8

Organically Grown Architectures: Creating Decentralized, Autonomous Systems by Embryomorphic Engineering

René Doursat

Institut des Systèmes Complexes (ISC),
Centre de Recherche en Epistémologie Appliquée (CREA),
CNRS and Ecole Polytechnique, 57-59, rue Lhomond, 75005 Paris, France.
doursat@shs.polytechnique.fr

Summary. Exploding growth in computational systems forces us to gradually replace rigid design and control with decentralization and autonomy. Information technologies will progress, instead, by"meta-designing" mechanisms of system self-assembly, self-regulation and evolution. Nature offers a great variety of efficient complex systems, in which numerous small elements form large-scale, adaptive patterns. The new engineering challenge is to recreate this self-organization and let it freely generate innovative designs under guidance. This article presents an original model of artificial system growth inspired by embryogenesis. A virtual organism is a lattice of cells that proliferate, migrate and self-pattern into differentiated domains. Each cell's fate is controlled by an internal gene regulatory network. *Embryomorphic* engineering emphasizes hyperdistributed architectures, and their development as a prerequisite of evolutionary design.

Key words: complex systems, artificial development, evolutionary computation, systems design, embryomorphic engineering

8.1 Introduction: designing complexity

8.1.1 Toward decentralized, autonomous systems

Today's information and communication systems are characterized by exploding growth in the number of components and complexity of their interactions. Systems engineers are confronted with an insatiable demand for functional innovation, robustness, scalability and security. This upward trend is accelerating at all levels of organization, whether hardware (integrated components), software (program modules) or networks (applications and users). Famously anticipated by Moore's law, the number of transistors on a microprocessor has climbed five orders of magnitude in the past 35 years. Similarly, operating systems and other very large computer programs commonly contain hundreds of

R.P. Würtz (ed.), *Organic Computing*. Understanding Complex Systems,
doi: 10.1007/978-3-540-77657-4_8,

millions of source lines of code (SLOC). Over one billion people routinely use the Internet, which connects half a billion hosts. In sum, an increasing number of users with greater mobility are constantly requiring more sophisticated functionality from larger applications running on faster architectures.

Consequently, computer scientists and engineers are gradually led to rethink the traditional perspective on systems design, i.e., the dogma of a totalistic act of creation imposing order and organization *exogenously*. The growth in complexity has already been accompanied by a de facto segmentation and distribution of the traditionally centralized control over systems design. This march toward decentralization is somewhat discernible in the fields of integrated circuit design and software development, where engineers collaborate in large teams around relatively independent components and modules. It has become even more apparent with the advent of leaderless open source communities, and most striking in the spontaneous growth of the Internet and World-Wide Web. To some degree, information architects and engineers are already beginning to lose grip on their creation, which exceeds the capacity of a single human mind. Therefore, rather than insisting on rigidly designing computational systems or system parts in every detail, the trend should be to "step back" even further and concentrate more on establishing the *generic conditions* that will allow and encourage those systems to self-assemble, self-regulate and evolve. In fact, future progress in information and communication technologies could ultimately depend on our ability to foster systems that *endogenously* grow, function, repair themselves and, more importantly, adapt and improve. This need is probably most acute in software development, which is currently less an exact science than a skillful art — the accumulation by trial and error of a corpus of design patterns and numerical recipes. Ironically, machines that are perfectly logical and regular still rely entirely on intuitive and fallible human instructions. The burden is fully on the side of our human cognitive system to instruct artificial systems, but this ability is now reaching its limits with very large architectures, as attested by the extremely high cost (in effort, time and money) devoted to source code development, maintenance and debugging.

A major challenge will thus be for information systems to step beyond their current state of heteronomy, where they are fully subjected to a designer, toward states of increasing organizational and functional autonomy. Biological organisms, which might give the *illusion* of deliberate design, are in fact the product of *undesigned evolution* through random variation and non-random natural selection, excluding the need to invoke any form of intelligent design for them (which would also necessarily be, by recursive reasoning, of a supernatural kind). By contrast, artificial structures will always possess a causal design link originating from their human makers, while at the same time this link promises to become less and less clear or direct. In the design versus evolution spectrum (figure 8.1), classical engineering is currently set on the "intelligent design" (ID) notch, with the opposite end occupied by biology under "undesigned evolution" (UE). The expectation is that the engineering

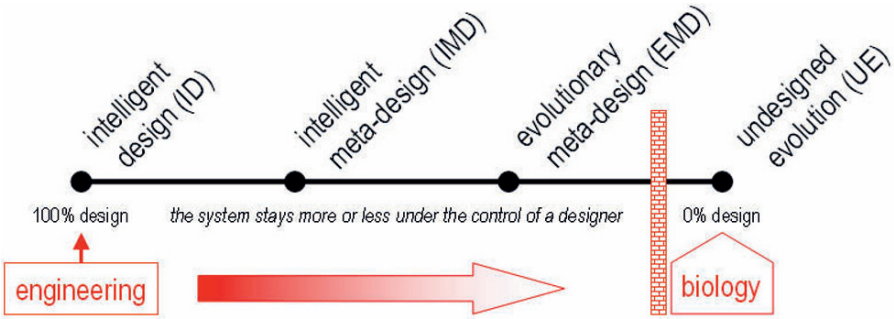

Fig. 8.1. Four stops and one brick wall on the design-evolution line (see text).

paradigm should progressively shift in the direction of biology through intermediate stages, but without fully reaching UE. In a first stage that we could refer to as "intelligent *meta-design*" (IMD), designers will focus on creating generative mechanisms rather than the systems themselves (Table 8.1). If we metaphorically imagined, on the contrary, biology drifting toward more design, ID would be the equivalent of directly assembling an animal's organs and limbs together, whereas IMD would correspond to creating the laws of cellular development, preparing the zygote's DNA and let it grow. Still closer to UE, in a stage we could call "evolutionary meta-design" (EMD), an even more disengaged meta-architect could also create the laws of variation and selection, prepare some primitive ancestor system (in the reverse biological metaphor, a prokaryote, for example) and step back to let evolution invent the rest. Applied to artificial systems, this paradigm shift is the inspiration of the present work. It emphasizes the importance of constituting fundamental laws of *development* and developmental *variations* in the IMD stage, before these variations can be selected upon in the EMD stage. In the framework of genetic algorithms and evolutionary computation, this means an *indirect* or *implicit* mapping (as opposed to direct or explicit) from genotype to phenotype.

8.1.2 Harnessing complex systems

Looking around, we observe an abundance of autonomous, emergent systems in the environment, whether in nature (geological patterns, biological cells, organisms, animal societies, ecosystems) or spontaneously emerging human super-structures (cities, markets, the Internet). Naturally decentralized, unplanned systems are robust and efficient, and constitute the overwhelming majority of system types. It is our artificially centralized and planned systems that are fragile, costly to build and rare, as they require a higher intelligence to arise. Our cognitive viewpoint, accustomed to the illusion of a central consciousness, traditionally refers to such decentralized systems as "complex",

systems design	systems "meta-design"
heteronomous order	autonomous order
centralized control	decentralized control
manual, extensional design	automated, intentional design
the engineer as a micromanager	the engineer as a lawmaker
rigidly placing components	allowing fuzzy self-placement
tightly optimized systems	hyperdistributed and redundant systems
sensitive to part failures	insensitive to part failures
need to control	prepare to adapt and self-regulate
need to redesign	prepare to learn and evolve

Table 8.1. Some contrastive features of systems design and "meta-design".

whereas in fact they might be simpler than our familiar contraptions with their uniquely hierarchical arrangement.

Complex systems are composed of a great number of small, repeated elements that interact locally and produce collective behavior at a macroscopic scale. They are characterized by a high degree of decentralization and self-organization, exhibiting spontaneous pattern formation (self-assembly) and homeostatic autonomy (self-regulation). Most complex systems are also adaptive, in the sense that they are able to learn and/or evolve through feedback from their external fitness to their internal architecture. The elements composing the system are themselves often internally structured as networks of smaller entities at a finer scale. For example, one cell can be modeled as a self-regulatory network of genetic switches, one social agent (insect) as a network of decision rules, or one neural unit as a local assembly of neurons (oscillator system). Conversely, agents can also interact collectively at the level of clusters or subnetworks (organs, assemblies, cliques) to combine in a modular fashion and form larger collectives. Thus, from both perspectives, complex systems can often be described as "networks of networks" on several hierarchical levels. The higher levels connecting elements or clusters of elements are generally spatially extended (cell tissues, cortical areas, ant colonies), whereas the lower levels inside elements are generally nonspatial (gene nets, neurons, rule sets). Elements follow the dynamics dictated by their inner networks and also influence neighboring elements through the emission and reception of signals (chemical, electrical). The attractors of the internal dynamics are fixed-point states or limit cycles, and the behavior of the whole connected system can be rephrased in terms of synchronization among autonomous dynamical subsystems. The work presented in this chapter is an instance of this paradigm based on a 2-D lattice of coupled gene regulatory networks.

Such natural adaptive systems, biological or social, could become a new and powerful source of inspiration for emerging information and communica-

tion technologies in their transition toward autonomous systems. This joins recent trends advocating and announcing the convergence of four scientific disciplines: nanoscience, biotechnology, information technology and cognitive science. Called NBIC in the US [2], these fields of investigation combine all the components of bio-inspired complex systems engineering, i.e., swarms of small components (Nano), biological complexity (Bio), systems design (Info) and artificial intelligence (Cogno). Also described by the Future and Emerging Technologies (FET) program of the European Union [5], this scientific perspective is close to several other initiatives, such as organic computing ([37], and this volume, in particular chapters 1 and 2), amorphous computing [1, 25, 41], natural computation, e.g., [28], complex systems engineering [23], ambient intelligence, and pervasive or ubiquitous computing [40].

As indicated above, however, a major difference with biological systems is that human-made systems will (and hopefully should), by definition, always remain under the guidance of a human designer to some degree, never breaking the barrier to the absolute UE stage that characterizes biology (figure 8.1). While we want to achieve meta-designing artificial systems that can grow (IMD stage) and evolve (EMD stage), it is obviously our intent to keep a partially "visible hand" on these systems, i.e., (a) some meta-control over their execution and (b) certain requirements about their structure or function. The important questions of control and optimization of complexity will be respectively addressed in sections 8.3 and 8.4 below.

8.1.3 Artificial development

The field of Artificial Life (AL) is chiefly concerned with the simulation of life-like or organismal processes through computer programs or robotic devices that generally are of a distributed nature and operate on a multitude of interacting components. Researchers in AL attempt to design and construct systems that have the characteristic of living organisms or societies of organisms out of non-living parts, whether virtual (software agents) or physical (electromechanical components, chemical materials). AL is, therefore, a "bottom up" attempt to recreate or synthesize biological phenomena with the goal of producing adaptive and intelligent systems. In this sense, it can be contrasted with the traditional "top down" analytical approach of Artificial Intelligence based on symbolic systems. Although not all AL systems are "complex", in the sense of a multitude of elements, AL is one of the most important and rapidly developing domains within the federation of complex systems research. In particular, it actively promotes biology-inspired engineering as a new paradigm complementing or replacing classical physics-based engineering.

AL opens entirely new perspectives in software, robotic, electrical, mechanical or even civil engineering. Can a sophisticated device or building architecture construct itself from a large reservoir of small components? Can a robot rearrange its parts and evolve toward better performance without explicit

instructions? Can a swarm of software agents self-organize and collectively innovate in problem-solving tasks? Among the great variety of biological systems that inspire and guide AL research, three broad areas can be identified according to the scale of their elementary components: (a) molecular or cellular systems, (b) anatomical or functional systems, and (c) individual or societal systems. Artificial molecular and cellular models find inspiration in the spontaneous organization of complex chemical and organic structures, such as protein self-assembly or organism development (e.g., [22], and chapter 9 of this volume). Applications are linked to nanotechnologies for biomedical or integrated electronic purposes ("smart materials", MEMS). On the anatomical and functional level, robotic parts (limbs, sensors, actuators) and local behavioral modules are integrated and put in interaction to produce emergent action in a single autonomous device, aiming toward adaptivity and nonsymbolic intelligence. This is the scope of "reactive" or "embodied" robotics, exemplified by insect-like robots and evolving mechanical morphologies (e.g., [20]). Finally, entire colonies of virtual or robotic creatures also constitute important objects of interest because of their unique properties of collective self-organization and diversity-inducing evolution (e.g., [41]). Generically termed "swarm intelligence", new methodologies such as ant colony optimization or particle swarm optimization are derived from the observation of animal societies and applied to problem-solving tasks.

The preferred computational tools of AL are cellular automata, multiagent networks and genetic algorithms. Complex networks form the natural structural backbone of AL models. Their topology can vary from regular lattices with nearest neighbor connectivity (cellular automata) to irregular graphs (random, small-world) containing long-range interactions. The first kind is spatially extended, in 2-D or 3-D, while the second generally does not rely on a background notion of space or Euclidean distance. This chapter presents an original AL study of the spatially explicit kind. With respect to the above classification, it addresses level (a) of system organization, specifically the computational modeling and simulation of the fundamental principles of self-patterning and self-assembly during embryogenesis. The development of an entire organism from a single cell is the most striking example of self-organization *guided* by information — in this case, genetic. In the present model, a virtual organism is represented by a mass of cells that proliferate, migrate and self-pattern into differentiated domains. Each cell contains an internal gene regulatory network and acquires a specific expression identity by interaction with positional information transmitted through neighboring cells. Different identities trigger different cell behaviors, which in turn induce new identities. In sum, development is driven by only a few fundamental laws of cell division and movement, propagation of genetic expression and positional information. The final architecture of the organism depends on the detailed interplay between the various rates of these processes.

Ultimately, on this score of "theme and variations" (laws and parameters), evolution is the player. Most importantly, the link between the genetic

parameters and the morphological features of the system is not arbitrary, as is generally the case in many evolutionary algorithm techniques, but expresses itself through a genuine developmental approach at the microscopic cellular level. The phenotype is a macroscopic *emergence* of the unfolding genotype, not an ad hoc one-to-one mapping. Possible future hardware applications of this model include systems in which nano-units containing the same instructions are mass-produced at low cost and mixed in a homogeneous material, where they self-organize without the need for reliability or precise arrangement as in traditional VLSI [1, 25]. Software or network applications (servers, security) could involve a swarm of small-footprint software agents that diversify and self-deploy to achieve a desired level of functionality. In all cases, embryo-inspired architectures suggest a "fine-grain" approach to systems design, i.e., one based on hyperdistributed collectives of a great number of very simple and relatively ignorant, cloned elements. This approach is called here *embryomorphic* engineering.

The remainder of the chapter is organized as follows. After preliminary remarks on the genetic causality of biological development in section 8.2, a virtual embryogenesis model, "the organic canvas", is described in section 8.3 in four incremental steps. Section 8.3.1, "the self-painting canvas", introduces the concept of genetically guided *self-patterning*. Section 8.3.2, "the growing canvas", adds a *multiscale* and *modular* dimension to this pattern formation process. Section 8.3.3, "the deformable canvas", brings in *self-assembly* through three critical mechanisms of cell movement: adhesion, division and migration. Finally, section 8.3.4, "the excitable canvas", briefly explores the possible *computing* capabilities of a fully developed organism. The purpose of section 8.3 is, thus, to lay out the IMD foundations of the model, by showing an example of *lawmaking* of artificial system development with inspiration from biology. Section 8.4 then discusses the transition to the EMD stage, i.e., the role that *evolution* could play in shaping the genome at the basis of the developmental process, and inventing new architectures. Specifically, it addresses the paradox of "planning the autonomy" at the center of the complex systems engineering enterprise.

8.2 The genetic causality of biological development

8.2.1 Free versus guided morphogenesis

Complex patterns produced by nature have always been a source of great fascination for philosophers and scientists. Ripples in sand dunes, spots in animal coats, geometric figures in plants, and a multitude of meanders, spirals, branches, lattices, and others, can be observed everywhere. Whether inanimate structures or living organisms, all these processes are instances of decentralized morphological self-organization and, as such, were not easily amenable to analysis and explanation. For a long time, in fact, the old cross-disciplinary and abstract problem of the "form" was deemed non-objective

by mainstream physics and relegated to the phenomenological realm of experience. Only relatively recently was it revived as a valid object of scientific inquiry, under rising needs for a better understanding, prediction and control of geophysical and biological systems. New computer technologies and numerical simulations created dramatic new advances in the understanding of objectively measurable complex forms and their emergent properties of order and chaos.

A taxonomy of all emergent patterns would contain many dimensions: inert vs. living, natural vs. human-induced, small-scale vs. large-scale, and so on. The present study focuses on a major distinction between what will be referred to as "*free* forms" and "*guided* forms" (figure 8.2). Free forms essentially result from the amplification of unstable fluctuations, as proposed by Turing in his now classical reaction-diffusion model of morphogenesis [35], which was further developed and popularized by Gierer and Meinhardt [14]. A pigmented medium, such as an animal coat, undergoes a symmetry-breaking due to positive feedback based on the short-range diffusion of an activator substance, reacting with negative feedback based on the long-range diffusion of an inhibitor substance [44]. In 2-D domains, this typically generates spots or stripes of alternating color (figure 8.2a-b). Setting aside questions about the actual existence of activator and inhibitor "morphogen" agents, it remains that the pattern formation phenomena covered by this model are fundamentally random and unpredictable. Are there going to be four, five, or six spots? Although the patterns are often statistically homogeneous and can be described by a characteristic scale or order parameter (diameter of the spots, width of the stripes), morphological details such as position, orientation and number are *not* invariants of the system. Another example of free patterning is given by convection cells in a heated fluid, such as the ones observed in the well-known Rayleigh-Bénard instability. Given the temperature gradient and other parameters of the fluid, it is possible to calculate the typical size of the polygonal convection domains but, again, *not* their precise shape and spatial arrangement.

Turing-like reaction-diffusion principles might be able to account for some pattern formation effects in biological development, such as mammal coat [44], butterfly wing spots [26], angelfish stripes [18], or seashell motifs [21], yet these effects seem only secondary — literally "superficial" — compared to the overall form of an organism. The precisely arranged body shape of animals, made of articulated segments and subparts (figure 8.2c-d), is not the result of free-forming random instabilities. It is a fundamentally *guided* morphogenesis process that plays out under deterministic control from the genome. Except for very rare cases of malformation, all members of a pentadactyl mammalian species reliably develop five digits, not sometimes four or sometimes six. All healthy embryos of *Drosophila* exhibit exactly seven bands of differentiated gene expression along the anteroposterior axis, which then give rise to 14 segments. Each one of these mammal digits or insect segments is independently controlled by a specific combination of genes. At every time step in the de-

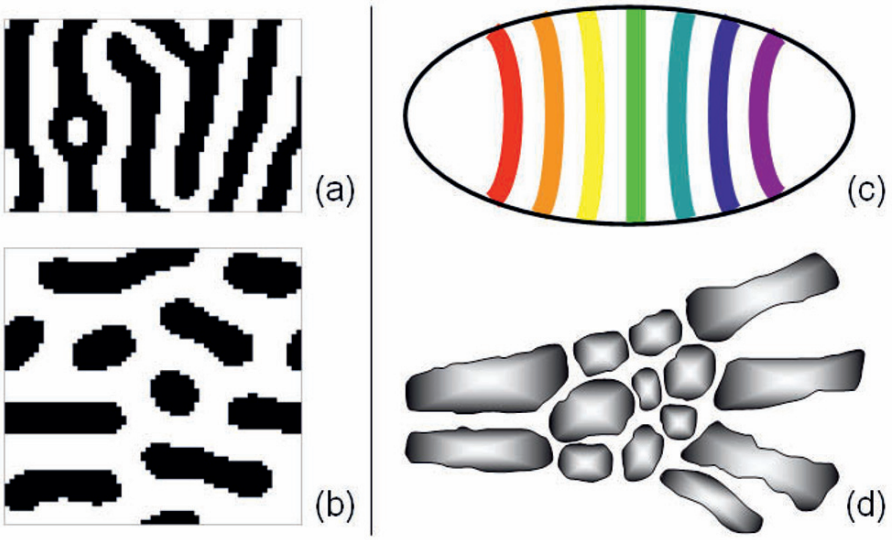

Fig. 8.2. Free vs. guided morphogenesis. A simple activator-inhibitor cellular automata model, such as Young's [44], creates stripes (a) and spots (b) in variable positions and unpredictable numbers. By contrast, the stripes (c) and spots (d) of developing animal segments are tightly controlled by multiple sets of genes, leaving very little room for chance arrangements.

velopment of an embryo, a homogeneous region of the overall embryo pattern is defined as a local group of cells that have the same gene expression profile, i.e., the same dynamic regime of RNA and protein concentrations.

In summary, biological forms are not statistically uniform. They are rich in morphological information and cannot be reduced to one characteristic scale like reaction-diffusion patterns. Some free pattern motifs (spots, stripes) can be embedded in a guided form (leopard, angelfish). Conversely, a guided form can be duplicated and distributed in free patterns (e.g., hundreds of copies of the same flower shape on the branches of one tree). Biological forms can thus combine a little free patterning with a lot of guided morphogenesis. It is the latter kind that the present work aims at modeling and reproducing as a possible paradigm of information-driven systems growth.

8.2.2 Development: the missing link of the modern synthesis

Darwin discovered the evolution of species, based on random variation and nonrandom natural selection, and established it as a central fact of biology. During the same period, Mendel brought to light the laws of inheritance of traits. In the twentieth century, his work was rediscovered and became the foundation of the science of genetics, which culminated with the revelation of DNA's role in heredity by Avery and its double-helix structure by Watson and

Crick. By integrating evolution and genetics together, the "modern synthesis" of biology has demonstrated the existence of a fundamental correlation between genotype and phenotype. Mutation in the first is causally related to variation in the second. Yet, 150 years after Darwin's and Mendel's era, the nature of the link from genes to organismal forms, i.e., the actual molecular and cellular basis of the *mechanisms* of development, are still unclear. To quote Kirschner and Gerhart [16, page ix]:

> "When Charles Darwin proposed his theory of evolution by variation and selection, explaining selection was his great achievement. He could not explain variation. That was Darwin's dilemma [...] To understand novelty in evolution, we need to understand organisms down to their individual building blocks, down to their deepest components, for these are what undergo change."

Understanding variation by comparing the actual *developmental* processes of different species is the primary concern of the field of evolutionary development biology, or "evo-devo". The genotype-phenotype link cannot remain an abstraction if we want to unravel the *generative* laws of development and evolution. The goal is to unify what Darwin called the "endless forms most beautiful" of nature [7], and reduce them to *variants* around a common theme [39]. The variants are the specifics of genetic information; the common theme is the developmental dynamics that this information guides. Modern synthesis postulates this reduction in principle but has never truly explained it physically.

How does a static, nonspatial genome dynamically unfold in time and 3-D space [13]? *How* are morphological changes correlated with genetic changes? Looking at the full evolutionary *and* developmental picture should also be a primary concern of systems engineering and computer science when venturing in the new arena of autonomous architectures. Optimization techniques inspired by biology in its traditional modern synthesis form have principally focused on evolution, giving rise to evolutionary computation and genetic algorithms based on metaphorical "genes", "reproduction", "mutation" and "selection". However, the great majority of these approaches rely on a direct mapping from artificial genomes to artificial phenotypes, which includes very few or no elements of morphodynamics. The present work's ambition is to contribute to restore the balance between evo and devo by shifting the emphasis on developmental meta-design as a *prerequisite* of evolutionary meta-design.

8.3 Description of the model: the organic canvas

This part describes a model of embryomorphic system development partly introduced in [11].

8.3.1 The self-painting canvas: gene-guided patterning

8.3.1.1 Gene regulatory networks

The central dogma of molecular biology states that a segment of DNA representing a gene is transcribed into messenger RNA, and then mRNA is translated into proteins by ribosomes and transfer RNA. The present model adopts a highly simplified one-to-one view of the gene-protein processing chain, ignoring additional effects such as post-transcriptional RNA splicing or post-translational protein modifications. However, it still retains and places at its core the concept of *gene regulation*. DNA contains non-coding sequences that play a critical regulatory role in the expression of genes. Various proteins can selectively bind to regions of the DNA strand upstream of a gene domain and interfere, positively or negatively, with the RNA polymerase responsible for gene transcription. The two main classes of transcription factors are "activators" and "inhibitors" that respectively encourage and hinder gene expression. In a binary view, the regulatory sites are "switches" that literally turn genes on and off. Regulatory proteins bind to regulatory sites as keys fit into locks, which can cluster and combine to form complex regulatory functions. Lock-key pairs are reused for different genes or even the same gene at different times and places of the developing organism. Since regulatory proteins are themselves the product of gene expression, the cell's total biosynthetic activity can be approximately represented by a *gene regulatory network* (GRN), where proteins are considered hidden variables (figure 8.3a,b,e). In sum, gene expression is controlled by regulatory switches, which are themselves controlled by gene expression.

8.3.1.2 Patterning from gene regulation in embryo space

How does this complex web of many-to-many regulatory interactions unfold in 3-D space and time to create pattern formation? The pattern domains of embryogenesis are differentiated regions of gene expression, or *identity domains*. They represent the "hidden geography" of the embryo [9]; at any period of its development, the organism is segmented into multiple compartments of approximately homogeneous gene expression levels (see figure 8.4 for a preview). These compartments can be visualized, for example, by *in situ* hybridization methods (i.e., complementary RNA strands that recognize specific mRNA and are labeled with fluorescent or radioactive compounds). Based on these facts, simple recursive reasoning yields the following patterning rule. First, it is assumed that gene expression levels in each cell (or, equivalently, mRNA or protein concentration levels) can be represented by quasi-static variables, because their reaction kinetics quickly converges to constant attractor values. Then, the combined regulatory action of three genes A, B, and C upon one gene I can be denoted by $I = f(A, B, C)$, where I, A, B, and C represent the stable expression levels of those four genes within a given time

interval (figure 8.3b). Denoting by $\mathbf{r} = (x, y, z)$ the coordinates of a cell, $I(\mathbf{r})$ represents the spatial landscape of gene I's expression level across the embryonic cell population. Therefore, through the dependency in f, the basic patterning rule states that a gene landscape $I(\mathbf{r})$ results from a geometric interaction between several earlier gene landscapes $A(\mathbf{r})$, $B(\mathbf{r})$, and $C(\mathbf{r})$ (figure 8.3c,d,g). Typically, in a simplified binary format, gene levels are coded by two values, 1 for "high" and 0 for "low", and $I(\mathbf{r})$ defines a geometrical domain D_I such that $\mathbf{r} \in D_I \Leftrightarrow I(\mathbf{r}) = 1$. In this case, function f has a logical type, e.g., $I = (\neg A \wedge B \wedge C) = (1 - A)BC$, and the domain of high I expression is simply the intersection of high B, high C and low A expression, i.e., $D_I = (D - D_A) \cap D_B \cap D_C$, where D denotes the entire domain of the organism.

Thus, combinations of switches can create new patterns by union and intersection of precursor patterns. This principle was demonstrated in the periodic striping of the *Drosophila* embryo along its anteroposterior (A/P) axis. The dorsoventral (D/V) and proximodistal (P/D) axes are also segmented into distinct bands or layers and, by intersection with the A/P stripes, give rise to smaller domains such as the organ primordia and "imaginal discs". These groups of cells mark the location and identity of the fly's future appendages (legs, wings, antennae). Going back in time, the whole process starts with the establishment of concentration gradients due to the diffusion of various maternal proteins across the initial cluster of cell nuclei, the syncytium. These gradients are the functional equivalent of a coordinate system. The particular combination of protein concentration in each point becomes the first regulatory trigger in a cascade of gene expression. Let X, Y and Z represent the concentration levels of three hypothetical proteins that vary anisotropically along the three dimensions of an abstract embryo. For example, assuming uniform gradients $\nabla X = (\alpha, 0, 0)$, $\nabla Y = (0, \beta, 0)$ and $\nabla Z = (0, 0, \gamma)$, we obtain three linear concentration landscapes $X(\mathbf{r}) = \alpha(x - x_0)$, $Y(\mathbf{r}) = \beta(y - y_0)$ and $Z(\mathbf{r}) = \gamma(z - z_0)$. The first set of genes will be expressed in domains defined by regulatory functions of the type $A(\mathbf{r}) = g(X(\mathbf{r}), Y(\mathbf{r}), Z(\mathbf{r}))$, and $B(\mathbf{r}) = h(X(\mathbf{r}), Y(\mathbf{r}), Z(\mathbf{r}))$, etc. Then, these primary domains will intersect to give rise to secondary domains such as $I(\mathbf{r}) = f(A(\mathbf{r}), B(\mathbf{r}), C(\mathbf{r}))$, etc., as explained above. In summary, embryogenesis consists of a cascade of morphological refinements supported by a cascade of gene regulation reactions. Molecular gradients provide *positional information* [43] that is integrated along several spatial dimensions, in each cell nucleus, through a chain reaction of keys and locks.

8.3.1.3 The positional-boundary-identity gene network model

The principle of recursive morphological refinement suggests that, despite numerous feedback loops and an overall complex topology, developmental GRNs seem to be broadly organized in successive gene groups that correspond to successive growth stages and anatomical modules of the embryo. The early

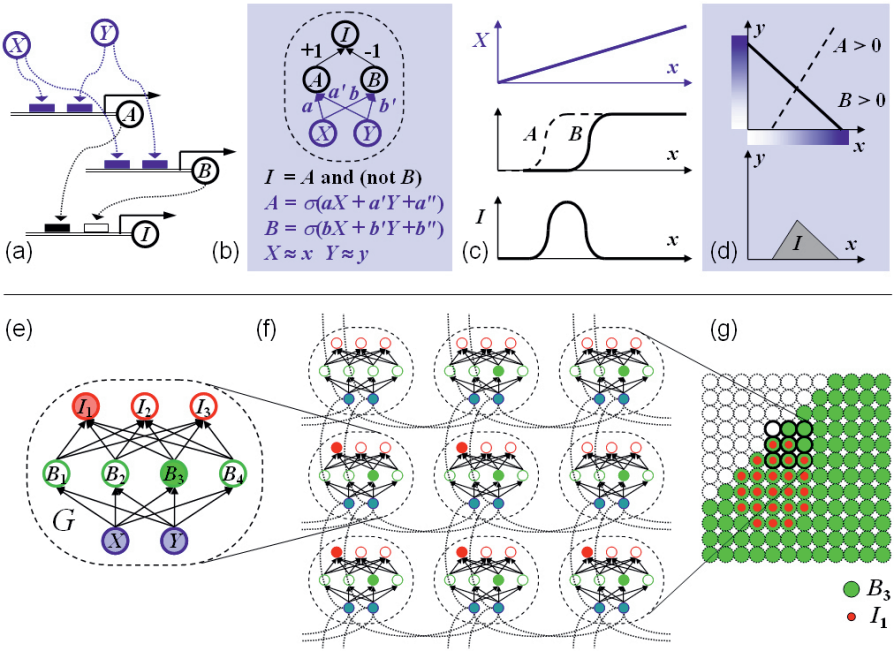

Fig. 8.3. Principles of spatial patterning from a lattice of positional-boundary-identity (PBI) gene networks. (a) Schematic top-down view of gene regulatory interactions on DNA strands. Proteins X and Y combine to promote the transcription of genes A and B by binding to their upstream regulatory sites, which produces proteins A and B (assuming a simple one gene-one protein relationship). Thereafter, A promotes, but B represses, the synthesis of I. (b) Formal bottom-up view of the same GRN. (c) Variation of expression levels on one spatial axis, construed as a chain of GRNs. The concentration of X follows a gradient created by diffusion. This gradient triggers a gain response in A and B at two different thresholds, thus creates boundaries at two different x coordinates (for a given Y level). These domains in turn define the domain of identity gene I, where A levels are high but B levels are low. (d) Same spatial view in 2-D. The domain of I covers the intersection between high A and low B. (e) Same type of PBI gene regulatory network as (a-b) with more nodes, denoted by G. (f) Detailed view of the architecture underlying the 2-D patterning of (d). Network G is repeated inside every cell of a lattice. (g) Local coupling of positional nodes creates gradients that create patterning. While G's structure and weights are cloned, node *activities* vary from cell to cell.

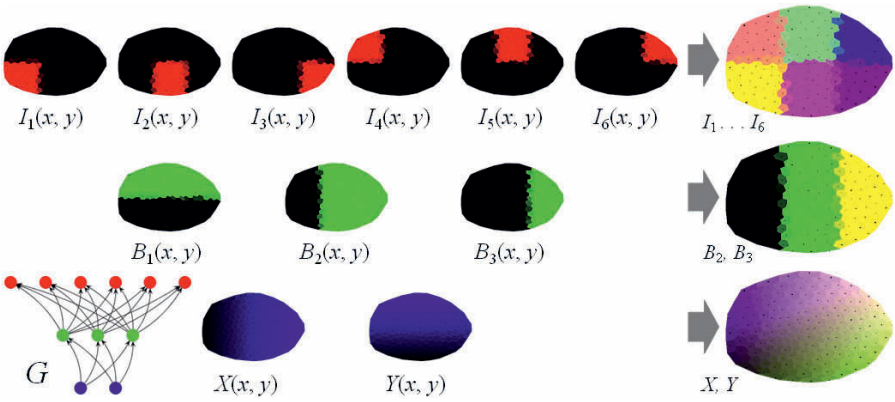

Fig. 8.4. Checkered self-patterning (top right) created by a simple 2P-3B-6I gene regulatory network G (as in figure 8.3b) in a 200-cell oval-shaped embryo. Each embryo view is selectively "dyed" for the expression map of one of the 11 genes, or a partial combination of these genes. With $X = x/x_{max}$, $Y = y/y_{max}$, weights are such that: $B_1 = \sigma(Y - 1/2)$, $B_2 = \sigma(X - 1/3)$, $B_3 = \sigma(X - 2/3)$; $I_5 = B_1 B_2 (1 - B_3)$, $I_6 = B_1 B_3$, etc.

striping process of *Drosophila* is controlled by such a regulatory hierarchy containing five main tiers of genes [8]. The present model relies on a three-tier caricature of the same idea, the *positional-boundary-identity* (PBI) network (figure 8.3b,e). In a 2-D virtual embryo, the bottom layer contains two "positional" nodes, X and Y; the middle layer, n "boundary" nodes $\{B_i\}_{i=1\ldots n}$; and the top layer, m identity nodes $\{I_k\}_{k=1\ldots m}$. Variables X, Y, B_i and I_k denote the gene expression levels of each of the $2 + n + m$ nodes. First, the positional activities follow gradients across the embryo, e.g., $X(\mathbf{r}) = \alpha(x - x_0)$ and $Y(\mathbf{r}) = \beta(y - y_0)$ as above. Then, in each cell, the boundary nodes compute linear discriminant functions of these positional nodes through the equation $B_i = \sigma(L_i(X,Y)) = \sigma(w_{ix}X + w_{iy}Y - \theta_i)$, where $\{w_{ix}, w_{iy}\}_{i=1\ldots n}$ are the regulatory weights from X and Y to B_i, parameter θ_i is B_i's threshold value, and sigmoid function σ is defined by $\sigma(u) = 1/(1 + e^{-\lambda u})$. The effect of a boundary node is, thus, to diagonally segment the embryo's plane into two half-planes of strong and weak expression levels (1 and 0). Finally, the identity gene levels are given by logical combinations of the near-binary expression levels of the boundary genes, for example, by calculating the products $I_k = \prod_i |w'_{ki}|(w'_{ki}B_i + (1 - w'_{ki})/2)$, where $w'_{ki} \in \{-1, 0, +1\}$ represent ternary weights from B_i to I_k. Thus, the contributing factor coming from B_i can take three possible values: $(1 - B_i)$, 0, or B_i. In the PBI model, the "identity domains", i.e., the regions of high I expression, are made of polygons at the intersection of multiple straight boundary lines (figures 8.3g and 8.4).

8.3.1.4 A lattice of gene regulatory networks

The full architecture of the virtual embryo is a network of networks. It consists of a lattice of cells, where each cell contains a gene regulatory network (GRN) [24, 30, 36] that can be, for example, of the PBI type just described (figure 8.3f). This lattice, however, is not necessarily rectangular or even regular. In the most general case, it is a swarm of nodes $c = 1 \dots N$ representing the cells' nuclei, with arbitrary coordinates $\{\mathbf{r}^c\}_{c=1\dots N}$, where $\mathbf{r}^c = (x^c, y^c)$ for an embryo in 2-D space or $\mathbf{r}^c = (x^c, y^c, z^c)$ for one in 3-D space. The nuclei are connected by edges that represent neighborhood relationships and dynamic coupling between GRNs (figure 8.8). The existence of an edge between two nodes c and d is established with respect to their Euclidean distance $\|\mathbf{r}^c - \mathbf{r}^d\|$, typically using a nearest-neighbor rule or the Delaunay triangulation that avoids gaps and crossings. The cell membranes can be either round or defined by their Voronoi region; in this model, no difference is made between a full cell volume and its nucleus. Denoting by G^c the GRN of cell c, a macroscopic edge $c \leftrightarrow d$ generally represents a complex coupling link between multiple gene nodes of G^c and G^d. In the experiments presented here, the embryo is a 2-D quasi-hexagonal Delaunay lattice and G^c is a PBI network. Intercellular coupling is restricted to X and Y positional nodes, where coupling strength between concentration levels X^c and X^d is symmetrical and depends on only $|x^c - x^d|$, and similarly for Y and y. As described above, this causes protein X to anisotropically diffuse on the x axis, following a gradient $X(\mathbf{r}^c) \cong \alpha(x^c - x_0)$, which is interpreted by the B and I layer inside each cell c and creates a pattern of gene expression $\{I_k(\mathbf{r}^c)_{c=1\dots N}\}_{k=1\dots m}$ on the lattice (figure 8.3g). In the binary approximation, where I_k is a product of near-binary B_i activities, the pattern consists of a patchwork of polygonal domains $\{D_{Ik}\}_{k=1\dots m}$ that can be partially overlapping. The embryo's partitioning into D_{Ik} territories is similar to the colorful compartments between lead cames in stained-glass works.

8.3.1.5 The feed-forward dynamics of gene network topology

In summary, under the simple feed-forward hypothesis, developmental genes are roughly organized in tiers, or "generations". Earlier genes map the way for later genes, and gene expression propagates in a directed fashion. First, positional morphogens create half-plane domains, and then domains intersect. Naturally, this is a crude caricature of real developmental GRNs. Although biological research has not fully unraveled the complex webs of regulatory gene-protein-gene interactions in any species, two important topological features of these webs have long been recognized, namely: (1) the existence of recurrent loops [15] and (2) gene multivalence. These features are not taken into account in the feed-forward network model, but small improvements should suffice to accommodate them. Concerning feature (1), recurrent loops and "feedback"

interactions can be added *within* each layer, while keeping the general multilayered architecture and prohibiting feedback from a higher layer to a lower layer. Feature (2) means that the same developmental genes can be reused at different periods and locations in the organism. Such genes, also called "toolkit genes", are in fact triggered by different switch combinations (multiple clusters of upstream regulatory sites on the DNA). Therefore, they can be formally represented by *duplicate* nodes placed in different tiers of the feed-forward network. In graph topological formalism, toolkit genes look like "hub" nodes that receive and send out numerous links. From the *dynamical* and activity propagation viewpoint, however, this resemblance disappears by segregating specific functional combinations of incoming and outgoing links from each other and duplicating the node. This suggests that developmental GRNs might not be so "complex" after all, in the scale-free [4] or small-world sense [38], but might essentially retain a directed acyclic graph (DAG) topology at the macroscopic level, filled with smaller cliques of recurrently connected nodes at a finer level. A longer discussion about the realism of GRN topology will not be pursued here. The main thrust of the present study is to motivate new ways of designing artificial systems by drawing inspiration from biological development, not to give a faithful account of biological reality.

8.3.2 The growing canvas: multiscale, recursive patterning

8.3.2.1 Hierarchical subpatterning

The primitive PBI network architecture used here is similar to the multilayered "perceptron" model of artificial neural networks. Its *generalization* power, i.e., ability to generate a wide variety of patterns, is problematic. A three-tier perceptron is theoretically universal, as it can produce any desired segmentation of the input space — here, the 2-D space of the input nodes X and Y, respectively, equivalent to coordinates x and y. This is in contrast to a two-tier PB perceptron without hidden layer, which can accomplish only linear partitions. In 2-D, again by analogy with stained-glass techniques, it means that any scene motif or embryo map in principle can be completely delineated by boundary lines (the "hidden units"), however fine its details may be. The homogeneous identity domains I (the "classes") then appear at the intersection of the half-planes defined by the B lines. Additionally, other types of B contours than straight lines can be employed; instead of the linear kernel L_i in $B_i = \sigma(L_i(X, Y))$, the boundary discriminant functions can be polynomial kernels $B_i = \sigma(P_i(X, Y))$ or other kernel types. However, an important question remains about the cost of this versatility. How many boundary nodes are necessary and sufficient to cover all the components of a segmentation pattern? As long as the different identity domains are not too numerous and remain reasonably connected, only a few B nodes are needed. On the other hand, as the identity domains become smaller and more fragmented, the number of B nodes increases rapidly, eventually tending to infinity in the limit of discontinuous points.

Thus, although theoretically versatile, the PBI network is in practice limited by scaling. As discussed earlier, biological embryo patterns develop in numerous incremental stages. An adult organism is produced through gradual morphological refinement, following a cascade of gene expression regulation from precursor gene tiers to secondary gene tiers, from secondary gene tiers to tertiary gene tiers, and so on. To account for this effect, the present model GRN is extended to include a pyramidal *hierarchy* of PBI networks, referred to as H-PBI (figure 8.5) and denoted by Γ. A pattern is now generated in a recursive fashion. First, the base PBI network G_0 at the root of Γ establishes the largest pre-identity domains (figure 8.5b). Then, in the next stage, another set of PBI subnetworks G_1, G_2, etc., partition these pre-identity domains into smaller identity compartments at a finer scale (figure 8.5c), and so on. The onset of a later PBI subnetwork $G_{\mu'} = \{X', Y', B'_{i'}, I'_{k'}\}$ is always controlled by one or several of the I nodes of an earlier PBI subnetwork $G_\mu = \{X, Y, B_i, I_k\}$. Formally, this can be written: $X'(\mathbf{r}) = \alpha'(x - x'_0)I_k(\mathbf{r})$, and same for Y'. It means that X' and Y' follow local gradients only inside, precisely, the domain D_{Ik} delimited by one of the identity nodes $I_k = 1$, and that they are zero everywhere else. This causal relationship is similar to the imaginal discs of *Drosophila*; once a territory D_{Ik} has been marked to be the future site of a leg or a wing (high I_k activity), a local coordinate system arises inside D_{Ik} in the form of gradients (such as X' and Y'), which then trigger the formation of a new subpattern $\{I'_{k'}\}_{k'=1\ldots m'}$ inside this territory, and so on [8]. Form details are added in a hierarchical or "fractal" fashion, analogous to the local inclusion of small stained glass motifs into bigger ones. Fractal patterning has also been explored in "map L-systems" [32], but using symbolic rules in an explicit geometrical representation.

8.3.2.2 Expansion

Simultaneously, the embryo grows as cells continue to divide and proliferate (figure 8.5a'-c'). Hence, multiscale patterning actually consists of two fundamental processes playing out in parallel: (1) the partitioning of identity domains into smaller identity domains, and (2) the continuous expansion of identity domains. During cell division or mitosis, the two daughter cells inherit the current expression state of the mother cell. Biologically, this state corresponds to mRNA, protein and metabolic concentrations; in the present model, it is represented by the set of values of the GRN nodes. Domains thus preserve their identity during expansion (the I nodes of G_0), while they are also occupied by new local gradients of positional information, i.e., new regional coordinate systems (X', Y') that activate the next PBI module in the hierarchy (G_1 or G_2) Biologically, these new local gradients might emerge from a diffusion process similar to the original diffusion of the global coordinates X, Y. For example, during proliferation a small number of daughter cells asymmetrically inherit more signaling proteins than their siblings, and then these proteins start diffusing from one border of the domain where the cells

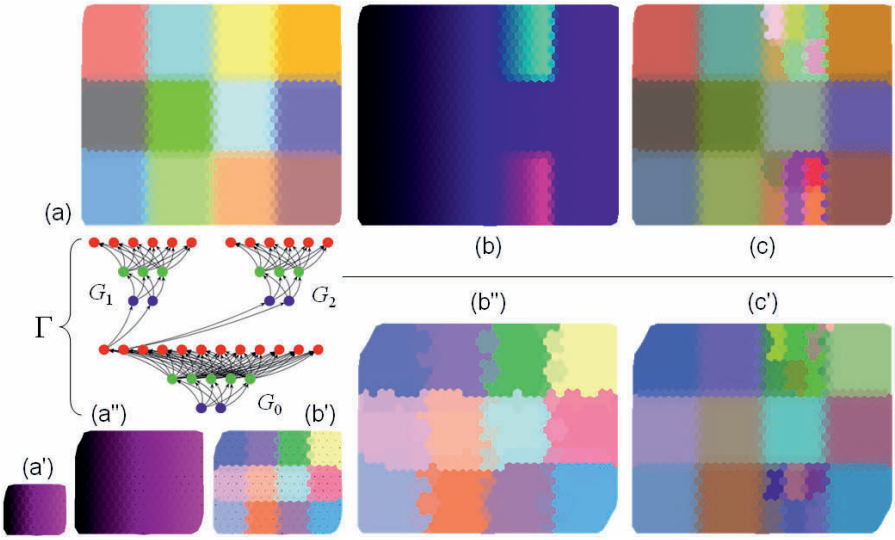

Fig. 8.5. Static and growing multiscale canvas. On a 32×32 hexagonal lattice of cells, an H-PBI gene network Γ gives rise to a "fractal" pattern in two steps: first, the base subnet G_0 (5B-12I) marks 12 rectangular segments (a) as in figure 8.4; then, two secondary subnets G_1 and G_2 (3B-6I) triggered by I_1 and I_2 create local gradients in two of those segments (b), subsegmenting them into six smaller domains (c). An equivalent pattern is obtained by a cell mass uniformly expanding from 8×8 (a') to 16×16 (a"-b') to 32×32 cells (b"-c'), while patterns continue to form and gradients continue to diffuse, as above.

gathered. However, in the present version, the gradients of P node activity are not modeled as diffusion processes but directly calculated from the geometrical shape of the identity domain boundaries according to $X'(\mathbf{r}) = \alpha'(x - x'_0)$ and $Y'(\mathbf{r}) = \beta'(y - y'_0)$. This shortcut is a slight violation of the localized dynamics, and is only aimed at replacing the lengthy convergence process of heat-like diffusion with its already-known final state (an approximately linear gradient, depending on the boundaries).

In this section, it is also assumed that all cells divide equally and simultaneously, i.e., the medium expands uniformly according to a geometrical law. For example, in a regular planar hexagonal lattice containing N cells, there is an average of $3N$ intercellular edges. At each expansion step, one new cell is added in the center of every $c \leftrightarrow d$ edge, and edge lengths are doubled to restore the original intercellular distance. The result is a new regular hexagonal lattice that has $4N$ cells and a surface area four times larger, but still has the same shape as the previous lattice (figure 8.5a", b"). Note that, since the medium is expanding, at the same time as new gradients emerge and finer details are added, the typical scale of patterning is not diminishing but in fact remains approximately constant, being on the order of magnitude of an aver-

age cell size. Following an artistic metaphor, the "growing canvas" continues to paint itself using the same "brush size" [9].

8.3.2.3 Modularity

Ordering genes in a multiscale hierarchy of the H-PBI type is a convenient way to guide self-patterning. Instead of trying to render all morphological details at once, these details arise in successive waves of expression from the broadest territories down to the finest patches. However, an even greater benefit intrinsic to a hierarchical network is *modularity*. As soon as layers and subnetworks have been defined, they can be reused as units of local computation. Modularity is a fundamental principle of genotype-phenotype economics in development and evolution [31]. Biological organisms often contain numerous repeated or "serially homologous" parts in their body plan [8]. This is most striking in the segments of arthropods (several hundreds in millipedes) or the vertebrae, teeth and digits of vertebrates. After duplication, these parts tend to diversify and evolve more specialized structures (lumbar vs. cervical vertebrae, canines vs. molars). Homology exists not only within individuals but also between different species, as classically shown by comparing the forelimbs of tetrapods from the bat to the whale. Recently, genetic sequencing has revealed that many stretches of DNA are in fact identical or highly similar. This came to support the idea that homology is the evolutionary result of *duplication* followed by *divergence* through mutation (and, sometimes, loss again).

Beyond biology, modularity is also a pervasive trait of many other natural complex systems [6]. In systems engineering, it means not only copying and reusing a partial design in different locations of an architecture, but also being able to independently modify these copies. In the present H-PBI model, this corresponds to connecting several identity genes $I_1 \dots I_k$ of a base network G_0 to a *unique* subnetwork G_1 (figure 8.6a) or, alternatively, *multiple copies* of the same subnetwork G_1, G_2, etc. (figure 8.6d). In the first case, the local pattern generated by G_1 is always identical in all primary domains $D_{I1} \dots D_{Ik}$, whether appearing in its original form (the eight +-shaped subdivisions in figure 8.6b) or in a mutated form (X-shaped subdivisions in figure 8.6c). In the second case, the local pattern can be initially identical (as in figure 8.6b-c), but then it has the possibility of evolving independently in each location and produce dissimilar variants of itself (the miscellaneous inclination angles θ of figure 8.6e). Additional mutations in the base network G_0 can change the whole body map (thinner center row d and thicker borders of figure 8.6f) without affecting the individual motifs G_1, G_2, etc. Modularity therefore plays at all levels of the GRN. The demonstration of figure 8.6 is similar to the "arthromorph" program [10] that generates chains of limbed segments representing artificial arthropods. In that simulation, "genes" control mutations at three levels: globally (same variation in all segments), in groups (same variation in a few adjacent segments only) or individually (distinct variation in every

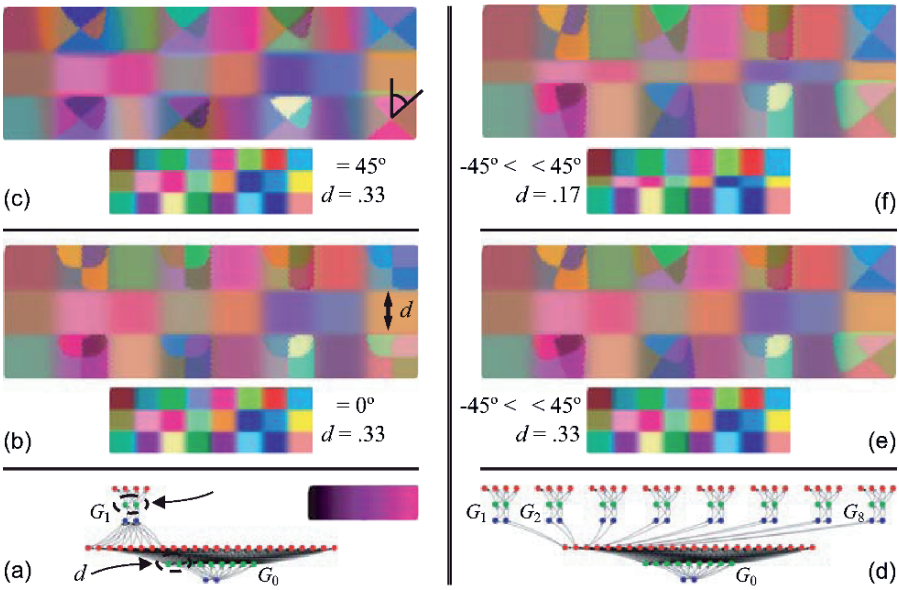

Fig. 8.6. Modularity in a segmented embryo (see text). A cell mass uniformly grows from 12×32 (inset) to 24×64 to 48×128. (a)-(c) A mutation in the unique subnetwork G_1 generates the same angle in all eight "limb" patterns. (d)-(e) Independent mutations in duplicated subnetworks $G_1 \ldots G_8$ create different angles. (f) A mutation in the base network G_0 modifies the limbs' height, without affecting their internal patterns.

segment). However, the virtual genes of arthromorphs code directly and arbitrarily for their macroscopic geometrical features. The example of figure 8.6 achieves a similar effect through the decentralized emergence of a myriad of microscopic states in a multicellular developmental model.

8.3.3 The deformable canvas: cell adhesion, division and migration

The growing canvas model based on the hierarchical gene regulation network H-PBI (section 8.3.2) is more powerful in generating a wide range of patterned images than the fixed canvas using a single three-tier PBI (section 8.3.1). With a growing canvas, it should be possible to meta-design a generative algorithm that could reconstruct any given image in a multiscale, fractal-like fashion. Such an algorithm would automatically "reverse compile" an image to produce the correct pyramidal GRN to be placed in every cell of the expanding lattice (figure 8.7). However, this is not the primary object of this work. What was achieved so far is only a model of genetically guided *patterning*, not morphogenesis in the sense of *shape* formation. The canvas' growth presented in the previous section is geometrically homothetic, i.e.,

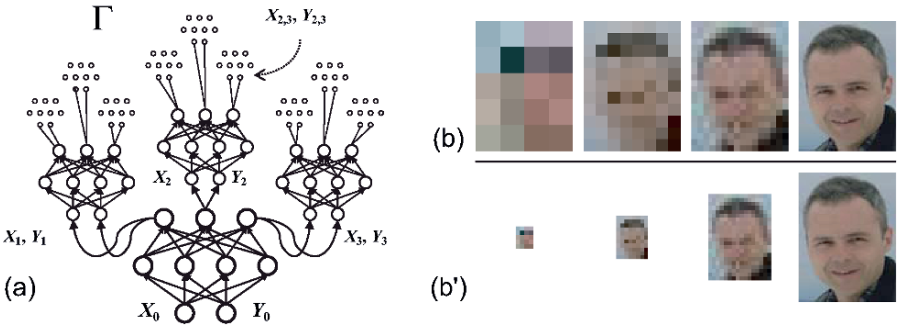

Fig. 8.7. The universal growing canvas (conceptual illustration, not actual simulation). As in figure 8.5, a generalized hierarchical GRN (a) could in principle reconstruct any image in a multiscale iterative fashion, in a fixed (b) or uniformly expanding (b') mass of cells (self-portrait metaphor after [9]).

an initial rectangular sheet of cells in 2-D remains roughly rectangular, even though its internal patterning can become very intricate.

Continuing to explore the principles of multicellular development as an inspiration for the self-organization of artificial systems, the model will be further improved to incorporate elements of cellular *biomechanics*. What is missing from the previous homothetic canvas is a topological *deformation* dynamics, or "morphodynamics", that can confer a nontrivial shape to the organic system. To this purpose, three principles are added in schematic formulations: (1) *elastic* cell rearrangement under differential *adhesion*, (2) inhomogeneous cell *division*, and (3) tropic cell *migration*. Practically for the model, all these mechanistic principles have the effect of varying the cells' coordinates in 2-D or 3-D space. Lastly and more importantly, we need to add a rule that relates those principles to the original self-patterning process. In this new "deformable canvas" model, a critical part of our meta-design effort is the establishment of a *functional dependency between cell identities and mechanical cell behaviors*. Just as the identity nodes I_k can propagate gene expression activity into subordinate PBI modules to create local segmentation patterns, the same I_k nodes now also trigger behavioral changes of the above (1), (2), (3) types in the cells where they are highly active.

8.3.3.1 Differential cell adhesion and elasticity

Lattice edges connecting cell centers are modeled as springs with force constant k and length r_0. Viscous resistance is also included with coefficient η. Thus, the equation of movement reads

$$m \cdot \ddot{\mathbf{r}}^{cd} = -k \left(1 - \frac{r_0}{\|\mathbf{r}^{cd}\|}\right) \mathbf{r}^{cd} - \eta \dot{\mathbf{r}}^{cd} \,. \tag{8.1}$$

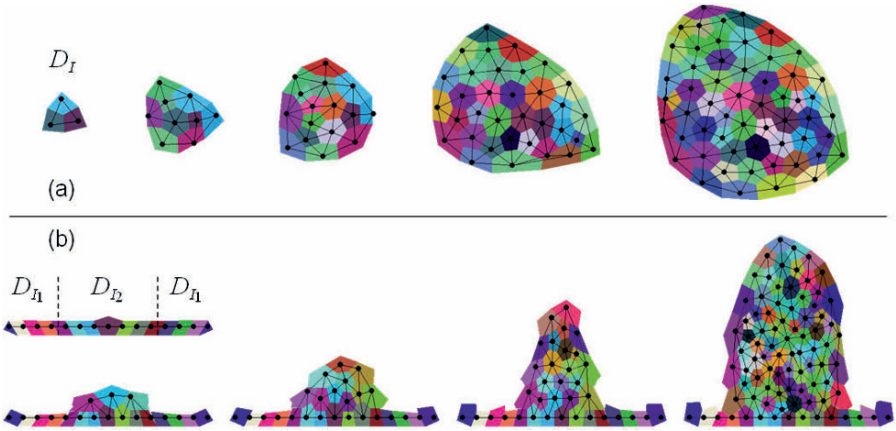

Fig. 8.8. Cell adhesion and elasticity. A simple mesh model illustrates the biomechanical behavior of a growing cell mass. No genetic network is used here; cells have arbitrary colors. Lattice edges and polygons result from a Delaunay-Voronoi tessellation (corrected on the periphery). (a) Isotropic "blob" of identical type-I cells dividing at 1% rate, in which nearby daughter cells rearrange under elastic forces (see text). (b) Anisotropic "limb" growth: from the initial 2-type cell sheet, only the center domain D_{I2} and its offspring divide (upward stretch modeled by 2x:y anisotropic rescaling). The eight lateral cells have different identity I_1 and no adhesion to the I_2 lineage.

Neglecting the effect of inertia, the coordinate update rule at each time step $\Delta t = 1$ becomes:

$$\Delta \mathbf{r}^c = -\Delta \mathbf{r}^d = \frac{\Delta \mathbf{r}^{cd}}{2} = -\frac{k}{2\eta}\left(1 - \frac{r_0}{\|\mathbf{r}^{cd}\|}\right)\mathbf{r}^{cd}\,. \tag{8.2}$$

Under this simple model of elastic rearrangement, each cell tends to optimize the distance $\|\mathbf{r}^{cd}\|$ with its neighbors' nuclei to reach r_0, i.e., occupy a convex volume of typical diameter r_0 (figure 8.8). Biological cells also stick to each other by means of adhesion proteins that cover their membrane. The great diversity of adhesion proteins gives them the ability to selectively recognize each other, thereby modulating the intercellular adhesion force or "stickiness". Some cells slide along one another without attaching, while other form tight, dense clumps. In the elastic force model, differential adhesion can be modeled by allowing the spring constant to vary from edge to edge, which means replacing k with k^{cd} in the above equation. For the limb-like growth illustrated in figure 8.8b), there is no adhesion between domains D_{I1} and D_{I2}, so $k^{cd} = 0$ between any cell c of identity I_1 and any cell d of identity I_2. In figure 8.9c'-d', adhesion is zero between either D_{I1} or D_{I2} and the rest of the embryo.

8.3.3.2 Inhomogeneous cell division

This mechanism is similar to cell proliferation at the basis of the homothetic expansion seen in section 8.3.2.2. The new aspect here is that cells divide according to a *non-uniform* probability that essentially depends on their *genetic identity*, i.e., the D_I domains to which they belong (figure 8.9). This means that the probability of division of cell c located in $\mathbf{r}^c$, denoted by $p(\mathbf{r}^c)$, depends at any time only on the current state of activity of the I nodes in the cell's H-PBI gene network: $p(\mathbf{r}^c) = p(I_1(\mathbf{r}^c), I_2(\mathbf{r}^c), \ldots)$.

Two daughter cells c and d resulting from the division of cell c under this probability are initially positioned next to each other at a small distance δ and a cleavage angle θ, also drawn at random (possibly from a non-uniform distribution in the case of anisotropic proliferation). By denoting $\mathbf{r}^{cd} = \mathbf{r}^c - \mathbf{r}^d$ the vector on directed edge $c \leftarrow d$, this means $\mathbf{r}^{cd} = (\delta \cos\theta, \delta \sin\theta)$. Then, the positions of c, d and the cells in their neighborhood are rearranged under elastic constraints implemented by the edges (explained in the previous section). Differential proliferation rates based on genetic identities produce bulges and deformations in the embryo shape, as some compartments expand faster than others (figure 8.9a-d), resembling organogenesis. Using anisotropic cleavage planes and anisotropic rescaling transformation $x : y \rightarrow ax : by$, this model can also generate directional offshoot akin to limb development (figures 8.8b and 8.9a'-d').

8.3.3.3 Tropic cell migration

A specificity of animal development, largely absent from plant development, is cell migration. Individual cells or groups of cells burrow their way through the cellular mass and extracellular matrix to colonize remote locations of the developing embryo. Depending on the adhesion properties of the migrating cells, they can either globally preserve their neighborhood relationships by "flocking" together or, on the contrary, detach from their immediate surroundings to create new intercellular bonds elsewhere. In the first case, migration happens *en masse* and takes the aspect of elastic sheet deformation. The most striking example of collective crowd movement is gastrulation, a complex folding event that forms the fundamental germ layers of the embryo in its earliest stages. The second case is best illustrated by neural crest cells that leave the dorsal neural tube to form other structures far from their source. However, it is often unclear which type of migration is predominant and most biomechanical deformation processes involve a mix of collective and individual dynamics. Often, an "active" group of cells entrains a more "passive" mass in its trajectory, like a locomotive. The existence of cells that act like singularities makes especially difficult a description purely based on continuous surfaces and differential geometry and requires discrete multi-agent simulations. The locomotion mechanisms responsible for cell migration are not fully understood. Generally, a cell is motivated to migrate by attraction toward specific

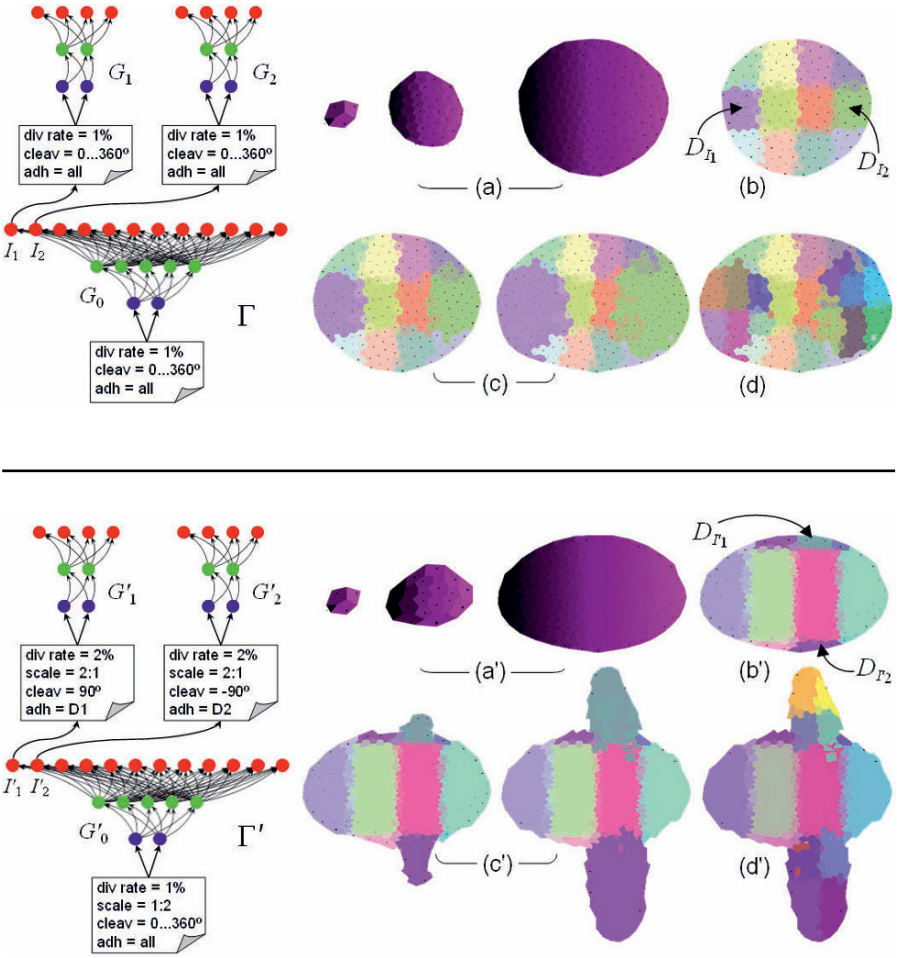

Fig. 8.9. "Organogenesis" by non-uniform cell proliferation. As in figure 8.5, a checkered embryo emerges from an H-PBI gene regulatory network Γ. Here, new cellular behavioral rules are added. Cells with high levels of identity genes I_1 and I_2 are prompted to further divide at the rate of 1% (c) (while others have stopped), before expressing subpatterns G_1 and G_2 in their newly formed anterior and posterior territories (d). Different weights in base module G'_0 of Γ' make a thicker central row and place $D_{I'1}$ and $D_{I'2}$ on the dorsal and ventral sides (b'). Moreover, different values of cleavage angles, anisotropic rescaling and adhesion coefficients ($k^{cd} = 0$ between $D_{I'1,2}$ and the rest) provoke I'_1 and I'_2 cells to grow "limbs" (c'), which are also subpatterned by G'_1 and G'_2 (d').

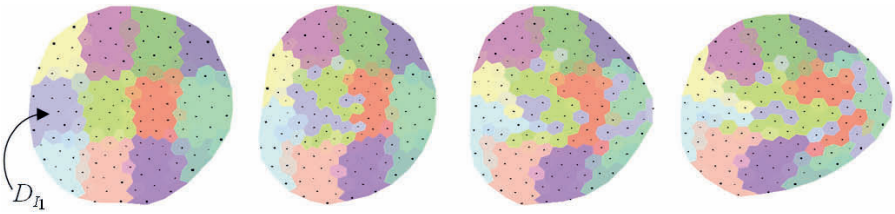

Fig. 8.10. Cell migration. Using the same gene regulatory network Γ as figure 8.9, the behavioral parameters of cells I_1 are replaced with a migration rule. Before proliferating, these cells push their way across the embryo toward increasing X concentration (here, eastward). This is modeled by adding to $\Delta\mathbf{r}$ a bias $\mathbf{u}$ that depends on genetic identity.

chemical signals that it recognizes ("chemotaxis"). These signals trigger its motion and guide it on its route toward its target. In the present model, this behavior can be simply implemented by adding an identity-dependent bias vector $\mathbf{u}^c = \mathbf{u}^c(I_1(\mathbf{r}^c), I_2(\mathbf{r}^c), \ldots)$ to the $\Delta\mathbf{r}^c$ equation of section 8.3.3.1 (figure 8.10).

8.3.4 The excitable canvas: organic computing

Meta-designing laws of artificial development with inspiration from biology, as the present model attempts to demonstrate, is a challenging engineering task. It combines schematic modeling of natural complex systems, such as embryogenesis, and departure from the natural model toward free invention. Yet, developmental modeling is only the first half of the IMD effort. Another important problem is *functional* meta-design. Once an "organic system" is mature, what could be its computing capabilities? What do the artificial cells and organs (identity domains) of an embryomorphic system *represent* in practice?

In biology, it is difficult to establish a distinction between a purely developmental and a purely physiological regime. Real cells are already "functional" as soon as they are produced by mitosis and this function partakes in development. For example, the bioelectrical signals endogenously generated by neurons play a critical role in establishing synaptic contacts in the brain. While connectivity obviously supports the exchange of activity, there also exists an important feedback from activity to connectivity dynamics. Nodes start communicating before edges are fully built, leading to self-structuring of the network [12].

In artificial embryomorphic computing systems, it is conceivable to keep these two phases relatively separate by distinguishing between "activity for development" and "activity for function". The intercellular transmission of positional information by regulatory coupling between genes (figure 8.3b) represents activity-for-development. Beyond a certain degree of maturity of the growing system, this type of activity would gradually or abruptly diminish

and give way to the other type of activity, serving a different purpose toward functional computing. It does not mean that the system would entirely stop developing; the morphodynamics would still be active, mostly to fulfill important self-repair tasks and provide robustness to perturbations. If one or several cellular components fail, they could be quickly replaced by the still-active growth potential of their neighbors. (Self-repair properties have not been verified yet in the current embryomorphic model.) Yet developmental activity would be mostly dormant during functional maturity.

Now, after finishing the self-assembling stage and while constantly under self-repairing mode, what *type* of computation could be carried out by the embryomorphic system? As a speculative proposal, by its very 2D or 3D spatial nature, the organism could become the substrate of *excitable media* dynamics. After creating slow, quasi-static developmental patterns, cells could form fast and transient *dynamical* patterns. Depending on their identity domain, local groups of cells would synchronize in different ways and enter various regimes of collective spatiotemporal order (figure 8.11). Computation in the organic canvas would consist of emerging patches of moving and shimmering spots, stripes, target or spiral waves. We find again the types of reaction-diffusion patterns of figure 8.2a, which played only a limited role in development, now appearing and disappearing on the short time scale, and on top of genetically guided development. These phenomena are common in nonlinear chemical reactions or multicellular structures [42, 34], such as slime mold aggregation, heart tissue activity, neural networks, etc. These systems all have in common the ability to position themselves in a critical state, from which they can rapidly bifurcate between chaos and order. In this perspective, the "self-made tapestry" [3] would become a self-made screen or "sensitive plate" of coupled oscillatory units in the sense that certain external patterns of initial conditions can quickly trigger internal patterns of collective response from the units.

This is already the case in neural computation. New spiking network models that take into account the fine temporal structure of neural signals have revealed a great diversity of collective spatiotemporal regimes: synchronization and phase locking, delayed correlations and traveling waves, rhythms and chaos. Through recurrent (and plastic) synaptic connections, neural cells transiently interact as dynamical subnetworks that promise an immense richness of coding expression and computational power, combining the discrete and the continuous. What could still be missing from the current embryomorphic model are long-range connections. The inspiration from embryogenesis is currently limited to geometrical 2-D or 3-D lattices but might be complemented with complex "N-D" networks arising in ways similar to neurogenesis and synaptogenesis.

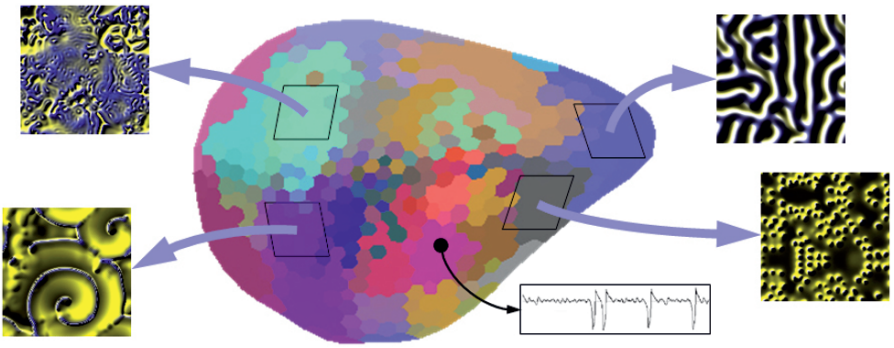

Fig. 8.11. The excitable canvas. A hypothetical view of an embryomorphic system in which genetic identity domains (colored patches) could support excitable media (zoom-in square insets), through coupling between quasi-oscillatory units (example of one temporal signal in the rectangle inset). Various regimes of dynamical activity could emerge on these substrates. Such spatiotemporal patterns might hold a great potential for representational and computing properties. (The square insets are snapshots of simulations run on Tim Tyler's demo applet http://texturegarden.com/java/rd, which implements the Gray-Scott model described in [29] under various sets of parameters.)

8.4 Discussion: planning the autonomy

At the core of the complex systems engineering enterprise lie paradoxical objectives: How can decentralization be controlled? How can autonomy be planned? How can we expect specific characteristics from systems that are otherwise free to invent themselves? The challenge is not only to *allow* self-organization and emergence but, more importantly, to *guide* them. First, it consists of preparing the conditions and mechanisms favorable to a robust and *reproducible* — as opposed to random — pattern formation process, under genetic control (IMD). Second, it consists of steering this process toward desired *goals*, while simultaneously leaving the door open to the spontaneous generation of innovative designs by evolution of the genes (EMD).

8.4.1 Growth, function, evolution

When meta-designing an embryomorphic artificial system, the engineer faces three main tasks: IMD1: How does the system *grow*? IMD2: How does the system *function*? EMD: How does the system *evolve*? The goal of the IMD phases is to define the developmental and computing mechanisms. The goal of the EMD phase is to define the rules of their evolution by variation and selection (section 8.4.2).

8.4.1.1 Developmental mechanisms

Growth results from a combination of elementary mechanisms, described and elaborated in sections 8.2 and 8.3. At the microscopic level of the embryomorphic model, it is grounded in a repertoire of basic cellular behaviors: cells change state (genetic expression), communicate (positional signals), cluster or detach (differential adhesion), travel (migration), create offspring (mitosis) or die (apoptosis). At the mesoscopic level of cell populations, several morphogenetic processes emerge: *guided patterning* through GRN-controlled expression maps, *organogenesis* through differential adhesion and domain-specific division rates, folding and *deformation* through elastic constraints and *sculpting* through cell removal. Additionally, superficial *free patterning* (spots, stripes) can also arise by reaction-diffusion (figure 8.2a). Starting from a single cell, a complex and organized architecture develops through the repeated application of a series of these principles, identically programmed inside each cell. Task IMD1 consists of choosing among these principles and designing their dynamics and parameters.

8.4.1.2 Functional mechanisms

Function roughly starts after growth (see section 8.3.4). Task IMD2 is about defining the *nature* of the cells and the type of computation that they carry out at the microscopic level. It also involves defining the range of macroscopic abilities of the system and its input/output interfaces with the environment. Do cells represent some kind of hardware components on a board, taking part in global digital-analog electric or optical activity patterns? Are they small modules of software logic that execute symbolic instructions on more conventional architectures in a "virtual machine" fashion? Can they actually be physical parts and blocks, joined together to support sensing, planning and acting in a robot? Or are they even full-fledged robots that coordinate in swarm formations for collective performance?

8.4.1.3 Evolutionary mechanisms

Evolution of both growth and function is the responsibility of the evolutionary meta-designer (EMD), who must define how the system *varies* (randomly) and how it is *selected* (nonrandomly). These points are discussed in the next section.

8.4.2 Selection without expectations

Three degrees of constraints that drive the fitness criteria and the artificial selection process can be identified, in decreasing intensity: (1) selecting for a specific system *architecture*, (2) selecting for a specific system *function*, and (3) selecting the *unexpected*.

8.4.2.1 Selecting for architecture

On the first level, the EMD engineer imposes tight requirements to obtain particular organismal patterns and shapes from the development process. Here, a reverse engineering problem must be addressed: Given a desired phenotype, what should be the genotype that will reliably reproduce this phenotype? One solution, if available, is the deterministic reverse compilation of the genotype from the phenotype. This is what Nagpal [25] has achieved in her virtual origami model. Given a macroscopic folding recipe (based on an ordered set of lines), she can automatically generate the exact microscopic developmental rules that each identical point of the medium must follow to reproduce the shape (based on wave propagation and state change). It is possible that such a method is also within reach in the present embryomorphic model, but this has not been investigated to-date.

In most cases, however, it is safe to assume that no algorithm for the reverse compilation of the genotype from the phenotype is available. Depending on the number of dimensions along which the system may vary, the search in parameter space can then appear nearly impossible. In the embryomorphic model presented in section 8.3, these parameters are the set of gene-to-gene regulatory weights together with the local behavioral rates of division, migration, gradient diffusion, and so on. (typically, the Γ network of figure 8.9, branching out like figure 8.7a). A naive fitness function rewarding only the final shape would create one (or possibly several) "narrow" peaks limited to local neighborhoods of those parameters, which would be unreachable for all practical purposes. However, this is the fallacy of "Mount Improbable" explained by Dawkins [10]. Biological evolution does not create complex organisms in one shot, but through a multitude of successive steps, randomly generated and nonrandomly selected. Each favored step brings a small "improvement", adding to the body plan a little more complexity, which gets rewarded by successful interaction with the physical environment. A famous example is how the eye has evolved multiple times from mere photosensitive cells, through gradual inward curvature of the epithelial tissue, condensation of the lens, etc. These changes were probably encouraged by an ever-increasing quality of the projected image (hence, an increasing survival probability under co-evolution pressure), as modeled by Nilsson and Pelger [27]. Behind the daunting cliff of a high fitness peak, hides a long and gentle upward slope.

Similarly, in the artificial systems challenge, an evolutionary meta-designer should also replace a jagged final-shape fitness landscape with a smooth transitional-shape fitness landscape. The best method to accomplish this is to define a score value or "distance" to the desired shape. This value must be an increasing function of favorable mutations, i.e., mutations that bring the developed system closer to the ideal template. It is conjectured here that this well-known principle of gradual search might actually *benefit*, not suffer, from the high parameter dimensionality offered by a true underlying *embryomorphic* model, such as the one presented in this chapter. A hierarchical gene

regulatory network of the H-PBI type hyperdistributed in a large cell mass might be in a better position to provide the necessary fine-grain mutations required by the gentle slope approach than the more direct genotype-phenotype mapping of traditional genetic algorithms.

8.4.2.2 Selecting for functionality

Why make the effort to devise a sophisticated self-organizing system only to force this system to produce a specific shape? Why not directly build the final shape in the first place? One important benefit would be robustness through homeostasis. Even though the final plan is known in advance, a genuine developmental dynamics is also expected to be intrinsically "self-repairing" (as mentioned in section 8.3.4). Nonetheless, requiring a specific architecture somehow defeats the idea of "stepping back" and meta-designing. Intervening in the microscopic placement of cells, whether by reverse compilation or fine-tuning, literally reintroduces the micromanagement attitude of classical systems design.

On the contrary, EMD engineers should abstract themselves even further from structural details and concentrate on selecting for the *functionality* of their system, otherwise leaving it complete freedom of morphology. Here, the same gradual optimization strategy as described above can be employed, except that the continuous distance quantity would not measure morphological resemblance but rather the closeness of *performance* to predefined goals. Given a task or repertoire of tasks to accomplish, it means ranking candidate systems according to their *partial* success in fulfilling these tasks, then allowing the most successful ones to reproduce faster and mutate, and so on. Afterwards, it is always instructive for the meta-designer to open the "black box" of the winning architectures and try to understand how they have come about and which specific subsystems or modules related them to success. The solutions "invented" by spontaneous evolution are often surprising and convoluted, in other words, remote from what a human designer would have conventionally designed.

Functional selection under free organization is the strategy adopted by most evolutionary computation works that also contain elements of distributed architectures or (small-size) complex systems. For example, this is the case of the logical functions computed by randomly composed multi-instruction programs in Avida [19], the locomotion abilities created by randomly articulated multi-segment robots in Golem [20] or Framsticks [17], or the shooting skills of intelligent video game agents emerging from randomly assembled multi-neuron networks in Nero [33]. Again, it is argued here, although not proven, that an even larger number of agents, such as in multicellular embryogenesis, would be even more favorable to a successful evolutionary search.

8.4.2.3 Selecting the unexpected

The increase in system size, however, might also require the EMD engineer to "let go" one step further and give up on specific selection requirements altogether. In summary, it is likely that the ultimate reconciliation between the antagonistic poles of planning and autonomy would be based on two complementary aspects. In order for an evolutionary process to successfully find "good" regions in systems state space, (a) the variation-by-mutation mechanisms must be fine-grained and rich enough to offer a large number of search paths (system size) and, at the same time, (b) the selection criteria must be loose enough to allow a large number of fitness maxima ("letting go"). With more search paths covering more fit regions, evolution is more likely to find good matches. Point (a) concerns the intrinsic ability of complex systems to create a *solution-rich* space [23] by combinatorial tinkering on highly redundant parts. Variational power is the most critical aspect of evolutionary processes; developmental systems made of a great number of small self-assembling components have the unique ability among all systems to generate behavior-rich variations. Point (b) concerns the ability of the meta-designers to relax their specifications, within reasonable limits of what is permitted by the system's environment and the problem at hand, and accept to be surprised — hopefully, in a pleasant way — by the outcome. Organic systems engineers will probably need to learn how to greatly diversify their demands and rather stand on the side, ready to harvest possibly "interesting" or "useful" organisms from a free-range menagerie.

Acknowledgments

This work was made possible in part by the EU projects *Embryomics* (FP6-NEST-2003-1-12916) and *BioEmergences* (FP6-NEST-2004-Path-IMP-28892). I thank Paul Bourgine for his warm support and encouragements. I am also grateful to Francisco Vico, Daniel Lobo, Jan Gecsei, Gregory Horman, and Rolf Würtz for their useful comments on the manuscript and editing help.

References

1. H. Abelson, D. Allen, D. Coore, C. Hanson, G. Homsy, T. Knight Jr., R. Nagpal, E. Rauch, G. Sussman, and R. Weiss. *Amorphous Computing.* MIT Artificial Intelligence Laboratory memo no. 1665, Aug. 1999.
2. W. S. Bainbridge, and M. C. Roco, eds. *Managing Nano-Bio-Info-Cogno Innovations: Converging Technologies in Society.* Berlin: Springer Science and Business Media, 2006.
3. P. Ball. *The Self-Made Tapestry: Pattern Formation in Nature.* Oxford University Press, 1999.

4. A.-L. Barabási, and R. Albert. Emergence of scaling in random networks. *Science* **286**: 509-512, 1999.
5. *Beyond the Horizon: Anticipating Future and Emerging Information Society Technologies* Final Report, ERCIM, FET, IST, EU, 2006.
6. W. Callebaut, and D. Rasskin-Gutman, eds. *Modularity*. The MIT Press, 2005.
7. S. B. Carroll. *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom*. W. W. Norton & Company, 2005.
8. S. B. Carroll, J. K. Grenier, and S. D. Weatherbee. *From DNA to Diversity*, Blackwell Scientific (Malden, MA), 2001.
9. E. Coen. *The Art of Genes*. Oxford University Press, 2000.
10. R. Dawkins. *Climbing Mount Improbable*. W. W. Norton & Company, 1996.
11. R. Doursat. The growing canvas of biological development: Multiscale pattern generation on an expanding lattice of gene regulatory networks. *InterJournal: Complex Systems* **1809**, 2006.
12. R. Doursat, and E. Bienenstock. Neocortical self-structuration as a basis for learning. *5th International Conference on Development and Learning (ICDL)*, Indiana University, Bloomington, Indiana, May 31-June 3, 2006.
13. G. M. Edelman. *Topobiology: An Introduction to Molecular Emrbyology*. Basic Books, 1988.
14. A. Gierer, and H. Meinhardt. A theory of biological pattern formation, *Kybernetik* **12**: 30-39, 1972.
15. S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* **22**: 437–467, 1969.
16. M. W. Kirschner, and J. C. Gerhart. *The Plausibility of Life: Resolving Darwin's Dilemma*. New Haven and London: Yale University Press, 2005.
17. M. Komosinski, and S. Ulatowski. Framsticks: Towards a simulation of a nature-like world, creatures and evolution. In D. Floreano, J.-D. Nicoud, and F. Mondada, eds., *5th European Conference on Advances in Artificial Life (ECAL-99)*, pp261-265, Lausanne, Sept. 13-17, 1999.
18. S. Kondo, and R. Asai. A reaction-diffusion wave on the skin of the marine angelfish *Pomacanthus*. *Nature* **376**: 765-768, 1995.
19. R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami. The evolutionary origin of complex features. *Nature* **423**: 139-144, 2003.
20. H. Lipson, and J. B. Pollack. Automatic design and manufacture of robotic lifeforms. *Nature* **406**: 974-978, 2000.
21. H. Meinhardt. *The Algorithmic Beauty of Sea Shells*. Springer-Verlag, 1998.
22. J. F. Miller, and W. Banzhaf. Evolving the program for a cell: from French flags to Boolean circuits. In S. Kumar and P. J. Bentley, eds., *On Growth, Form and Computers*. Academic Press, 2003.
23. A. A. Minai, D. Braha, and Y. Bar-Yam. Complex engineered systems. In D. Braha, Y. Bar-Yam and A. A. Minai, eds., *Complex Engineered Systems: Science Meets Technology*. Springer Verlag, 2006.
24. E. Mjolsness, D. H. Sharp, and J. Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152: 429–453, 1991.
25. R. Nagpal. Programmable self-assembly using biologically-inspired multi-agent control. *1st Int Conf on Autonomous Agents*, Bologna, Italy, July 15-19, 2002.
26. H. F. Nijhout. A comprehensive model for colour pattern formation in butterflies. *Proc. R. Soc. Lond. B* **239**: 81-113, 1990.
27. D. E. Nilsson, and S. Pelger. A pessimistic estimate of the time required for an eye to evolve. *Proceedings of the Royal Society of London B* **256**: 53-58, 1994.

28. L. Nunes de Castro. *Fundamentals of Natural Computing: Basic Concepts, Algorithms, and Applications*. Chapman & Hall/Crc Computer and Information Sciences, 2006.
29. J. E. Pearson. Complex patterns in a simple system. *Science* **261**: 189-192, 1993.
30. I. Salazar-Ciudad, J. Garcia-Fernández, and R. Solé. Gene networks capable of pattern formation. *Journal of Theoretical Biology* **205**: 587–603, 2000.
31. G. Schlosser, and G. P. Wagner, eds. *Modularity in Development and Evolution*. The University of Chicago Press, 2004.
32. P. Siero, G. Rozenberg, and A. Lindenmayer. Cell division patterns: syntactical description and implementation. *Computer Graphics and Image Processing* **18**: 329-346, 1982.
33. K. Stanley, B. Bryant, and R. Miikkulainen. Real-time evolution in the NERO video game. *IEEE Symposium on Computational Intelligence and Games*, pp182-189, Essex University, Colchester, UK, April 4-6, 2005.
34. H. L. Swinney, and V. I. Krinsky, eds. *Waves and patterns in chemical and biological media*. The MIT Press, 1991.
35. A. M. Turing. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. London B* **237**: 37-72, 1952.
36. G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature* **406**: 188–192, 2000.
37. C. von der Malsburg, R. P. Würtz, and A. Schäfer. *The Organic Computing Group*, http://www.organic-computing.org, 2006.
38. D. J. Watts, and S. H. Strogatz. Collective dynamics of "small-world" networks. *Nature* **393**: 440-442, 1998.
39. G. Webster, and B. Goodwin. *Form and Transformation: Generative and Relational Principles in Biology*. Cambridge University Press, 1996.
40. M. Weiser. Some computer science issues in ubiquitous computing. *Communications of the ACM* **36**: 75-84, 1993.
41. J. Werfel, and R. Nagpal. Extended stigmergy in collective construction. *IEEE Intelligent Systems* **21**(2): 20-28, 2006.
42. A. Winfree. *The geometry of biological time*. Springer-Verlag, 1980, 2001.
43. L. Wolpert. Positional information and the spatial pattern of cellular differentiation development. *Journal of Theoretical Biology* **25**: 1–47, 1969.
44. D. Young. A local activator-inhibitor model of vertebrate skin patterns. *Mathematical Biosciences* **72**: 51-58, 1984.