Adaptive Resonance Theory

Gail A. Carpenter and Stephen Grossberg[1]
Department of Cognitive and Neural Systems
Center for Adaptive Systems
and
Center of Excellence for Learning in Education, Science, and Technology
Boston University
677 Beacon Street, Boston, MA 02215, USA

May, 2009

Technical Report CAS/CNS-TR-2009-008

Encyclopedia of Machine Learning
C. Sammut & G. Webb, Eds.
Invited chapter

Adaptive Resonance Theory

Definition

Adaptive Resonance Theory, or ART, is both a cognitive and neural theory of how the brain quickly learns to categorize, recognize, and predict objects and events in a changing world, and a set of algorithms which computationally embody ART principles and are used in large-scale engineering and technological applications where fast, stable, incremental, learning about complex changing environments is needed. ART clarifies the brain processes from which conscious experiences emerge. It predicts a functional link between processes of Consciousness, Learning, Expectation, Attention, Resonance, and Synchrony (CLEARS), including the prediction that "all conscious states are resonant states." This connection clarifies how brain dynamics enable a behaving individual to autonomously adapt in real time to a rapidly changing world. ART predicts how top-down attention works and regulates fast stable learning of recognition categories. In particular, ART articulates a critical role for "resonant" states in driving fast stable learning; hence the name *adaptive resonance*. These resonant states are bound together, using top-down attentive feedback in the form of learned expectations, into coherent representations of the world. ART hereby clarifies one important sense in which the brain carries out predictive computation. ART has explained and successfully predicted a wide range of behavioral and neurobiological data, including data about human cognition and the dynamics of spiking laminar cortical networks. ART algorithms have been used in large-scale applications such as medical data base prediction, remote sensing, airplane design, and the control of autonomous adaptive robots.

Motivation and Background

Many current learning algorithms do not emulate the way in which humans and other animals learn. The power of human and animal learning provides high motivation to discover computational principles whereby machines can learn with similar capabilities. Humans and animals experience the world on the fly, and carry out incremental learning of sequences of episodes in real time. Often such learning is unsupervised, with the world itself as the teacher. Learning can also proceed with an unpredictable mixture of unsupervised and supervised learning trials. Such learning goes on successfully in a world that is non-stationary; that is, whose rules can change unpredictably through time. Moreover, humans and animals can learn quickly and stably through time. A single important experience can be remembered for a long time. ART proposes a solution of this *stability-plasticity dilemma* [1] by showing how brains learn quickly without forcing catastrophic forgetting of already learned, and still successful, memories.

Thus, ART autonomously carries out fast, yet stable, incremental learning under both unsupervised and supervised learning conditions in response to a complex non-stationary world. In contrast, many current learning algorithms are use batch learning in which all the information

about the world to be learned is available at a single time. Other algorithms are not defined unless all learning trials are unsupervised. Yet other algorithms become unstable in a non-stationary world, or become unstable if learning is fast; that is, if an event can be fully learned on a single learning trial. ART overcomes these problems.

Some machine learning algorithms are feed-forward clustering algorithms that undergo catastrophic forgetting in a non-stationary world. The ART solution of the stability-plasticity dilemma depends upon feedback, or top-down, expectations that are matched against bottom-up data and thereby focus attention upon critical feature patterns. A good enough match leads to resonance and fast learning. A big enough mismatch leads to hypothesis testing or memory search that discovers and learns a more predictive category. Thus, ART is a self-organizing expert system that avoids the brittleness of traditional expert systems.

The world is filled with uncertainty, so probability concepts seem relevant to understanding how brains learn about uncertain data. This fact has led some machine learning practitioners to assume that brains obeys Bayesian laws. However, the Bayes rule is so general that it can accommodate any system in Nature. Additional computational principles and mechanisms must augment Bayes to distinguish a brain from, say, a hydrogen atom or storm. Moreover, probabilistic models often use non-local computations. ART shows how the brain embodies a novel kind of real-time probability theory, hypothesis testing, prediction, and decision-making whose local computations adapt to a non-stationary world. These ART principles and mechanisms go beyond Bayesian analysis, and are embodied parsimoniously in the laminar circuits of cerebral cortex. Indeed, the cortex embodies a new kind of Laminar Computing that reconciles the best properties of feedforward and feedback processing, digital and analog processing, and data-driven bottom-up processing and hypothesis-driven top-down processing [2].

Structure of Learning System

How CLEARS Mechanisms Interact

Humans are *intentional* beings who learn expectations about the world and make predictions about what is about to happen. Humans are also *attentional* beings who focus processing resources upon a restricted amount of incoming information at any time. Why are we both intentional and attentional beings, and are these two types of processes related? The stability-plasticity dilemma and its solution using resonant states provides a unifying framework for understanding these issues.

To clarify the role of sensory or cognitive expectations, and of how a resonant state is activated, suppose you were asked to "find the yellow ball as quickly as possible, and you will win a $10,000 prize". Activating an expectation of a "yellow ball" enables its more rapid detection, and with a more energetic neural response. Sensory and cognitive top-down expectations hereby lead to *excitatory matching* with consistent bottom-up data. Mismatch between top-down expectations and bottom-up data can suppress the mismatched part of the bottom-up data, to focus attention upon the matched, or expected, part of the bottom-up data.

Excitatory matching and attentional focusing on bottom-up data using top-down expectations generates resonant brain states: When there is a good enough match between bottom-up and top-down signal patterns between two or more levels of processing, their positive feedback signals amplify and prolong their mutual activation, leading to a resonant state. Amplification and prolongation of activity triggers learning in the more slowly varying adaptive weights that control the signal flow along pathways from cell to cell. Resonance hereby provides a global context-sensitive indicator that the system is processing data worthy of learning, hence the name *Adaptive* Resonance Theory.

In summary, ART predicts a link between the mechanisms which enable us to learn quickly and stably about a changing world, and the mechanisms that enable us to learn expectations about such a world, test hypotheses about it, and focus attention upon information that we find interesting. ART clarifies this link by asserting that, in order to solve the stability-plasticity dilemma, only resonant states can drive rapid new learning.

It is just a step from here to propose that those experiences which can attract our attention and guide our future lives by being learned are also among the ones that are conscious. Support for this additional assertion derives from the many modeling studies whose simulations of behavioral and brain data using resonant states map onto properties of conscious experiences in those experiments.

The type of learning within the sensory and cognitive domain that ART mechanizes is *match learning*: Match learning occurs only if a good enough match occurs between bottom-up information and a learned top-down expectation that is read out by an active recognition category, or code. When such an approximate match occurs, previously learned knowledge can be refined. Match learning raises the concern about what happens if a match is not good enough? How does such a model escape perseveration on already learned representations?

If novel information cannot form a good enough match with the expectations that are read-out by previously learned recognition categories, then a memory search, or hypothesis testing, is triggered that leads to selection and learning of a new recognition category, rather than catastrophic forgetting of an old one. Figure 1 illustrates how this happens in an ART model; it will be discussed in greater detail below. In contrast, learning within spatial and motor processes is proposed to be *mismatch learning* that continuously updates sensory-motor maps or the gains of sensory-motor commands. As a result, we can stably learn what is happening in a changing world, thereby solving the stability-plasticity dilemma, while adaptively updating our representations of where objects are and how to act upon them using bodies whose parameters change continuously through time. Brain systems that use inhibitory matching and mismatch learning cannot generate resonances; hence, their representations are not conscious.
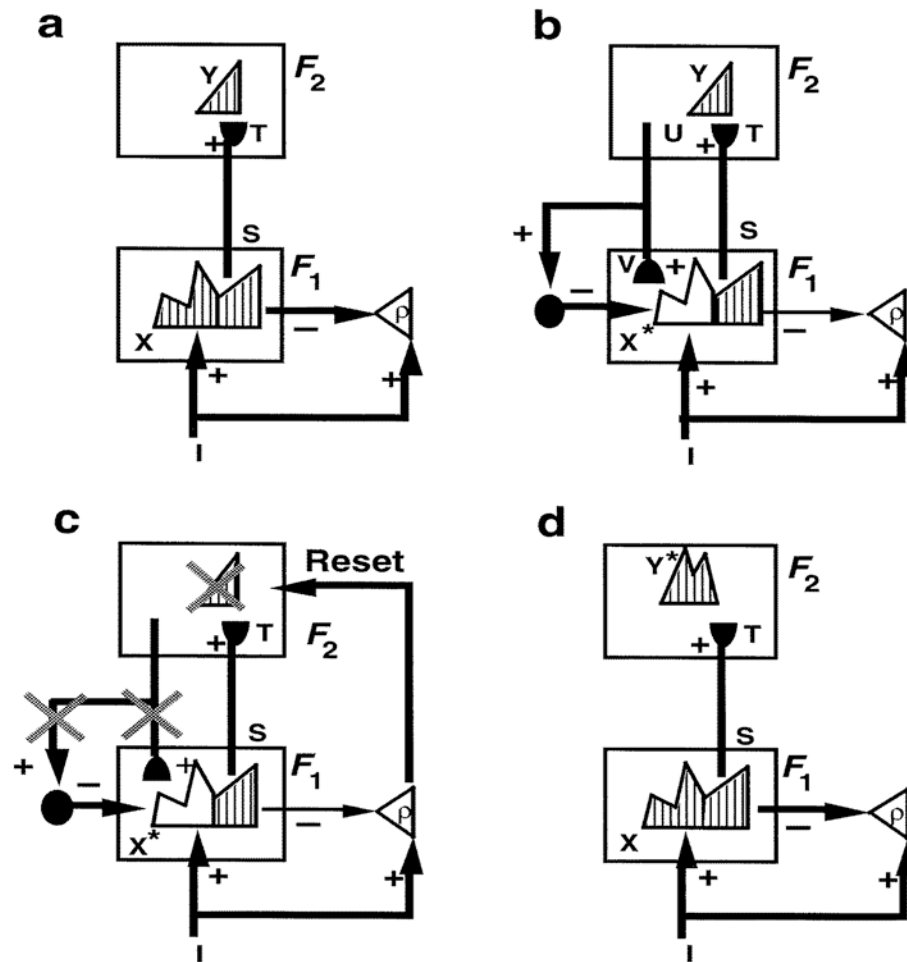
Figure 1. Search for a recognition code within an ART learning circuit: (a) The input pattern I is instated across the feature detectors at level F_1 as a short term memory (STM) activity pattern X. Input I also nonspecifically activates the orienting system with a gain that is called vigilance (ρ); that is, all the input pathways converge with gain ρ onto the orienting system and try to activate it. STM pattern X is represented by the hatched pattern across F_1. Pattern X both inhibits the orienting system and generates the output pattern S. Pattern S is multiplied by learned adaptive weights, also called long term memory (LTM) traces. These LTM-gated signals are added at F_2 cells, or nodes, to form the input pattern T, which activates the STM pattern Y across the recognition categories coded at level F_2. (b) Pattern Y generates the top-down output pattern U which is multiplied by top-down LTM traces and added at F_1 nodes to form a *prototype* pattern V that encodes the learned expectation of the active F_2 nodes. Such a prototype represents the set of commonly shared features in all the input patterns capable of activating Y. If V mismatches I at F_1, then a new STM activity pattern X^* is selected *at F_1*. X^* is represented by the hatched pattern. It consists of the features of I that are confirmed by V. Mismatched features are inhibited. The inactivated nodes corresponding to unconfirmed features of X are unmatched. The reduction in total STM activity which occurs when X is transformed into X^* causes a decrease in the total inhibition from F_1 to the orienting

system. (c) If inhibition decreases sufficiently, the orienting system releases a nonspecific arousal wave to F_2; that is, a wave of activation that equally activates all F_2 nodes. This wave instantiates the intuition that "novel events are arousing". This arousal wave resets the STM pattern Y at F_2 by inhibiting Y. (d) After Y is inhibited, its top-down prototype signal is eliminated, and X can be reinstated at F_1. The prior reset event maintains inhibition of Y during the search cycle. As a result, X can activate a different STM pattern Y at F_2. If the top-down prototype due to this new Y pattern also mismatches I at F_1, then the search for an appropriate F_2 code continues until a more appropriate F_2 representation is selected. Such a search cycle represents a type of nonstationary hypothesis testing. When search ends, an attentive resonance develops and learning of the attended data is initiated. [Adapted with permission from [3].]

Complementary Computing in the Brain: Resonance and Reset

It has been mathematically proved that match learning within an ART model leads to stable memories in response to arbitrary list of events to be learned [4]. However, match learning also has a serious potential weakness: If you can only learn when there is a good enough match between bottom-up data and learned top-down expectations, then how do you ever learn anything that you do not already know? ART proposes that this problem is solved by the brain by using an interaction between complementary processes of *resonance* and *reset,* that are predicted to control properties of attention and memory search, respectively. These complementary processes help our brains to balance between the complementary demands of processing the familiar and the unfamiliar, the expected and the unexpected.

Organization of the brain into complementary processes is predicted to be a general principle of brain design that is not just found in ART [5]. A complementary process can individually compute some properties well, but cannot, by itself, process other complementary properties. In thinking intuitively about complementary properties, one can imagine puzzle pieces fitting together. Both pieces are needed to finish the puzzle. Complementary brain processes are more dynamic than any such analogy: Pairs of complementary processes interact to form emergent properties which overcome their complementary deficiencies to compute complete information with which to represent or control some aspect of intelligent behavior.

The resonance process in the complementary pair of resonance and reset is predicted to take place in the What cortical stream, notably in the inferotemporal and prefrontal cortex. Here top-down expectations are matched against bottom-up inputs. When a top-down expectation achieves a good enough match with bottom-up data, this match process focuses attention upon those feature clusters in the bottom-up input that are expected. If the expectation is close enough to the input pattern, then a state of resonance develops as the attentional focus takes hold.

Figure 1 illustrates these ART ideas in a simple two-level example. Here, a bottom-up input pattern, or vector, I activates a pattern X of activity across the feature detectors of the first level F_1. For example, a visual scene may be represented by the features comprising its boundary and surface representations. This feature pattern represents the relative importance of different

features in the inputs pattern I. In Figure 1a, the pattern peaks represent more activated feature detector cells, the troughs less activated feature detectors. This feature pattern sends signals S through an adaptive filter to the second level F_2 at which a compressed representation Y (also called a recognition category, or a symbol) is activated in response to the distributed input T. Input T is computed by multiplying the signal vector S by a matrix of adaptive weights that can be altered through learning. The representation Y is compressed by competitive interactions across F_2 that allow only a small subset of its most strongly activated cells to remain active in response to T. The pattern Y in the figure indicates that a small number of category cells may be activated to different degrees. These category cells, in turn, send top-down signals U to F_1. The vector U is converted into the top-down expectation V by being multiplied by another matrix of adaptive weights. When V is received by F_1, a matching process takes place between the input vector I and V which selects that subset X* of F_1 features that were "expected" by the active F_2 category Y. The set of these selected features is the emerging "attentional focus".

Binding Distributed Feature Patterns and Symbols during Conscious Resonances

If the top-down expectation is close enough to the bottom-up input pattern, then the pattern X* of attended features reactivates the category Y which, in turn, reactivates X*. The network hereby locks into a resonant state through a positive feedback loop that dynamically links, or binds, the attended features across X* with their category, or symbol, Y.

Resonance itself embodies another type of complementary processing. Indeed, there seem to be complementary processes both within and between cortical processing streams [5]. This particular complementary relation occurs between distributed feature patterns and the compressed categories, or symbols, that selectively code them:

Individual features at F_1 have no meaning on their own, just like the pixels in a picture are meaningless one-by-one. The category, or symbol, in F_2 is sensitive to the global patterning of these features, and can selectively fire in response to this pattern. But it cannot represent the "contents" of the experience, including their conscious qualia, due to the very fact that a category is a compressed, or "symbolic" representation. Practitioners of Artificial Intelligence have claimed that neural models can process distributed features, but not symbolic representations. This is not, of course, true in the brain. Nor is it true in ART.

Resonance between these two types of information converts the *pattern* of attended features into a coherent context-sensitive state that is linked to its category through feedback. This coherent state, which binds together distributed features and symbolic categories, can enter consciousness while it binds together spatially distributed features into either a stable equilibrium or a synchronous oscillation. The original ART article [6] predicted the existence of such synchronous oscillations, which were there described in terms of their mathematical properties as "order-preserving limit cycles". See [7] and [8] for reviews of confirmed ART predictions, including predictions about synchronous oscillations.

Resonance Links Intentional and Attentional Information Processing to Learning

In ART, the resonant state, rather than bottom-up activation, is predicted to drive learning. The resonant state persists long enough, and at a high enough activity level, to activate the slower learning processes in the adaptive weights that guide the flow of signals between bottom-up and top-down pathways between levels F_1 and F_2 in Figure 1. This viewpoint helps to explain how adaptive weights that were changed through previous learning can regulate the brain's present information processing, without learning about the signals that they are currently processing unless they can initiate a resonant state. Through resonance as a mediating event, one can understand from a deeper mechanistic view why humans are intentional beings who are continually predicting what may next occur, and why we tend to learn about the events to which we pay attention.

More recent versions of ART, notably the Synchronous Matching ART (SMART) model [8] show how a match may lead to fast gamma oscillations that facilitate spike-timing dependent plasticity (STDP), whereas mismatch can lead to slower beta oscillations that lower the probability that mismatched events can be learned by a STDP learning law.

Complementary Attentional and Orienting Systems Control Resonance vs. Reset

A sufficiently bad mismatch between an active top-down expectation and a bottom-up input, say because the input represents an unfamiliar type of experience, can drive a memory search. Such a mismatch within the attentional system is proposed to activate a complementary *orienting system*, which is sensitive to unexpected and unfamiliar events. ART suggests that this orienting system includes the nonspecific thalamus and the hippocampal system. See [8] for a summary of data supporting this prediction. Output signals from the orienting system rapidly reset the recognition category that has been reading out the poorly matching top-down expectation (Figures 1b and 1c). The cause of the mismatch is hereby removed, thereby freeing the system to activate a different recognition category (Figure 1d). The reset event hereby triggers memory search, or hypothesis testing, which automatically leads to the selection of a recognition category that can better match the input.

If no such recognition category exists, say because the bottom-up input represents a truly novel experience, then the search process automatically activates an as yet uncommitted population of cells, with which to learn about the novel information. In order for a top-down expectation to match a newly discovered recognition category, its top-down adaptive weights initially have large values, which are pruned by the learning of a particular expectation.

This learning process works well under both unsupervised and supervised conditions [9]. Unsupervised learning means that the system can learn how to categorize novel input patterns without any external feedback. Supervised learning uses predictive errors to let the system know whether it has categorized the information correctly. Supervision can force a search for new categories that may be culturally determined, and are not based on feature similarity alone. For example, separating the featurally similar letters E and F into separate recognition categories is

culturally determined. Such error-based feedback enables variants of E and F to learn their own category and top-down expectation, or prototype. The complementary, but interacting, processes of attentive-learning and orienting-search together realize a type of error correction through hypothesis testing that can build an ever-growing, self-refining internal model of a changing world.

Controlling the Content of Conscious Experiences: Exemplars and Prototypes

What combinations of features or other information are bound together into conscious object or event representations? One view is that exemplars, or individual experiences, are learned, because humans can have very specific memories. For example, we can all recognize the particular faces of our friends. On the other hand, storing every remembered experience as exemplars can lead to a combinatorial explosion of memory, as well as to unmanageable problems of memory retrieval. A possible way out is suggested by the fact that humans can learn prototypes which represent general properties of the environment [10]. For example, we can recognize that everyone has a face. But then how do we learn specific episodic memories? ART provides an answer to this question that overcomes problems faced by earlier models.

ART prototypes are not merely averages of the exemplars that are classified by a category, as is typically assumed in classical prototype models. Rather, they are the actively selected *critical feature patterns* upon which the top-down expectations of the category focus attention. In addition, the generality of the information that is codes by these critical feature patterns is controlled by a gain control process, called *vigilance* control, which can be influenced by environmental feedback or internal volition [11]. Low vigilance permits the learning of general categories with abstract prototypes. High vigilance forces a memory search to occur for a new category when even small mismatches exist between an exemplar and the category that it activates. As a result, in the limit of high vigilance, the category prototype may encode an individual exemplar.

Vigilance is computed within the orienting system of an ART model (Figures 1b-d). It is here that bottom-up excitation from all the active features in an input pattern I are compared with inhibition from all the active features in a distributed feature representation across F_1. If the ratio of the total activity across the active features in F_1 (that is, the "matched" features) to the total activity due to all the features in I is less than a *vigilance parameter* ρ (Figure 1b), then a reset wave is activated (Figure 1c), which can drive the search for another category with which to classify the exemplar. In other words, the vigilance parameter controls how bad a match can be before search for a new category is initiated. If the vigilance parameter is low, then many exemplars can all influence the learning of a shared prototype, by chipping away at the features that are not shared with all the exemplars. If the vigilance parameter is high, then even a small difference between a new exemplar and a known prototype (e.g., F vs. E) can drive the search for a new category with which to represent F.

One way to control vigilance is by a process of *match tracking*. Here a predictive error (e.g., D is predicted in response to F), the vigilance parameter increases until it is just higher than the ratio of active features in F_1 to total features in I. In other words, vigilance "tracks" the degree of

match between input exemplar and matched prototype. This is the minimal level of vigilance that can trigger a reset wave and thus a memory search for a new category. Match tracking realizes a Minimax Learning Rule that conjointly *maximizes* category generality while it *minimizes* predictive error. In other words, match tracking uses the least memory resources that can prevent errors from being made.

Because vigilance can vary across learning trials, recognition categories capable of encoding widely differing degrees of generalization or abstraction can be learned by a single ART system. Low vigilance leads to broad generalization and abstract prototypes. High vigilance leads to narrow generalization and to prototypes that represent fewer input exemplars, even a single exemplar. Thus a single ART system may be used, say, to learn abstract prototypes with which to recognize abstract categories of faces and dogs, as well as "exemplar prototypes" with which to recognize individual views of faces and dogs. ART models hereby try to learn the most general category that is consistent with the data. This tendency can, for example, lead to the type of overgeneralization that is seen in young children until further learning leads to category refinement.

Memory Consolidation and the Emergence of Rules: Direct Access to Globally Best Match

As sequences of inputs are practiced over learning trials, the search process eventually converges upon stable categories. It has been mathematically proved [11] that familiar inputs directly access the category whose prototype provides the globally best match, while unfamiliar inputs engage the orienting subsystem to trigger memory searches for better categories until they become familiar. This process continues until the memory capacity, which can be chosen arbitrarily large, is fully utilized. The process whereby search is automatically disengaged is a form of *memory consolidation* that emerges from network interactions. Emergent consolidation does not preclude structural consolidation at individual cells, since the amplified and prolonged activities that subserve a resonance may be a trigger for learning-dependent cellular processes, such as protein synthesis and transmitter production.

It has also been shown that the adaptive weights which are learned by some ART models can, at any stage of learning, be translated into fuzzy IF-THEN rules [9]. Thus the ART model is a self-organizing rule-discovering production system as well as a neural network. These examples show that the claims of some cognitive scientists and AI practitioners that neural network models cannot learn rule-based behaviors are as incorrect as the claims that neural models cannot learn symbols.

How the Laminar Circuits of Cerebral Cortex Embody ART Mechanisms

More recent versions of ART have shown how predicted ART mechanisms may be embodied within known laminar microcircuits of the cerebral cortex. These include the family of LAMINART models (Figure 2; see [7]) and the Synchronous Matching ART, or SMART, model (Figure 3, see [8]). SMART, in particular, predicts how a top-down match may lead to fast

gamma oscillations that facilitate spike-timing dependent plasticity (STDP), whereas a mismatch can lead to slower beta oscillations that lower the probability that mismatched events can be learned by a STDP learning law. At least three neurophysiological labs have recently reported data consistent with the SMART prediction.

Figure 2. LAMINART circuit clarifies how known cortical connections within and across cortical layers join the layer 6 → 4 and layer 2/3 circuits to form a laminar circuit model for the interblobs and pale stripe regions of cortical areas V1 and V2. Inhibitory interneurons are shown filled-in black. (a) The LGN provides bottom-up activation to

layer 4 via two routes. First, it makes a strong connection directly into layer 4. Second, LGN axons send collaterals into layer 6, and thereby also activate layer 4 via the 6 → 4 on-center off-surround path. The combined effect of the bottom-up LGN pathways is to stimulate layer 4 via an on-center off-surround, which provides divisive contrast normalization [12][13][14] of layer 4 cell responses. (b) Folded feedback carries attentional signals from higher cortex into layer 4 of V1, via the modulatory 6 → 4 path. Corticocortical feedback axons tend preferentially to originate in layer 6 of the higher area and to terminate in layer 1 of the lower cortex [15], where they can excite the apical dendrites of layer 5 pyramidal cells whose axons send collaterals into layer 6. The triangle in the figure represents such a layer 5 pyramidal cell. Several other routes through which feedback can pass into V1 layer 6 exist (see [16] for a review). Having arrived in layer 6, the feedback is then "folded" back up into the feedforward stream by passing through the 6 → 4 on-center off-surround path [17]. (c) Connecting the 6 → 4 on-center off-surround to the layer 2/3 grouping circuit: like-oriented layer 4 simple cells with opposite contrast polarities compete (not shown) before generating half-wave rectified outputs that converge onto layer 2/3 complex cells in the column above them. Just like attentional signals from higher cortex, as shown in (b), groupings that form within layer 2/3 also send activation into the folded feedback path, to enhance their own positions in layer 4 beneath them via the 6 → 4 on-center, and to suppress input to other groupings via the 6 → 4 off-surround. There exist direct layer 2/3 → 6 connections in macaque V1, as well as indirect routes via layer 5. (d) Top-down corticogeniculate feedback from V1 layer 6 to LGN also has an on-center off-surround anatomy, similar to the 6 → 4 path. The on-center feedback selectively enhances LGN cells that are consistent with the activation that they cause [18], and the off-surround contributes to length-sensitive (endstopped) responses that facilitate grouping perpendicular to line ends. (e) The entire V1/V2 circuit: V2 repeats the laminar pattern of V1 circuitry, but at a larger spatial scale. In particular, the horizontal layer 2/3 connections have a longer range in V2, allowing above-threshold perceptual groupings between more widely spaced inducing stimuli to form [19]. V1 layer 2/3 projects up to V2 layers 6 and 4, just as LGN projects to layers 6 an 4 of V1. Higher cortical areas send feedback into V2 which ultimately reaches layer 6, just as V2 feedback acts on layer 6 of V1 [20]. Feedback paths from higher cortical areas straight into V1 (not shown) can complement and enhance feedback from V2 into V1. Top-down attention can also modulate layer 2/3 pyramidal cells directly by activating both the pyramidal cells and inhibitory interneurons in that layer. The inhibition tends to balance the excitation, leading to a modulatory effect. These top-down attentional pathways tend to synapse in layer 1, as shown in Figure 2b. Their synapses on apical dendrites in layer 1 are not shown, for simplicity. (Reprinted with permission from [16].)
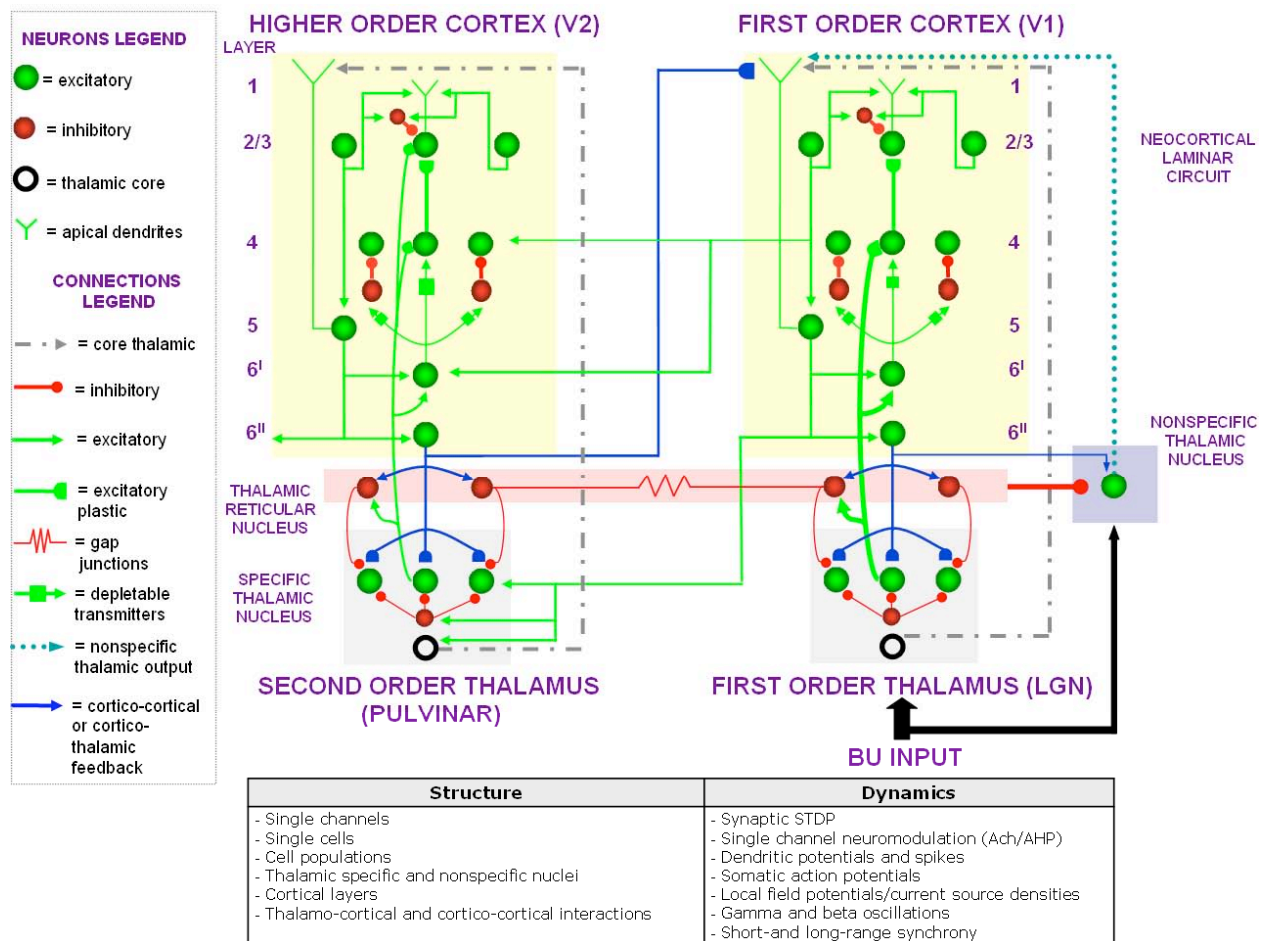
NEURONS LEGEND
- = excitatory
- = inhibitory
- = thalamic core
- = apical dendrites

CONNECTIONS LEGEND
- = core thalamic
- = inhibitory
- = excitatory
- = excitatory plastic
- = gap junctions
- = depletable transmitters
- = nonspecific thalamic output
- = cortico-cortical or cortico-thalamic feedback

HIGHER ORDER CORTEX (V2)
FIRST ORDER CORTEX (V1)
LAYER
NEOCORTICAL LAMINAR CIRCUIT
NONSPECIFIC THALAMIC NUCLEUS
THALAMIC RETICULAR NUCLEUS
SPECIFIC THALAMIC NUCLEUS
SECOND ORDER THALAMUS (PULVINAR)
FIRST ORDER THALAMUS (LGN)
BU INPUT

Structure	Dynamics
- Single channels	- Synaptic STDP
- Single cells	- Single channel neuromodulation (Ach/AHP)
- Cell populations	- Dendritic potentials and spikes
- Thalamic specific and nonspecific nuclei	- Somatic action potentials
- Cortical layers	- Local field potentials/current source densities
- Thalamo-cortical and cortico-cortical interactions	- Gamma and beta oscillations
	- Short-and long-range synchrony

Figure 3. SMART model overview. A first-order and higher-order cortical area are linked by corticocortical and corticothalamocortical connections. The thalamus is subdivided into specific first-order, second-order, nonspecific, and thalamic reticular nucleus (TRN). The thalamic matrix (one cell population shown as an open ring) provides priming to layer 1, where layer 5 pyramidal cell apical dendrites terminate. The specific thalamus relays sensory information (first-order thalamus) or lower-order cortical information (second-order thalamus) to the respective cortical areas via plastic connections. The nonspecific thalamic nucleus receives convergent BU input and inhibition from the TRN, and projects to layer 1 of the laminar cortical circuit, where it regulates reset and search in the cortical circuit (see text). Corticocortical feedback connections link layer 6^{II} of the higher cortical area to layer 1 of the lower cortical area, whereas thalamocortical feedback originates in layer 6^{II} and terminates in the specific thalamus after synapsing on the TRN. Layer 6^{II} corticothalamic feedback matches the BU input in the specific thalamus. V1 receives two parallel BU thalamocortical pathways. The LGN→V1 layer 4 pathway and the modulatory LGN→V1 layer 6^{I}→4 pathway provide divisive contrast normalization of layer 4 cell responses. The

intracortical loop V1 layer $4 \rightarrow 2/3 \rightarrow 5 \rightarrow 6^I \rightarrow 4$ pathway (*folded feedback)* enhances the activity of winning layer 2/3 cells at their own positions via the $6^I \rightarrow 4$ on-center, and suppresses input to other layer 2/3 cells via the $6^I \rightarrow 4$ off-surround. V1 also activates the BU V1→V2 corticocortical pathways (V1 layer 2/3→V2 layers 6^I and 4) and the BU corticothalamocortical pathways (V1 layer $5 \rightarrow$ PULV→V2 layers 6^I and 4), where the layer $6^I \rightarrow 4$ pathway provides divisive contrast normalization to V2 layer 4 cells analogously to V1. Corticocortical feedback from V2 layer $6^{II} \rightarrow$ V1 layer $5 \rightarrow 6^I \rightarrow 4$ also uses the same modulatory $6^I \rightarrow 4$ pathway. TRN cells of the two thalamic sectors are linked via gap junctions, which provide synchronization of the two thalamocortical sectors when processing BU stimuli. [Reprinted with permission from [8].]

Review of ART and ARTMAP Algorithms

From Winner-Take-All to Distributed Coding

As noted above, ART networks serve both as models of human cognitive information processing [21][22][23] and as neural systems for technology transfer [24][25][26]. Design principles derived from scientific analyses and design constraints imposed by targeted applications have jointly guided the development of many variants of the basic networks, including fuzzy ARTMAP [9], ART-EMAP [27], ARTMAP-IC [28], and Gaussian ARTMAP [29]. Early ARTMAP systems, including fuzzy ARTMAP, employ *winner-take-all (WTA) coding*, whereby each input activates a single category node during both training and testing. When a node is first activated during training, it is mapped to its designated output class.

Starting with ART-EMAP, subsequent systems have used *distributed coding* during testing, which typically improves predictive accuracy, while avoiding the computational problems inherent in the use of distributed code representations during training. In order to address these problems, distributed ARTMAP [30][31] introduced a new network configuration, in addition to new learning laws.

Comparative analysis of the performance of ARTMAP systems on a variety of benchmark problems has led to the identification of a *default ARTMAP* network [32], which features simplicity of design and robust performance in many application domains. Default ARTMAP employs winner-take-all coding during training and distributed coding during testing within a distributed ARTMAP network architecture. With winner-take-all coding during testing, default ARTMAP reduces to a version of fuzzy ARTMAP.
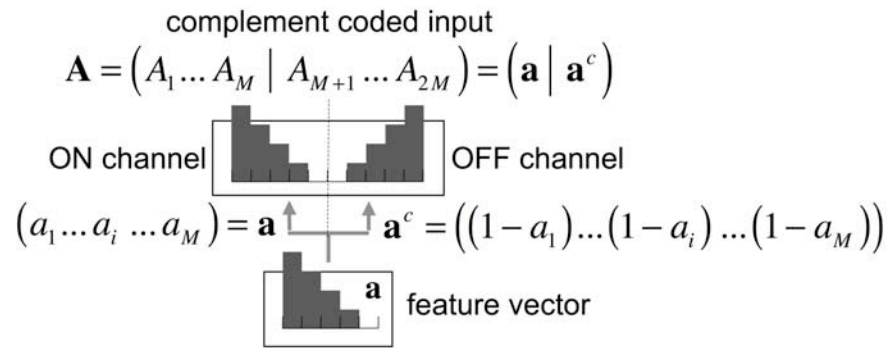
Figure 4. Complement coding transforms an *M*-dimensional feature vector **a** into a 2*M*-dimensional system input vector **A**. A complement-coded system input represents both the degree to which a feature *i* is present (a_i) and the degree to which that feature is absent $(1-a_i)$.

Complement coding: Learning both absent and present features

ART and ARTMAP employ a preprocessing step called *complement coding* (Figure 4), which models the nervous system's ubiquitous use of the computational design known as *opponent processing* [33][34]. Balancing an entity against its opponent, as in agonist-antagonist muscle pairs, allows a system to act upon relative quantities, even as absolute magnitudes may vary unpredictably. In ART systems, complement coding [35] is analogous to retinal ON-cells and OFF-cells [36]. When the learning system is presented with a set of input features $\mathbf{a} \equiv (a_1...a_i...a_M)$, complement coding doubles the number of input components, presenting to the network both the original feature vector and its complement.

Complement coding allows an ART system to encode within its critical feature patterns of memory features that are consistently *absent* on an equal basis with features that are consistently *present*. Features that are sometimes absent and sometimes present when a given category is learning become regarded as uninformative with respect to that category. Since its introduction, complement coding has been a standard element of ART and ARTMAP networks, where it plays multiple computational roles, including input normalization. However, this device is not particular to ART, and could, in principle, be used to preprocess the inputs to any type of system.

To implement complement coding, component activities a_i of a feature vector **a** are scaled so that $0 \le a_i \le 1$. For each feature *i*, the ON activity a_i determines the complementary OFF activity $(1-a_i)$. Both a_i and $(1-a_i)$ are represented in the 2*M*-dimensional system input vector $\mathbf{A} = (\mathbf{a} \mid \mathbf{a}^c)$ (Figure 4). Subsequent network computations then operate in this 2*M*-dimensional input space. In particular, learned weight vectors $\mathbf{w}_J$ are 2*M*-dimensional.

ARTMAP Search and Match Tracking in fuzzy ARTMAP

As illustrated by Figure 1, the ART matching process triggers either learning or a parallel memory search. If search ends at an established code, the memory representation may either remain the same or incorporate new information from matched portions of the current input. While this dynamic applies to arbitrarily distributed activation patterns, the F_2 search and code for fuzzy ARTMAP (Figure 5) describes a winner-take all system.
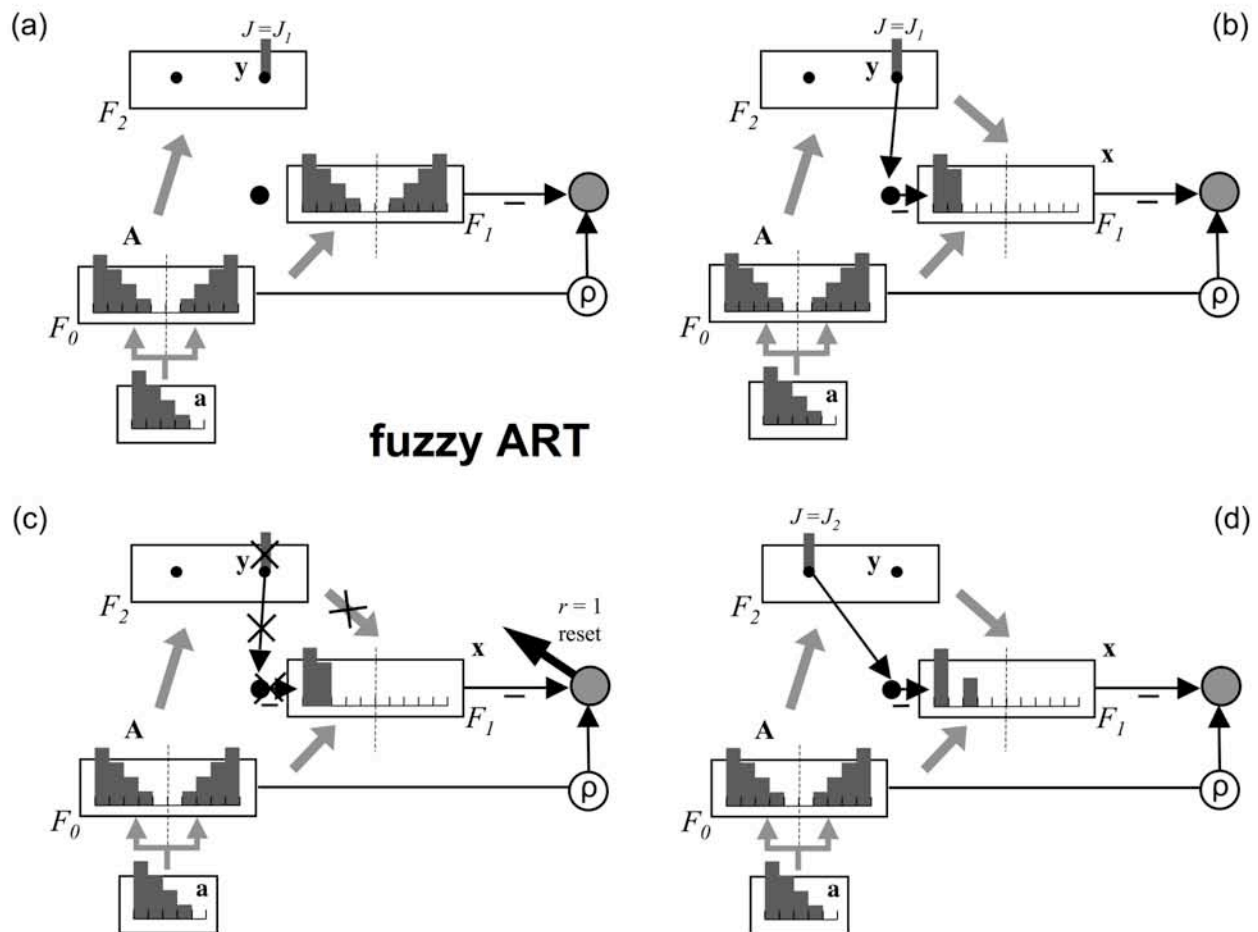

Figure 5. A fuzzy ART search cycle [35], with a distributed ART network configuration [23]. The ART 1 search cycle [11] is the same, but allows only binary inputs and did not originally feature complement coding. The match field F_1 represents the matched activation pattern $\mathbf{x} = \mathbf{A} \wedge \mathbf{w}_J$, where $\wedge$ denotes the component-wise minimum, or fuzzy intersection, between the bottom-up input $\mathbf{A}$ and the top-down expectation $\mathbf{w}_J$. If the matched pattern fails to meet the matching criterion, then the active code is reset at F_2, and the system searches for another code $\mathbf{y}$ that better represents the input. The match / mismatch decision in the ART orienting system. Each active feature in the input pattern

$\mathbf{A}$ excites the orienting system with gain equal to the vigilance parameter ρ. Hence, with complement coding, the total excitatory input is $\rho|\mathbf{A}| = \rho \sum_{i=1}^{2M} A_i = \rho M$. Active cells in the matched pattern $\mathbf{x}$ inhibit the orienting system, leading to a total inhibitory input equal to $-|\mathbf{x}| = -\sum_{i=1}^{2M} x_i$. If $\rho|\mathbf{A}| - |\mathbf{x}| \leq 0$, then the orienting system remains quiet, allowing resonance and learning to occur. If $\rho|\mathbf{A}| - |\mathbf{x}| > 0$, then the reset signal $r=1$, initiating search for a better matching code.

Before ARTMAP makes a class prediction, the bottom-up input $\mathbf{A}$ is matched against the top-down learned expectation, or critical feature pattern, that is read out by the active node (Figure 5b). The matching criterion is set by a *vigilance* parameter ρ. As noted above, low vigilance permits the learning of abstract, prototype-like patterns, while high vigilance requires the learning of specific, exemplar-like patterns. When a new input arrives, vigilance equals a baseline level, $\bar{\rho}$. Baseline vigilance is set equal to zero by default, in order to maximize generalization. Vigilance rises only after the system has made a predictive error. The internal control process that determines how far it must rise in order to correct the error is called *match tracking* [37]. As vigilance rises, the network is required to pay more attention to how well top-down expectations match the current bottom-up input.

Match tracking (Figure 6) forces an ARTMAP system not only to reset its mistakes, but to learn from them. With match tracking and fast learning, each ARTMAP network passes the Next Input Test, which requires that, if a training input were re-presented immediately after a learning trial, it would directly activate the correct output class, with no predictive errors or search. Match tracking thus simultaneously implements the design goals of maximizing generalization and minimizing predictive error, without requiring the choice of a fixed matching criterion. ARTMAP memories thereby include both broad and specific pattern classes, with the latter typically formed as exceptions to the more general "rules" defined by the former. ARTMAP learning typically produces a wide variety of such mixtures, whose exact composition depends upon the order of training exemplar presentation.

Unless they have already activated all their coding nodes, ARTMAP systems contain a reserve of nodes that have never been activated, with weights at their initial values. These *uncommitted* nodes compete with the previously active *committed* nodes, and an uncommitted node will be chosen over poorly matched committed nodes. An ARTMAP design constraint specifies that an active uncommitted node should not reset itself. Weights initially begin with $w_{iJ} = 1$. Thus, when the active node J is uncommitted, $\mathbf{x} = \mathbf{A} \wedge \mathbf{w}_J = \mathbf{A}$ at the match field. Then, $\rho|\mathbf{A}| - |\mathbf{x}| = \rho|\mathbf{A}| - |\mathbf{A}| = (\rho - 1)|\mathbf{A}|$. Thus $\rho|\mathbf{A}| - |\mathbf{x}| \leq 0$ and an uncommitted node does not trigger a reset, provided that $\rho \leq 1$.
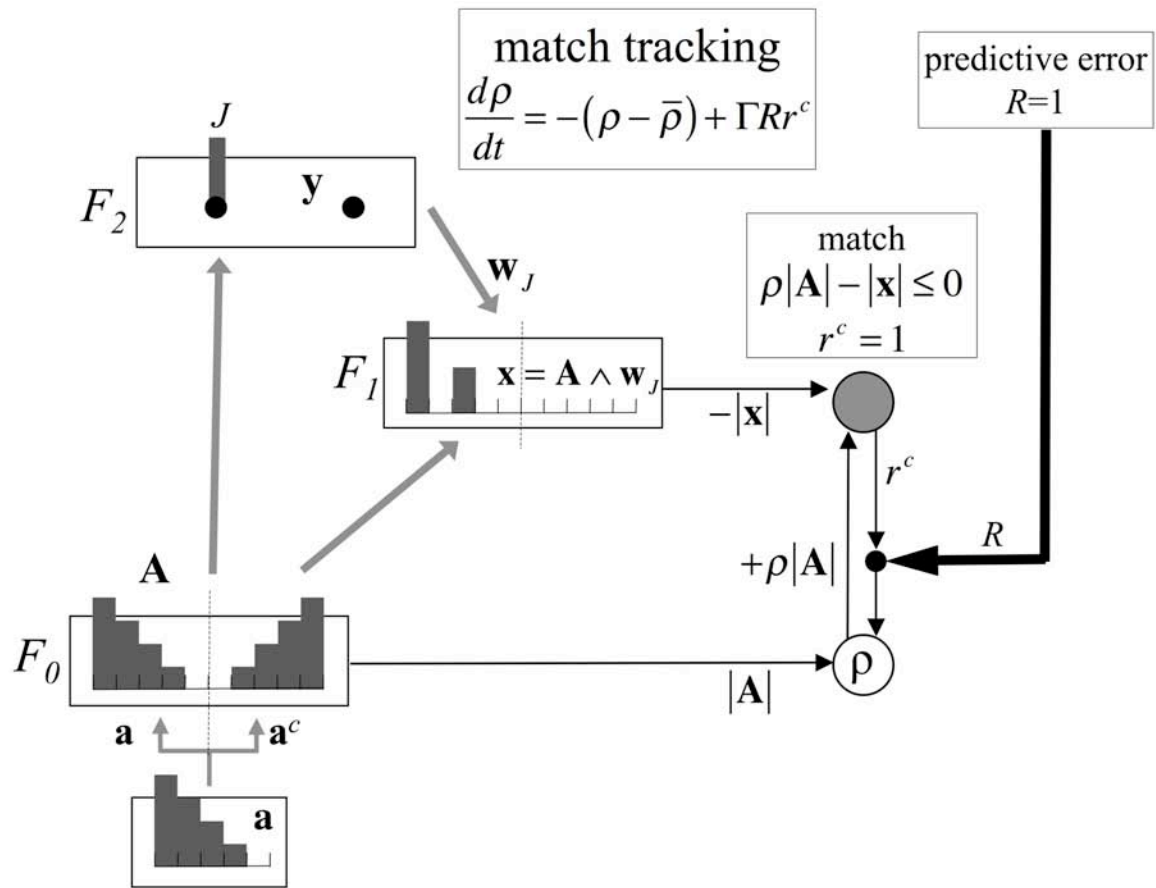
Figure 6.   ARTMAP match tracking [37]. When an active node J meets the matching criterion $\left(\rho|\mathbf{A}| - |\mathbf{x}| \leq 0\right)$, the reset signal r=0 and the node makes an prediction. If the predicted output is incorrect, the feedback signal R=1. While $R = r^c = 1$, r increases rapidly. As soon as $\rho > \dfrac{|\mathbf{x}|}{|\mathbf{A}|}$, r switches to 1, which both halts the increase of r and resets the active F_2 node. From one chosen node to the next, r decays to slightly below $\dfrac{|\mathbf{x}|}{|\mathbf{A}|}$ (MT–: [28]. On the time scale of learning r returns to $\bar{\rho}$.
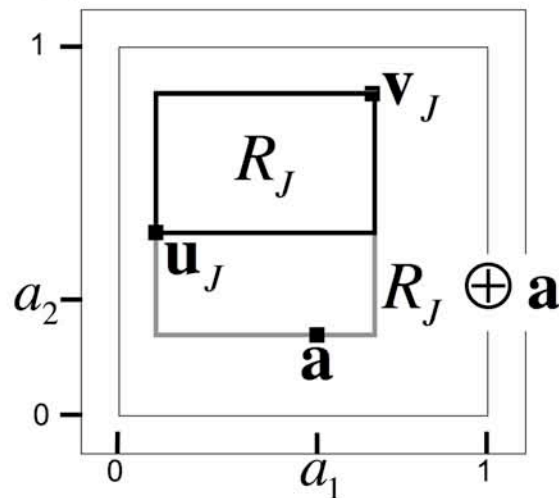
Figure 7. Fuzzy ART geometry. The weight of a category node J is represented in complement-coding form as $\mathbf{w}_J = \left(\mathbf{u}_J \mid \mathbf{v}_J^C\right)$, and the M-dimensional vectors $\mathbf{u}_J$ and $\mathbf{v}_J$ define the corners of the category box R_J. When $M=2$, the size of R_J equals its width plus its height. During learning, R_J expands toward $R_J \oplus \mathbf{a}$, defined as the smallest box enclosing both R_J and $\mathbf{a}$. Node J will reset before learning if $|R_J \oplus \mathbf{a}| > M\left(1 - \rho\right)$.

ART Geometry

Fuzzy ART long-term memories are visualized as hyper-rectangles, called *category boxes*. The weight vector $\mathbf{w}_J$ is interpreted geometrically as a box R_J whose ON-channel corner $\mathbf{u}_J$ and OFF-channel corner $\mathbf{v}_J$ are, in the format of the complement-coded input vector, defined by $\left(\mathbf{u}_J \mid \mathbf{v}_J^C\right) \equiv \mathbf{w}_J$ (Figure 7). For fuzzy ART with the choice-by-difference $F_0 \rightarrow F_2$ signal function T_J [38], an input $\mathbf{a}$ activates the node J of the closest category box R_J, according to the L$_1$ (city-block) metric. In case of a tie, as when $\mathbf{a}$ lies in more than one box, the node with the smallest R_J is chosen, where $|R_J|$ is defined as the sum of the edge lengths $\sum_{i=1}^{M} |v_{iJ} - u_{iJ}|$. The chosen node J will reset if $|R_J \oplus \mathbf{a}| > M\left(1 - \rho\right)$, where $R_J \oplus \mathbf{a}$ is the smallest box enclosing both R_J and $\mathbf{a}$. Otherwise, R_J expands toward $R_J \oplus \mathbf{a}$ during learning. With fast learning, $R_J^{new} = R_J^{old} \oplus \mathbf{a}$.

Biasing Against Previously Active Category Nodes and Previously Attended Features during Attentive Memory Search

Activity $\mathbf{x}$ at the ART field F_1 continuously computes the match between the field's bottom-up and top-down input patterns. A reset signal r shuts off the active F_2 node J when $\mathbf{x}$ fails to meet the matching criterion determined by the value of the vigilance parameter r. Reset alone does not, however, trigger a search for a different F_2 node: unless the prior activation has left an enduring trace within the F_0-to-F_2 subsystem, the network will simply reactivate the same node as before. As modeled in ART 3 [39], biasing the bottom-up input to the coding field F_2 to favor previously inactive nodes implements search by allowing the network to activate a new node in response to a reset signal. The ART 3 search mechanism defines a medium-term memory (MTM) in the F_0-to-F_2 adaptive filter which biases the system against re-choosing a node that had just produced a reset. A presynaptic interpretation of this bias is transmitter depletion, or habituation (Figure 8).

Medium-term memory in all ART models allows the network to shift attention among learned categories at the coding field F_2 during search. The new biased ART network [40] introduces a second medium-term memory that shifts attention among input features, as well as categories, during search.
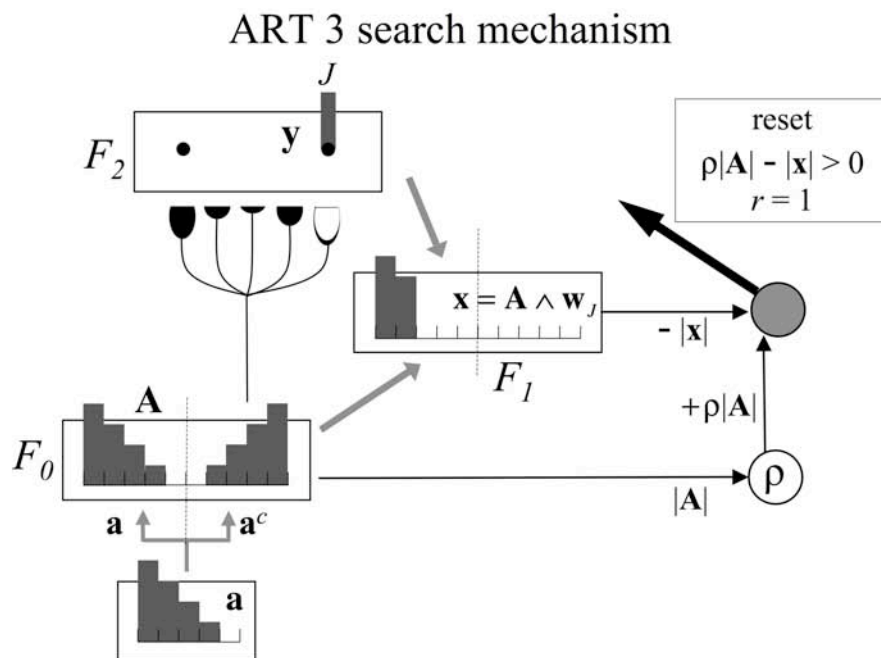

Figure 8. ART 3 search implements a medium-term memory within the F_0-to-F_2 pathways, which biases the system against choosing a category node that had just produced a reset.

Self-Organizing Rule Discovery

This foundation of computational principles and mechanisms has enabled the development of an ART information fusion system that is capable of incrementally learning a cognitive hierarchy of rules in response to probabilistic, incomplete, and even contradictory data that are collected by multiple observers [41][42].

References

[1]   Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review. 87*, 1-51.

[2]   Grossberg, S. (2007). Consciousness CLEARS the mind. *Neural Networks, 20*, 1040-1053.

[3]   Carpenter, G.A. & Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences, 16*, 131-137.

[4]   Carpenter, G.A. & Grossberg, S. (1991). *Pattern recognition by self-organizing neural networks*. Cambridge, MA: MIT Press.

[5]   Grossberg, S. (2000). The complementary brain: Unifying brain dynamics and modularity. T*rends in Cognitive Sciences, 4*, 233-246.

[6]   Grossberg, S. (1976).   Adaptive pattern classification and universal recoding, II: Feedback, expectation, olfaction, and illusions. *Biological Cybernetics, 23*, 187-202.

[7]   Raizada, R.D.S. & Grossberg, S. (2003). Towards a Theory of the Laminar Architecture of Cerebral Cortex: Computational Clues from the Visual System. *Cerebral Cortex, 13*, 100-113.

[8]   Grossberg, S. & Versace, M. (2008). Spikes, synchrony, and attentive learning by laminar thalamocortical circuits. *Brain Research, 1218*, 278-312.

[9]   Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H. & Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks, 3*, 698-713.

[10]   Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353-363.

[11]   Carpenter G. A., & Grossberg S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37*, 54–115.

[12]   Grossberg, S. (1973) Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics, 52,* 217-257.

[13]   Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review. 87*, 1-51.

[14]   Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience. 9*, 181-197.

[15]   Salin, P.A. & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiological reviews, 75*, 107-154.

[16] Raizada, R. D. S., & Grossberg, S. (2001). Context-sensitive binding by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, *8*, 431-466.

[17] Bullier, J., Hup´e, J. M., James, A., & Girard, P. (1996). Functional interactions between areas V1 and V2 in the monkey. *Journal of Physiology Paris, 90*(3-4), 217-220.

[18] Sillito, A. M., Jones, H. E., Gerstein, G. L.,&West, D. C. (1994). Feature-linked synchronization of thalamic relay cell firing induced by feedback from the visual cortex. *Nature*, *369*, 479-482.

[19] Amir, Y., Harel, M., & Malach, R. (1993). Cortical hierarchy reflected in the organization of intrinsic connections in macaque monkey visual cortex. *J Comp Neurol*, *334*, 19-46.

[20] Sandell, J. H., & Schiller, P.H. (1982). Effect of cooling area 18 on striate cortex cells in the squirrel monkey. *J Neurophysiol*, *48*(1), 38-48.

[21] Grossberg, S. (1999). The link between brain, learning, attention, and consciousness. *Consciousness and Cognition*, *8*, 1-44.

[22] Grossberg, S. (2003). How does the cerebral cortex work? Development, learning, attention, and 3D vision by laminar circuits of visual cortex. *Behavioral and Cognitive Neuroscience Reviews*, *2*, 47-76.

[23] Carpenter G. A. (1997). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks*, *10*, 1473–1494.

[24] Caudell, T.P., Smith, S.D.G., Escobedo, R., & Anderson, M. (1994). NIRS: Large scale ART 1 neural architectures for engineering design retrieval. *Neural Networks*, *7*, 1339-1350.

[25] Lisboa, P. (2001). Industrial use of saftey-related artificial neural netoworks. *Contract Research Report 327/2001,* Liverpool John Moores University.

[26] Parsons, O., & Carpenter, G. A. (2003). ARTMAP neural networks for information fusion and data mining: map production and target recognition methodologies. *Neural Networks*, *16*(7), 1075-1089.

[27] Carpenter. G.A., & Ross, W.D. (1995). ART-EMAP: A neural network architecture for object recognition by evidence accumulation. *IEEE Transactions on Neural Networks*, *6*, 805-818, 1995.

[28] Carpenter, G.A., & Markuzon, N. (1998). ARTMAP-IC and medical diagnosis: Instance counting and inconsistent cases. *Neural Networks*, *11*, 323-336.

[29] Williamson, J.R. (1998). Gaussian ARTMAP: A neural network for fast incremental learning of noisy multidimensional maps. *Neural Networks*, *9*, 881-897.

[30] Carpenter, G.A. (1998). Distributed learning, recognition, and prediction by ART and ARTMAP neural networks. *Neural Networks, 10*, 1473-1494.

[31] Carpenter, G.A., Milenova, B.L., & Noeske, B.W. (1998). Distributed ARTMAP: a neural network for fast distributed supervised learning. *Neural Networks, 11*, 793-813.

[32] Carpenter, G. A. (2003). Default ARTMAP. *Proceedings of the international joint conference on neural networks (IJCNN'03)*, 1396–1401.

[33] Hurvich, L. M., & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review*, *64*, 384- 390.

[34] Grossberg, S (1972). A neural theory of punishment and avoidance, II: Quantitative theory. *Mathematical Biosciences, 15*, 253-285.

[35] Carpenter, G. A., Grossberg S., & Rosen D. B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks, 4*, 759–771.

[36] Schiller, P.H. (1982). Central connections of the retinal ON and OFF pathways. *Nature, 297*, 580-583.

[37] Carpenter G. A., Grossberg S., & Reynolds J. H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks, 4*, 565–588.

[38] Carpenter, G.A., & Gjaja, M.N. (1994). Fuzzy ART choice functions. *Proceedings of the World Congress on Neural Networks (WCNN-94)*, Hillsdale, NJ: Lawrence Erlbaum Associates, **I**-713-722.

[39] Carpenter, G. A., & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks, 4*, 129-152.

[40] Carpenter, G.A., & Gaddam, S.C. (2009). Biased ART: A neural architecture that shifts attention toward previously disregarded features following an incorrect prediction. CAS/CNS Technical Report.

[41] Carpenter, G.A., Martens, S., & Ogas, O.J. (2005). Self-organizing information fusion and hierarchical knowledge discovery: a new framework using ARTMAP neural networks. *Neural Networks, 18*, 287-295.

[42] Carpenter, G.A., & Ravindran, A. (2008). Unifying multiple knowledge domains using the ARTMAP information fusion system. *Proceedings of the 11th International Conference on Information Fusion*, Cologne, Germany, June 30 – July 3, 2008.