defined as the set of nodes separated from the release site by r edges. N_r is the number of nodes in ring r and E_r is the number of edges connecting members of ring $r-1$ with members of ring r. For a honeycomb lattice, $N_r = 3r$ and $E_r = (9r-6)/2$ for r even or $(9r-3)/2$ for r odd.

Every node in the lattice is connected to ground through a decay nFET M_{d}, and to E_1 of its neighbors through horizontal pFETs M_1 (Figure B.1(a)). Current, representing neurotropin, is injected into a node through a pair of series pFETs M_{f} and M_{p} that connect the node to the supply rail. These injection transistors are gated by time-dependent voltage

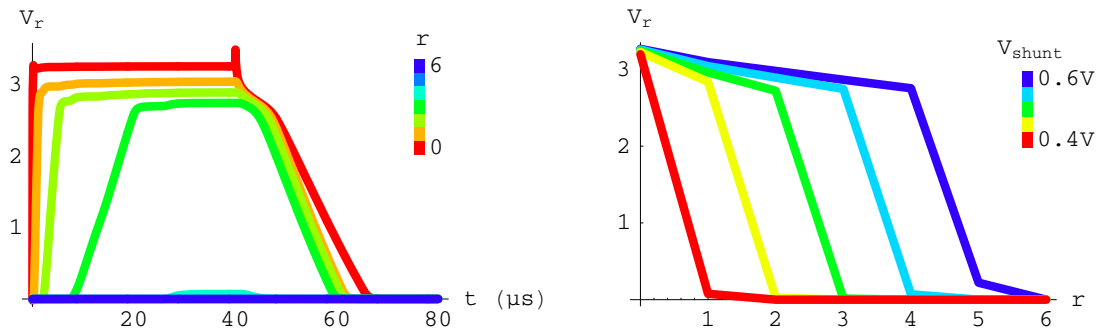
pulses.

Nodes in the same ring r have the same node voltage V_r, which is set by the lateral current $E_{r+1}I_{r+1}$ exiting ring r through the E_{r+1} horizontal transistors M_{r+1}. This exit current is equal to the current entering ring r from ring $r-1$, less the N_r decay currents sunk by member nodes through transistors M_{d}. Current injected at the release site flows outward through concentric node rings until all of the injected current has been sunk.

The horizontal transistors M_r share the common gate bias V_{spread}, and conduct current only when their source voltage V_r exceeds $\kappa_{\mathrm{p}}V_{\mathrm{spread}}$. We typically set $V_{\mathrm{spread}} \approx 1.8V$, so if $V_k > \kappa_{\mathrm{p}}V_{\mathrm{spread}}$ then the decay transistor M_{d} is driven well into saturation, since $\kappa_{\mathrm{p}}V_{\mathrm{spread}} \gg 4V_{\mathrm{T}}$, where the thermal voltage $V_{\mathrm{T}} \approx 26mV$. A transistor in saturation passes an approximately constant current whose amplitude is set by its gate voltage. Accordingly, we can treat the decay transistor M_{d} at a node in ring r as a constant current sink I_{d} if there is significant current flow through M_{r+1}. We make similar assumptions to replace the series pFETs M_{f} and M_{p} with a time-dependent current source $I_{\mathrm{in}}(t)$. The resulting current source representation is shown in Figure B.1(b).

In steady-state, all current entering the lattice through $I_{\mathrm{in}}(t)$ must exit the lattice through a combination of current sinks I_{d}. Current spreads outward from the release site up to some radius R at which all current has been sunk. We can solve for R by expressing the sourced current $I_{\mathrm{in}}(t)$ in terms of its current sinks:

$$\begin{aligned} I_{\mathrm{in}}(t) &= I_{\mathrm{d}}\left[1 + \sum_{r=1}^{R} 3r\right] \\ &= I_{\mathrm{d}}\left[1 + \frac{3}{2}R(R-1)\right] \end{aligned}$$



(a) (b)

Figure B.2: Neurotropin pulse shape

SPICE simulation of neurotropin release pulse. (a) Temporal pulse shape. A current pulse $I_{\text{in}}(t)$ is injected at $r = 0$ beginning at $t = 0\mu s$ and ending at $t = 40\mu s$. $V_{\text{shunt}} = 0.4V$. (b) Spatial pulse shape, measured at $t = 38\mu s$. Pulse range is limited by the constant decay current I_{d}, as controlled by V_{shunt}.

Inverting,

$$R = \frac{1}{6}\left[-3 + \sqrt{3}\sqrt{8\frac{I_{\text{in}}(t)}{I_{\text{d}}} - 5}\right] \tag{B.1}$$

Current injected at a release site spreads up to R nodes away before being shunted to ground. Increasing the constant decay current I_{d} decreases R, since more current gets shunted closer to the release site.

B.0.5.2 Pulse shape

During a release pulse, current $I_{\text{in}}(t)$ is injected at a release site and radiates outward through concentric rings of lattice nodes until all of the injected current has been sunk through the distributed decay currents I_{d}. Current enters a node in ring r from adjacent nodes in the next innermost ring $r - 1$ through the horizontal transistors M_r. From this horizontal input current I_r, the node's decay transistor M_{d} subtracts a constant I_{d}. The

excess current $I_r - I_{\mathrm{d}}$ charges the node capacitance until the node voltage V_r exceeds $\kappa_{\mathrm{p}} V_{\mathrm{spread}}$, at which point the next set of horizontal transistors M_{r+1} turn on, arresting the rise in V_r by diverting the excess current outward to charge adjacent nodes in the next ring $(r+1)$.

Figure B.2(a) shows an example of this sequential ring charging, as simulated in SPICE. A $40\mu s$ pulse of current is injected at the release site at $r = 0$. For $r = 1..4$, node voltages V_r rise in order toward $V_{\mathrm{dd}} = 3.3V$. Finally, at $r = 5$, enough current sinks I_{d} have been recruited to balance the input current $I_{\mathrm{in}}(t)$ that V_5 remains only partially charged. Upon removal of the input current at $t = 40\mu s$, all nodes decay toward ground. Immediately after input removal, node voltages near the release site converge, coupled by the initially above-threshold operation of their horizontal transistors, before sinking in concert due to the collective load of their decay currents I_{d}. Once node voltages drop below $\kappa_{\mathrm{p}} V_{\mathrm{spread}}$, horizontal current flow ceases and individual node voltages decay linearly and independently to ground.

Larger decay currents I_{d} reduce current spread, as simulated in Figure B.2(b). I_{d} increases exponentially with the gate bias V_{shunt}. The same injected current can be balanced by fewer nodes using larger decay currents, so increasing V_{shunt} reduces the effective range r of the release pulse. We will use V_{shunt} in the future as a knob to control pulse range.

For the purposes of growth cone guidance on the chip, we can characterize the pulse gradient as being shallow near the release site, steep at some intermediate range, and flat again at long distances. For convenience, we will approximate this pulse shape by using a peak-normalized Gaussian function for the spreading kernel.

$$G(\mathbf{x}, \mathbf{y}) = \rho_{\mathrm{NT}} \exp\left[-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma_{\mathrm{NT}}^2}\right] \tag{B.2}$$

where the standard deviation σ_{NT} reflects the spatial extent R of the pulse and is controlled by V_{shunt}.

Bibliography

[1] B. Alsina, T. Vu, and S. Cohen-Cory. Visualizing synapse formation in arborizing optic axons in vivo: dynamics and modulation by BDNF. *Nature Neuroscience*, 4(11):1093–1101, 2001.

[2] J. Arthur and K. Boahen. Recurrently connected silicon neurons with active dendrites for one-shot learning. In *International Joint Conference on Neural Networks*, pages 1699–1704, Budapest, Hungary, 2004. IEEE Press.

[3] J.J. Atick and A.N. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4(2):196–210, 1992.

[4] R.J. Balice-Gordon and J.W. Lichtman. Long-term synapse loss induced by focal blockade of postsynaptic receptors. *Nature*, 372(6506):519–524, 1994.

[5] A. Balkowiec and D.M. Katz. Cellular mechanisms regulating activity-dependent release of native brain-derived neurotrophic factor from hippocampal neurons. *Journal of Neuroscience*, 22(23):10399–10407, 2002.

[6] K.A. Boahen. A retinomorphic vision system. *IEEE Micro*, 16(5):30–39, 1996.

[7] M. Bracco, S. Ridella, and R. Zunino. Digital implementation of hierarchical vector quantization. *Neural Networks, IEEE Transactions on*, 14(5):1072–1084, 2003.

[8] M. Buffelli, R.W. Burgess, G.P. Feng, C.G. Lobe, J.W. Lichtman, and J.R. Sanes. Genetic evidence that relative synaptic efficacy biases the outcome of synaptic competition. *Nature*, 424(6947):430–434, 2003.

[9] R.J. Cabelli, A. Hohn, and C.J. Shatz. Inhibition of ocular dominance column formation by infusion of NT-4/5 or BDNF. *Science*, 267(5204):1662–1666, 1995.

[10] R.J. Cabelli, D.L. Shelton, R.A. Segal, and C.J. Shatz. Blockade of endogenous ligands of TrkB inhibits formation of ocular dominance columns. *Neuron*, 19(1):63–76, 1997.

[11] H.C. Card, G.K. Rosendahl, D.K. McNeill, and R.D. McLeod. Competitive learning algorithms and neurocomputer architecture. *Computers, IEEE Transactions on*, 47(8):847–858, 1998.

[12] G.A. Carpenter and S. Grossberg. The Art of Adaptive Pattern-Recognition by a Self-Organizing Neural Network. *Computer*, 21(3):77–88, 1988.

[13] G. Cauwenberghs and V. Pedroni. A low-power CMOS analog vector quantizer. *Solid-State Circuits, IEEE Journal of*, 32(8):1278–1283, 1997.

[14] B. Chapman and I. Godecke. Cortical cell orientation selectivity fails to develop in the absence of ON-center retinal ganglion cell activity. *Journal of Neuroscience*, 20(5):1922–1930, 2000.

[15] C.F. Chen and W.G. Regehr. Developmental remodeling of the retinogeniculate synapse. *Neuron*, 28(3):955–966, 2000.

[16] C.Y. Chiu and M. Weliky. Spontaneous activity in developing ferret visual cortex in vivo. *Journal of Neuroscience*, 21(22):8906–8914, 2001.

[17] T.Y.W. Choi, B.E. Shi, and K.A. Boahen. An ON-OFF orientation selective address event representation image transceiver chip. *Circuits and Systems I-Regular Papers, IEEE Transactions on*, 51(2):342–353, 2004.

[18] S. Cohen-Cory. BDNF modulates, but does not mediate, activity-dependent branching and remodeling of optic axon arbors in vivo. *Journal of Neuroscience*, 19(22):9996–10003, 1999.

[19] S. Cohen-Cory. The developing synapse: construction and modulation of synaptic structures and circuits. *Science*, 298(5594):770–776, 2002.

[20] S. Cohen-Cory and S.E. Fraser. Effects of brain-derived neurotrophic factor on optic axon branching and remodeling in-vivo. *Nature*, 378(6553):192–196, 1995.

[21] M.C. Crair, D.C. Gillespie, and M.P. Stryker. The role of visual experience in the development of columns in cat visual cortex. *Science*, 279(5350):566–570, 1998.

[22] J.C. Crowley and L.C. Katz. Early development of ocular dominance columns. *Science*, 290(5495):1321–1324, 2000.

[23] E.W. Dent and F.B. Gertler. Cytoskeletal dynamics and transport in growth cone mobility and axon guidance. *Neuron*, 40:209–227, 2003.

[24] J. Du, L.Y. Feng, F. Yang, and B. Lu. Activity- and Ca2+-dependent modulation of surface expression of brain-derived neurotrophic factor receptors in hippocampal neurons. *Journal of Cell Biology*, 150(6):1423–1433, 2000.

[25] J. Du, L.Y. Feng, E. Zaitsev, H.S. Je, X.W. Liu, and B. Lu. Regulation of TrkB receptor tyrosine kinase and its internalization by neuronal activity and Ca2+ influx. *Journal of Cell Biology*, 163(2):385–395, 2003.

[26] T. Elliott and J. Kramer. Coupling an aVLSI neuromorphic vision chip to a neurotrophic model of synaptic plasticity: The development of topography. *Neural Computation*, 14(10):2353–2370, 2002.

[27] T. Elliott and J. Kramer. Developing topography and ocular dominance using two avlsi vision sensors and a neurotrophic model of plasticity. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems*, volume 15, pages 1155–1162, Vancouver, 2003. MIT Press.

[28] T. Elliott, A.C. Maddison, and N.R. Shadbolt. Competitive anatomical and physiological plasticity: a neurotrophic bridge. *Biological Cybernetics*, 84(1):13–22, 2001.

[29] T. Elliott and N.R. Shadbolt. Competition for neurotrophic factors: mathematical analysis. *Neural Computation*, 10(8):1939–1981, 1998.

[30] T. Elliott and N.R. Shadbolt. A neurotrophic model of the development of the retinogeniculocortical pathway induced by spontaneous retinal waves. *Journal of Neuroscience*, 19(18):7951–7970, 1999.

[31] E. Erwin and K.D. Miller. Correlation-based development of ocularly matched orientation and ocular dominance maps: Determination of required input activities. *Journal of Neuroscience*, 18(23):9870–9895, 1998.

[32] E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps - ordering, convergence properties and energy functions. *Biological Cybernetics*, 67(1):47–55, 1992.

[33] E. Erwin, K. Obermayer, and K. Schulten. Self-organizing maps - stationary states, metastability and convergence rate. *Biological Cybernetics*, 67(1):35–45, 1992.

[34] E. Erwin, K. Obermayer, and K. Schulten. Models of orientation and ocular dominance columns in the visual cortex - a critical comparison. *Neural Computation*, 7(3):425–468, 1995.

[35] W.-C. Fang, B.J. Sheu, O.T.-C. Chen, and J. Choi. A VLSI neural processor for image data compression using self-organization networks. *Neural Networks, IEEE Transactions on*, 3(3):506–518, 1992.

[36] M.B. Feller. The role of nAChR-mediated spontaneous retinal activity in visual system development. *Journal of Neurobiology*, 53(4):556–567, 2002.

[37] M.B. Feller, D.A. Butts, H.L. Aaron, D.S. Rokhsar, and C.J. Shatz. Dynamic processes shape spatiotemporal properties of retinal waves. *Neuron*, 19(2):293–306, 1997.

[38] A.G. Gartner and V. Staiger. Neurotrophin secretion from hippocampal neurons evoked by long-term-potentiation-inducing electrical stimulation patterns. *Proceedings of the National Academy of Science of the United States of America*, 99(9):6386–6391, 2002.

[39] D.C. Gillespie, M.C. Crair, and M.P. Stryker. Neurotrophin-4/5 alters responses and blocks the effect of monocular deprivation in cat visual cortex during the critical period. *Journal of Neuroscience*, 20(24):9174–9186, 2000.

[40] Y. Goda and G.W. Davis. Mechanisms of synapse assembly and disassembly. *Neuron*, 40(2):243–264, 2003.

[41] G.J. Goodhill. A theoretical model of axon guidance by the Robo code. *Neural Computation*, 15(3):549–64, 2003.

[42] G.J. Goodhill and H. Baier. Axon guidance: stretching gradients to the limit. *Neural Computation*, 10(3):521–7, 1998.

[43] G.J. Goodhill and J.S. Urbach. Theoretical analysis of gradient detection by growth cones. *Journal of Neurobiology*, 41(2):230–41, 1999.

[44] T. Graepel, M. Burger, and K. Obermayer. Phase transitions in stochastic self-organizing maps. *Physical Review E*, 56(4):3876–3890, 1997.

[45] Y. Guo and S.B. Udin. The development of abnormal axon trajectories after rotation of one eye in Xenopus. *Journal of Neuroscience*, 20(11):4189–97, 2000.

[46] M. Hartmann, R. Heumann, and V. Lessmann. Synaptic secretion of BDNF after high-frequency stimulation of glutamatergic synapses. *Embo Journal*, 20(21):5887–5897, 2001.

[47] Y. He and U. Cilingiroglu. A charge-based on-chip adaptation Kohonen neural network. *Neural Networks, IEEE Transactions on*, 4(3):462–469, 1993.

[48] D.O. Hebb. *The Organization of Behavior*. Wiley, New York, 1949.

[49] D.C. Hendry, A.A. Duncan, and N. Lightowler. IP core implementation of a self-organizing neural network. *Neural Networks, IEEE Transactions on*, 14(5):1085–1096, 2003.

[50] H.G.E. Hentschel and A. van Ooyen. Models of axon guidance and bundling during development. *Proceedings of the Royal Society of London B: Biological Sciences*, 266:2231–2238, 1999.

[51] T.M. Heskes. Transition times in self-organizing maps. *Biological Cybernetics*, 75(1):49–57, 1996.

[52] E.J. Huang and L.F. Reichardt. Neurotrophins: roles in neuronal development and function. *Annual Reviews Neuroscience 2001*, 24:677–736, 2001.

[53] A.B. Huber, A.L. Kolodkin, D.D. Ginty, and J.-F. Cloutier. Signaling at the growth cone: ligand-receptor complexes and the control of axon growth and guidance. *Annual Reviews Neuroscience 2003*, 26:509–63, 2003.

[54] A.D. Huberman, G.Y. Wang, L.C. Liets, O.A. Collins, B. Chapman, and L.M. Chalupa. Eye-specific retinogeniculate segregation independent of normal neuronal activity. *Science*, 300(5621):994–998, 2003.

[55] K. Hynna and K. Boahen. Space-rate coding in an adaptive silicon neuron. *Neural Networks*, 14(6-7):645–56, 2001.

[56] L.C. Katz and C.J. Shatz. Synaptic activity and the construction of cortical circuits. *Science*, 274(5290):1133–1138, 1996.

[57] B. Knoll and U. Drescher. Ephrin-As as receptors in topographic projections. *Trends in Neurosciences*, 25(3):145–149, 2002.

[58] E.I. Knudsen. Supervised learning in the brain. *J Neurosci*, 14(7):3985–97, 1994.

[59] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.

[60] V. Lessmann, K. Gottmann, and M. Malcangio. Neurotrophin secretion: current facts and future prospects. *Progress in Neurobiology*, 69(5):341–374, 2003.

[61] R. Linsker. From basic network principles to neural architecture: emergence of orientation columns. *Proceedings of the National Academy of Science of the United States of America*, 83(22):8779–83, 1986.

[62] R. Linsker. From basic network principles to neural architecture: emergence of orientation-selective cells. *Proceedings of the National Academy of Science of the United States of America*, 83(21):8390–4, 1986.

[63] R. Linsker. From basic network principles to neural architecture: emergence of spatial-opponent cells. *Proceedings of the National Academy of Science of the United States of America*, 83(19):7508–12, 1986.

[64] B. Lu. BDNF and activity-dependent synaptic modulation. *Learning and Memory*, 10:86–98, 2003.

[65] D.J.C. Mackay and K.D. Miller. Analysis of Linsker's simulations of Hebbian rules. *Neural Computation*, 2:173–187, 1990.

[66] D. Macq, M. Verleysen, P. Jespers, and J.-D. Legat. Analog implementation of a Kohonen map with on-chip learning. *Neural Networks, IEEE Transactions on*, 4(3):456–461, 1993.

[67] M. Mahowald. *An Analog VLSI Stereoscopic Vision System*. Kluwer Academic Publishers, Boston, 1994.

[68] T. McLaughlin, C.L. Torborg, M.B. Feller, and D.D.M. O'Leary. Retinotopic map refinement requires spontaneous retinal waves during a brief critical period of development. *Neuron*, 40(6):1147–1160, 2003.

[69] M. Melton, T. Phan, D.S. Reeves, and D.E. Van den Bout. The TInMANN VLSI chip. *Neural Networks, IEEE Transactions on*, 3(3):375–384, 1992.

[70] P. Merolla and K. Boahen. A recurrent model of orientation maps with simple and complex cells. In S. Thrun, L. Saul, and B. Sholkopf, editors, *Advances in Neural Information Processing Systems*, volume 16, Vancouver, 2004. MIT Press.

[71] K. Miller. A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between ON- and OFF-center inputs. *Journal of Neuroscience*, 14(1):409–441, 1994.

[72] K.D. Miller, J.B. Keller, and M.P. Stryker. Ocular dominance column development: analysis and simulation. *Science*, 245(4918):605–615, 1989.

[73] G. Ming, J. Henley, M. Tessier-Lavigne, H. Song, and M. Poo. Electrical activity modulates growth cone guidance by diffusible factors. *Neuron*, 29(2):441–52, 2001.

[74] G.L. Ming, H.J. Song, B. Berninger, C.E. Holt, M. Tessier-Lavigne, and M.M. Poo. cAMP-dependent growth cone guidance by netrin-1. *Neuron*, 19(6):1225–35, 1997.

[75] G.L. Ming, S.T. Wong, J. Henley, X.B. Yuan, H.J. Song, N.C. Spitzer, and M.M. Poo. Adaptation in the chemotactic guidance of nerve growth cones. *Nature*, 417(6887):411–8, 2002.

[76] S. Nakagawa, C. Brennan, K.G. Johnson, D. Shewan, W.A. Harris, and C.E. Holt. Ephrin-B regulates the ipsilateral routing of retinal axons at the optic chiasm. *Neuron*, 25(3):599–610, 2000.

[77] C.M. Niell, M.P. Meyer, and S.J. Smith. In vivo imaging of synapse formation on a growing dendritic arbor. *Nature Neuroscience*, 7(3):254–260, 2004.

[78] M. Nishiyama, A. Hoshino, L. Tsai, J.R. Henley, Y. Goshima, M. Tessier-Lavigne, M.M. Poo, and K.S. Hong. Cyclic AMP/GMP-dependent modulation of Ca2+ channels sets the polarity of nerve growth-cone turning. *Nature*, 423(6943):990–995, 2003.

[79] K. Obermayer, G.G. Blasdel, and K. Schulten. Statistical-mechanical analysis of self-organization and pattern-formation during the development of visual maps. *Physical Review A*, 45(10):7568–7589, 1992.

[80] A.A. Penn, P.A. Riquelme, M.B. Feller, and C.J. Shatz. Competition in retinogeniculate patterning driven by spontaneous activity. *Science*, 279(5359):2108–2112, 1998.

[81] K.E. Personius and R.J. Balice-Gordon. Loss of correlated motor neuron activity during synaptic competition at developing neuromuscular synapses. *Neuron*, 31(3):395–408, 2001.

[82] M. Porrmann, U. Witkowski, and U. Ruckert. A massively parallel architecture for self-organizing feature maps. *Neural Networks, IEEE Transactions on*, 14(5):1110–1121, 2003.

[83] D.L. Ringach. Haphazard wiring of simple receptive fields and orientation columns in visual cortex. *Journal of Neurophysiology*, 92(1):468–76, 2004.

[84] S. Rovetta and R. Zunino. Efficient training of neural gas vector quantizers with analog circuit implementation. *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, 46(6):688–698, 1999.

[85] E.S. Ruthazer, C.J. Akerman, and H.T. Cline. Control of axon branch dynamics by correlated activity in vivo. *Science*, 301(5629):66–70, 2003.

[86] E.S. Ruthazer and H.T. Cline. Insights into activity-dependent map formation from the retinotectal system: a middle-of-the-brain perspective. *Journal of Neurobiology*, 59(1):134–146, 2004.

[87] R. Segev and E. Ben-Jacob. Self-wiring of neural networks. *Physics Letters A*, 237(4-5):307–313, 1998.

[88] F. Sengpiel, P. Stawinski, and T. Bonhoeffer. Influence of experience on orientation maps in cat visual cortex. *Nature Neuroscience*, 2(8):727–732, 1999.

[89] T. Serrano-Gotarredona and B. Linares-Barranco. A real-time clustering microchip neural engine. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 4(2):195–209, 1996.

[90] J. Sharma, A. Angelucci, and M. Sur. Induction of visual orientation modules in auditory cortex. *Nature*, 404(6780):841–847, 2000.

[91] C.J. Shatz and M.P. Stryker. Prenatal tetrodotoxin infusion blocks segregation of retinogeniculate afferents. *Science*, 242(4875):87–89, 1988.

[92] H.Z. Shouval, D.H. Goldberg, J.P. Jones, M. Beckerman, and L.N. Cooper. Structured long-range connections can provide a scaffold for orientation maps. *Journal of Neuroscience*, 20(3):1119–1128, 2000.

[93] D. Stellwagen and C.J. Shatz. An instructive role for retinal waves in the development of retinogeniculate connectivity. *Neuron*, 33:357–367, 2002.

[94] G.S. Stent. Physiological mechanism for Hebb's postulate of learning. *Proceedings of the National Academy of Science of the United States of America*, 70(4):997–1001, 1973.

[95] M.P. Stryker and W.A. Harris. Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex. *Journal of Neuroscience*, 6(8):2117–2133, 1986.

[96] M. Sur and C.A. Leamey. Development and plasticity of cortical areas and networks. *Nature Reviews Neuroscience*, 2(4):251–262, 2001.

[97] N.V. Swindale. The development of topography in the visual cortex: a review of models. *Network-Computation in Neural Systems*, 7(2):161–247, 1996.

[98] A. Tashiro, A. Dunaevsky, R. Blazeski, C.A. Mason, and R. Yuste. Bidirectional regulation of hippocampal mossy fiber filopodial motility by kainate receptors: a two-step model of synaptogenesis. *Neuron*, 38(5):773–784, 2003.

[99] P. Thiran, V. Peiris, P. Heim, and B. Hochet. Quantization effects in digitally behaving circuit implementations of Kohonen networks. *Neural Networks, IEEE Transactions on*, 5(3):450–458, 1994.

[100] A. van Ooyen. Competition in the development of nerve connections: a review of models. *Network: Computation in Neural Systems*, 12:R1–R47, 2001.

[101] L. von Melchner, S.L. Pallas, and M. Sur. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, 404(6780):871–876, 2000.

[102] M. Weliky and L.C. Katz. Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature*, 386(6626):680–685, 1997.

[103] M. Weliky and L.C. Katz. Correlational structure of spontaneous neuronal activity in the developing lateral geniculate nucleus in vivo. *Science*, 285(5427):599–604, 1999.

[104] L.E. White, D.M. Coppola, and D. Fitzpatrick. The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex. *Nature*, 411(6841):1049–1052, 2001.

[105] R.O.L. Wong. Retinal waves and visual system development. *Annual Reviews Neuroscience 1999*, 22:29–47, 1999.

[106] C.R. Yu, J. Power, G. Barnea, S. O'Donnell, H.E. Brown, J. Osborne, R. Axel, and J.A. Gogos. Spontaneous neural activity is required for the establishment and maintenance of the olfactory sensory map. *Neuron*, 42(4):553–566, 2004.

[107] K.A. Zaghloul and K. Boahen. Optic nerve signals in a neuromorphic chip I: outer and inner retina models. *Biomedical Engineering, IEEE Transactions on*, 51(4):657–666, 2004.

[108] K.A. Zaghloul and K. Boahen. Optic nerve signals in a neuromorphic chip II: testing and results. *Biomedical Engineering, IEEE Transactions on*, 51(4):667–675, 2004.

[109] X.H. Zhang and M.M. Poo. Localized synaptic potentiation by BDNF requires local protein synthesis in the developing axon. *Neuron*, 36(4):675–688, 2002.