Chapter 17

ART Networks

§ 1. Basic Anatomy of an ART Network

With this chapter we arrive at what is in many ways the pinnacle of theoretical neuroscience in regard to large scale neural network systems as it stands today: the ART network. Figure 17.1 is the block diagram representation of the general network anatomy we consider in this chapter. It consists of an afferent convergence port, $GN^{(2)}$, two ART resonator fields, F_1 and F_2, and a control network system, which we will call the ***attentional/orienting*** sub-system.

It is appropriate to mention here at the outset that the terminology we will be using here differs slightly from the standard terminology found in the archival literature on ART. The standard terminology is the product of Grossberg's theory of ***embedding fields*** [GROS2-3,16,18], and the main difference is that the resonator subnetwork is there said to be *part of* an attentional sub-system. Insofar as one is speaking of *psychological function* the standard terminology is wholly proper. But in this text it is pedagogically desirable to make a distinction between data path signal

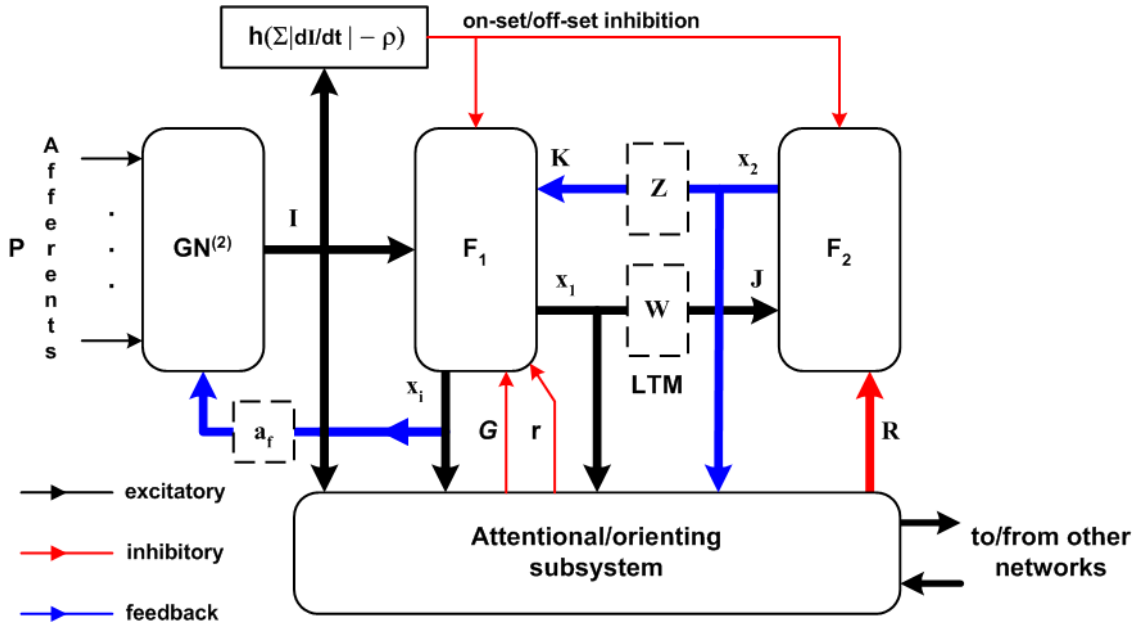


Figure 17.1: Basic anatomy of an ART network. In this textbook a terminology distinction is made between the resonator $F_1 \leftrightarrow F_2$ and the attentional/orienting subsystem. In the standard ART literature the resonator is regarded as belonging to an attentional subsystem. The distinction drawn here is made in order to cleanly separate data path function (resonator) from adaptation control function (attentional/orienting subsystem). Thick lines in the figure denote vector signals. Thin lines denote scalar signals. Signal processing efferents (outputs) of the system would typically be taken from F_2. Control pathway inputs and outputs are made via the attentional/orienting subsystem. Normalizer $GN^{(2)}$ operates as a convergence network for afferent inputs.

and control path signal functions, and this is why the non-standard terminology is adopted here.

The data path signal processing carried out in the F_1-F_2 resonator does serve to support what one may call an attention function from the perspective of psychology inasmuch as it is in the resonator where some signals are "ignored" (quenched) and others are "attended to" (enhanced), but this goes merely to the *matter* of *what* is being "attended to" and *what* is being "ignored." The control function, on the other hand, is the mathematical mechanism determining the *form* of what one can call "the attention function" at the local scale of the network. Aside from this mild difference in points of view, everything we will discuss in this chapter is isomorphic to the standard treatments of ART networks and remains faithful to Grossberg's theory.

The previous two chapters have presented the details of all but the attentional/orienting subsystem and the mechanics of adaptation. Our primary task in this chapter is to supply those details yet to be explained. After this task has been completed the reader will find he can draw new box boundaries in a more detailed block diagram and recover the "boxes" as they are usually presented in the archival literature.

§1.1 Concepts of Embedding Field Theory

It is a common if lamentable habit for papers on neural network theory to present the form of the neural network system being discussed with little more than a vague statement to the effect that "such and such corresponds to this or that anatomical brain structure." Little or, more often, no real justification is given for why this alleged correspondence should be regarded as actual. Often the "justification" seems little more than "in the brain the forward data path goes through *this* structure first and *that* structure next." Why the particular mathematical forms of the layers in the network system were put together in the specific ways they are often seems more a matter of mathematical or computational convenience than of biological or psychological fidelity. Those in neuroscience who are skeptical of the usefulness of neural network theory – and there are many – sometimes level the criticism that mathematical neural networks are little more than fancy curve-fitting exercises and/or that they lack the depth of cumulative research. The neural networks presented in the theoretical literature often have more of the flavor of being inventions than discoveries. Lamentably, this criticism is true more often than it is not. But *informed* criticism of this sort cannot in fairness be leveled at ART models and networks. In this section we discuss why this is so. The reader will find that one important fruit of this discussion is a deeper understanding of and appreciation for the *meaning* Grossberg's terminology. It is through the window of this meaning that one can see the connection between ART constructs and neuroscience.

The foundations of adaptive resonance theory were laid down in the years from 1969 to 1976. Arguably, the most important chapters in this development are found in [GROS2-3,5-7,11-14,16,18-20]. We will not review the entire theory of embedding fields in detail here. That would be quite a lengthy undertaking. Nonetheless, it is important to know something of the flavor of this theory and the long road Grossberg took in developing it.

When one remembers ART networks describe very large scale neural systems, it will come as no surprise that ART's foundational starting point is psychology and psychological phenomena. One of the key stepping-stone-to-ART papers begins with the words,

> This paper describes a psychophysiological model aimed at discussing how animals pay attention to and discriminate among certain cues while ignoring others, based on criteria of relevance derived from past experience or innately preprogrammed in their neural apparatus. The model builds on previous results . . . that introduce some psychophysiological mechanisms of classical and instrumental learning, and of pattern discrimination. These results include network mechanisms of drive, reward, punishment, serial learning, arousal, expectation, and various perceptual constancies . . . This collection of mechanisms comprises the theory of Embedding Fields.
>
> A central theme in the present model will be that two systems are continually readjusting each other. One system (an attentional system) strives toward an ever more stable response to patterns of fluctuating cues by focusing attention on important subclasses of cues. This system is incapable of adapting to unexpected environmental changes. The second system (an arousal system) overcomes the rigidity of the attentional system when unexpected events occur, and allows the network to adapt to new reinforcement contingencies [GROS16].

The last paragraph quoted above gives us the reason why a psychological perspective regards the F_1-F_2 resonator as part of the attentional subsystem. The resonator is "rigid" insofar as its basic dynamics are concerned and in the sense that, left entirely on its own, it succumbs to the problem of the stability-plasticity dilemma. The arousal system – which is commonly called the orienting system in later ART network papers – is the mathematical mechanism by which the stability-plasticity dilemma is addressed by ART. Embedding field theory recognizes two general classes of signal activities. ***Specific activities*** concern the details of specific signal patterns (vectors). The individual components of $\mathbf{x}_1$ and $\mathbf{x}_2$ are examples of specific activities when they specifically target distinct inputs to SNI maps. ***Nonspecific activities*** are those which involve signals that target multiple SNI maps without discrimination. The nonspecific inhibition (15.7) and attentional gain control pathway of figure 15.23 are examples of activities of this class.

> [We] will suggest that the nonspecific neural activity generated by a novel event filters through all internal drive representations. The effect of this activity on behavior will depend on the pattern, or context, of activity in all these representations when a novel event occurs. Sometimes the novel event can enhance the effect of an ongoing drive, sometimes it can cause a reversal in sign (as in the frustration reaction), and sometimes it can introduce and enhance the effect of a different drive. We will be led to assume that every novel event has the capacity to activate orienting reactions, but whether or not it does depends on competition from the drive loci which the event also activates. The nonspecific activity generated by the novel event will also be

> assumed to reach internal sensory representations, where it helps determine which cues will enter short-term memory to influence the pattern of internal discriminatory and learning processes [GROS16].

We have seen Grossberg describe embedding field theory as a psychophysiological theory. This does not mean he dogmatically claims specific ART network structures *are* models of specific anatomical regions of the brain. (To do so would indeed have been recklessly brash in 1975 given the state of knowledge available at the time). Nonetheless, the *psychological* functions served by different parts of a network do suggest in a general if equivocal way certain correspondences with known facets of the roles of general anatomical regions of the brain. He refers to these correspondences as ***mock structures***.

> The networks will contain several functionally distinct regions. The interactions between these regions call to mind familiar anatomical facts. It will be apparent that the network regions are not presumed to be exact replicas of real anatomical fragments. Nonetheless, the anatomical relationships between the network regions, as well as their functional roles in total network processing, suggest natural analogs with real anatomies. These analogs will be pointed out both to suggest possible new insights about the functioning of real anatomies, and to serve as an interpretive marker for the networks that will arise in the future from additional postulates. The psychological validity of formal network interactions is, however, independent of how well we guess neuroanatomical labels for network components at this stage of theorizing, since the formal anatomy is still, at best, a lumped version of a real anatomy.
>
> A network region of particular interest is reminiscent of the hippocampus. This region supplies motivational feedback to several other network areas . . . This feedback is determined by a competition between channels responding to different drives. Each channel is influenced by sensory and drive inputs. The sensory pathways can be strengthened or weakened by reinforcing events ("conditioned reinforcers"). If a given channel has a prepotent combination of input from conditioned reinforcers and drive, it will suppress other channels using its on-center off-surround anatomy . . . Thus the mock-hippocampus receives input from a region that is implicated in reinforcement, and delivers feedback to this region. We therefore (undogmatically) interpret this second region as a mock-septum . . . The mock-hippocampus also supplies conditionable non-specific feedback, in the form of late, slow potential shift, to sensory processing areas (e.g., mock-neocortex) of the network. This feedback, which is related to the network's arousal, drive, reinforcement, and motivational mechanisms, helps to determine which cues will be attended to by the network [GROS16].

In figure 17.1, the $\mathbf{x}_1$ pathway to F_2 is a correspondent to what Grossberg above calls a conditioned reinforcer. The $\mathbf{x}_2$ pathway back to F_1 corresponds to what he calls a "contingent negative variation" [GROS6, 16]. "Cues" correspond to the afferent pattern inputs **P** shown in the block diagram. "Drive" inputs, which are not yet depicted by figure 17.1, are inputs coming into F_1-F_2 from elsewhere in the overall system. One thing such inputs do, by stimulating $\mathbf{x}_2$ in the F_2 field, is provide an anticipatory biasing to F_1 *without also* stimulating $\mathbf{x}_1$ into high levels of activity. The ability to do the one without also doing the other is a capability provided by the attentional/orienting subsystem of figure 17.1 [GROS6].

> The mock-septum is influenced by a source of drive input (mock-hypothalamus) and of non-specific arousal (mock-reticular formation). The level of nonspecific arousal is modulated by the

> degree of unexpectedness of external events. . . Thus, although the arousal itself is nonspecific, its regulation can be dependent upon specific sensory cues. The nonspecific arousal filters through the drive-representing channels, and can either contrast enhance their activity, or cause a positive (negative) motivational bias to flip into a negative (positive) motivational bias. Thus nonspecific arousal can have specific effects on the pattern of motivational feedback. The non-specific arousal also feeds into sensory processing areas (e.g., mock-neocortex), where it influences which cues will generate enough neural activity to reverberate in short-term memory, and thereupon be able to influence processes of learning and discrimination. The nonspecific arousal that is triggered by unexpected events differs from the nonspecific conditionable feed-back that is related to network drive, reinforcement, and motivational levels. Indeed, these two sources can compete with each other [GROS16].

The hypothalamus is a deep-lying subcortical structure in the cerebrum notable for its use of blood-born signaling chemicals (hormones), its central role in integrating autonomic and endocrine functions with behavior, and its control of homeostasis (body temperature, metabolism, blood pressure, stress responses, etc.). By likening the source of drive input in his model to the hypothalamus ("mock-hypothalamus"), Grossberg is indirectly telling us that the signals for drive representation are slow compared to timeframe of $\mathbf{I}$, $\mathbf{x}_1$, and $\mathbf{x}_2$. (In [GROS20] he says this explicitly).

The reticular formation is part of the brain stem and is involved with the coordination of reflexes and simple, stereotyped behaviors. It also contains neurons that project to almost every part of the cerebrum with metabotropic neurotransmitters. These projections modulate arousal, wakefulness, and vigilance. By likening the nonspecific signals to the "mock-reticulum" we see that these signals are, likewise, relatively slow signaling processes because of the slow-acting nature of metabotropic signaling.

The septum is part of the limbic system, and it is thought the septum and hippocampus, acting together, performance a comparison function between expected and actual stimuli. Their joint actions are sometimes called a "stop system" and sometimes called a "needs checker" by motivational psychologists. The theory holds that if an actual stimulus does not match what is expected, or if the expected stimulus is aversive, the septal-hippocampal system immediately inhibits any motor behavior in the process of execution and identifies this behavior as "faulty." In addition, it initiates exploratory behaviors for identifying environmental stimuli associated with punishment, non-reward, or failure. A large fraction of cells in the septum that project to the hippocampus use acetylcholine as their neurotransmitters, which is another relatively slow-acting modulator neurotransmitter. Thus, the mock-septum, mock-hypothalamus, and mock-reticulum combine to implicate a relatively slow (long duration) control mechanism modulating the actions of the main signal processing pathways.

> In summary, at least two major feedback loops exist in the network. One feeds between external sensory and internal sensory (e.g., drive) processing areas (cortex $\rightarrow$ hippocampus $\rightarrow$

> cortex). The other feeds within the internal sensory processing areas (septum → hippocampus → septum).
>
> The drive representations are organized into dipoles, such that each dipole controls a positive and a negative incentive motivational channel; e.g., relief and fear, hunger and frustration. The regulation of motivational output from the dipoles, and of learning based on this output, has been interpreted as using two distinct transmitter systems, which are presumed to be analogous to adrenergic and cholinergic transmitters[1] . . . The need to synchronize the activity of the two parallel channels in a given dipole, and to sample the resultant activity in both dipole channels, suggests that the two transmitter systems are also organized in parallel across the two channels [GROS16].

Grossberg's ***dipole*** terminology refers to affective opposites such as "fear and relief." It is widely – but not universally – accepted in psychology today that "emotions" come in pairs of opposites, although there is widespread disagreement over precisely which emotions pair up to form opposites.[2] In Grossberg's theory, he refers to such things as "fear" as a "drive," which is one view of "what an emotion is/does" but is not a universally-accepted view. (When it comes to the current state of emotion-motivation theory, there is *no* universally-accepted model in psychology). In this book we will not embroil ourselves in the present-day controversies raging in emotion- and motivational- psychology. For our purposes, we will regard a Grossberg "drive" as a signal that tends to stimulate some particular behavior or action. Drive signals project to maps (cell populations) he calls ***arousal populations***. In the embedding field theory, the network subsystems that contain arousal maps are organized into behaviorally-opposed pairs called dipoles. Projections from the arousal subsystem to the sensory system are called ***incentive/motivation*** signals.

§1.2 Dipole Networks and Orientations

We can see that Grossberg's model of attention, reinforcement and discrimination learning is a model at a high level of abstraction inasmuch as its foundation started with psychological phenomena and their associated behavioral consequences. His research was not aimed at settling controversies in psychology that embroil the theories of these phenomena, but rather was aimed at clarifying what the mathematical implications are for neural network systems that the empirical findings of psychology were reporting. Nonetheless, what emerged was a general principle for neurological organization *in vivo*. Almost everything that goes into making the transition from the basic resonator to a full-blown ART network can be found in [GROS16].

The principle of the dipole network is one of the key ideas involved with control of adaptation

[1] Adrenergic implies a neurotransmitter such as noradrenaline, which plays a metabotropic modulatory role in central systems. Cholinergic, of course, refers to acetylcholine as the neurotransmitter.

[2] Indeed, there is currently no agreement as to what "an emotion" even is. Reber's *Dictionary of Psychology* tells us that the different "definitions" of "emotion" are really so many different mini-theories.

in ART networks and holds a prominent place within what here we are calling the attentional/orienting subsystem of figure 17.1. As we have just seen, key parts of the attentional/orienting functions play out over a very different time scale than that of the data path signal processing in the basic resonator. This presents certain computational challenges in putting together network simulators, and one consequence of these challenges is that often the simulation of these processes is greatly simplified in ART network computer models, e.g. [CARP2]. In this chapter, we will likewise make such a simplification. But before we do, it is a good idea to understand what it is that is being represented in abstract form, and that is the goal of this section.

The basic idea of the dipole network was developed in [GROS20] in 1972 as part of research into neural network mechanisms capable of reproducing observed psychological effects in classical conditioning experiments. The basic Grossberg conditioning network is shown in figure 17.2 below. Afferents C stimulating a "sensory" layer S_1 are called the "conditioned stimuli" of a

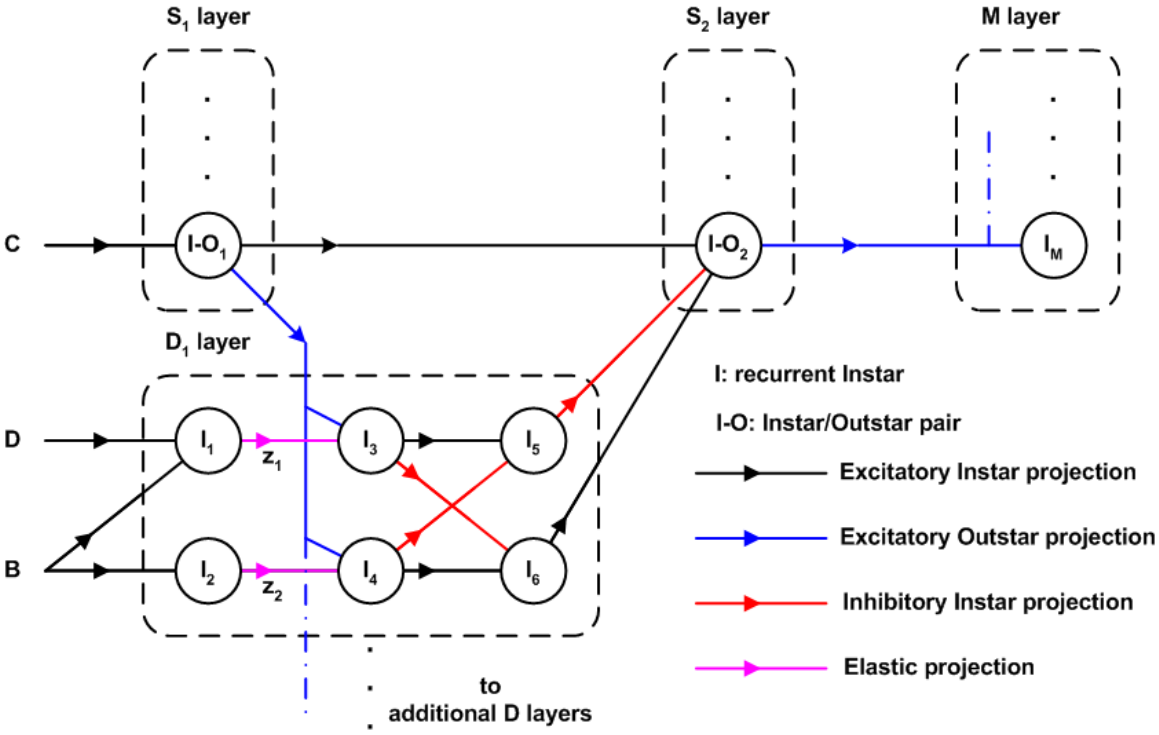


Figure 17.2: Grossberg conditioning network. Layers S_1 and S_2 are "sensory" layers of Instar/Outstar nodes. S_2 makes Outstar projections to all the nodes in a destination layer, the M layer, originally intended to depict conditionable "motor" responses. S_1 receives conditional stimulus inputs, C, and makes Instar projections to S_2 and Outstar projections into one or more dipole layers. D_1 is one such nonrecurrent dipole layer. A dipole layer receives "drive" inputs, D, and tonic biasing-drive inputs B. In [GROS20] the Outstar projections undergo adaptation according to an Outstar adaptation rule (OAR). Within the dipole layer, the projections from I_1 and I_2 to the "arousal nodes" I_3 and I_4 have elastic weights z_1 and z_2 which undergo short-term depression when the activity levels of I_1 and I_2 are high, and recovery from depression when I_1 and I_2 activity levels are low or zero. The I_3-I_4 projections to I_5 and I_6 create a "rebound effect" in which removal of drive stimulus D causes I_5 to become inactive and I_6 to become active, whereas application of drive D causes I_5 to become active and I_6 to become inactive.

conditioning experiment. Afferents D and B represent various "drive" signals, such as hunger or fear, that serve as the unconditioned stimuli in a classical conditioning experiment. "Drive" signals activate "arousal" nodes, I_5 and I_6, in one or more dipole layers. These project to one or more downstream sensory layers S_2. The dipole layer is able to produce what is called a ***rebound effect*** in which the removal of a drive D results in a short-term enhancement of the opposite drive response. For example, removal of a "fear" drive signal causes a short-term "relief" rebound. Outstar projections from S_1 into the dipole layer "learn" to stimulate the same drive response behavior when the conditioned stimulus C is applied thereafter. In this way, responses of layer S_2 to stimulus C become "conditioned by experience" so that afterward C becomes "associated" with the unconditioned drive D. Signal D is called a "phasic" drive signal. Drive signal B is called a "tonic" signal. "Behaviors" are represented by the projections from S_2 to a "motor" layer M that represents "habit" responses to sensory stimuli. The work in [GROS20] was aimed particularly at aversive types of arousals, such as the fear-relief dipole, and so the primary response I_5 projected to S_2 with an inhibitory signal and the rebound response I_6 projected with an excitatory signal. A drive-rebound mechanism having the nature of a positive incentive rather than an aversive disincentive would reverse these projections, making I_5 excitatory and I_6 inhibitory.

The psychological hypotheses upon which this network is based were presented and justified in [GROS19]. The work in [GROS20] was aimed primarily at the rebound mechanism and the development of the dipole layer network. Grossberg found that in order to produce a rebound effect it was necessary to introduce ***elastic*** projection mechanisms in which connection weights z_1 and z_2 undergo a short-term depression modulation. In the presence of zero or low activity in I_1 and I_2, elastic weights z_1 and z_2 slowly build up to equal maximum strengths. At high levels of activity in I_1 and I_2, z_1 and z_2 undergo a rapid decay in synaptic strength. Grossberg draws an analogy between this mechanism and neurotransmitter depletion (this is what he is referring to in the previous quote when he speaks of "two distinct transmitter systems"). In the presence of a long-lasting tonic input B, z_1 and z_2 decay at equal rates and remain at more or less the same non-zero numerical values. Then when "drive" D is applied to I_1, the decay of z_1 becomes faster, with the result that z_1 becomes less than z_2. If drive stimulus D is sufficiently large, the activity of I_3 remains higher than that of I_4, with the result that I_5 remains active and I_6 is inhibited. But when drive D is small or is removed, I_4 receives a higher level input than I_3, causing I_5 to be inhibited and I_6 to become active until z_1 builds back up to equal z_2 in magnitude. This is what produces the transient rebound effect Grossberg was after. When $z_1 = z_2$ and activities $x_1 = x_2$ in I_1 and I_2, then in the absence of other stimuli I_3 and I_4 have equal activities $x_3 = x_4$ and the cross-inhibitions to I_5 and I_6 keep both dipole outputs at zero.

The operation of this dipole layer is described by a set of eight coupled differential equations for the state of the dipole layer plus two output equations describing the projections from I_5 and I_6. Using lower case Greek letters to denote parametric constants, the dipole layer is described by the nonlinear state equation [GROS20]

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \\ \dot{x}_6 \\ \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} -\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\alpha & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\varepsilon & 0 & 0 & 0 & \zeta \cdot f(x_1(t-\tau)) & 0 \\ 0 & 0 & 0 & -\varepsilon & 0 & 0 & 0 & \zeta \cdot f(x_2(t-\tau)) \\ 0 & 0 & \kappa & -\kappa-\eta & 0 & 0 & 0 & 0 \\ 0 & 0 & -\kappa & \kappa & 0 & -\eta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\beta-\delta \cdot f(x_1(t-\tau)) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\beta-\delta \cdot f(x_2(t-\tau)) \end{bmatrix} \cdot \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \\ x_6(t) \\ z_1(t) \\ z_2(t) \end{bmatrix} + \begin{bmatrix} \mathrm{D}+\mathrm{B} \\ \mathrm{B} \\ Z_{1d1}S_1 \\ Z_{1d2}S_1 \\ 0 \\ 0 \\ \beta \cdot \gamma \\ \beta \cdot \gamma \end{bmatrix}.$$

Here $Z_{1dj}S_1$ denotes the total Outstar excitation into I_3 or I_4 from the S_1 layer. The activation function $f(x)$ is the Heaviside extractor $h(x - \Gamma)$ where Γ is a threshold. The various system parameters are non-negative.

Grossberg gave some steady-state parametric bounds for the parameters of this system in [GROS20] but no specific numerical values. This is understandable given that the state equation is nonlinear and the availability of computing resources for simulation studies of such a system was limited in 1972. What is important for our purposes here are the following observations on this system.

The first observation is that the Instars in this system are simple Instars with self-recurrent feedback rather than shunting-node Instars. The same is true of the Instars in S_1, S_2, and M. The inhibitory feedback an Instar supplies to itself makes these nodes a class of "leaky integrator" type Instar. Output projections from the Instars are made via a thresholding Heaviside extractor activation function.

The second observation is that a principal nonlinearity in the state equation is produced by the presence of *delayed* $x_1(t - \tau)$ and $x_2(t - \tau)$ terms in the nonlinear state matrix. Recall from the previous section that delays of relatively long duration are implicated in the mock anatomies that are associated with the attentional and orienting functions of figure 17.1. The same is true for the conditioning network of figure 17.2 and, indeed, a more general version of this system introduces additional delay lags in the transmission of signals from I_3 and I_4 to I_5 and I_6 such that the stimulation of the dipole layer via the Outstar connections from S_1 induces a delayed response from the dipole layer to S_2. The putative physiology represented by the dipole layer is modulatory and metabotropic in its nature, and this implicates the introduction of signal processing delays of

long duration relative to the time scale of the direct data processing pathway S_1-S_2-M. This is important in preventing spurious dipole action from interfering with the normal signal processing carried out in the data pathway in the absence of "drive" and "conditioned response" factors.

The third observation concerns the overall general form of the dynamics of the elastic weights z_1 and z_2. Let $y_i = h(x_i(t - \tau) - \Gamma)$ with $i = 1$ or 2. If we assume x_i is approximately constant except at its onset or offset, the differential equation describing z_i reduces to a linear differential equation with solution

$$z_i(t) = z_i(0) \cdot \exp\left[-\left(\beta + \delta \cdot y_i\right) \cdot t\right] + \frac{\beta \cdot \gamma}{\beta + \delta \cdot y_i} \cdot \left(1 - \exp\left[-\left(\beta + \delta \cdot y_i\right) \cdot t\right]\right).$$

When $y_i = 0$, z_i asymptotically approaches the steady-state value γ at rate β. (β has units of inverse time constant; it is a neper frequency). When $y_i > 0$, the steady-state value of z_i is less than γ, and if $\beta \ll \delta \cdot y_i$ then the steady-state weight value is much less than γ. Thus the amount of elastic short-term depression is controlled by the δ parameter and the level of activation y_i. The neper frequency for non-zero y_i also increases, denoting that the build-up to maximum z_i occurs slowly compared to the run-down to a depressed level of z_i. Thus, a "depressed" elastic pathway remains depressed over a length of time determined by β following the offset of y_i. Examining the dipole layer of figure 17.2, this tells us that the duration of the rebound signal from I_6 to S_2 is determined by the recovery time of z_1 following the offset of drive D.

How the system of figure 17.2 will react to stimulus C after conditioning depends on the state of the system at the time the Outstar weights from S_1 to D_1 undergo adaptation. If this learning takes place when stimuli C and D are active, such that $x_3 > x_4$, adaptation under the OAR will lead to an Outstar weight pattern such that subsequent activations of C will stimulate I_5 and inhibit I_6 even in the absence of the drive stimulus D. However, "rebound" does not then occur upon cessation of C because the Outstar pathways to D_1 are not elastic. Grossberg notes that to produce a "higher-order" rebound effect from C requires modifications to the dipole layers which introduce feedback from I_3 to I_1 and I_4 to I_2 in order to bring the elastic pathways into operation. However, he discusses this only qualitatively in [GROS20].

§ 2. Matching and Reset in the Orienting Subsystem

Although [GROS19-20] predated Grossberg's discovery of ART, the ideas contained in them turned out to be important contributing factors for solving the stability-plasticity dilemma in ART networks. Indeed, in "the ART paper" [GROS6], Grossberg introduced another recurrent form of dipole network (figure 9 of that paper). He proposed that something like this network would be

needed to achieve stable learning behavior in an ART resonator system while maintaining network plasticity sufficient to "recognize and learn" novel events in the data pathway. The discussion of this new form of dipole in [GROS6] was qualitative rather than quantitative, and it is fair to say that Grossberg's idea serves to this day in the role of an inspiration for developing efficient algorithms for ART networks rather than an active model. For example, the ART 1 network of [CARP2] alludes to the existence of "dipoles" in the F_2 layer, but merely models the *end effect* such a properly designed network should have on the resonator function. ART 1 then employs a merely mathematical function in its algorithm to mechanize this end effect, skipping the biological details of how this might happen in a biological ART model.

This is a perfectly legitimate tactic for an artificial ART network to employ, and it is likewise a tactic that has computational advantages for neuroscience theory in modeling biological systems using adaptive resonance theory. However, unless one is well acquainted with ART such a tactic can often seem ad hoc and open to charges that the model is merely a curve fit or that it employs non-biological mathematical chicanery. This tends to reduce its credibility in the eyes of the wider neuroscience community. Our task in this section is to examine step-by-step the model evolution from the ideas latent in figure 17.2 to the incorporation of these ideas in the attentional/orienting subsystem of figure 17.1.

The first step is very easy. The S_2 layer makes Outstar projections to the M layer; the F_2 layer of a resonator makes Outstar projections back to the F_1 layer. Suppose in figure 17.2 that the M and S_1 layers were actually one and the same. This amounts to "folding M back into S_1" [CARP6]. If we do this and, in addition, make the straight-forward changes of making S_2 a competitive network of shunting-node Instars, replace the simple Instars of S_1 with shunting-node Instars, and introduce the Instar fan-in matrix **W** between S_1 and S_2, then we convert the data pathway of figure 17.2 into a basic ART resonator.

The next step is not so trivially obvious. The rebounding dipole layer of figure 17.2 does not play precisely the same role as the orienting function in figure 17.1, but the orienting function in its basic operation is cast in a role *analogous* to that played by D_1 in Grossberg's psychology model. It is this analogy we must examine.

In figure 17.2 the projections from the dipole layers to S_2 bear such names as "incentives" and "motivations" serving to arouse or inhibit the excitability of S_2 in response to upstream signals from sensory layer S_1. More specifically, the projections from the dipole field are intended to alter the behavior of S_2 on the basis of "past experiences." The projections **R** in figure 17.1 have a similar function except, in their case, what is to be modified is the F_2 response to F_1 in so far as this modulation of F_2 response serves to stabilize the learning of classification codes by F_2. Let us

suppose F_2 has (by means we have not yet discussed) "learned" to classify certain input vectors **I** such that some degree of partitioning of the input space of **I** has been accomplished. Let us further suppose that now a novel input pattern $\mathbf{I}_{new}$ is presented for which F_2 responds with a 0-1 distribution, classifying that input with, say, node v_{21}.

The key question so far as the stability-plasticity dilemma (SPD) is concerned is this: Does $\mathbf{I}_{new}$ actually present the same ***features*** represented in the prototype exemplar v_{21}, so that differences between $\mathbf{I}_{new}$ and this exemplar are merely irrelevant "noise"? Or does $\mathbf{I}_{new}$ actually present important *different* features such that it should properly be classified as something else? Or, at the extreme, are the features of $\mathbf{I}_{new}$ so different that the resonator *should not classify it at all?* These questions are particularly critical during the early development of feature codes by the resonator, when not all the rows of **W** and columns of **Z** have achieved stable LTM patterns.

Consider what ensues if no orienting function exists to modulate the behavior of F_2. In this case, F_2 will attempt to classify $\mathbf{I}_{new}$ and it will adapt **W** and **Z** according to the strength of the responses of the F_2 nodes, whatever these may be. Put another way, the ***default mode of operation*** for F_2 favors ***plasticity***. Left to its own devices, the resonator falls under the conditions of the Grossberg sparse pattern theorem [GROS5] and *stable* coding will not be achievable. Grossberg tells us,[3]

> The following difficulty must be overcome. Suppose that two patterns $\mathbf{I}_1$ and $\mathbf{I}_2$ would ordinarily be encoded by the same population v_{21} in F_2 . . . If $\mathbf{I}_1$ is presented sufficiently often before $\mathbf{I}_2$ is presented, how can $\mathbf{I}_2$ be prevented from being encoded by v_{21} and yet be allowed to search for and find an as yet unpracticed population in F_2? An adaptive resonance between F_1 and F_2 does not suffice. . . Somehow presentation of $\mathbf{I}_2$ must inhibit v_{21} – including the large excitatory F_1-to-F_2 signal generated by $\mathbf{I}_2$ – until $\mathbf{I}_2$ can find an uncommitted population among the uninhibited, or *renormalized* populations of F_2. In particular, there must be at least two sources of input to F_2: the excitatory signals that code the patterns at F_1, and the signals that are elicited by a mismatch of patterns. The latter signals differentially inhibit populations which are currently active in STM. These input are nonspecific because the STM code is opaque[4]. How does nonspecific arousal interact with current STM activity to differentially inhibit active populations? [GROS6].

Within the block diagram divisions Grossberg made in the 1976 paper, and the later ART network models developed from it, the signal processing responsible for "differentiating" a non-specific input to selectively disable v_{2j} nodes places a recurrent dipole field within the same "box" of the block diagram as the v_{2j} nodes are located. (The overall "box" is called the F_2 layer in

[3] The mathematical symbols in the quotation appearing here have been altered to fit the notation used in this textbook.

[4] By "opaque" Grossberg means STM_2 does not contain sufficient information to make an evaluation of whether or not a specific activity pattern in v_2 is a proper association with the input pattern. Simply put, if the resonator is learning to encode inputs, it cannot know in advance what the right answers are. Therefore the second (dipole) input source cannot be specific in targeting populations in v_2. Therefore it is non-specific.

standard ART network terminology). For the block diagram of figure 17.1, we are going to place this function inside the attentional/orienting "box" and this is why **R** is a vector signal in that diagram. The differentiation of a nonspecific arousal signal to produce specific inhibitions will all take place within that "box," which is merely a different way of describing the same function.

Let us consider the functional requirements of the orienting function Grossberg has just described. ***First***, we have need of some means to measure and evaluate a "mismatch." The first question to ask here is: Mismatch of *what*? In ART, "mismatch" refers to a mismatch between the bottom-up input pattern **I** and the top-down "expectation" $\mathbf{Z} \cdot \mathbf{x_2}$. The only information available to measure and evaluate "mismatch" is the information contained in **I** and STM. We will need a ***mismatch function*** and a threshold, called the ***vigilance parameter***, for deciding when the measured mismatch is too much and for generating a ***nonspecific inhibition***, *A*. ***Next***, the nonspecific inhibition *A* must lead to discriminating inhibitions (**R**) that ***reset*** nodes in F_2 to permit a ***search*** for a different node to use in classifying **P**. We can see from this that we will have to come up with new definitions for the signals to/from the dipole field of figure 17.2.

Third, the arousal signal must not happen too quickly – or else it would activate before the establishment of the initial resonance and before a mismatch can properly be evaluated. Once aroused, inhibition must be of sufficiently long-lasting duration to permit the search for a new v_{2j} to complete and to permit adaptation to take place in **W** and **Z** as the new node "learns" **P**. This requirement tells us that, whatever the form our dipole layer eventually takes on, the *function* served by the elastic modulation mechanism of the D_1 layer in figure 17.2 must be retained. Furthermore, the signal produced by the mismatch function plays a role analogous to that played by the nonspecific "drive" inputs going into the dipole layer in figure 17.2.

Fourth, we must allow for the possibility that the second v_{2j} node – let us call it v_{22} – might also not make a proper match for **P** and a third, fourth, fifth, etc. v_2 search might be required. Thus, the durations of the specific inhibitions that ***reset*** specific nodes in v_2 must be sufficiently long that, if necessary, the search process can try *every node in the F_2 layer* before the specific reset inhibition of v_{21} can "wear off." ***Finally***, we must allow the possibility that *all* v_2 nodes could have been previously "committed" to specific encodings and that the ***encoding capacity*** of the resonator is used up before the pattern **P** was every applied to the resonator for the first time. In this case, the orienting function must prevent *any* attempt to encode **P** at the expense of "erasing" the previously learned pattern classifications. Thus, the orienting function serves learning ***stability*** at the expense of learning plasticity. Resolution of the SPD depends on making the proper tradeoffs between the "plasticity inclinations" of the resonator and the "stability inclinations" of the orienting subsystem.

We will see later that there are some additional considerations that come into play in solving the stability-plasticity dilemma. We will come to these in due course. For the present, we have the general outline of requirements we need to begin filling in some details regarding the orienting subsystem.

§ 2.1 Mismatch Measurement

When does an active category x_{2j} mismatch the pattern input **P**? Different published ART networks have employed different answers to this question. For example, ART 1 [CARP2], which is designed to operate on binary-valued signals, compares a vector length measure, known as a Hamming weight, of input vector **P** against a similar Hamming weight measure for $\mathbf{x}_1$. ART 2 [CARP3] employs various derived unit vectors to, in effect, compare the vector direction of **P** against the vector direction of **K**. Figure 17.3 illustrate a third method for detecting mismatch, which we will use in the example network presented in this chapter. This method is similar to the mismatch detection function of ART 2.

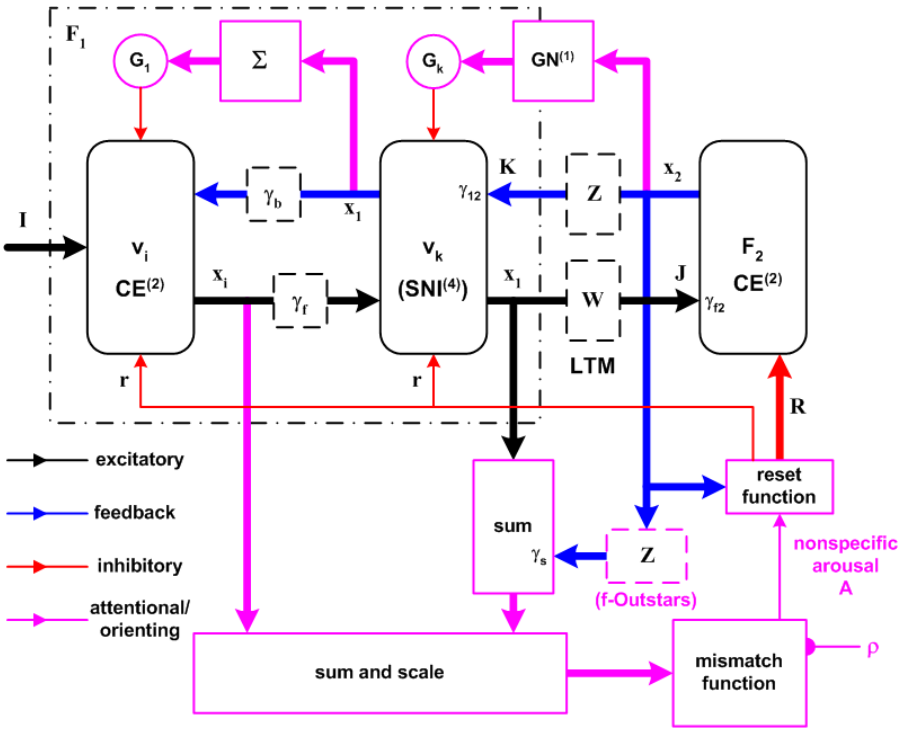


Figure 17.3: Mismatch network used in the example network. The signal processing functions for mismatch detection and attentional gain are identified by the pink boxes and signaling lines. The mismatch function is similar to the function used by ART 2. Vector terms are summed by a simple nonrecurrent Instar and scaled. If the result exceeds the ***vigilance threshold*** ρ, a non-specific arousal signal A is produced and sent to the subnetwork for implementing the reset function. The attentional gain functions G_1 and G_k are similar to the one used in figure 15.23. G_1 is the same as described in chapter 16, G_k differs only in the use of a $GN^{(1)}$ prior to summing. The mismatch arousal function enforces *equal stimulus levels among non-zero stimuli* as a principal part of its feature detection scheme. v_k is a non-interacting layer of $SNI^{(4)}$-type Instars.

Figure 17.3 also supplies more details on the construction of field F_1. F_1 is a resonator of a type we will call $R^{(3)}$. This resonator differs from the $R^{(2)}$ resonator described in chapter 16. In $R^{(2)}$ the structure consisted of the cascade of a non-interacting layer of $SNI^{(4)}$ Instars followed by a $CE^{(2)}$ layer. The F_1 field in figure 17.3 reverses this order. Another difference is that the input to layer v_k is the sum of the forward-pathway signal $\mathbf{x_i}$ (weighted by a scalar path gain γ_f) and the top-down expectation signal vector **K**. Layer v_k has *n* nodes (the same as the number of afferent inputs, and the same as layer v_i) and its nodes feed back to their corresponding nodes in layer v_i via a scalar feedback weight γ_b. The v_i layer is designed to contrast-enhance $\mathbf{x_i}$ if the top-down pattern **K** is non-zero and mismatches the bottom-up pattern $\mathbf{x_i}$. Layers v_i and v_k are also given a nonspecific inhibitory input *r*. Layer v_k and the F_2 field constitute a type-$R^{(2)}$ resonator similar to the resonator introduced in chapter 16.

§ 2.2 ART Network Functional Principles

The term "ART network" describes a general class of network system architectures rather than one particular instantiation of a system. What all ART network systems share in common is a set of functional principles [CARP3], all of which must be met by the organization of the network.

1. **Stability-plasticity tradeoff.** The principal difference between an ART network and the simpler competitive networks of the earlier chapters lies with the ART resolution of the stability-plasticity dilemma. An ART network always maintains the potential for adapting weights **W** and **Z** yet also has a finite capacity for input categorization by the F_2 layer. The task of the network is to learn a stable recognition code in response to arbitrary input pattern sequences (plasticity). At the same time, the network must be resistant to *recoding* of previously learned categorizations in response to subsequent input patterns regardless of how many different patterns are presented to it and regardless of how long the duration of these patterns may be (stability). The principle of active stability-plasticity tradeoff is the fundamental principle of ART network organization, and all other functional principles serve this one.

2. **Search-direct access tradeoff.** At the beginning of its operation, the *N* nodes of the F_2 layer are initially uncommitted, i.e. the network generally has no predisposition to respond to any particular set of input patterns. Rather, it develops differentiated responses to different classes of input patterns through adaptation according to whatever patterns it happens to be presented with during its early operation. Thus, over time the network progresses from a state of having no committed F_2 nodes to a state of having committed all *N* nodes to the categorization of *N* distinct classes of input patterns. When presented with a pattern belonging to an "already learned" classification, the network responds at once with the appropriate output pattern $\mathbf{x}_2$ for that classification. This is called *direct access*. On the other hand, when presented with an input pattern for which it has no already established classification, the network must automatically "commit" another F_2 node to the learning of this new classification. Thus, the network must *search* for an uncommitted F_2 node, and it must automatically disengage this search as an input pattern becomes "familiar" to the network. Furthermore, the network's *N* learned categories must remain distinct, i.e., ***dual-coding*** of one input pattern by two or more F_2 nodes must be

prevented.

3. **Match-reset tradeoff.** The Instar weight vectors W_j and Outstar weight vectors Z_j are exemplars for the N possible classification categories of the network. Whether or not an afferent input pattern "belongs" to a particular category depends on how well the vector $\mathbf{x}_i$ matches the top-down expectation **K** of figure 17.1. If the nonspecific mismatch arousal A is less than the network's **vigilance threshold** ρ (figure 17.3), the input pattern is said to "match" an established category. If A exceeds this threshold then the input pattern is said to mismatch expectation. In this case, the responding $\mathbf{x}_2$ must be squelched (reset) so that the network can either (1) learn a new classification (if uncommitted F_2 nodes are still available) or (2) signal that this input pattern cannot be assimilated by *any* of the network's recognition categories.

4. **Resonant state accommodation**. Adaptation is the equilibrium between assimilation and accommodation. In an ART network equilibrium is defined by the condition of STM resonance. Resonance in an ART network means that both $d\mathbf{x}_i/dt$ and $d\mathbf{x}_2/dt$ are, ideally, equal to zero (or, practically, that both are nearly equal to zero) and will remain so for as long as the input afferent pattern does not change. It follows from this practical definition of adaptation that accommodations (changes) in **W** and **Z** can only take place when the network is in a resonant state.

5. **Mismatch reset hold-off during accommodation.** Accommodations in **W** and **Z** alter the feedback from F_2 to F_1 and this, in turn, alters the feedforward signaling from F_1 to F_2. The "initial learning" of a new category by an F_2 node must be carried out without that node being reset by any spurious mismatches that may occur during the weight adaptation process. Inasmuch as both $\mathbf{x}_1$ and $\mathbf{x}_2$ undergo changes during weight adaptation the practical consequence of this principle is that mismatch resets can be allowed to take place only when the network is in a resonant state. This principle is met in different ways by different ART network architectures. [CARP3] describes one such method; the example network presented in this chapter uses a different method. Both, however, are aimed at the same final outcome, namely stable learning of new categories by the network up to the classification capacity of the network.

6. **No LTM recoding by superset inputs.** SNI competitive layers are feature detectors. In general, two afferent input patterns **P** encoded by the same F_2 node will contain some input signals common to both input patterns (the "features") and others that are significantly different. The latter are treated as "noise" by the network. An input pattern that contains all the "feature signals" encoded for by F_2 plus additional signals in excess of these is said to be a "superset input." Put in other terms, such a pattern contains signals that are not part of the top-down expectation **K**. These "extra" signals cannot be permitted to cause recoding of the LTM patterns that define a category because this leads to loss of stability in the network "learning" process. In ART network terminology, this principle is often referred to as "the 2/3 rule." Satisfaction of this principle in ART 2 and in the example network presented in this chapter is achieved by introducing contrast enhancement capability into the F_1 field.

7. **Stable-choice/re-arousal tradeoff.** The mismatch/search/resonance dynamics that support the principles outlined so far apply when the afferent input pattern is itself stable. Otherwise the network could not be said to be accommodating itself to ("learning") *an* input pattern. However, there is no prior constraint that can be placed on the input afferent patterns, and these can change at any time. The onset of a new pattern (with the

coincident offset of the prior input pattern) must interrupt whatever classification or accommodation activities may be in progress by the network. Put another way, resets in F_2 that may have been caused by mismatch signals from the orienting subsystem *endure* during the duration of a stable input pattern but are *overridden* by changes in the afferent input. This principle is satisfied in the example network presented here by the onset/offset inhibition subsystem in figure 17.1. Similarly, because the dynamics of contrast enhancement and choice in SNI competitive layers is dependent on the total excitation acting on the network – and because there is a limit to the degree to which Grossberg normalizers can compensate for variations in the total stimulation presented by afferent inputs – there must be a minimum afferent excitation threshold mechanism in place in order for network accommodation to occur. (This requirement, too, is part and parcel of "the 2/3 rule" for ART networks).

§ 2.3 Dipole Action in F_2

The central core of an ART network's ability to resolve the stability-plasticity dilemma lies with the mechanisms of mismatch reset by the orienting subsystem. This is accomplished by dipole actions incorporated into the F_2 layer of the network. In this section we will examine the biological signal processing underpinnings of the F_2 dipole structure. In the next section we will see how this action is mimicked in ART network simulators.

The need for and role of the dipole network was first put forth in [GROS6]. Figure 17.4 illustrates the basic concept. In this model the basic $CE^{(2)}$ layer of chapter 15 is replaced by a pair of three-node networks, the first of which is called the on-cell pole, and the other the off-cell pole.

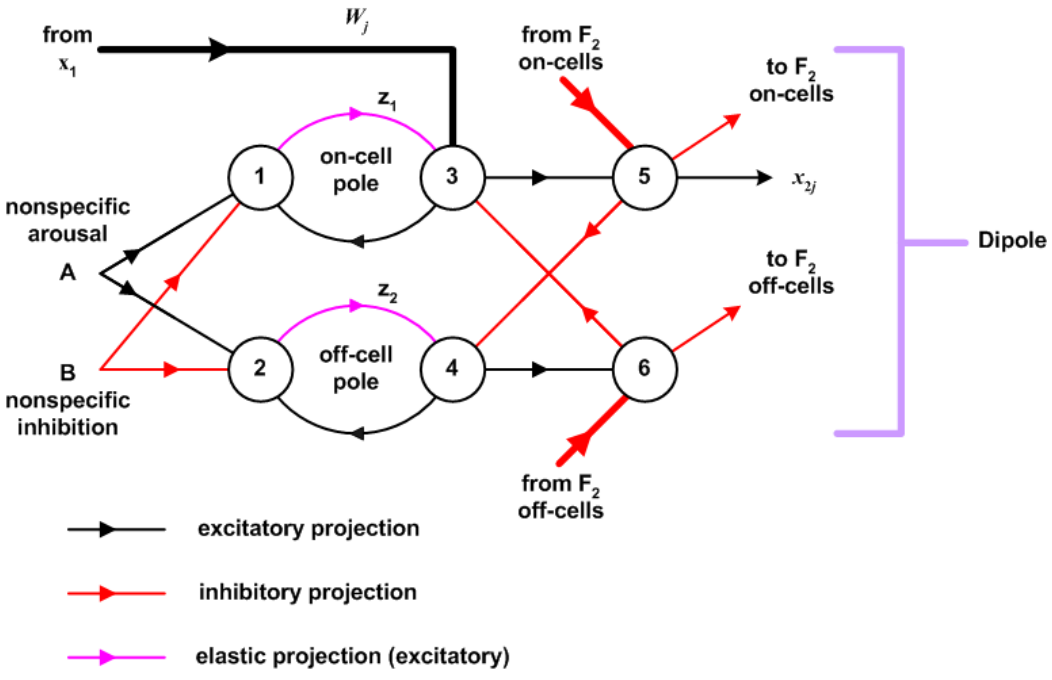


Figure 17.4: Dipole network model for an ART v_2 node. Instars 5 and 6 are type-$SNI^{(3)}$ nodes. The x_{2j} output projects to an Outstar node, and the $f(x)$ activities of 5 and 6 make inhibitory projections to other Instars in the 5-6 layer. Instar pairs 1-3 and 2-4 modulate the relative activities of the on-cell and off-cell dipole pair by means of elastic (short-term modulated) projections z_1 and z_2. See text for discussion.

In the terminology of this chapter, the *N* dipole networks constitute a v_2 layer. Instars 3 through 6 are associated with the F_2 layer of figure 17.1 while Instars 1 and 2 are associated with the reset function network (figure 17.3) of the orienting subsystem.

Instars 5 and 6 are the type-SNI$^{(3)}$ nodes introduced in chapter 15 for classifier CL$^{(2)}$. Instar 5 makes inhibitory $f(x_{2j})$ projections to other on-cell nodes in the third layer of the other dipoles in the v_2 layer. Instar 6 makes similar projections to the off-cell nodes. Instars 1 through 4 make up a rebound network, which is basically a recurrent version of the conditioning network shown in figure 17.2 earlier. The network pictured above is essentially the same that of figure 9 in [GROS6] except for the addition of nonspecific inhibition B and the explicit representation of the input $W_j^T\mathbf{x}_1$ as an input to Instar 3.

The purpose of this network is to implement the ART search process following a mismatch reset event. Grossberg qualitatively explains the function of this network as follows:

> [How] does nonspecific arousal, which is distributed *uniformly* across all populations in v_2, alter the balance of excitation in favor of previously inactive populations? This problem is particularly evident in adaptive resonances. Here a mismatch between the pattern [$\mathbf{x}_2$] coded by a population (say v_{21}) and a test pattern at [F_1] . . . suppresses the [F_1-to-v_{21}] signal, and causes x_{21} to decay, *before* nonspecific arousal arrives. Clearly a more slowly decaying trace must remain to indicate that v_{21} has just been active. This trace must also be slowly decaying to maintain inhibition of incorrect populations during a search routine. More precisely, STM activity at v_{21} depletes the slow trace in v_{21}'s arousal pathway, while the trace accumulates at inactive populations. Then equal arousal signals to all populations are gated, or shunted, by their slow traces, so that previously inactive populations receive larger arousal signals [GROS6].

The key to this recent-activity-dependent search process lies with the elastic projections z_1 and z_2 in figure 17.4. Let us assume figure 17.4 represents an active node in v_2 (call it v_{21}) with the on-cell Instar 5 corresponding to the active node of a 0-1 distribution in F_2. Let us further assume the non-specific inputs A and B are inactive. Then node 3 actively responds to the feedforward input signal $W_j^T\mathbf{x}_1$, exciting both Instars 5 and 1. At the same time, lateral F_2 projections to the other F_2 nodes hold their nodes 5 in the inactive state. The activity of node 1 causes a slow decay in the synaptic connection z_1 (an elastic short-term depression or STD), while inactivity by node 2 allows elastic connection z_2 to recover to full strength.

Now assume a mismatch-induced nonspecific arousal A is presented to the network. This signal excites nodes 1 and 2 but, owing to the greater magnitude of z_2, this arousal stimulates node 4 more strongly than node 3, assuming the level of stimulation is high enough to overcome node 4 inhibition from node 5. Consequently, the off-cell node 6 is activated and sends an inhibitory signal to node 3. This reduces the activity of node 5, which promotes greater activity in node 4 and leads to a positive feedback cascade action where node 5 is turned off and node 6 is turned on. Feedback from node 4 to node 2 maintains this state after arousal A is removed. The

lateral inhibition from the off-cell to other off-cells in F_2 allows these other nodes to respond to $\mathbf{x}_1$ and, thus, a new competition arises among the remaining F_2 nodes leading to a new 0-1 distribution. v_{21} remains inactive until the slow decay of z_2 reaches the point where $W_j^T\mathbf{x}_1$ excitation can again bring node 3 active. (Note that after removal of A, the inactivity of node 3 allows z_1 to begin rebuilding the strength of its elastic connection).

In the case where we do not have an initial 0-1 distribution, the most active F_2 node will also be the node where z_1 depletion is greatest. This is because: (1) the most active node 5 will also have the most active node 3; and (2) the lesser activities of Instars 5 in the other nodes is accompanied by a greater level of activity in their Instars 6 relative to the Instar 6 of the most active F_2 node. Therefore, a nonspecific arousal A will still selectively target the *most active* F_2 node for mismatch reset.

The elastic projections z_1 and z_2 instantiate a phenomenological model of short term depression effects. STD is a known biological phenomenon. Grossberg speculates [GROS6, CARP4] that the effect may be due to neurotransmitter depletion in the presynaptic terminals of upstream neurons within the populations represented by Instars 1 and 2. This is consistent with hypotheses made by various physiology researchers, although it is known today that this cannot be the only factor contributing to STD:

> There have been many attempts to model depression and recovery from depression. One approach is based on activity-dependent "depletion" of available release sites . . . Such approaches to modeling depression fail in several ways. This type of model predicts that pre-synaptic activity at rates faster than the time required to recover from depression would deplete available release sites, thereby making synapses extremely ineffective during high-frequency trains. Although depression occurs during trains, synapses are generally about ten times more effective during high-frequency trains than would be predicted by such a depletion model. This observation gives rise to the hypothesis that recovery from depression might be more rapid during stimulus trains . . .
>
> Recent studies have provided new insight into depression and recovery from depression. For example, at the climbing fiber synapse, although recovery from depression follows a single exponential in 1 mM external Ca^{2+}, in 4 mM Ca^{2+} a rapid phase of recovery is apparent. This suggests the hypothesis that high levels of residual calcium accelerate recovery from depression.
>
> Ca_{res} accelerates recovery from depression at the climbing fiber-Purkinje cell synapse . . ., an effect that is also observed at hippocampal synapses . . . and at the calyces of Held . . . The mechanisms responsible for depression and recovery from depression differ for various types of synapses. They include inhibition or neuromodulation of Ca^{2+} channels through metabotropic receptors, depletion of docked vesicles, desensitization of postsynaptic receptors, and other as of yet unidentified mechanisms [REGE].

Because Instar map models are very high-level models, encompassing the abstract representation of thousands of individual neurons, it is mere speculation to try to assign any particular physiological mechanism to the elastic behavior of z_1 and z_2 without first undertaking more detailed anatomical and physiological research to seek out more detailed neural network

models – perhaps at the level of Rulkov network models paired with even lower-level models (Linvill models, Hodgkin-Huxley models) of STD and recovery-from-STD phenomena. What we *can* say is that the elastic modulation mechanism Grossberg proposed fulfills a necessary and crucial role in the proper function of ART networks, and to the degree that ART networks provide an accurate model of psychophysical phenomena, the elastic projection hypothesis gains support from the success of ART network theory. Here the ever-mounting body of evidence year after year places the ART model on ever-increasingly-solid grounds as the best description computational neuroscience currently possesses of high-level brain system behavior.

This brings us to the second nonspecific input, input B, depicted in figure 17.4. It is equally a vital part of ART network dynamics that significant changes in the network's afferent input pattern must trigger a global recovery-from-mismatch-reset as well as also effecting a general reset of the ART competitive layers within the network (property 7 above). Failure to include the recovery-from-STD effect leads to misclassification errors by the network and, under some signaling conditions, can even lead to recoding of LTM patterns. Elastic projection STD in the ART model is quick in onset and, left to itself, slow in offset. Significant input pattern changes, on the other hand, produce rapid offset of this STD, i.e., rapid recovery-from-STD. As we saw in the earlier quote from Regehr and Stevens [REGE], rapid recovery-from-STD is a known feature of the short term depression phenomenon. Nonspecific input B, in addition to producing inhibition in Instars 1 and 2 in figure 17.4, also produces rapid recovery of both z_1 and z_2 to a state of full-strength connection. Because physiologists are in agreement that rapid recovery-from-STD does exist but the jury is still out on the physiological explanation of this phenomenon, we will not speculate here on a mechanism for the effect of input B. It remains for the time being a phenomenological mechanism within the overall functioning of an ART network.

§ 3. Functional Subsystems of Example ART Networks

The operational characteristics of ART are illustrated in this chapter by means of an example network. In this section the parameters for the subsystems within this network are described. The dynamical behaviors of the network are illustrated in the following sections. The network has n = 25 input signals in the input afferents pattern vector **P** and N = 3 classification nodes (dipoles) in the F_2 layer.

§ 3.1 The Data Path Subsystems

The network contains two layers within F_1 comprising a type $R^{(3)}$ resonator with 25 nodes in each layer (v_i and v_k in figure 17.3). The feedforward and feedback weights connecting v_i and v_k are fixed and make one-to-one projections of the form

$$\gamma_f \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \cdot \mathbf{x}_i \quad \text{and} \quad \gamma_b \cdot \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} \cdot \mathbf{x}_1 .$$

In the example network, $\gamma_f = 0.60$ and $\gamma_b = 0.20$.

Input afferents are preprocessed by normalizer $GN^{(2)}$ with parameters $B = 1$, $A = 0.5$. The contrast enhancing parameter is $C = B/(n - 1)$. The total input to $GN^{(2)}$ is $\mathbf{P} + a_f \cdot \mathbf{x}_i$ where $\mathbf{P}$ is the afferent inputs pattern vector and $a_f = 0.50$. The principal purpose of this feedback is to provide a degree of filtering for uncorrelated high-frequency noise that might be present in $\mathbf{P}$.

The attentional priming signal G_1 (figure 17.3) is given by $\mu \cdot \sum (\mathbf{x}_1)$ where the sum is taken over the n elements of $\mathbf{x}_1$ in the feedback path from v_k to v_i in figure 17.3. Here $\mu = 0.05$. The attentional priming signal G_k (figure 17.3) is similar and is given by $\sum GN^{(1)}(\mathbf{x}_2)$. This signal is weighted by the term B_3 (equation 15.5). For the example network, $B_3 = 0.02$. The sum is taken over the N signals of $\mathbf{x}_2$ in the F_2 layer. The $GN^{(1)}$ parameters are $A = 0.90$ and $B = 1$.

Layer v_k (figure 17.3) is a nonrecurrent layer of type-$SNI^{(4)}$ nodes (chapter 16, section 5). The parameters of this layer are $\tau \cdot B = 1$, $\tau = 0.001$, $B_3 = 0.02$. The output of this layer is $h(\mathbf{x}_1)$ where h is the Heaviside extractor activation function. This output is denoted $\mathbf{x}_1$ in the previous figures 17.1 and 17.3. v_k receives top-down feedback (expectations) from F_2 given by $\gamma_{12} \cdot \mathbf{Z} \cdot \mathbf{x}_2$, where $\mathbf{Z}$ is the Outstar feedback matrix and $\mathbf{x}_2$ is an abbreviation for $h(\mathbf{x}_2)$, h being the Heaviside extractor function. $\gamma_{12} = 0.50$ for the example network. It receives bottom-up signals from layer v_i weighted by the scalar forward gain factor $\gamma_f = 0.60$.

Layer v_i is an $n = 25$ node layer of type-$SNI^{(3)}$ (chapter 16) with parameters $B = 1$, $A = 0.65$, $u^{(1)} = 0.85$, $D = 2.0$, $u^{(2)} = 0.98$, and $g_{max} = 1$. These parameters yield QT = 2.43 and, as can be seen from the parametric values, v_i is set to produce a relatively high level of contrast enhancement (CE). v_i receives top-down feedback from layer v_1 weighted by scalar feedback weight $\gamma_b = 0.20$ for the example network.

The F_2 subnetwork consists of a single competitive layer of N dipoles. In actual simulations this layer is approximated using a competitive layer of type-$SNI^{(3)}$ nodes augmented by a dipole action approximation (rather than an explicit simulation of the network of figure 17.4). The dipole function is formally placed within the orienting subsystem's reset function block in figure 17.3 and is described below. The F_2 SNI layer (layer v_2 with outputs $h(\mathbf{x}_2)$) has parameters $B = 1$, $A = 0.8$, $D = 9.0$, $u^{(1)} = 0.95$, $u^{(2)} = 0.98$, and $g_{max} = 1$, giving QT = 4.75 for v_2. Layer v_2 receives bottom-up inputs from v_k of the form $\gamma_{f2} \cdot \mathbf{W} \cdot \mathbf{x}_1$, where $\mathbf{W}$ is the Instar weight matrix and the

forward path gain is $\gamma_{f2} = 30.0$. This large forward path gain is required because, unlike resonator type-$R^{(2)}$ in chapter 16, the elements of weight matrix **W** are determined by IAR adaptation and are much smaller than the fixed values used in the simulations in chapter 16. Note that neither γ_{12} nor γ_{f2} affect the values in **Z** or **W** obtained under the IAR method. Layers v_k and v_2 collectively form a type-$R^{(2)}$ resonator.

§ 3.2 Initial Weight Settings and Adaptation

The networks employ the Instar Adaptation Rule (IAR) for the Instar weights and the classic Outstar Adaptation Rule (c-OAR) for the top-down feedback weights. However, the orienting subsystem places certain restrictions on when adaptation will take place. These restrictions are key to the ART network's resolution of the stability-plasticity dilemma, which is the hallmark of ART networks generally. There are two general cases or modes of operation that must be taken into account to resolve the SPD: (1) the case when v_2 contains so-called *uncommitted* nodes, i.e. F_2 nodes that have not yet "learned" the LTM encoding for a particular input pattern; and (2) when all v_2 are *committed*, i.e. every node in F_2 has encoded LTM patterns.

An ART reset-and-search action is based on the occurrence of a mismatch between the bottom-up input pattern and the top-down expectation pattern. However, this reset must be inhibited while a new category LTM is first being established. This is accomplished by making the top-down weights **Z** equal to zero prior to any adaptation establishing the top-down LTM. At the same time, however, v_2 must be able to respond to $\mathbf{x}_1$ regardless of whether or not an initial LTM is established. This is accomplished by initializing the **W** matrix to small random values.

The ART network examples in this chapter respond to input patterns by producing a 0-1 distribution in $\mathbf{x}_2$ for patterns that are recognized and learned. In this way, the IAR and OAR actions are "gated" by the choice made at F_2. However, certain restrictions must be enforced on when adaptation is allowed in order to stabilize LTM learning and prevent spurious recoding of previously learned LTM category codes. These restrictions are as follows.

1. Adaptation is only allowed to occur when the system is in resonance. Formally, this means $d\mathbf{x}_i/dt$ and $d\mathbf{x}_2/dt$ are both equal to zero for all elements of these vectors. As a practical matter, the restriction requires that the sum of the absolute values of the derivatives for non-zero values of the x_i and x_{2j} does not exceed some critical threshold. (For zero-valued elements the derivatives are automatically zero by the clipping action of the Heaviside extractor function). Because the magnitude of a derivative depends on the magnitude of the signal – i.e., $d(ax)/dt = a \cdot dx/dt$ – the derivative vectors are normalized by $GN^{(1)}$ normalizers prior to comparison against the resonance threshold. For our example network, these normalizers have parameters $B = 1$ and $A = 0.90$.
2. Adaptation is only allowed to occur when the degree of mismatch arousal is less than the threshold for mismatch reset. In other words, no mismatched patterns are to be encoded

in LTM. Thus, the mismatch arousal signal exerts a global and nonspecific inhibition on LTP and LTD in layer v_2.

3. Adaptation is only allowed to occur if the total input excitation, $\sum_{i=1:n} x_{ki}$, exceeds a threshold for adaptation. This is tantamount to saying that weak stimulation of v_2 by v_k cannot stimulate LTP or LTD induction.

The biological plausibility of condition (1) can be understood by remembering that Instar nodes represent populations rather than neurons, and that the excitation and activation signals of a node are not "firing rates" but, rather, abstract representations of the total activity of the network. The existence of a resonance state implies a condition of cyclic equilibrium within the populations represented by Instar nodes. Such a condition implicates the establishment of tetanus signaling by the member neurons within the population, which is likewise a well-established condition under which LTP and LTD induction can take place.

The biological plausibility of condition (2) is understood by remembering that LTP and LTD are inherently metabotropic processes. Now, metabotropic second messenger cascades can interfere with one another, and thus it is entirely reasonable that nonspecific signals distributed over a neural population can block the metabotropic processes responsible for LTP and LTD.

The biological plausibility of condition (3) is understood by remembering that synapses capable of LTP are typically likewise capable of LTD. Although the physiology of LTP and LTD induction is far from completely understood, it is known that NMDA-mediated LTP and LTD depend on the amount of Ca^{2+} entering the cell. Physiologically reasonable models such as the calcium control model [GERS1: 377-383] incorporate a lower threshold of Ca^{2+} concentrations, below which no LTD or LTP induction takes place. As this depends on signaling activity, it is thus quite reasonable to conclude that LTM adaptation should likewise depend upon total stimulus activity converging at v_2 in at least the case of **W** adaptation. As for presynaptic forms of LTP and LTD, the physiology of this phenomenon is not so well understood, but it is hypothetically reasonable that this, too, should depend on the level of activity of the postsynaptic cells (perhaps by means of some retrograde second messenger such as NO; this is speculative but that merely reflects our current state of knowledge of presynaptic LTP and LTD). Thus, it is also reasonable that LTM for **Z** adaptation should depend on the total level of activity in $\mathbf{x}_1$.

When all three of the conditions stated above are satisfied, adaptation takes place by means of the IAR and c-OAR methods,

$$W_j(t+\Delta t) = W_j(t) + \eta \cdot \left[\mathbf{x}_1(t) - W_j(t)\right] \cdot x_{2j}(t)$$
$$Z_j(t+\Delta t) = Z_j(t) + \eta \cdot \left[\mathbf{x}_1(t) - Z_j(t)\right] \cdot x_{2j}(t)$$

where W_j and Z_j are the weight vectors chosen by the 0-1 distribution in v_2 and η is the learning

rate constant for adaptation.

§ 3.3 Input Onset/Offset Reset

Figure 17.1 depicts a nonspecific inhibitory signal aroused by changes in the input signal pattern **I**. Simulation studies carried out on ART networks reveal the functional need for changes in the input pattern – provided these changes are large – to trigger a reset of STM in F_1 and F_2. The requirement arises because STM patterns are "memory traces" (in the mathematical if not the psychological sense of that word), and changes in STM tend to lag changes in **I**. This is especially the case for STM_2 which, as our earlier simulations of the Grossberg resonators has shown, is significantly slower than either **I** or STM_1. If STM is not reset by the occurrence of large changes in **I**, misclassifications and even spurious recoding of LTM happen. Carpenter and Grossberg refer to this as "coordinating STM processing with an input presentation rate" [CARP3].

Of the several biological hypotheses that go into ART networks, the onset/offset reset mechanism is perhaps the hypothesis most vulnerable to criticism by physiologists. The issue is not one of whether or not such a reset can be implemented by biologically-realistic network models; it is easy enough to do so through, e.g., one of Grossberg's dipole networks such as that of figure 17.2. Rather, the issue is whether or not such a reset mechanism exists *at all* in biological systems. For example, subdural probe data collected by Bruns and Eckhorn [BRUN] show a considerable time lag taking place between onset or offset of a visual stimulus and the wave-like firing activities recorded from a human subject. Similar lags between stimulus onset and changes in signaling waveforms are observed in visual areas of awake monkeys [FREI1].

This does not mean that fast onset/offset resets do not occur in biological systems. There is a vast distance between the level of phenomena studied by the works just cited and the level of information processing modeled by ART networks. The ART models show that such a reset mechanism performs an indispensable function in ART networks. This is tantamount to a *prediction* coming out of adaptive resonance theory. It is a fundamental role of computational neuroscience to make such predictions, else the field languishes as a mere sinecure for mathematical exercises with no assignable contribution to biological science. Biologists are, rightfully, suspicious of predictions that arise solely from mathematical models, and this is to say no more than that all hypotheses are suspect until confirmed. Here we have a prediction for experimentalists to either confirm or refute (the latter being the more challenging since it is not possible to prove a negative). The experimental challenge here is great because such an effect, if it exists at all, would only be registered on a fairly large scale of neuronal structure, and existing experimental methods are not quite up to the task of detecting fast-acting (small lag) changes

following stimulus onset or offset at this scale.

For the models used in this chapter, onset/offset reset is implemented by the simple expedient of comparing the difference of the sum of the terms in $| \mathbf{I}(t + \Delta t) - \mathbf{I}(t) |$ against a fixed threshold.

§ 3.4 Dipole Reset Function

Figure 17.4 and the accompanying text explained the idea of an F_2 dipole layer in ART networks. In practice, ART network simulators rarely employ this level of detail in modeling the network. In part this is to save computation time. In part it is to avoid certain numerical issues that can arise when modeling a system in which the subsystems have a great deal of difference in the time constants among different parts of the system. What is usually done instead is that the overall *function* of the dipole action is approximated without going through the interior details of calculating how the network implements this function. Such a model is called an **input-output** model of the subsystem. We use such a model here for the example ART network.

The dipole field has four distinct modes of operation: (1) a **normal mode** in which the inhibitory pathways of the dipole field are inactive and F_2 acts merely as the second contrast enhancing layer of an ART resonator; (2) a **reset mode** in which an arousal signal, arising from a mismatch between the network's input pattern and the top-down expectation from a learned category, causes the inhibition of the most active v_2 node; (3) an **inhibition persistence mode** modeling the time span over which the dipole field inhibitions remain active; and (4) an **inhibition reset mode** in which dipole resets are cleared by a second nonspecific signal arising from input onset/offset resets.

The normal mode corresponds to the case where Instars 2, 4, and 6 of figure 17.4 are all inhibited for each node in v_2. In this case, and in the case for every v_2 node that has not undergone a recent dipole reset action, F_2 acts as if it were merely a type-$CE^{(2)}$ layer consisting of $N - m$ nodes, where $m \leq N$ is the number of nodes of v_2 that are not in a dipole reset mode.

The reset mode corresponds to the activation of the nonspecific arousal signal A in figure 17.4. This arousal signal excites Instars 1 and 2 in figure 17.4, and if the x_{2j} node (Instar 5) has been active, this corresponds to the elastic connection z_1 having a lower efficacy than z_2. This would lead to inhibition of Instar 3 with consequent inhibition of Instar 5. However, we must also take into account the on-center/off-surround coupling that exists in F_2. It is possible that A could be activated at a time when more than one v_2 node is active (i.e. more than one x_{2j} is non-zero). In this case, the node with the *largest* excitation will have the smallest strength of connection z_1, and thus will also have the largest aroused excitation of its Instar 6 node. Thus, relative to the other nodes in v_2, this most-active node is the first to undergo reset via the dipole structure. (If two or

more nodes happen to have equal levels of activity – a "tie" in MAXNET terminology – all such maximally-active nodes are reset).

Nodes in v_2 with zero or small activation levels are unaffected by the arousal signal A unless this signal persists past the current time step in the simulation. Thus, with the formerly most-active node (or nodes, in the case of a tie) reset and held inactive by the dipole action, a new most-active node will develop through the normal $CE^{(2)}$ reverberation dynamics. This is called an **ART search**. If the new most-active node also produces a mismatch between the bottom-up signal and the new top-down expectation, arousal A will be reactivated and the process is repeated. This is termed an **ART search cycle**. If the new most-active node produces an adequate bottom-up to top-down match, A becomes inactive and the search cycle ends.

To model the input-output behavior of this process we define two vectors, **d** and $\mathbf{J}_{rst}$, each having N elements. Vector **d** is called the reset-state vector; vector $\mathbf{J}_{rst}$ is called the reset-charge vector. **d** has binary-valued elements with $d_j = 1$ denoting that node x_{2j} is being held in a dipole-reset state and $d_j = 0$ denoting that node x_{2j} is either not in a dipole reset state or else is recovering from a dipole reset inhibition. d_j is set equal to 1 if it is currently 0 when a mismatch arousal occurs and its associated x_{2j} node has the largest excitation level in $\mathbf{x}_2$. If $d_j = 1$ already when a mismatch arousal occurs, it remains in this state.

Elements d_j are returned to the 0 state when the mismatch arousal signal becomes inactive and their associated nodes then enter inhibition persistence mode. When a state variable d_j is set to 1, its corresponding reset-charge element J_j in $\mathbf{J}_{rst}$ is set to a value $J_j = J_{sat} \cdot x_{2j}$ where J_{sat} is a maximum charge parameter (equal to 0.5 in the simulations shown in this chapter). The weighting of J_j by x_{2j} reflects that the loss of connection strength z_1 increases with the activity level of x_{2j}. After d_j returns to zero, J_j undergoes a geometric decay such that $J_j(t + \Delta t) = a \cdot J_j(t)$, where a is the geometric decay ratio (equal to 0.9995 in the simulations shown in this chapter). This dynamic reflects the recovery of z_1 (and the decay in z_2) that occurs when x_{2j} is being held in a dipole reset inhibition.

In general, the inhibitory signal applied to the SNI node in v_2 is given by a saturating Heaviside extractor function, $\mathbf{J}_{ar} = h_d(\mathbf{J}_{rst})$, with

$$h_d(u) = \begin{cases} h_{\max}, & u > h_{\max} \\ u, & 0 < u \le h_{\max} \\ 0, & u \le 0 \end{cases}.$$

For the simulations used in this chapter, the numerical range of the variables is such that h_d is functionally equivalent to a simple non-saturating Heaviside extractor. $\mathbf{J}_{ar}$ is represented by the

inhibitory vector **R** in figure 17.1. The elements of this vector are applied specifically to their associated SNI nodes in v_2, and thus the dipole field representation in the ART network anatomy in this chapter is divided up between F_2 and the reset function block shown in figure 17.3. With the parameter values used in the simulations in this chapter, the effect of inhibition persistence mode typically lasts on the order of about 8000 simulation time steps.

Finally, when an input onset/offset reset occurs, this reset signal is applied nonspecifically to all dipoles, corresponding to activation of nonspecific input B in figure 17.4. In this inhibition reset mode, the dipole state variables are cleared, i.e. $\mathbf{d} = \mathbf{0}$ and $\mathbf{J}_{rst} = \mathbf{0}$.

This completes the description of the dipole reset input-output function used by the network in this chapter. To complete the description of the reset function block in figure 17.3 only one more detail is needed. When a mismatch arousal A is activated, the reset block also sends a nonspecific inhibition signal to all the SNI nodes in F_1. Because F_1 does not contain any dipole fields, this inhibitory signal is a simple impulse activated whenever the mismatch arousal is active. The function of this reset is merely to clear layer F_1 of the influence of the previous input pattern and top-down feedback signal $\mathbf{Z} \cdot \mathbf{x}_2$.

§ 3.5 The Mismatch Arousal Function

Different ART network models typically differ in the specific method by which a mismatch between bottom-up and top-down signals is determined. The specific way in which a mismatch is determined not only fundamentally affects the behavior of the network but also affects what the network will view as a feature to be matched.

Generally, the afferent input pattern will be accompanied by "noise" variations. We can identify two general classes of such variations. The first is relatively uncorrelated noise that changes from time step to time step during simulation. Biologically, there is little significance to be attributed such variations and so we will call this type of variation ***background noise***. The network in this chapter is relatively insensitive to background noise above some signal-to-noise ratio. The second class of "noise" is low frequency variation from one pattern presentation to the next in the values of the pattern elements. This is the type of "noise" we looked at in the previous chapters. Here these random fluctuations can persist unchanging so long as the presented input pattern lasts, and biologically these variations *can* carry significance (because they reflect a particular variation in the activity patterns of upstream network subsystems). We will call this type of "noise" ***pattern variation*** (PV).

The mismatch arousal function for the example network is based in part on NICE-induced variations caused by top-down vs. bottom-up differences. However, pattern variation *is also a*

source of NICE in the contrast-enhancing v_i layer. Therefore, it is likewise a source of mismatch arousal for this network *even if the top-down expectation is perfectly matched* to the statistical expected value of the input afferent pattern. The overall effect of this is that the network "prefers" input patterns where the non-zero signals (those not quenched by $GN^{(2)}$) have *equal amplitudes*. Put another way, the network favors input patterns that produce a locally uniform distribution in $\mathbf{x}_1$ in the absence of top-down NICE. Thus, the network ***treats amplitude uniformity as a feature*** to be detected. Loosely speaking, the network is a kind of "binary input" network, although it does tolerate a very small amount of pattern amplitude non-uniformity and, owing to $GN^{(2)}$, allows for a relatively wide range of overall pattern input amplitudes.

The top-down expectation signals are made explicitly available through the use of facilitating Outstars (*f*-Outstars). These *f*-Outstars use $\mathbf{x}_1$ in their *f*-OAR adaptation process and, thus, the matrix elements produced by adaptation in the *f*-Outstars match those of the direct feedback pathway from $\mathbf{x}_2$ to v_k, i.e. both $\mathbf{Z}$ matrices depicted in the figure have identical values. However, the output signals from these *f*-Outstars cannot be used directly to compare with $\mathbf{x}_i$. This is because when the network initially starts up no $\mathbf{x}_2$ node is yet "committed" to any particular pattern classification ($\mathbf{Z} = \mathbf{0}$) and so the initial top-down feedback *never* matches $\mathbf{x}_i$. It is necessary to sum $\mathbf{x}_1$ and $\mathbf{Z} \cdot \mathbf{x}_2$ before a mismatch comparison can be made. The mismatch comparand term is $\mathbf{x}_1 + \gamma_s \cdot \mathbf{Z} \cdot \mathbf{x}_2$. The gain factor used in the simulations is $\gamma_s = 10.0$.

A method not entirely dissimilar to this is used in Carpenter's and Grossberg's ART 2 network, although ART 2 is a very different anatomy than our example network and, it should be noted, generally has performance characteristics most researchers, including your author, would regard as superior to the example network. On the other hand, ART 2 is a step or two farther removed from biological signal processing tie-backs to physiology (owing to several abstractions and simplifications employed to produce superior computational performance). It is, in other words, a somewhat more abstract model and algorithm than is the example network used here. Your author thinks that to properly appreciate the thinking that went into ART 2, it is helpful and perhaps even necessary to see what dynamics develop in the network presented here.

The mismatch arousal network forms the sum $\mathbf{V} = \mathbf{x}_i + \mathbf{x}_1 + \gamma_s \cdot \mathbf{Z} \cdot \mathbf{x}_2$ and passes this on to the mismatch calculation function (the lower right-hand block in figure 17.3). It also passes on two mathematical metrics of the constituent vector lengths, namely $\|\mathbf{x}_i\|$ and $\| \mathbf{x}_1 + \gamma_s \cdot \mathbf{Z} \cdot \mathbf{x}_2 \|$ – the L_2 or Euclidean norms of these two vectors. The mismatch calculation block first forms a **reset vector**

$$\mathbf{Q} = \frac{\mathbf{x}_i + \mathbf{x}_1 + \gamma_s \cdot \mathbf{Z} \cdot \mathbf{x}_2}{\|\mathbf{x}_i\| + \|\mathbf{x}_1 + \gamma_s \cdot \mathbf{Z} \cdot \mathbf{x}_2\|}.$$

From this it forms an arousal signal $\alpha = 1 - \|\mathbf{Q}\|$. If $\alpha > \rho$ the mismatch arousal signal A is generated. For all simulations shown later in this chapter, $\rho = 0.02$ is used.

For all practical intents, this is the same mismatch function employed in ART 2. The basis of this method hinges on the fact that for any two vectors $\mathbf{q}_1$ and $\mathbf{q}_2$, $\|\mathbf{q}_1\| + \|\mathbf{q}_2\| \geq \|\mathbf{q}_1 + \mathbf{q}_2\|$ with equality if and only if $\mathbf{q}_1$ and $\mathbf{q}_2$ are co-linear (point in the same direction). This is a consequence of the triangle inequality for metric functions. When $\mathbf{x}_i$ and $\mathbf{x}_1 + \gamma_s \cdot \mathbf{Z} \cdot \mathbf{x}_2$ are co-linear (which happens when $\mathbf{x}_i$ does not do too much contrast enhancement and the top-down expectation likewise is a match to $\mathbf{x}_i$), then $\mathbf{Q}$ is a unit vector and $\alpha = 0$. Otherwise, $\alpha > 0$.

A principal criticism that can be leveled at this mismatch calculation is illustrated by the question, "What kind of neural structure produces L_2 norms?" There is little doubt that such a function can be approximated to some degree of accuracy and within some limited range of variables by network structures such as feedforward neural network systems of Adaline-type Instars. But here the attending questions are: (1) what specific network does this take? (2) how accurate is the approximation? (3) over how wide a dynamic range does the approximation remain accurate? (4) is the accuracy sufficient to produce the results needed for ART network operation? and (5) are such structures actually present in biological systems? These are legitimate questions and until they are addressed a gap exists between ART network models that use the L_2 norm and the lower levels of neural network theory. At present, the strongest reason for regarding L_2-norm-based models as legitimate is the indubitable success ART models have had in modeling a variety of large scale psychophysical phenomena. This is, of course, an empirical basis of trust and, at the end of the day, the only one that really counts for neuroscience. But it is understandable why not a few biological neuroscientists tend to be suspicious of purely mathematical arguments and why at present one cannot discard the suspicion that what we have here is merely a bit of convenient mathematical chicanery. This is an issue of science that ART models, and likewise most other models in computational neuroscience, must some day address.

What is the advantage gained by use of this method of mismatch arousal and reset? Simply put, it is greater tolerance to amplitude variations among the elements of the input pattern vector. Layer v_i still does contrast enhancement on afferent patterns containing significant pattern variation, and so the tolerance of the network to these variations is limited. However, when a mismatch occurs between the bottom-up signal and the top-down expectation in this network, the mismatch function response is generally larger and more dramatic than it is for mismatch due to low-level pattern variation. This gives the network more ability to handle "analog" input signals to a limited degree. It is to be noted in this context that ART 2 has a much better ability to handle analog inputs owing to its use of a different structure in the F_1 network that is not as sensitive to

NICE as the example network presented in this chapter.

§ 3.6 Reset Conditions

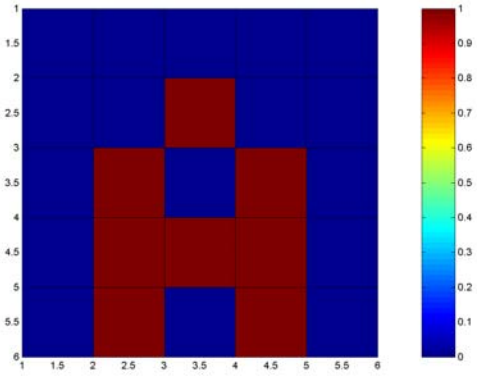
Various constraints are placed on when mismatch arousal resets can occur, in accordance with principles 4 and 5 of §2.2. For the example network, mismatch reset is inhibited when the network is not in the resonant state (§3.2). This inhibition function is not shown explicitly in the block diagrams presented earlier, but it is an implicit function within the attentional/orienting subsystem of figure 17.1. The network in this chapter also employs a reset arousal hold-off interval following the generation of the reset pulse A. This is to allow a brief time for the layers of the network to undergo the transient response that follows dipole reset of the maximum-valued v_2 node. This is not a function particularly critical to the operations of the networks since immediately following a reset the networks are not in a resonance condition, but the network model of this chapter uses it and so it is mentioned here. The reset hold-off interval is twenty time steps in the simulator.

Mismatch resets are also suppressed when the maximum activation in the v_2 layer of F_2 is below a critical activation level (0.15 for the example network). Biologically, this is tantamount to saying dipole Instar 3 (figure 17.4) must achieve some minimal level of activity before short term depression of z_1 appears. *Functionally* this inhibition precludes nodes in F_2 from being held in a dipole-reset state during or following a period where no significant afferent inputs are applied to the network and input signal activity ramps up at a rate slow enough to prevent the input onset/offset reset from being activated. This mode of mismatch arousal inhibition prevents spurious misclassifications as the network transitions from a "no input" to an "input presented" condition, and it guards against spurious re-coding of LTM. Like the other reset inhibition condition, the network structure for this inhibition mode is implicit in the attentional/orienting subsystem and is not explicitly shown in the previous diagrams.

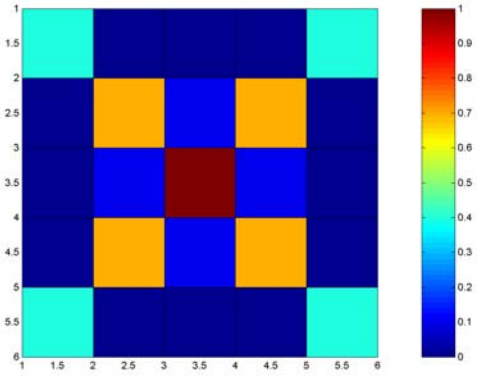
§ 4. Network Simulations

In this section we look at some examples of the behavior and performance of the example network described above. Although input patterns **P** are vectors, we continue our practice of displaying these patterns in the form of a 5×5 retina in order to make it easier to discuss the results. The nine basic input patterns used in the simulations are illustrated in figure 17.5 below. The names given to these patterns are identified below each one in the figure. The base patterns are binary (except for the analog X pattern) with pixel values 0 and 1. During the performance tests, pattern variance noise (PV), that lasts throughout the dwell time of the input pattern, and random noise, that varies from step to step, are added to the patterns.

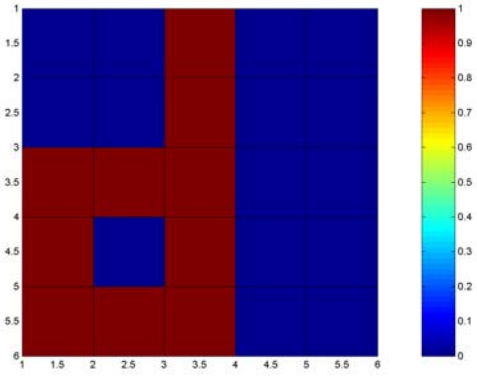
Each pattern presented to the network is given a dwell time of 2000 simulation time steps. The dwell time is the duration over which the pattern is presented. The simulations are broken into "runs" of groups of three patterns repeated three times (18,000 time steps per run). Except where noted otherwise, each new run is temporally continuous with the previous run and the network's initial conditions during any run are those established at the end of the previous run.



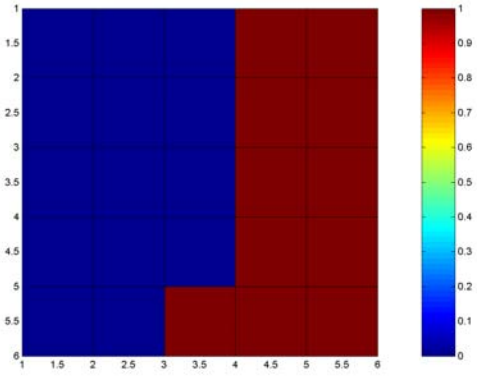
A pattern



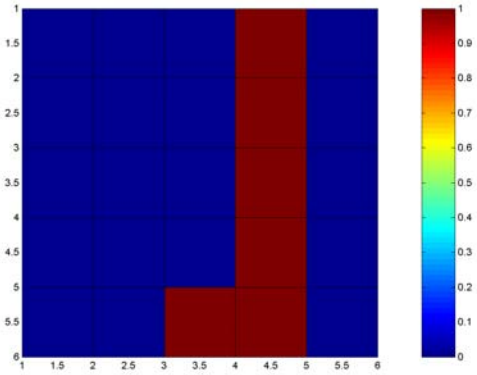
analog X pattern



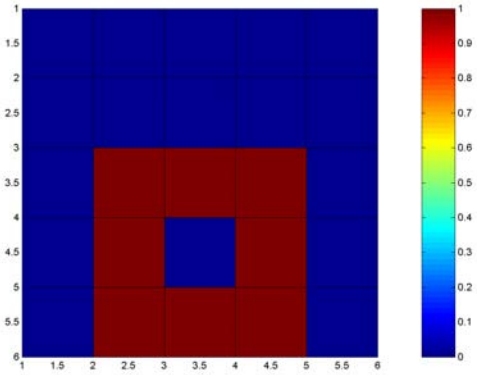
d pattern



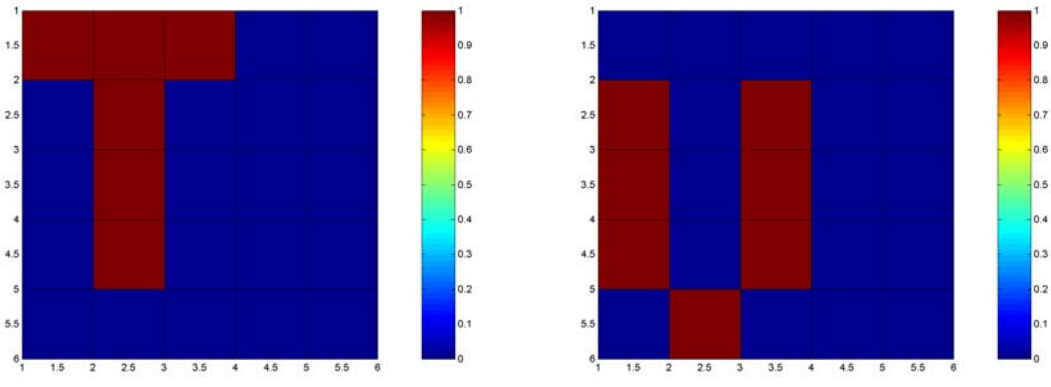
fat J pattern



J pattern

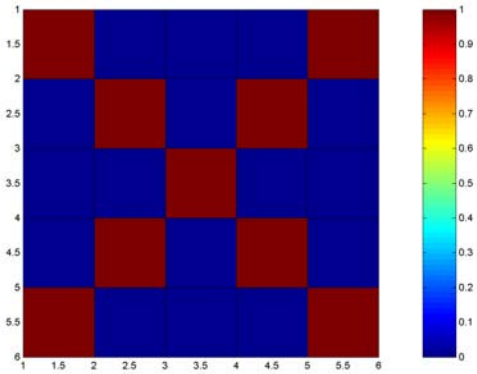


O pattern



T pattern

U pattern



X pattern

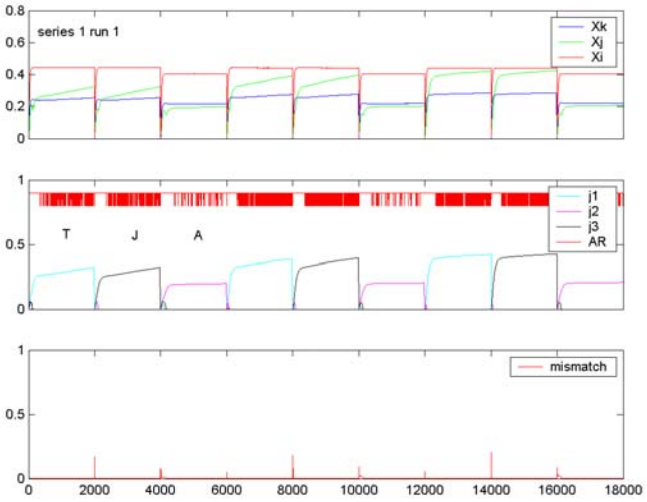
Figure 17.5: Patterns used for network simulations displayed in retina layout form. Names given to the patterns are identified below each pattern figure.

§ 4.1 Training Sequences and LTM Stability

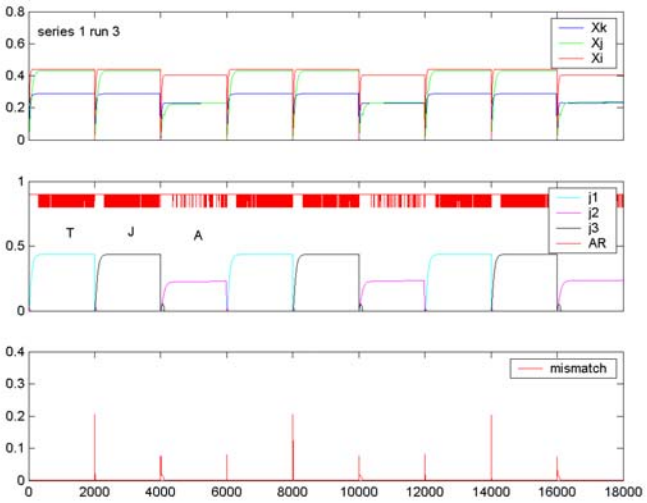
The first series of simulation runs illustrates the learning dynamics and LTM stability of the network. For this series the network begins in a relaxed initial state with LTM matrix **Z** initialized to all-zeros and LTM matrix **W** initialized to small, uniformly distributed random values. The network is to "learn" input patterns T, J, and A (see figure 17.5). This is accomplished by presenting these patterns sequentially, i.e. T-J-A-T-J-A-etc. until LTM stabilizes. This constitutes a form of unsupervised "training" for the network, which is said to "practice" these patterns.

The network required a total of three runs (nine presentations each of the T, J, and A patterns) to achieve solidly-established LTM values for each of its $N = 3$ available categories. The results of the first and third runs in this series are illustrated in figures 17.6 below.

Figure 17.6(A) illustrates the response of the network to the first three presentations of each pattern. The pattern order of presentation does not matter, nor is it necessary for the order to remain the same throughout the training sequence. Which particular node of F_2 will learn a given



(A) Series 1 run 1: first 3 presentations of T, J, A pattern sequence



(B) Series 1 run 3: 7th, 8th, and 9th presentations of T, J, A sequence

Figure 17.6: Results of initial training sequence for T, J, and A patterns. Pattern input pixel amplitudes were 0 and 1. Background noise uniformly distributed in the range (0, 0.01) was added to each pixel at each step in the simulation.

pattern depends on the random initial settings of **W**. For this particular simulation run, node x_{21} happened to encode the T pattern, x_{22} encoded the A pattern, and x_{23} encoded the J pattern. Referring to the center figures in (A) and (B), the red trace near the top of the figure (AR) displays when the network is in resonance (lower value) or not in resonance (higher value). Adaptation can only take place when this trace is "low" and the mismatch trace is below $\rho = 0.02$. As can be seen from the figures, adaptation is less frequent for the A pattern, which has the smaller total excitations for $\mathbf{x}_i$ and $\mathbf{x}_1 = \mathbf{x}_k$ in the figure. The reason the A pattern has the smaller

total excitation of the three, despite the fact that its Hamming weight w_H (the number of nonzero pixels in **P**) is 8 (compared to $w_H = 6$ for both T and J), is due to the actions of $GN^{(2)}$, which tends to penalize the pixel amplitudes of patterns with larger Hamming weights. The network actually succeeds in encoding the T and J patterns in **Z** and **W** after only three presentations of these patterns (i.e., by the end of run 1), whereas it was not until the end of run 3 that the A pattern was solidly encoded in **Z** and **W**.

Comparing the center figure of 17.6(A) with that of 17.6(B), it is seen that the peak levels of the $\mathbf{x}_2 = \mathbf{x}_j$ in (A) show an upward slope while those of (B) are relatively flat. The sloping action in the (A) figure is due to adaptation. As the network "learns" its categories, F_2 supplies increasing amounts of feedback to F_1, which raises the total network excitation to a steady-state peak value. By the end of run 3, the difference between the W_j, Z_j, and $\mathbf{x}_1$ vectors is quite small, which is why the $\mathbf{x}_2$ excitations no longer show appreciable slopes in the (B) figure despite rather vigorous adaptation being signaled by the red "adaptive resonance state" indicator trace AR.

The x_i trace in the top figures of (A) and (B) is the total excitation for the nodes of layer v_i. This trace is relatively flat throughout the entire simulation, which indicates that contrast enhancing layer v_i is performing no significant contrast enhancement on **P** even after feedback from F_2 (acting on v_i indirectly through v_1) becomes appreciable. The absence of contrast enhancement is due in part to the settings of the network parameters (which were designed to put the T and J patterns within the non-enhancing dynamic range of the network) and in part due to the absence of pattern variance PV in the input patterns **P**. The actions of $GN^{(2)}$ produce larger individual amplitudes in **I** for patterns with lower Hamming weights, and larger amplitudes tend to bring v_i closer to its contrast-enhancing threshold.

After the network had encoded the T, J, and A patterns at the end of run 3, a pattern sequence O, X, U was applied for run 4. The purpose of this simulation is to test the stability of the LTM encoding. As can be seen from figures 17.5, the U pattern overlaps the A pattern, while the O pattern overlaps with both A and J. The X pattern overlaps all five other patterns. Thus, the network is being presented with patterns for which F_2 must respond (since it has nonzero **W** and **Z** elements corresponding to some of the pixel elements in O, X, and U). The response of the network to this sequence is illustrated in figure 17.7 below.

The network LTM encoding was unchanged by this pattern sequence. Referring to the center figure in 17.7, application of the O pattern resulted first in a trial classification of the pattern as a J pattern. However, this classification was reset as soon as resonance was established and the network searched for a better match. Its second attempt was to classify O as the A pattern, but this, too, was reset as soon as resonance was established. Finally, the network classified pattern O

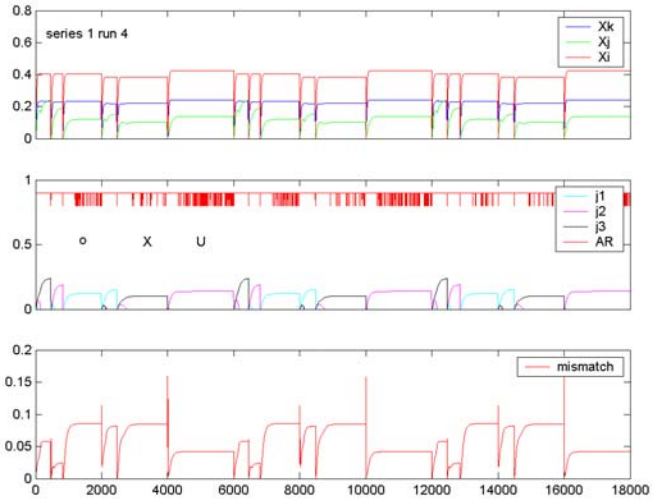


Figure 17.7: Run 4 response of the trained network to pattern sequence O, X, U.

as the T pattern. Note, however, the large mismatch arousal for this classification, which is well above the $\rho = 0.02$ threshold. The large arousal suppressed adaptation, and this is why the LTM encoding of T was unaffected by the O pattern. The network did not issue a reset during this classification because the amplitude of x_{21} falls below the "no pattern" threshold level of 0.15. Otherwise a reset would have taken place and, because the capacity of F_2 is only three nodes, this would have left the system with an all-zero output $\mathbf{x}_2$. As noted earlier in this chapter, the use of a "no pattern" threshold on $\mathbf{x}_2 = \mathbf{x}_j$ serves physiologically to model the threshold effect in synaptic short-term depression and functionally to prevent spurious resets from occurring when the **P** input to the network is quiescent.

Results are analogous for the O pattern input. The network initially attempts to classify this pattern as T, undergoes as reset search as soon as resonance is established, and ends up classifying this input as a J. Again, the mismatch arousal is above threshold (suppressing LTM adaptation) but the reset itself is suppressed by the low magnitude of x_{23}. Finally, the U pattern is classified as an A pattern but, again, mismatch arousal is well above the vigilance threshold and x_{22} falls below the "no pattern" threshold. Thus, no recoding takes place here either. Note that despite the fact the network "classifies" these input patterns, the fact that these patterns are different from T, J, and A is signaled by the level of the mismatch arousal signal.

Following run 4, a long-dwelling "fat J" pattern (figure 17.5) with a dwell of 18,000 time steps was applied to the network. This pattern is a superset of the encoded J pattern, as can be seen from the figure 17.5 figure patterns. It also has a one-pixel overlap with the A pattern. The simulation results are shown in figure 17.8 below.

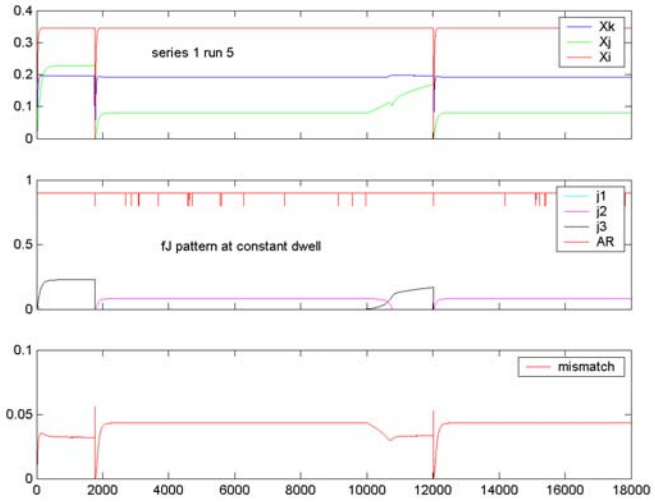


Figure 17.8: Response of the network to a long-dwell "fat J" input pattern.

Not surprisingly, the network initially attempts to classify this pattern as a J. However, the large mismatch between the top-down expectation and the "fat J" pattern causes a reset search as soon as resonance is established. The network then classifies the pattern as an A pattern, this being the only other pattern with which "fat J" shares a common feature. However, the arousal is still above threshold and so the LTM for A is protected from recoding. x_{22} is below the "no pattern" threshold, so a reset does not ensue. Starting at around time step 10,000 of this run, x_{23} comes out of its dipole-reset condition and quickly re-establishes itself as the classification for the input pattern. Again, however, the mismatch is too great and another reset ensues as soon as the network reaches the resonant state.

The outcomes of these simulations would have been analogous had other initial training patterns been used or other possible untrained patterns been applied. Put in other words, the results obtained here do not depend on special properties of any of the input patterns other than the fact that pattern variance PV was zero in this series. The simulations presented here illustrate that the network successfully accomplishes resolution of the stability-plasticity dilemma in the absence of pattern variance and in the presence of low levels of random background noise.

§ 4.2 Network Response to Pattern Variance

To a limited degree, afferents **P** containing pattern variance – unequal pixel amplitudes from one presentation to another – can be regarded as "analog" patterns, although "corrupted binary patterns" is probably a better description. With its high degree of contrast-enhancing built into the F_1 layer, the example network can be said to "prefer" or "be tuned" to binary patterns. However, a

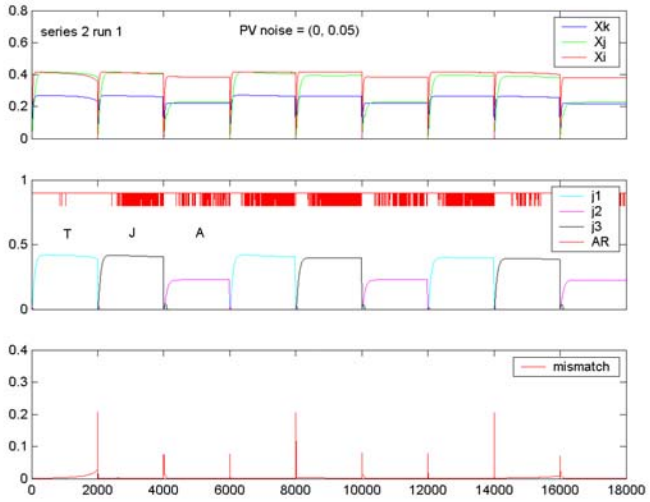


Figure 17.9: Series 2 run 1 with pattern variance in the T, J, A sequence. PV is randomly varied with each new pattern presentation in this run. The PV distribution is uniform over the range (0, 0.05), a 5% variation.

certain amount of robustness in the presence of pattern variance is important even for "binary" networks if only because one must assume a certain amount of PV is likely to be present in biological signaling. Such would be a reasonable interpretation of PET or fMRI scan measurements. The second series of demonstrations illustrates how the example network responds to pattern variance. In this series the afferent patterns were applied in the T, J, A sequence once again, this time with random pattern variance uniformly distributed in the range PV = (0, 0.05) – a range similar to our earlier discussion of NICE in chapter 16. The network's initial state at the beginning of the first run in the series was set by the final run results from the first series.

Figure 17.9 above illustrates the outcome of the first simulation run in this series. The three patterns are properly classified during each presentation, and weight adaptation is active in each presentation. Here the first presentation of the T pattern is interesting. After a few adaptation adjustments made by the network, the F_1 layer begins to contrast enhance the pattern. This is shown by the downward slope suddenly taken on by the x_k total excitation in the topmost figure and by the rise of the mismatch arousal signal in the bottom. As soon as contrast enhancement is underway, the system drops out of the resonance condition and further adaptation is disabled. If the T pattern had been given a longer dwell time, F_1 would have eventually contrast enhanced toward a 0-1 distribution until the mismatch reset kicked in.

A lesser amount of contrast enhancement can also be discerned in the third presentation of the J pattern in this run. This is evidenced by the visible rise in the level of the mismatch arousal signal. The contrast enhancement is sufficient to slow and finally halt adaptation through loss of a

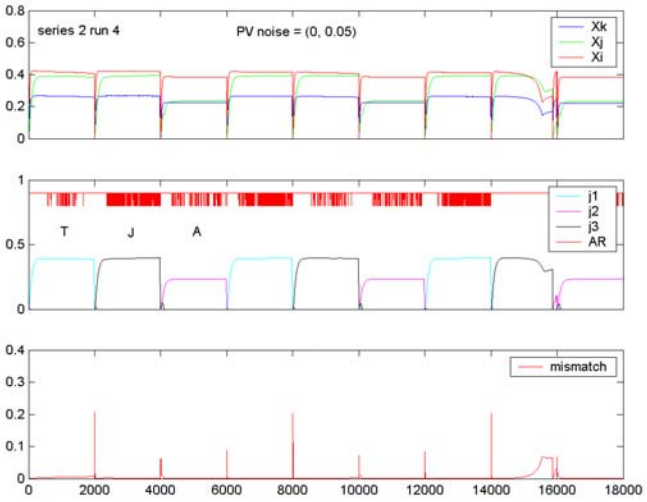


Figure 17.10: 4th run in the second (pattern variance) series. This run is noteworthy for the large amount of contrast enhancement in the third presentation of the J pattern. The CE is large enough in this case to trigger a mismatch reset of $\mathbf{x}_2$. No adaptation took place at all during this J presentation.

state of resonance in the system. Results from a second and third run, each continuations of the run immediately preceding it, showed very similar responses. In all three runs (nine presentations of each pattern), no sign of contrast enhancement shows up in the A pattern. The reason for the relative immunity of pattern A is its relatively lower level of overall excitation. This, in turn, is a consequence of the actions of $GN^{(2)}$, which penalizes the pixel amplitudes of higher Hamming weight patterns, and the normalizing action of the v_i layer of F_1, which produces a similar effect.

This does not mean that PV at this level is entirely benign. The fourth run, illustrated in figure 17.10, shows that PV enhancement, even at this 5% level, is capable of producing enough contrast enhancement to trigger mismatches with reset. This occurs in the third presentation of the J pattern in figure 17.10. Note, however, that no adaptation takes place during this presentation because the network never achieves resonance.

Adaptation does take place during most of the pattern presentations in this series. In effect, the network views the PV-altered presentations as somewhat different patterns, although not different enough for them to be classified differently in most cases. The onset of contrast enhancement in the lower Hamming weight patterns (T and J) limits the amount of total adaptation taking place. The effects of the pattern variance does show up in LTM in the form of random variations in the weight settings (driven by the random variations in **P** due to pattern variance). In adaptive signal processing, this random "jiggling" of **W** and **Z** is commonly termed ***weight noise***. This "random walk" or "Brownian-motion-like" variability in the weight settings is self-limiting through the mechanism of resonance loss from contrast enhancement. The result is the **W** and **Z** matrices drift

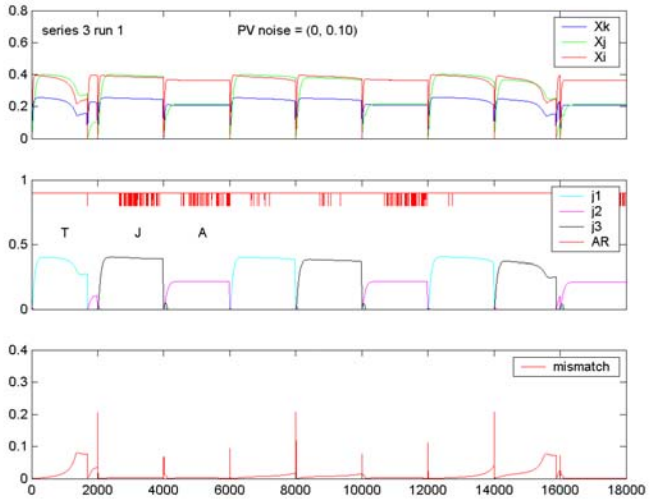


Figure 17.11: First run in series three, PV range = (0, 0.10). This run is a direct continuation from series 2 run 4 in figure 17.10.

about, centered on the statistical average of the **P** presentations for each F_2 node partition.

The conclusion to be drawn from the series 2 simulations is the following. At low (5%) levels of pattern variance, low Hamming weight afferent patterns are more susceptible to NICE brought about by the pattern variance. The onset of large contrast enhancement events abolishes the state of adaptive resonance, thereby shutting down adaptation and protecting LTM encoding. Smaller PV situations – those not productive of NICE – produce random walks in the LTM weight settings, but these variations remain centered about the expected value of the presented patterns within the particular partitions defined by the nodes of F_2.

Not surprisingly, larger amounts of PV exacerbate the effects just noted. Series 3 in the simulations carries on continuously from the end of series 2, but with the PV range increased to 10% of the base pixel amplitudes. Figure 17.11 illustrates the first run in series 3. Contrast enhancement can be noted in all three presentations of the T pattern and in the second and third presentations of the J pattern. In the first T and third J, the contrast enhancement is large enough to trigger mismatch resets (which would be a form of classification failure). Again, however, the misadaptation of the LTM encoding is self-limiting due to loss of resonance during contrast enhancing. Once again, the higher Hamming weight pattern (A) was more or less immune to NICE. Additional runs in series 3 produced similar network responses. Resets were not always produced during each run, but the rate of reset incidents does significantly increase.

The general conclusions drawn from series three, in addition to those drawn from series 2, are the following: (1) larger PV produces more NICE and more frequent mismatch resets in the lower Hamming weight patterns; (2) LTM encoding is not seriously compromised; and (3) larger levels

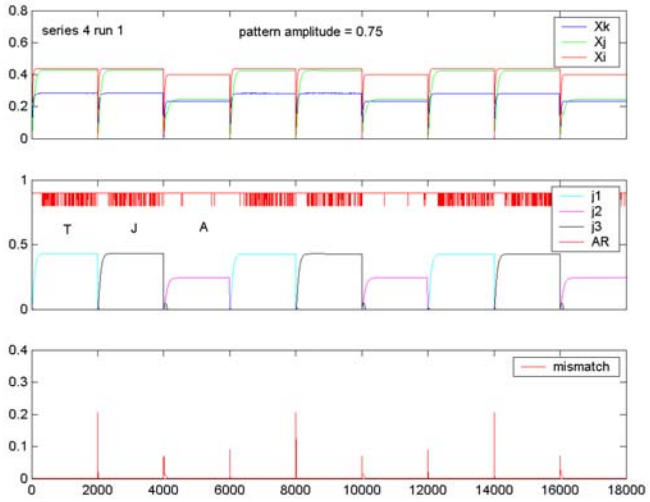


Figure 17.12: Series 4 run 1. Pattern pixel amplitudes here are reduced to 0.75 and pattern variance is set to zero. Background noise continues to be applied in the range (0, 0.01).

of PV produce a greater range of "weight walk" in **W** and **Z**, but the ability of the network to classify patterns is not seriously compromised inasmuch as those patterns with extreme amounts of PV can be regarded as significantly different from the low-PV patterns encoded in LTM.

§ 4.3 Network Sensitivity to Pattern Amplitude

Pattern variance is pixel-to-pixel amplitude variation that persists throughout the time a pattern is presented to the network. The final series of simulations in demonstrating the behavior of the network tests for the effects of flat-loss in amplitude, i.e. the reduction of *all* pixel amplitudes in a pattern by a fixed gain factor. This series is series 4 of the simulation.

Again, the patterns presented are the T, J, and A sequence. The first run of series 4 presents these patterns without PV but with nonzero pixel amplitudes reduced to 0.75. Figure 17.12 illustrates run 1. The network response is more or less indistinguishable from the series 1 (amplitude = 1.0) cases seen after the network is trained. The network is robust to flat pattern amplitude reductions to 75% of the original level. This is due in part to the actions of the $GN^{(2)}$ stage and in part to the normalizing actions of the v_i stage of F_1.

Run 2 of this series further reduced the nonzero pattern pixel amplitudes down to 0.20, i.e. 20% of their original values. The results of this run are illustrated in figure 17.13. Here all patterns were correctly classified, but we can observe that no adaptation ever took place at all. What has happened is that the high-frequency noise terms are now large enough, relative to the pixel amplitudes, to prevent the network from detecting a resonance state. (Note how AR always remains in the high state during this run). Evidence of the high-frequency noise can be seen riding

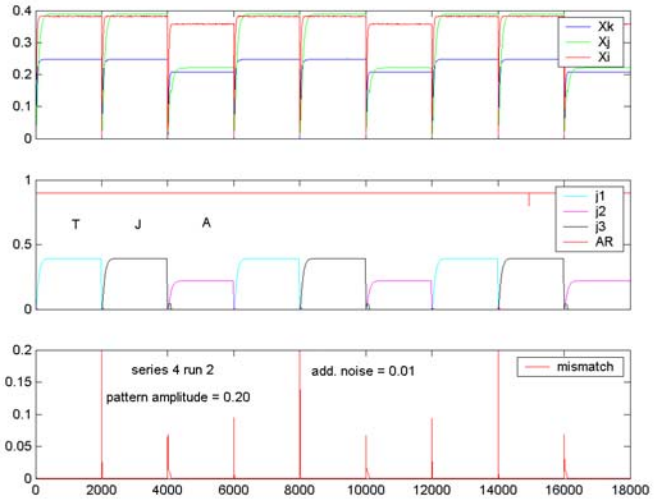


Figure 17.13: Series 4 run 2. Here the pattern amplitude is reduced to only 0.20. Pattern variance is set to zero, but the background noise continues to be applied in the uniform distribution from (0, 0.01).

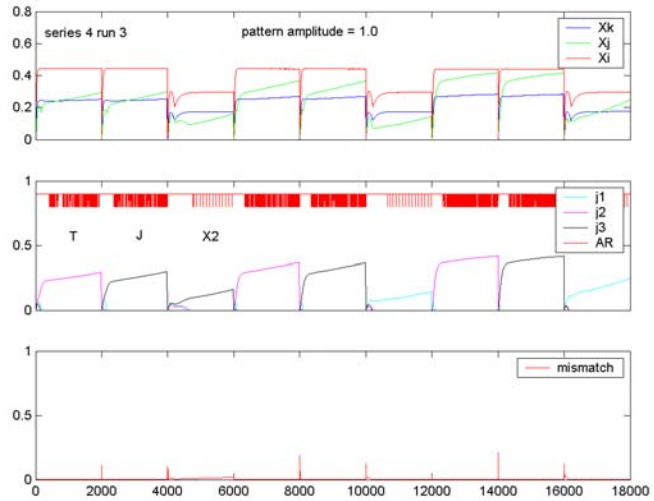


Figure 17.14: Training attempt for pattern sequence T, J, analog X.

atop the $\mathbf{x}_i$ total excitation in the top-most figure. Only at one point, in the third presentation of J, does the network briefly detect resonance and undertake a single adaptation step. The conclusion that can be reached from this analysis is that the LTM coding is stable in the face of flat-gain amplitude variation, but that low-level signals applied to an untrained network will not be learned and classified when accompanied by a sufficient level of background noise.

Run 3 of this series tests the network's response to more explicitly "analog" input patterns. Figure 17.14 illustrates the run. Here the applied sequence is T, J, and then the "analog X" pattern of figure 17.5. The network's initial state is untrained with randomized $\mathbf{W}$ and $\mathbf{Z} = \mathbf{0}$.

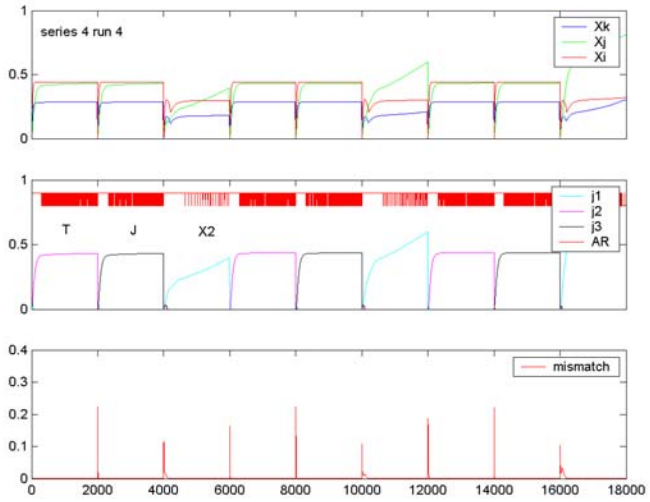


Figure 17.15: Continuation of the run of figure 17.14. Encoding for T and J is completed at the conclusion of the previous run, but the encoding for the contrast-enhanced analog X pattern does not come to completion until its third presentation in this run.

The network begins encoding the T and J patterns in the usual manner. However, the analog X pattern produces a quite different response. The F_1 layer quickly contrast-enhances the analog X pattern to a 0-1 distribution even before layer F_2 can begin to encode a pattern. The indicator for this is the rise, dip, and recovery of the total excitation of $\mathbf{x}_k$ in the top figure of 17.14 early on in the presentation of "analog X."

LTM encoding of the T and J patterns is essentially completed by the end of the run in figure 17.14. LTM encoding of the 0-1 distribution, which is all that remains of "analog X" after F_1 carries out its contrast enhancement, takes longer. Figure 17.15 shows the continuation of the sequence immediately following the figure 17.14 run. Note how with each successive presentation of analog X the amount of time the network spends in a resonance state increases. LTM encoding of analog X (as a single non-zero weight in W_j and a corresponding single non-zero weight in Z_j) completed in the next presentation of analog X following the completion of the figure 17.15 run.

In one sense, the network does succeed in encoding and classifying the patterns it is presented. However, in the case of the analog X pattern, only the single largest-amplitude pixel from figure 17.5 is encoded. In a broader sense, it can be rightly said the network fails to classify analog X *as an analog pattern*. As was observed earlier, this network is essentially a binary-input ART network with a limited ability to handle pattern amplitude spreads within **P**.

This, of course, brings up the question of how this network, once trained on T, J, and analog X, will respond to other patterns that share the single encoded pixel from analog X. Figure 17.16

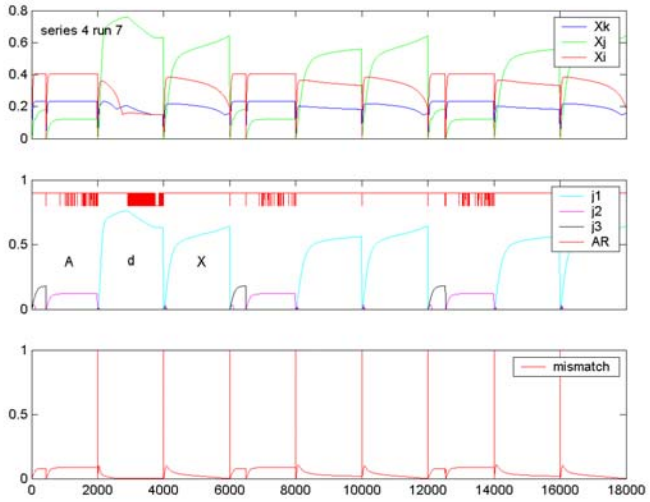


Figure 17.16: Response of the network trained on T, J, and analog X to the sequence A, d, X.

illustrates the response of the trained network to the sequence A, d, X (see figure 17.5). Here we can note that both the d and X patterns share the common pixel encoded by LTM from the analog X pattern.

The network first attempts to classify the A pattern as a J, suffers a mismatch reset, and then classifies it as a T pattern. There is no recoding of either the J or T LTM due to the mismatch arousal level. The d pattern becomes contrast-enhanced to a 0-1 distribution shortly after F_2 becomes active and provides its $\mathbf{Z}\cdot\mathbf{x}_2$ feedback to the v_1 layer of F_1. The same thing happens to the X pattern. This is revealed by the total excitation patterns for $\mathbf{x}_i$ and $\mathbf{x}_k$ in the topmost part of the figure. No *new* non-zero weights were produced in either **W** or **Z** during this process, i.e. the LTM encoding was stable. This run illustrates that there is indeed no recoding of subset patterns by superset patterns in this network. Contrast enhancement in F_1 is one mechanism for this, and mismatch arousal is the other protection mechanism.

§ 5. Summary

In this chapter we have introduced the basic concepts of the ART network and demonstrated them by means of an example network. Except for the use of the L_2 norm in the mismatch detect function and the introduction of embedding field dipoles, all the components of this example network are those developed in the previous chapters. However, it will not have escaped the notice of the reader that the simulations took a somewhat large number of time steps. Since the actions of classification and adaptation occur only in the resonant state, considerations of computational cost reduction push us to ask: Is there no more efficient way to implement an ART

network?

The answer to this question, as you might guess, is "yes." This is done by making further abstractions, not on the putative biological organization of the system but upon what is to be regarded as the "key" or "essential" outcomes of all these calculations. In other words, the next step in ART network modeling is ***algorithm development*** for more efficiently approximating the behaviors of networks such as the one demonstrated here. This brings us to the realm of the published ART network algorithms. We will take this up in more detail in the next chapter, where we will examine what is probably the most famous published ART network algorithm: ART 2.