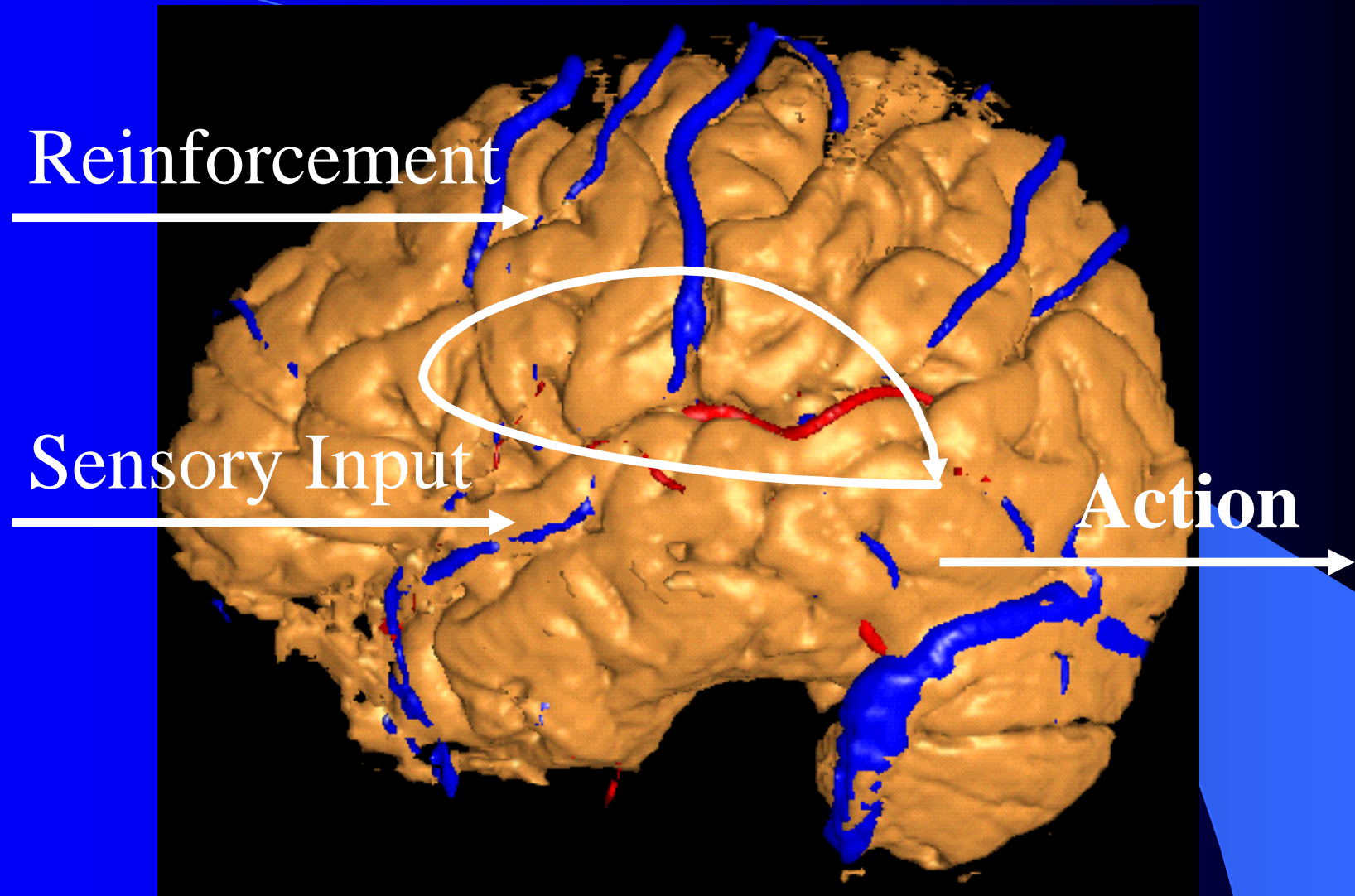


From Backpropagation to Brain-Like Intelligent systems: Current Status and Opportunities

- A roadmap for developing mathematical designs/models but also a conceptual theory already
- Why optimality? Basics, physics, issues
- Levels of Intelligence from Minsky to global mind –
 - Emergence of the 1st Generation ADP Theory of Mammal Brain
 - From Two-brain theory to “3 brain” to advanced ADP

Paul J. Werbos, www.werbos.com



**The Brain As a Whole System
Is an Intelligent Controller**

Important directions in the neural network field

Search on “**COPN**” at www.nsf.gov

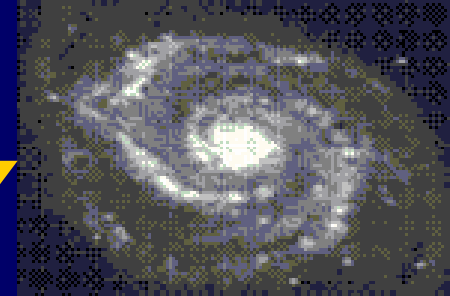
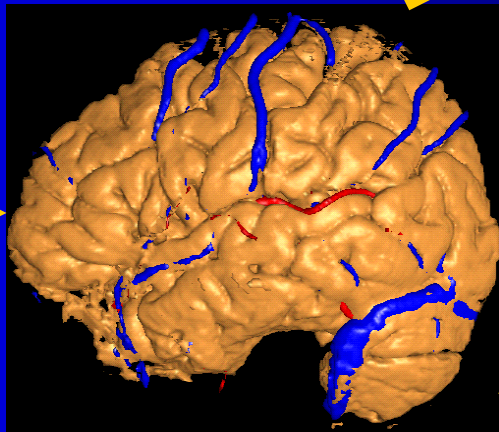
$$\frac{\Pr(A|B)}{\Pr(A)} = \frac{\Pr(B|A)}{\Pr(B)}$$

Prediction

Memory

...

Clustering



Optimization

$$J(t) = \text{Max} \langle J(t+1) + U \rangle$$

$$\frac{\partial^+ z_n}{\partial z_i} = \frac{\partial z_n}{\partial z_i} + \sum_{j=i+1}^{n-1} \frac{\partial^+ z_n}{\partial z_j} \frac{\partial z_j}{\partial z_i}$$

From Optimality to Brain & Behavior: a Roadmap to Build & Understand Brain-Like Intelligence

- A roadmap for developing mathematical designs/models but also a conceptual theory already
- Why optimality? Basics, physics, issues
- Levels of Intelligence from Minsky to global mind –
 - Emergence of the 1st Generation ADP Theory of Mammal Brain
 - From Two-brain theory to “3 brain” to advanced ADP

Paul J. Werbos, www.werbos.com



National Science Foundation

Engineering Directorate

Computer & Info. Science Directorate

ECCS

DMMI

IIS

EPDT:
Chips,
Optics,
Etc.

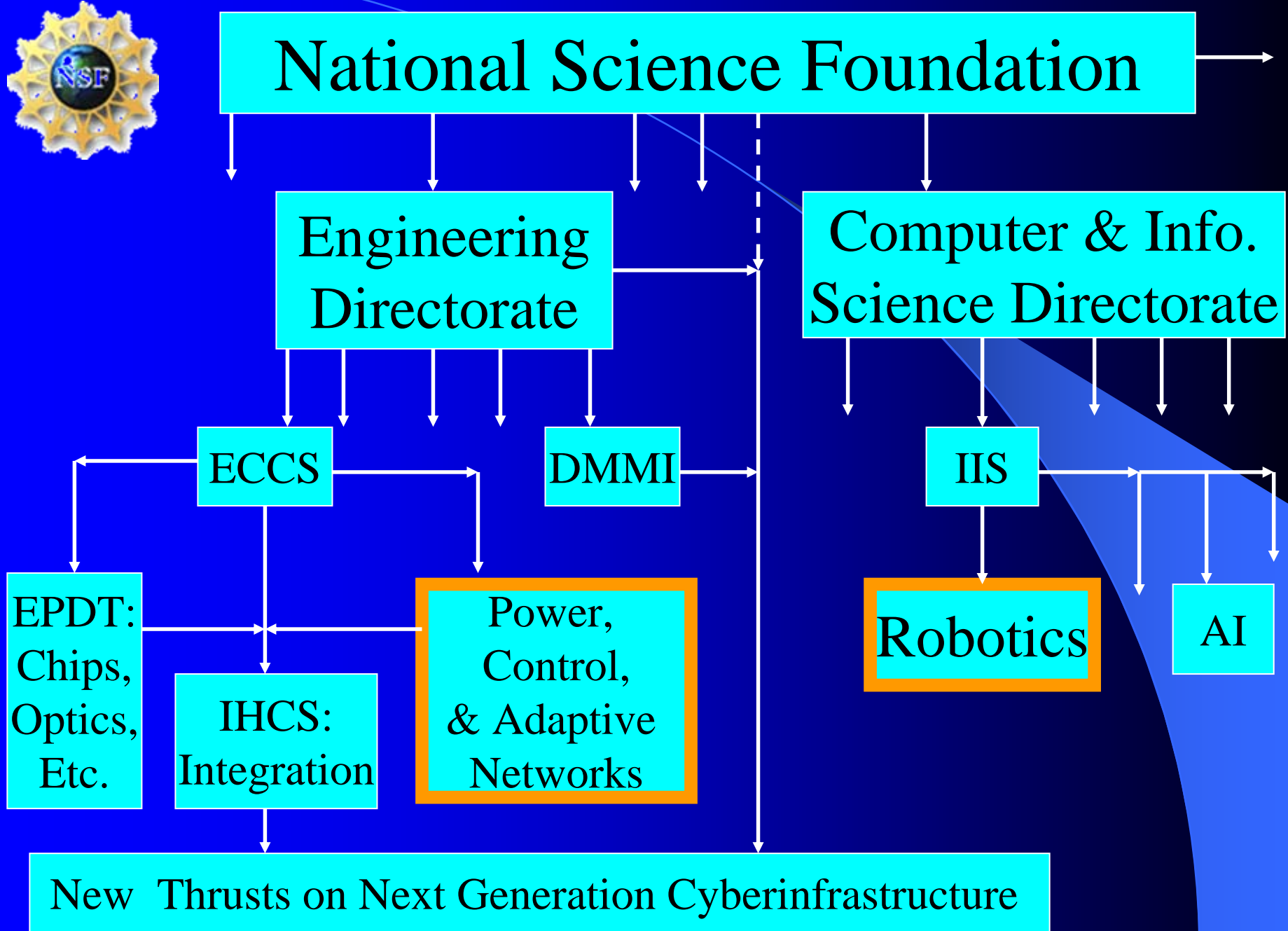
IHCS:
Integration

Power,
Control,
& Adaptive
Networks

Robotics

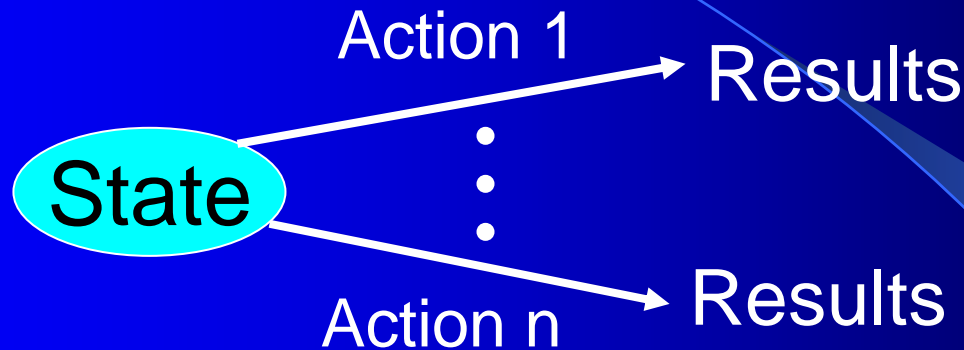
AI

New Thrusts on Next Generation Cyberinfrastructure



Why Optimality? (1) Basics

Behavior is About **Choices** (Cybernetica 1968)



- **Functionality** of the brain is making choices which yield **better results**
- **Intelligence** is about **learning** how to make better choices.
- To put this into mathematics, we need a way to compare results: Von Neumann's Cardinal Utility Function, foundation of decision analysis, game theory, risk analysis
- **"Goals" & "intentions"** are ways to define ultimate values or utility, or subgoals. For organisms, values at the top. Telos, happiness, Mill.

Why Optimality (2): Physics vs Brains

$$\begin{array}{c} \hline \varphi(\underline{x}, t) \quad t_- < t < t_+ \\ \hline \varphi(\underline{x}, t_+) \end{array}$$

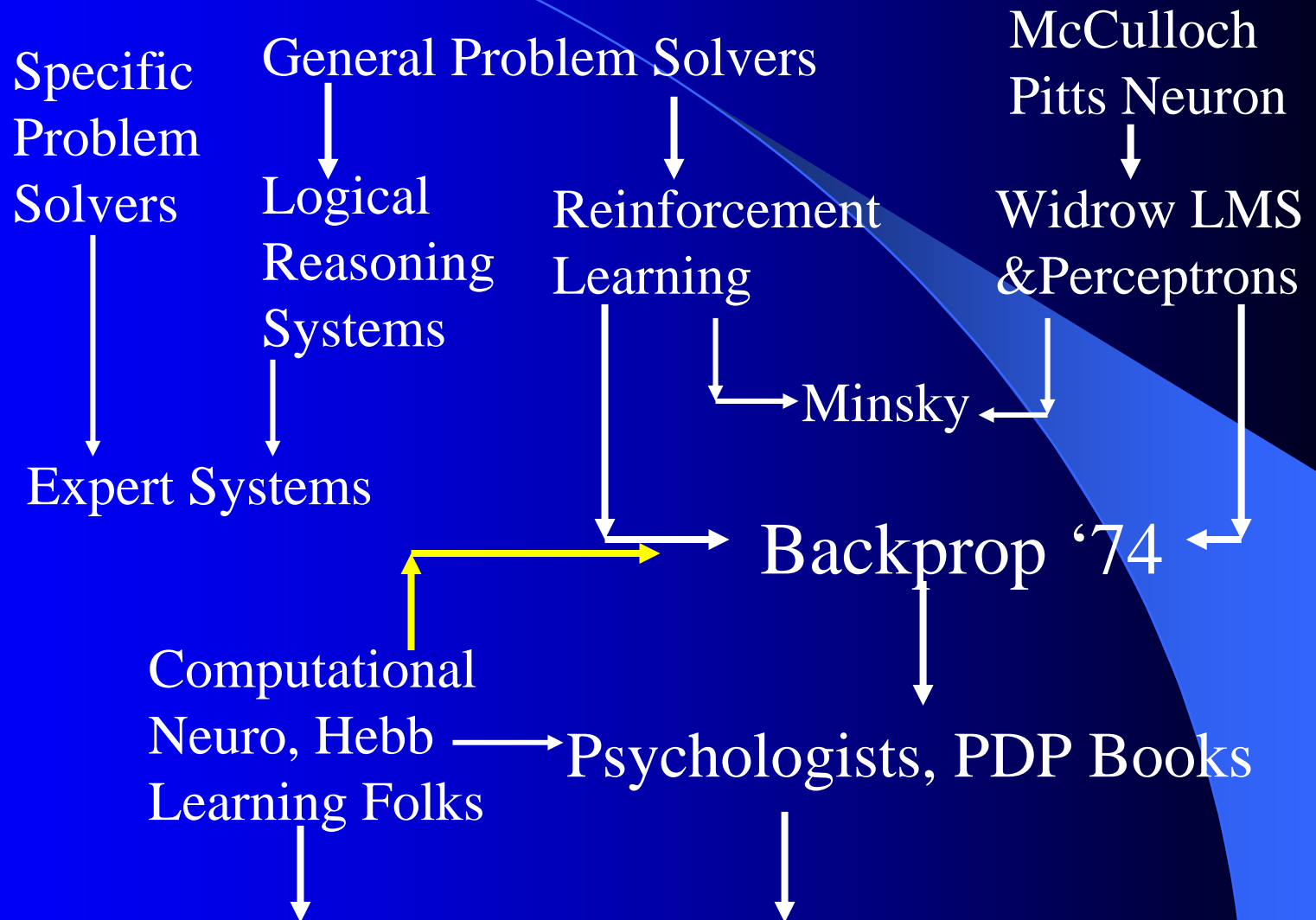
- Lettvin (~70): “Give us no anthropomorphic view of brain. **Even the universes maximizes ‘utility.’**” Lagrange proposed that universe “chooses” $\varphi(\underline{x}, t)$ so as to max the sum of $\mathcal{L}(\underline{x}, t)$, its utility function or Lagrangian, subject to fixed boundary conditions.
- “**Hamilton-Jacobi**” equation tells us what universe must do to maximize \mathcal{L} but:
 - Is it a max or min or minmax?
 - HJ is deterministic – no random disturbance or noise
 - Quantum theory adds noise in odd ways www.werbos.com/reality.htm
- **Bellman equation (1953)** was huge breakthrough, giving optimal strategy of action or “policy” for stochastic case, using Von Neumann’s Cardinal Utility Function and guidance.
- Optimality principle also important to thermodynamics, to physics of emergent properties

Optimality (III): Traditional Questions

(& See papers at www.werbos.com)

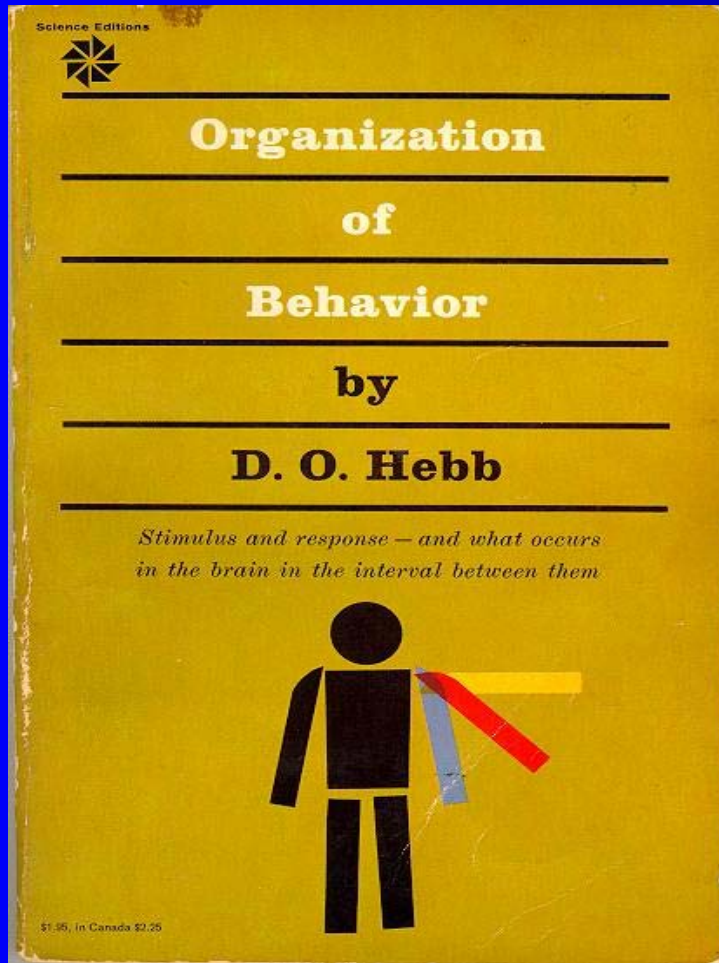
- If brains are so optimal, why do humans do so many stupid things?
 - Brains are designed to **learn** to approximate optimal policy, as effectively as possible with **bounded computational resources** (neural networks), starting from a **less optimal start**. They never learn to play a perfect game of chess (nor will our computers) because of resource constraints. We just **do the best we can**.
 - When one human criticizes another, we are comparing two highly intelligent systems. Some brains learn faster than others, and humans are an intermediate stage towards even higher/faster intelligence.
- If the optimization theory is right, wouldn't brains get stuck in local minima?
 - They sure do. Everyone on earth is in a “local minimum,” or a rut, to some degree. In other words, we could all do a bit better if we had more creativity. But look at those hairy guys (chimps) in the jungle, and the rut they are in!
 - The optimality theory says we combine an incremental learning ability with an ability to learn to be more “creative” – to do better and better “stochastic search” of the options available to us. (Widrow example.)

Where Did ANNs Come From?



IEEE ICNN 1987: Birth of a "Unified" Discipline

Hebb 1949: Intelligence As An Emergent Phenomenon or Learning



“The general idea is an old one, that any two cells or systems of cells that are especially active at the same time will tend to become ‘associated,’ so that activity in one facilitates activity in the other” -- p.70 (Wiley 1961 printing)

The search for the General Neuron Model (of Learning)

“Solves all problems”

Claim (1964) : Hebb's Approach Doesn't Quite Work As Stated

- Hebbian Learning Rules Are All Based on **Correlation Coefficients**
- Good Associative Memory: **one component** of the larger brain (Kohonen, ART, Hassoun)
- **Linear** decorrelators and predictors
- Hopfield $f(\underline{u})$ minimizers never scaled, **but**:
 - Gursel Serpen and SRN minimizers
 - Brain-Like Stochastic Search (Needs R&D)

Understanding Brain Requires Models Tested/Developed Using Multiple Sources of Info

- Engineering: Will it work? Mathematics understandable, generic?
- Psychology: Connectionist cognitive science, animal learning, folk psychology
- Neuroscience: computational neuroscience
- AI: agents, games (backgammon, go), etc.
- LIS and CRI

Maximizing utility over time

Model of reality

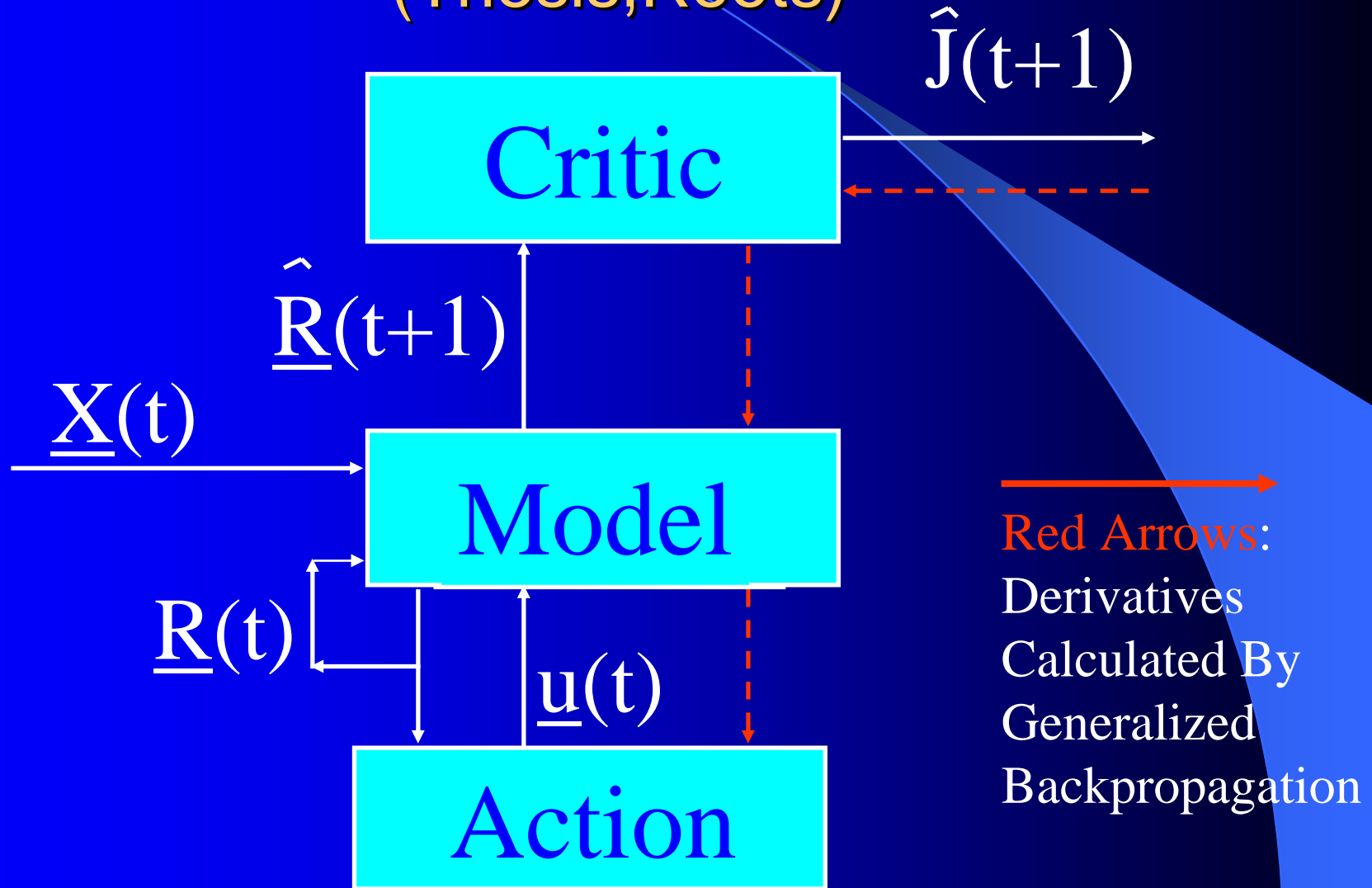
Utility function U

Dynamic programming

$$J(\mathbf{x}(t)) = \text{Max}_{\mathbf{u}(t)} \langle U(\mathbf{x}(t), \mathbf{u}(t)) + J(\mathbf{x}(t+1)) \rangle / (1+r)$$

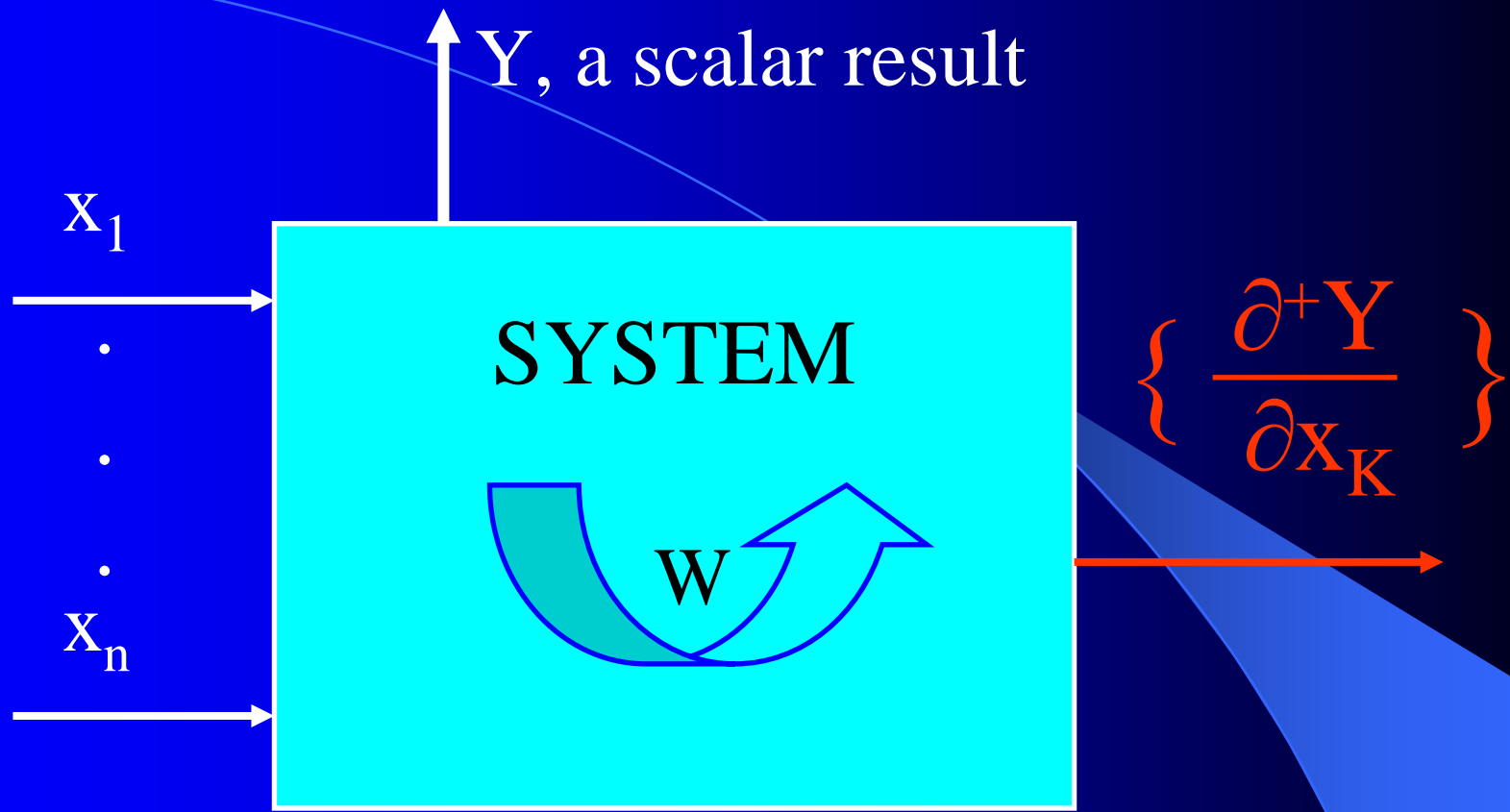
Secondary, or strategic utility function J

1971-2: Emergent Intelligence Is Possible If We Allow Three Types of Neuron (Thesis, Roots)



Harvard Committee Response

- We don't believe in neural networks – see Minsky (Anderson&Rosenfeld, Talking Nets)
- **Prove** that your backwards differentiation works. (That is enough for a PhD thesis.) The critic/DP stuff published in '77,'79,'81,'87..
- **Applied** to affordable vector ARMA statistical estimation, general TSP package, and robust political forecasting

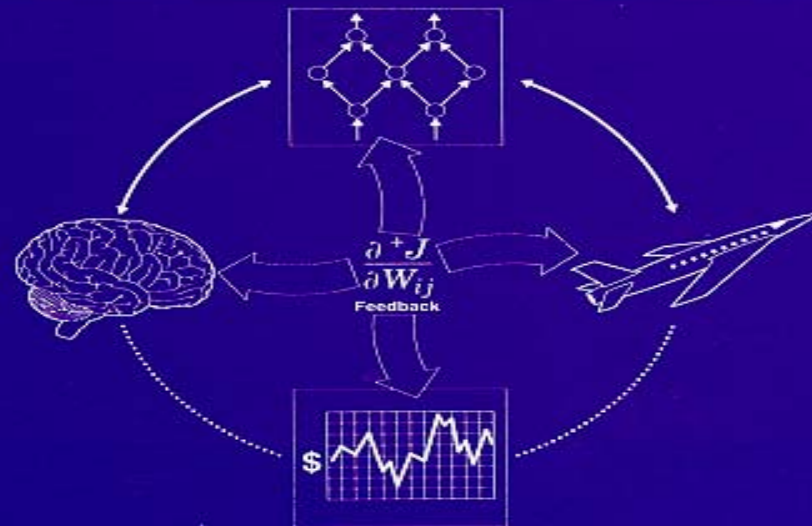


(Inputs x_k may actually come from many times)

Backwards Differentiation: But what kinds of SYSTEM can we handle? See details in AD2004 Proceedings, Springer, in press.

THE ROOTS OF BACKPROPAGATION

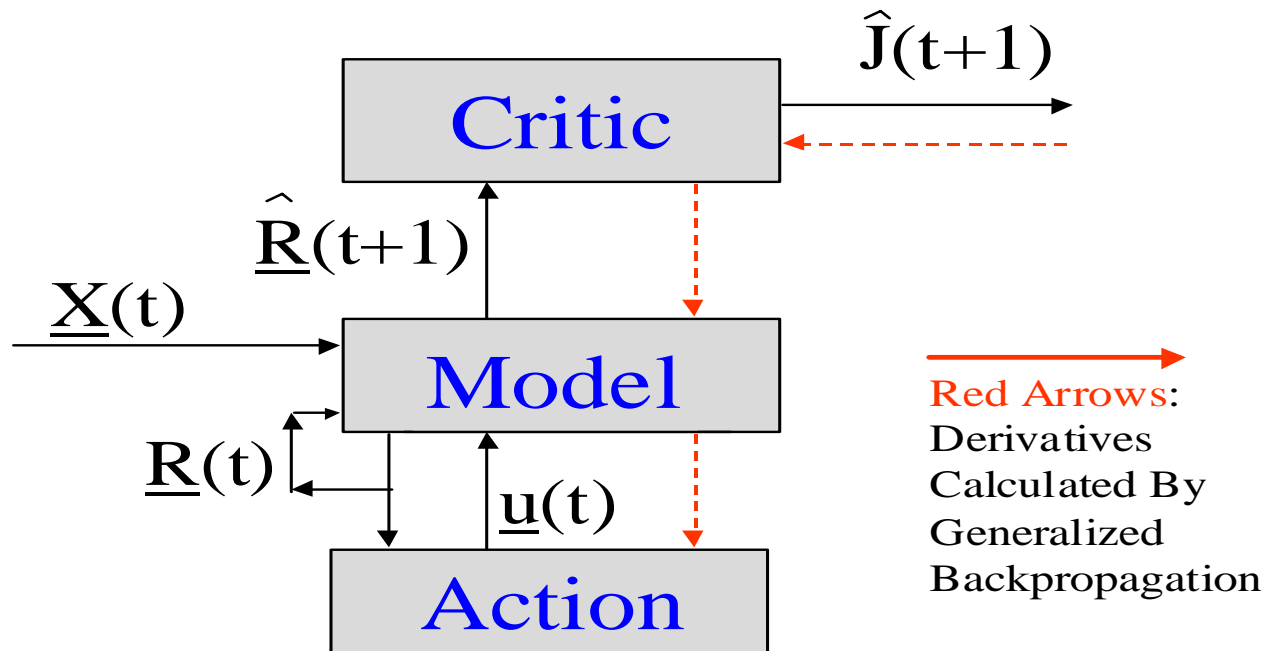
From Ordered Derivatives
to Neural Networks
and Political Forecasting



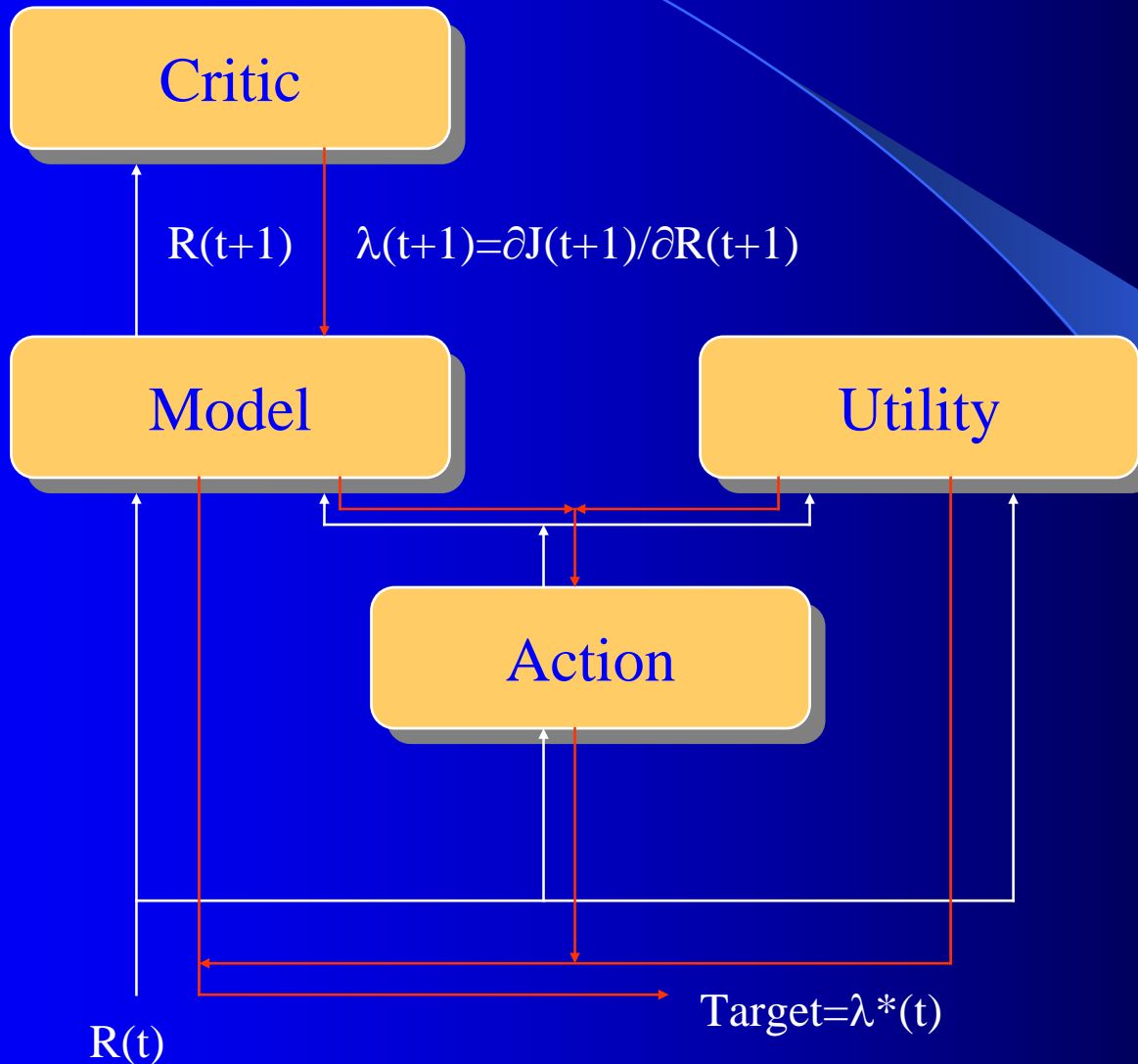
PAUL JOHN WERBOS

A Volume in the Wiley Series on ADAPTIVE AND LEARNING SYSTEMS
FOR SIGNAL PROCESSING, COMMUNICATIONS, AND CONTROL
SIMON HAYKIN, SERIES EDITOR

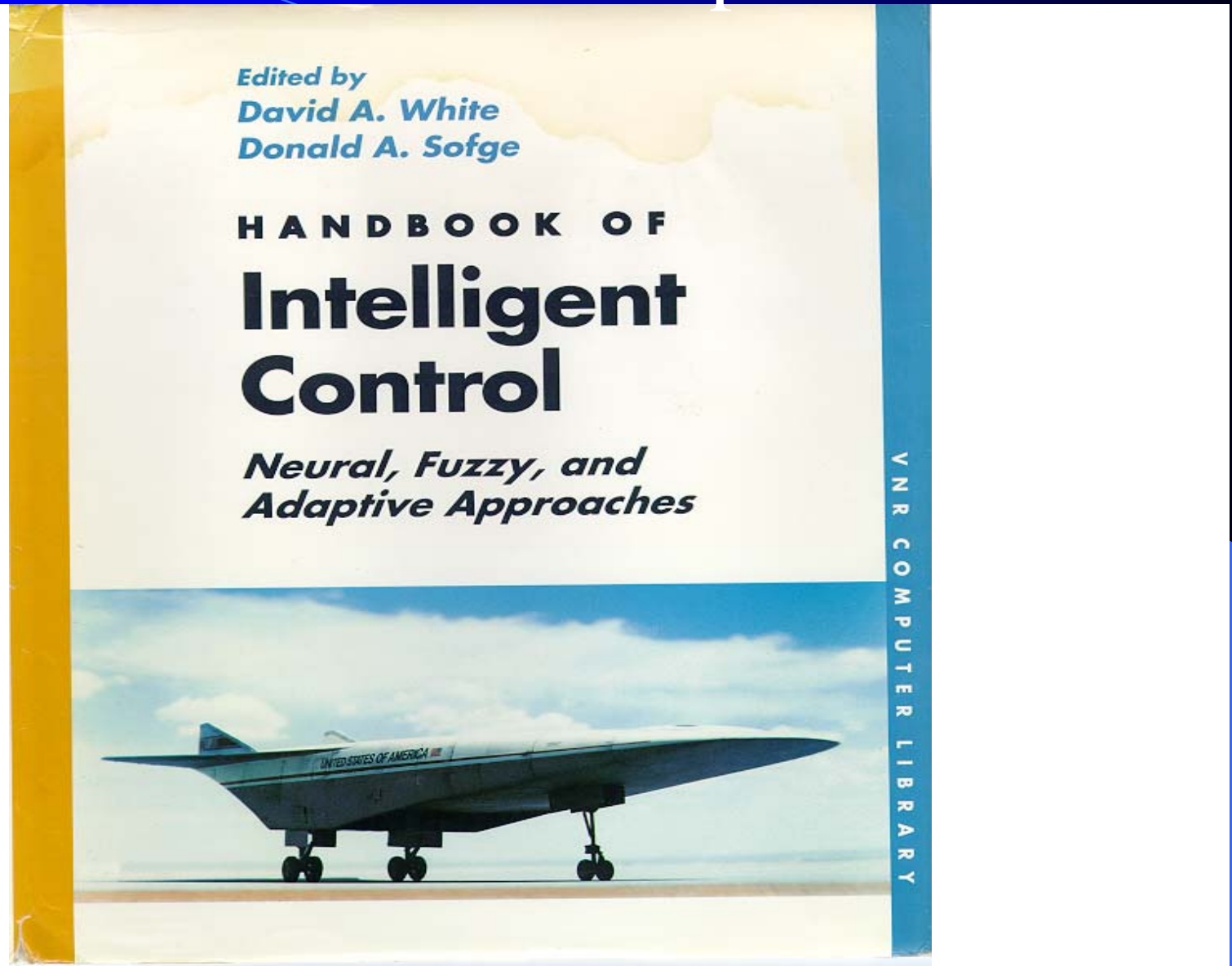
- # To Fill IN the Boxes:
- (1) NEUROCONTROL, to Fill in Critic or Action;
 - (2) System Identification or Prediction (Neuroidentification) to Fill In Model



Dual Heuristic Programming (DHP)



NSF/McAir Workshop 1990



White and Sofge eds, Van Nostrand, 1992

1st Generation Theory of Mammal Brain

- As in 71-72 proposal, **brain has 3 main parts**:
 - Cortex+thalamus: **Model to predict/impute reality**. See Nicolelis&Chapin, Science, rat whisker work.
 - Limbic system: **Critic gives “emotional” assessment** of what Freud called “objects” (Papez, James Olds)
 - Brain-stem: **action or “motor” system** (and inherited fixed preprocessors/postprocessors)
 - **Clock signals** from extracortical sources (Foote, Llinas)
 - Backprop unavoidable. (Bliss, Spruston, Sejnowski)
- **Technical level improvements and big runs enough to span gap form 1971-72 to mammal brain**:
 - Fill in “Model” with hybrid Simultaneous/Time-Lagged Recurrent Network trained by Error Critic (fully specified in Handbook of Intelligent Control)
 - Critic is sum of multiple “HDP” components each trained by GDHP, which gives power of DHP for continuous variables but handles continuous/discrete mix.
 - In each box, faster learning, per robust statistics, learning from memory, etc.
- **BUT IS IT ENOUGH? For what?**

Beyond Bellman: Learning & Approximation for Optimal Management of Larger Complex Systems

www.eas.asu.edu/~nsfadp

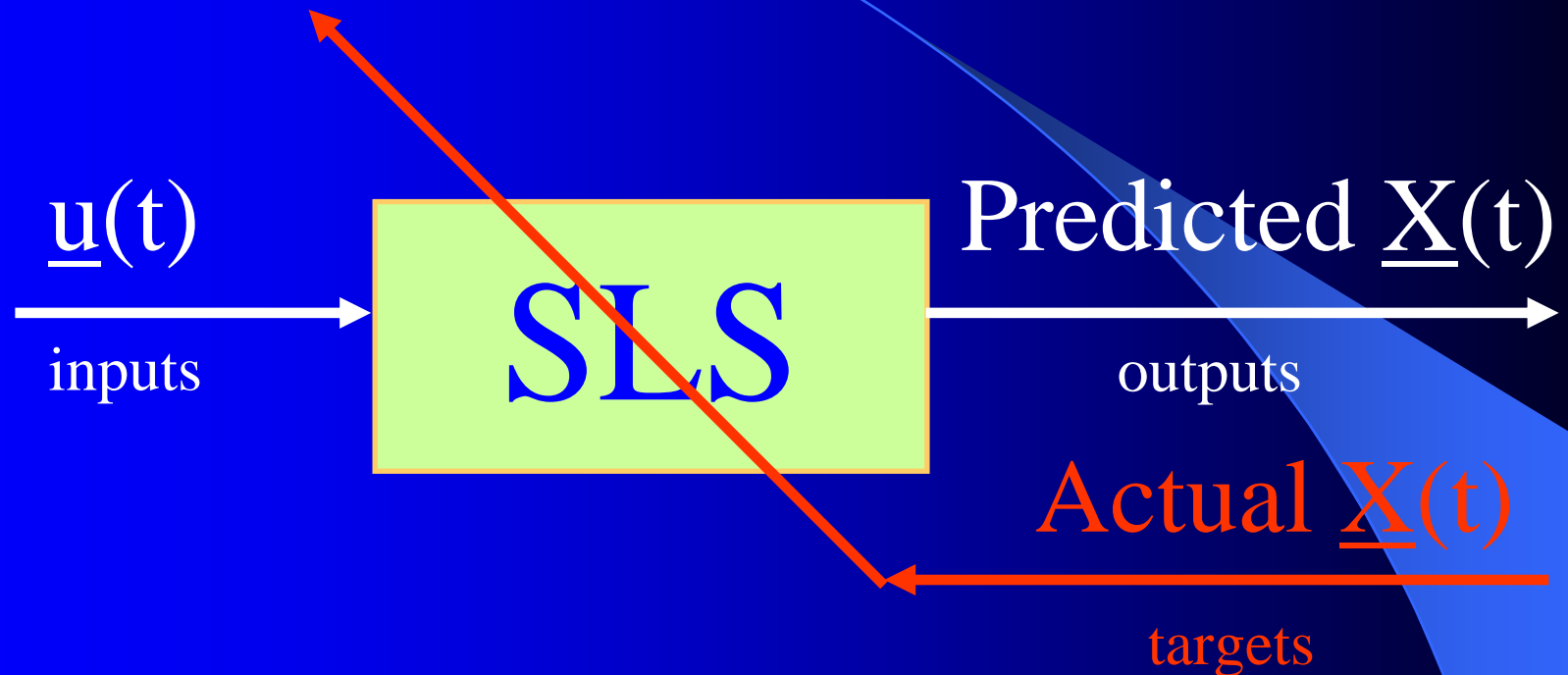
- Basic thrust is **scientific**. Bellman gives exact optima for 1 or 2 continuous state vars. New work allows 50-100 (thousands sometimes). Goal is to **scale up in space and time** -- the math we need to know to know how brains do it. And unify the recent progress.
- Low lying fruit -- missile interception, vehicle/engine control, strategic games
- Workshops: ADP02 & Dynamic Stochastic Grid testbed; ADP06 April 2006

AREAS OF NEED	SCALE UP ADP	SUBSYSTEMS: SUPERVISED LEARNING, PREDICTION, ETC
UNIFY TODAY'S "BEST STATE OF ART"	50 NL VARIABLES, CONTROL APPS, MULTI-METHODS	FOUNDATIONS, COMBINE RISK/LOSS, LOCAL/GLOBAL, ERROR CRITIC, ...
FROM SOA TO BRAIN SCALE	MULTI TIME SCALE, VAST DOMAIN (But Still 1 Brain)	Same As Left (e.g. images, Go, CNN, Network diagnostics)

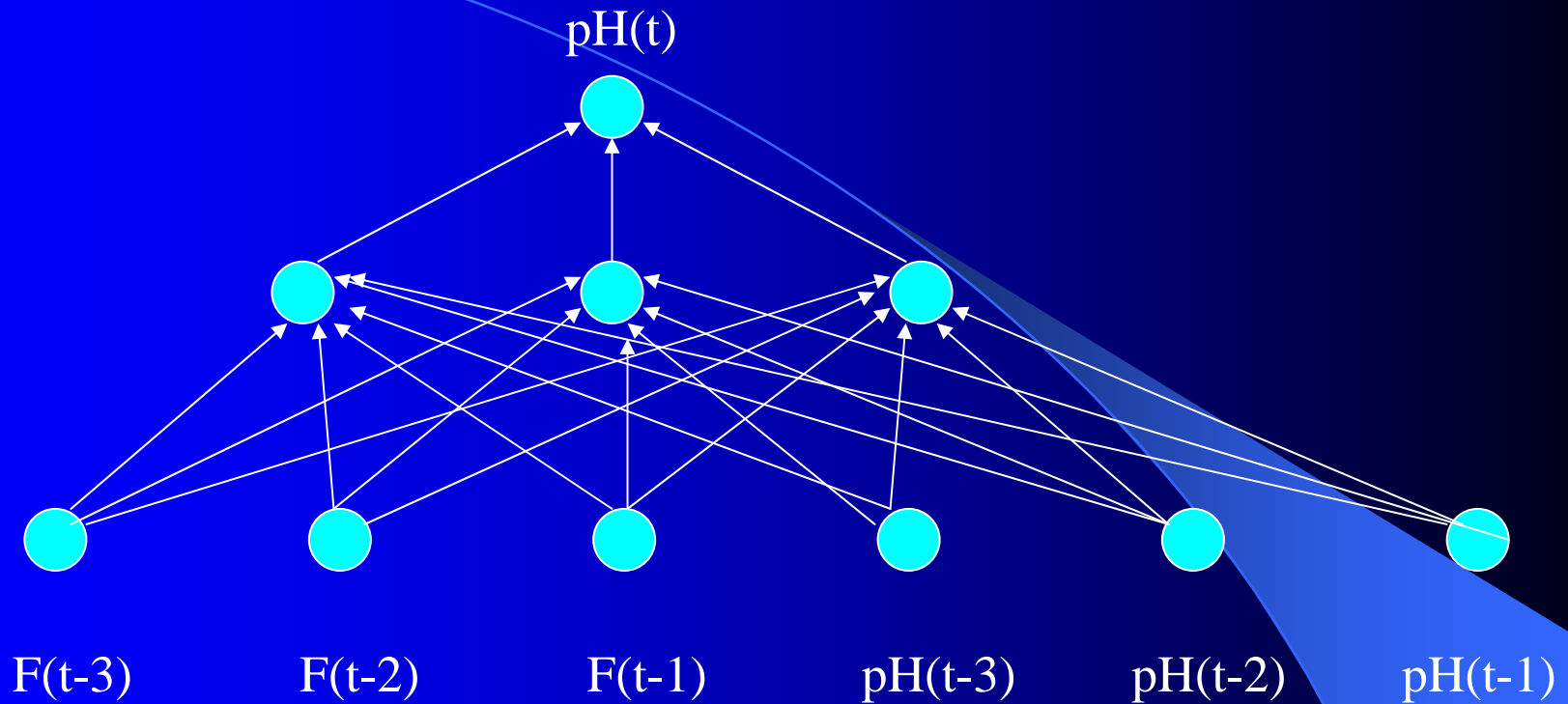
3 Types of Diagnostic System (A Major App of State-of-Art Prediction/Estimation)

- All 3 train **predictors**, use sensor data $\underline{X}(t)$, other data $\underline{u}(t)$, fault classifications F_1 to F_m
- Type 1: predict $F_i(t)$ from $\underline{X}(t)$, $\underline{u}(t)$, MEMORY
- Others: first train to predict $\underline{X}(t+1)$ from $\underline{X}, \underline{u}, \text{MEM}$
 - Type 2: when **actual** $\underline{X}(t+1)$ 6σ from prediction, ALARM
 - Type 3: if prediction net **predicts BAD** $\underline{X}(t+T)$, ALARM
- Combination best. See PJW in Maren, ed, *Handbook Neural Computing Apps*, Academic, 1990.

Supervised Learning Systems (SLS)



SLS may have internal dynamics but no “memory” of times $t-1$, $t-2$...

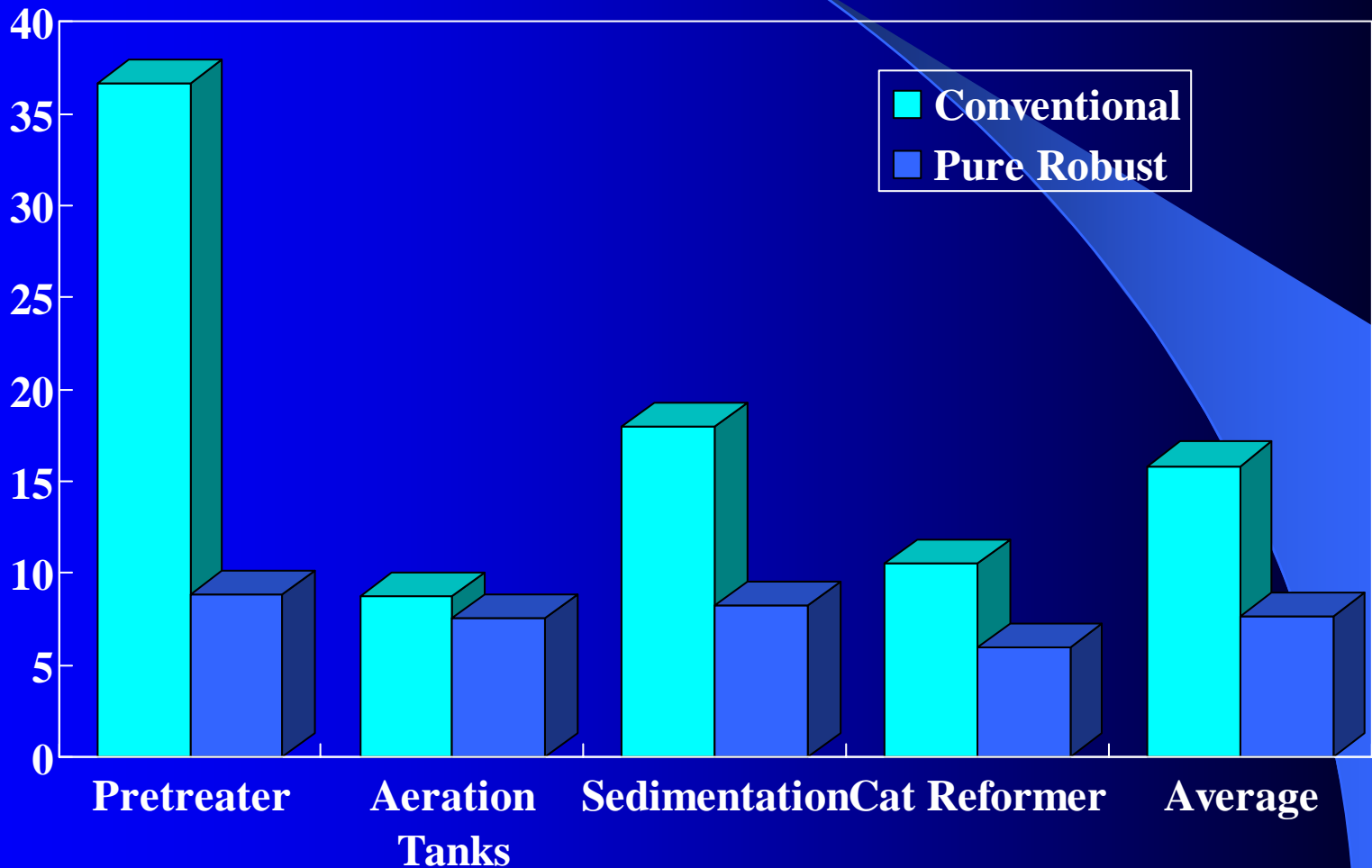


Example of TDNN used in HIC, Chapter 10

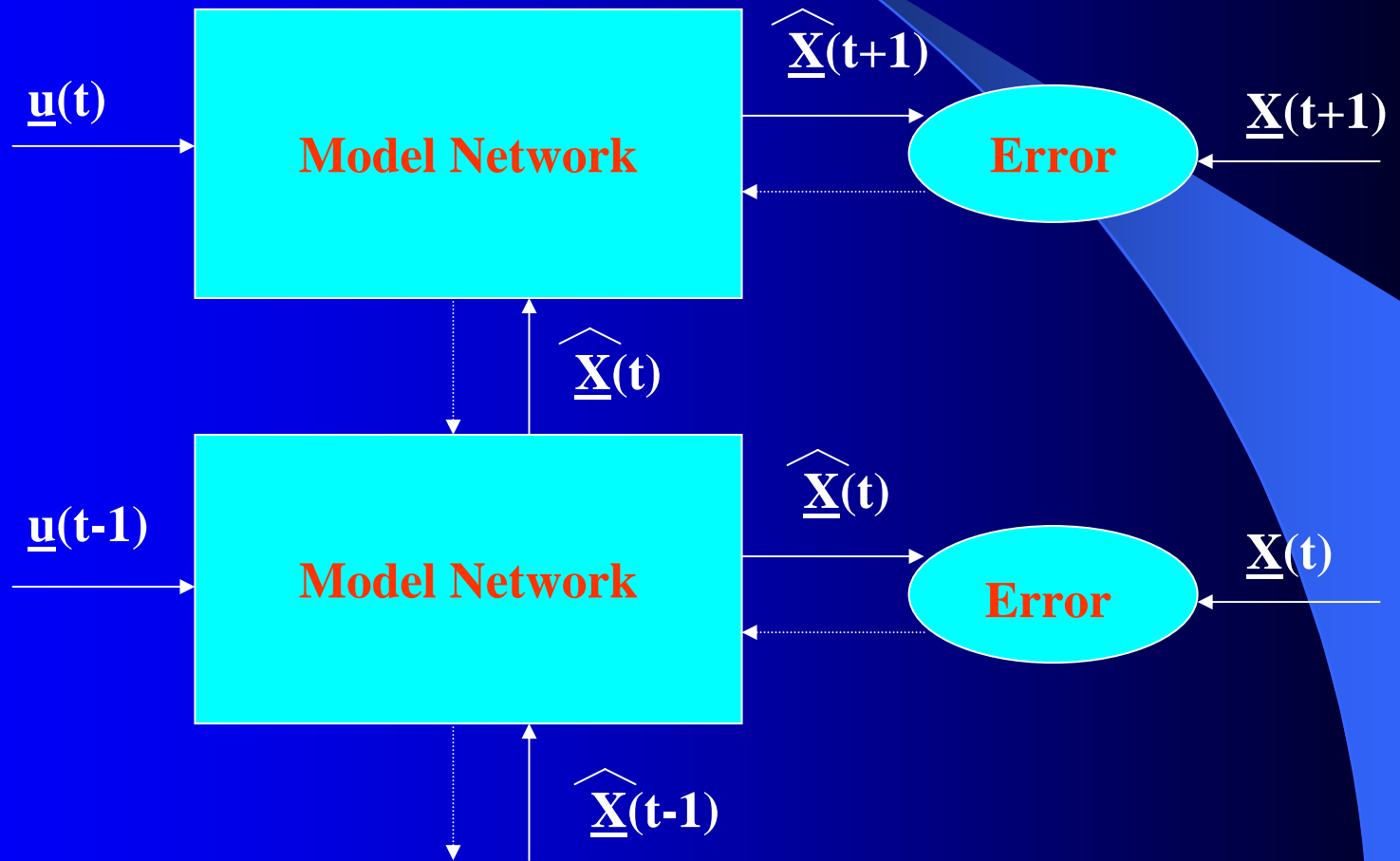
TDNNs learn NARX or FIR Models, not NARMAX or IIR

Prediction Errors (HIC p.319)

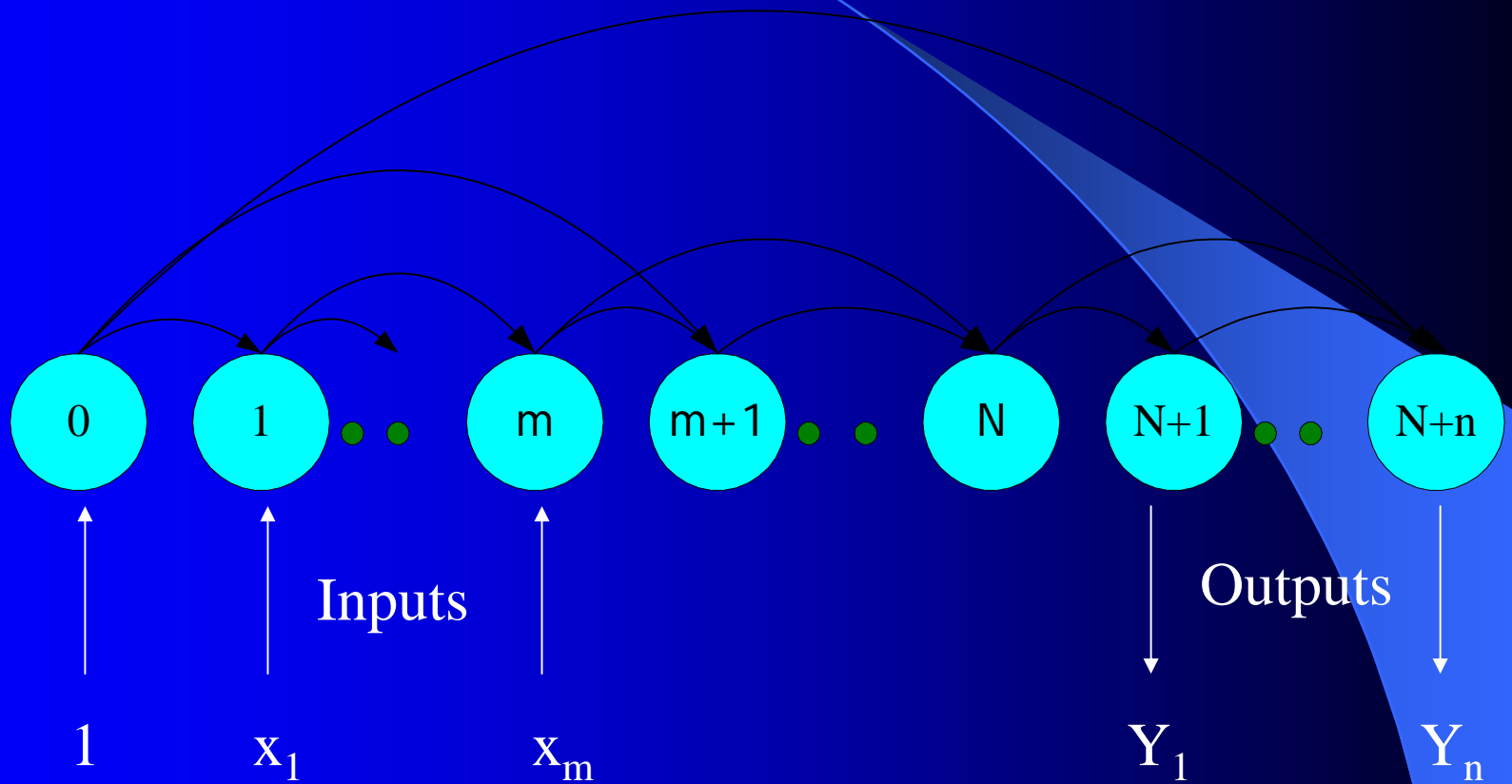
From chapter 10, posted at www.werbos.com



PURE ROBUST METHOD



Generalized MLP



EQUATIONS OF GENERALIZED MLP

$x_i = X_i$ $i=1$ to m , read-in

do for $i=m+1$ to $N+n$

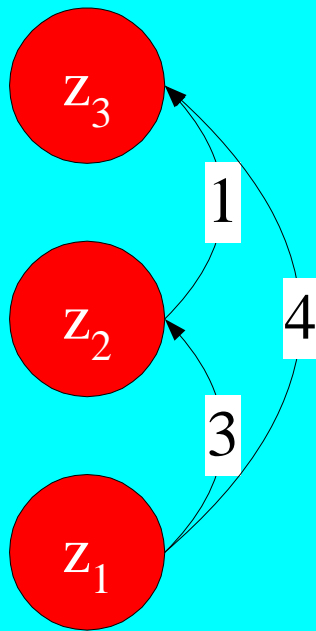
$$v_i = \sum_{j=0}^m W_{ij} x_j$$

$$x_i = s(v_i)$$

$Y_i = X_{i-N}$ $i=1$ to N , read-out

$$E = 1/2 \sum_{i=1}^n (Y_i - Y_i^*)^2$$

How calculate the derivatives?



$$\frac{\partial z_3}{\partial z_2} = 1$$

$$\frac{\partial z_3}{\partial z_1} = 4$$

$$\frac{\partial^+ z_3}{\partial z_1} = \frac{\partial z_3}{\partial z_1} + \frac{\partial^+ z_3}{\partial z_2} \frac{\partial z_2}{\partial z_1} = 7$$

$$\frac{\partial^+ z_n}{\partial z_i} = \sum_{j=i+1}^n \frac{\partial^+ z_n}{\partial z_j} \frac{\partial z_j}{\partial z_i}$$

A Chain Rule For Ordered Derivatives

Backpropagation for generalized ML

$$F_{-Y_i} = Y_i - Y_i^* \quad (i=1 \text{ to } n)$$

do for $j=N+n$ to 0

$$F_{-x_j} = F_{-Y_{j-N}} + \sum_{i=j+1}^{N+n} W_{ij} F_{-v_i}$$

$$F_{-v_j} = s'(x) F_{-x_j}$$

$$F_{-W_{ij}} = x_j F_{-v_i} \quad (\text{all weights})$$

BP/GMLP WITH DUAL SUBROUTINE

$$F_Y = \underline{Y} - \underline{Y}^*$$

$$F_W = F_GMLP_W(\underline{X}, W, F_Y)$$

Definition: If $\underline{Y} = \underline{f}(\underline{X})$, for any vector g :

$$F_f(X, g)_j = \sum_i g_i \frac{\partial f_i}{\partial x_j}$$

F_f “backpropagates through” f

CONVENTIONAL ANNS USED FOR FUNCTION APPROXIMATION IN CONTROL

- Global: Multilayer Perceptron (MLP)
 - Better Generalization, Slower Learning
 - Barron's Theorems: More Accurate Approximation of Smooth Functions as Number of Inputs Grows
- Local: RBF, CMAC, Hebbian
 - Like Nearest Neighbor, Associative Memory
 - Sometimes Called “Glorified Lookup tables”

No feedforward or associative memory net can give brain-like performance! Useful recurrence--

- For short-term memory, for state estimation, for fast adaptation – **time-lagged recurrence** needed. (**TLRN** = time-lagged recurrent net)
- For better $Y=F(X,W)$ mapping, **Simultaneous Recurrent Networks** Needed. For large-scale tasks, SRNs WITH SYMMETRY tricks needed – cellular SRN, Object Nets
- For robustness over time, “recurrent training”

Why TLRNs Vital in Prediction: Correlation \neq Causality!

- E.g.: law X sends extra \$ to schools with low test scores
- Does negative correlation of \$ with test scores imply X is a bad program? No! Under such a law, negative correlation is hard-wired. Low test scores cause \$ to be there! No evidence + or - re the program effect!
- Solution: compare \$ at time t with performance changes from t to t+1! More generally/accurately: train dynamic model/network – essential to any useful information about causation or for decision!

The Time-Lagged Recurrent Network (TLRN)



$$\underline{Y}(t) = \underline{f}(\underline{X}(t), \underline{R}(t-1)); \underline{R}(t) = \underline{g}(\underline{X}(t), \underline{R}(t-1))$$

\underline{f} and \underline{g} represent 2 outputs of one network

All-encompassing, NARMAX(1 \equiv n)

Felkamp/Prokhorov Yale03: \gg EKF, \approx hairy

4(5) Ways to Train TLRNs (SRN)

(arXiv.org, adap-org 9806001)

- “Simple BP” – incorrect derivatives due to truncated calculation, robustness problem
- BTT – exact, efficient, see Roots of BP ('74), but not brain-like (back time calculations)
- Forward propagation – many kinds (e.g, Roots, ch.7, 1981) – not brainlike, $O(nm)$
- **Error Critic** – see Handbook ch. 13, Prokhorov
- **Simultaneous BP** – SRNS only.

Neural Networks and Adaptive Dynamic Programming (ADP): The New Path to Building and Understanding Brain-Style Intelligence

pwerbos@nsf.gov

For technical details, please see the tutorial
at www.eas.asu.edu/~nsfadp.

- “Government public domain”: These slides may be copied, posted, or distributed freely, so long as they are kept together, including this notice

Venayagamoorthy/Wunsch/Harley ADP Turbogenerator Control



- Stabilized voltage & reactance under intense disturbance where neuroadaptive & usual methods failed
- Being implemented in full-scale experimental grid in South Africa
- Best paper award IJCNN99

4 Types of Adaptive Critics

- Model-free (levels 0-2)*
 - Barto-Sutton-Anderson (BSA) design, 1983
- Model-based (levels 3-5)*
 - Werbos Heuristic dynamic programming with backpropagated adaptive critic, 1977, Dual heuristic programming and Generalized dual heuristic programming, 1987
- Error Critic (TLRN, cerebellum models)
- 2-Brain, 3-Brain models

New Designs for Universal Stability in Classical Adaptive Control and Reinforcement Learning

<http://arXiv.org> adap-org/9810001

- Neurocontrol & tracking in general
- How foresight stabilizes adaptive control
- Key concepts in reinforcement learning

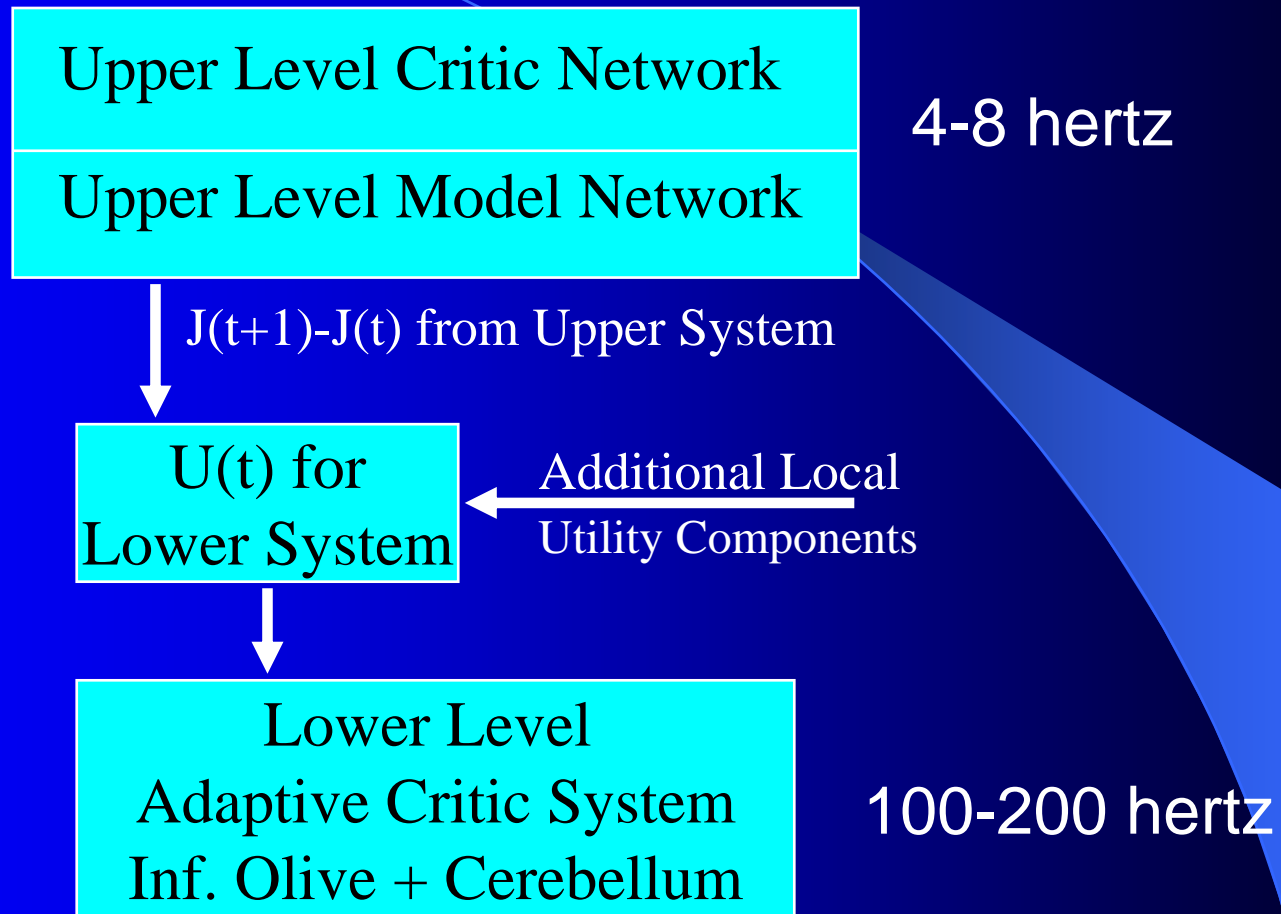
ADP

- New (in)stability results and method
- New work needed -- proofs, apps, brain

Gaps in the “SOA” level of ADP Proper: Where Is...?

- Whole system universal stability proof for linear MIMO adaptive control using HDPG, DHPG, GDHPG? (See arxiv.org 1998..)
- General-purpose tools in MatLab, etc.?
- Community knowledge, unification, tools?
- “Syncretism”: combine memory (local DEKF?) with prediction, to combine fast learning and slow generalization?
- ADP linked to good observers like TLRN? (e.g. see Feldkamp/Prokhorov paper posted at...)
- Good balance of online/offline iteration/learning, of model use vs robustness, discrete/continuous (e.g. GDHP)?
- Good “competition” example?
- Followup on best big application demonstrations?

2nd Generation “Two Brains in One Model”

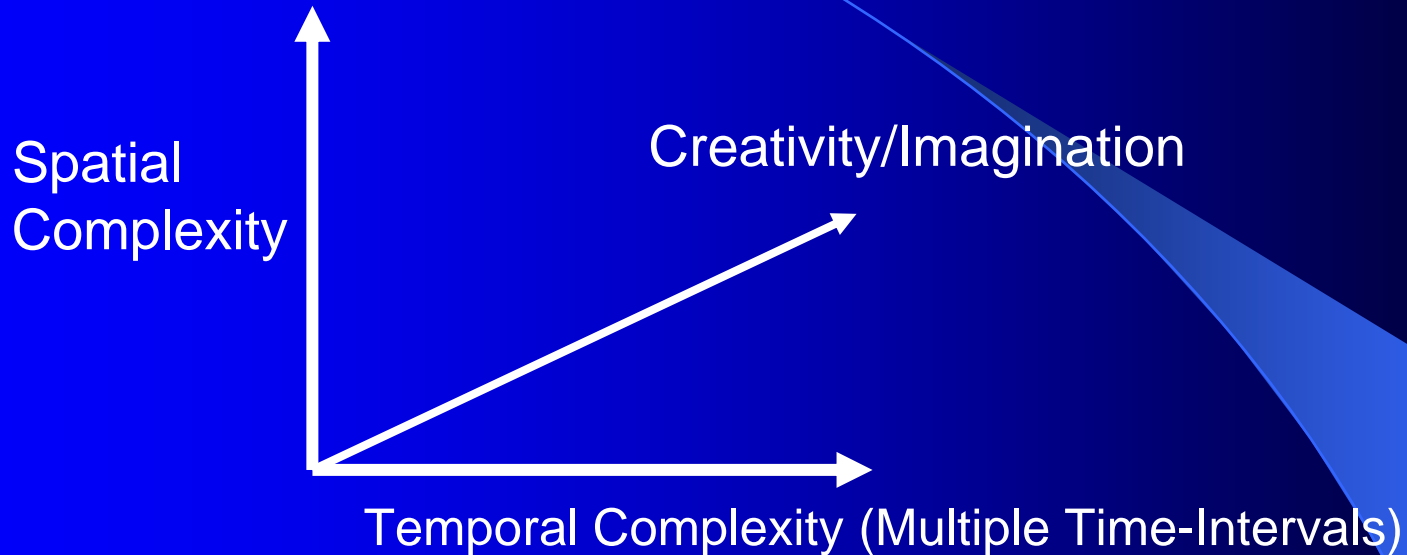


Concept in “Statistical/Numerical...”, Trans. SMC, 1987 (on web)
Joint papers with Pellionisz (experimental follow-on still warranted)
See equations in Handbook of Intelligent Control, Ch. 13 & Prokhorov

3rd Gen: 3 Brains in 1?

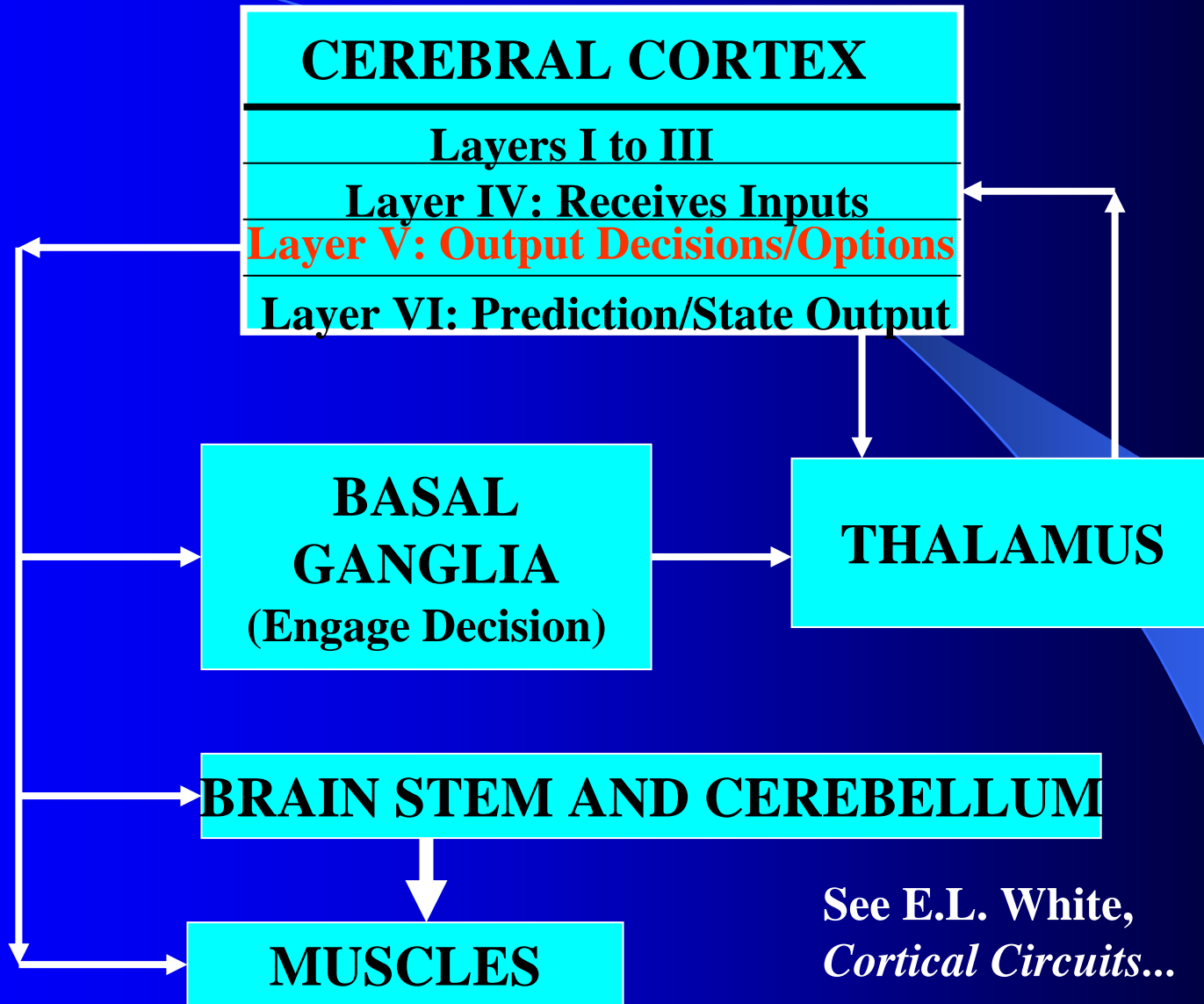
- Upper Brain: Values, Noise, Limbic Critic and Neocortex
- Middle: Basal Ganglia, AI-Like, Tasks, Mishkin, Houk, Brooks, Landing Intent
- Lower: Smoothing/Speed/LQG Like, Olive Critic and Cerebellum
- Complex 3rd Generation Theory over-responsive to AI (Albus) sketched in 1997 paper in Karny et al.

Key Issues in 3rd Generation Model



- Can we (and do brains) do better than 2nd gen brain in handling greater spatial & temporal complexity, by new designs & exploiting unspecialized but structured prior information (Kant) to get faster/better learning?
- What is our answer to AI's "spatial/temporal chunking" & stochastic search?
- All 3 demand more attention and work!!!

3rd Generation View of Creativity/Imagination: Layer V = "Option Networks"



See E.L. White,
Cortical Circuits...

- Challenge: www.werbos.com/WerbosCEC99.htm.
- Important work by Serpen, Pelikan, Wunsch, Thaler, Fu – but still wide open. Widrow testbed.

3rd Generation View of Time

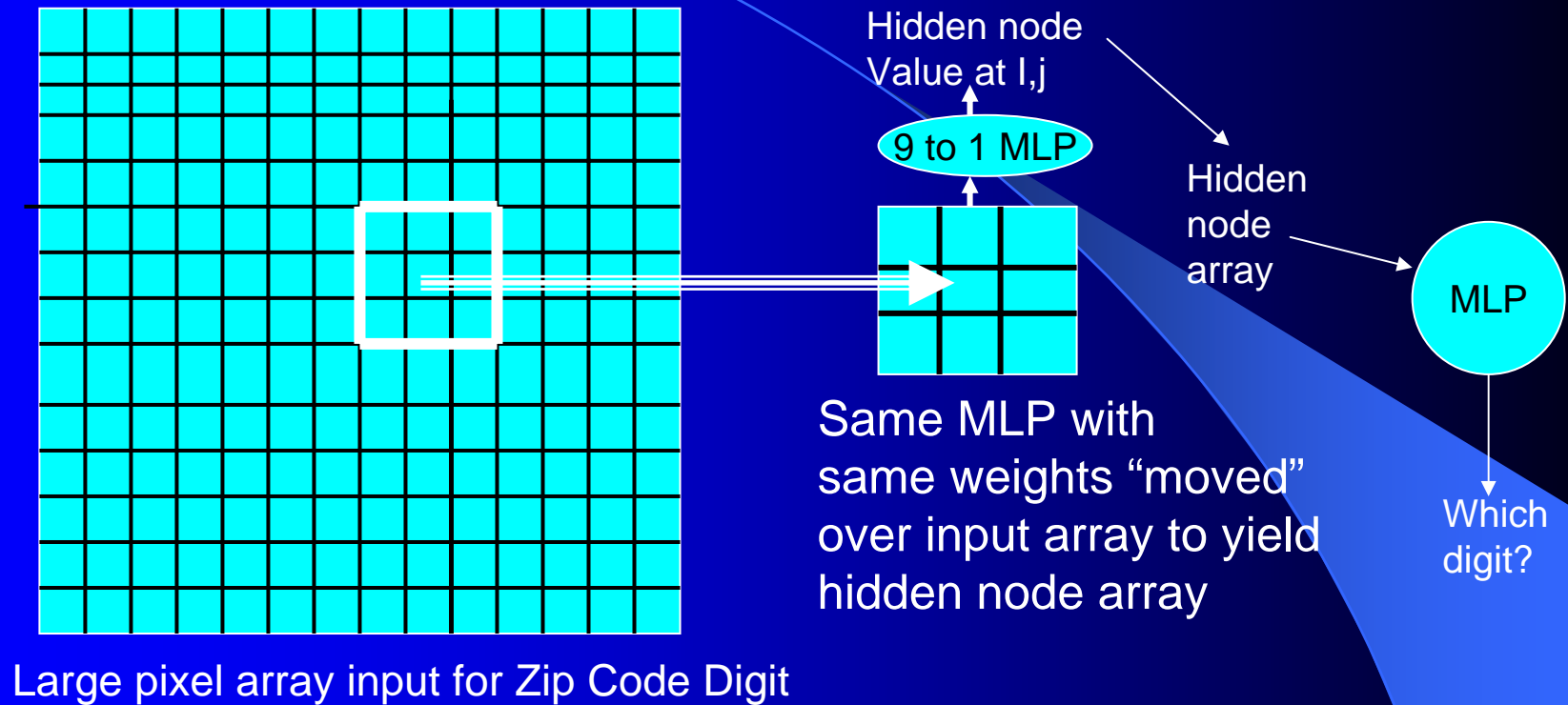
- Before 1997, under NSF\$, Sutton had modified “Bellman equation” for idea of “options” – chunks of action over time optimized at low level, to be selected by ADP at high level as discrete choices. (No object, no parameters)
- In 1987 paper, I reported more general Bellman equations for time structuring, e.g.:

$$J_i^T = (J_i^A)^T + \text{SUM (over } j \text{ in } N(i)) J_j^T (J^B)_{ij}$$

where J^A represents utility within valley i before exit, and J^B works back utility from the exits in New valleys j within the set of possible next valleys $N(i)$. Leads directly to a neural net approximator using “decision blocks” similar to then-current ideas re basal ganglia and “tasks”.

- Despite many discussions, **no apps except options** in robotics “behavior libraries” (e.g. Schaal) yet! Barriers: politics; my time; presence of spatial complexity also in many potential apps! Most “context” better handled by TLRNs.

Moving Window Net: Clue Re Complexity



- Best ZIP Code Digit Recognizer Used “Moving Window” or “conformal” MLP! (Guyon, LeCun, AT&T story, earlier...)
- Exploiting symmetry of Euclidean translation crucial to reducing number of weights, making large input array learnable, outcomes.

Cellular SRN: The Recurrent (SRN) Generalization of "Conformal MLP"

GENERALIZED MAZE PROBLEM

$J_{\text{hat}}(ix, iy)$ for all $0 < ix, iy < N+1$
(an N by N array)

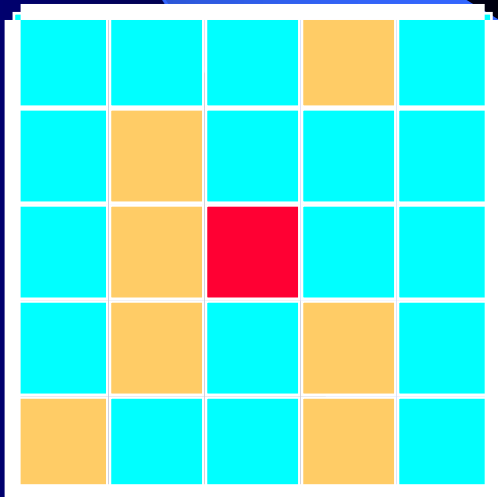
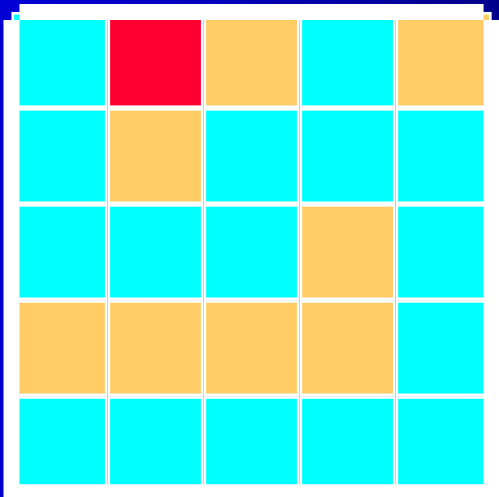
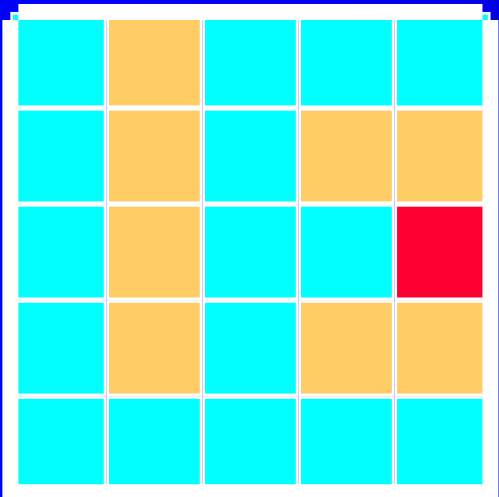
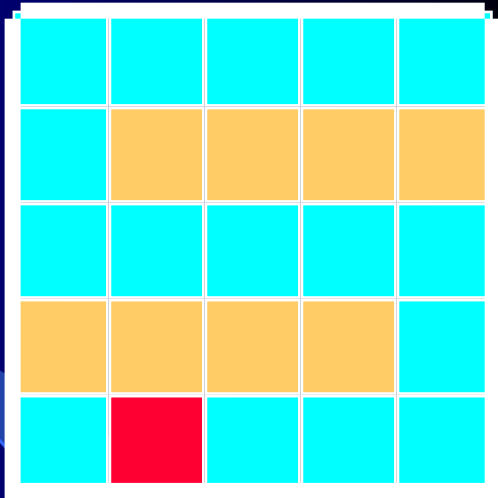
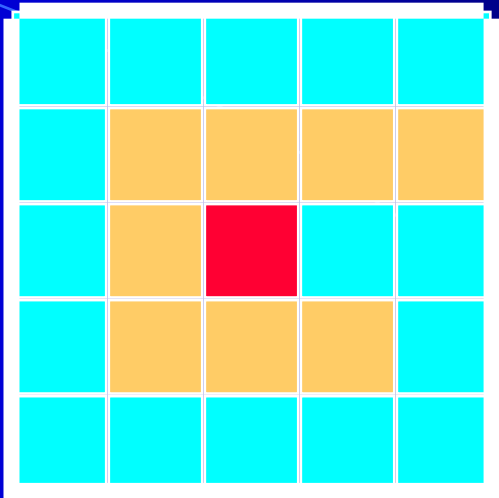
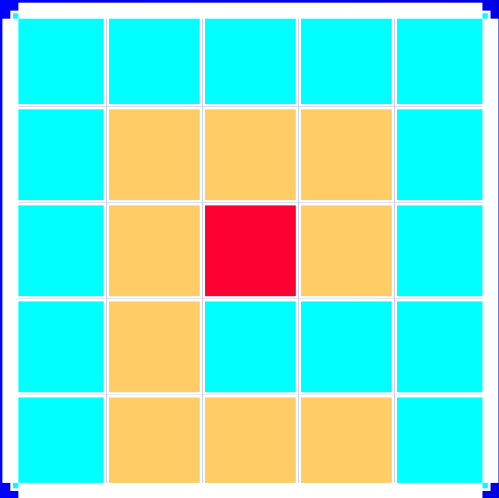
↑
NETWORK

↑
Maze Description

- Obstacle (ix, iy) all ix, iy
- Goal (ix, iy) all ix, iy

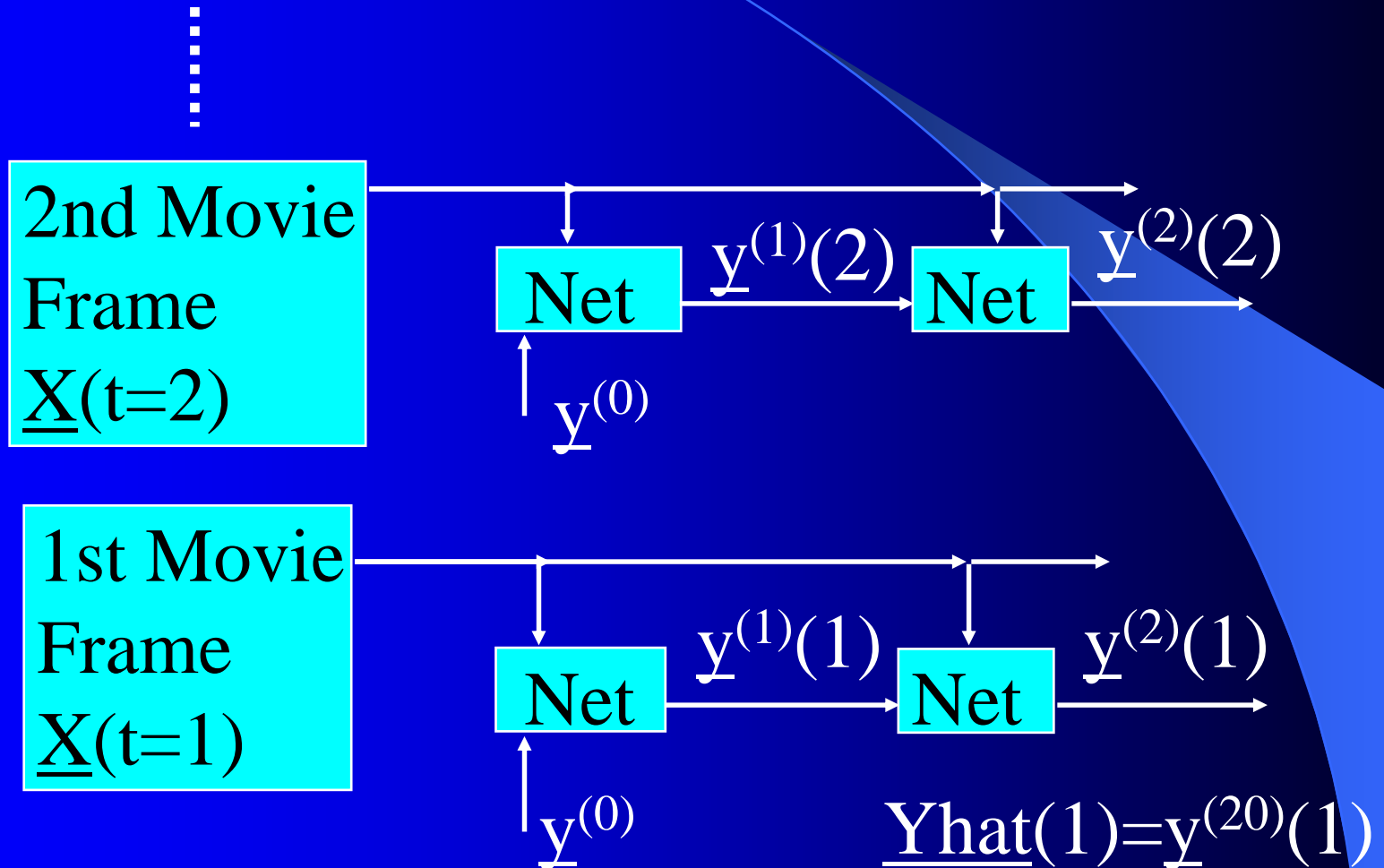
At [arXiv.org](https://arxiv.org), [nlin-sys](https://nlin-sys.com), see [adap-org](https://adap-org.com) 9806001

For rapid practical learning, Ilin, Kozma

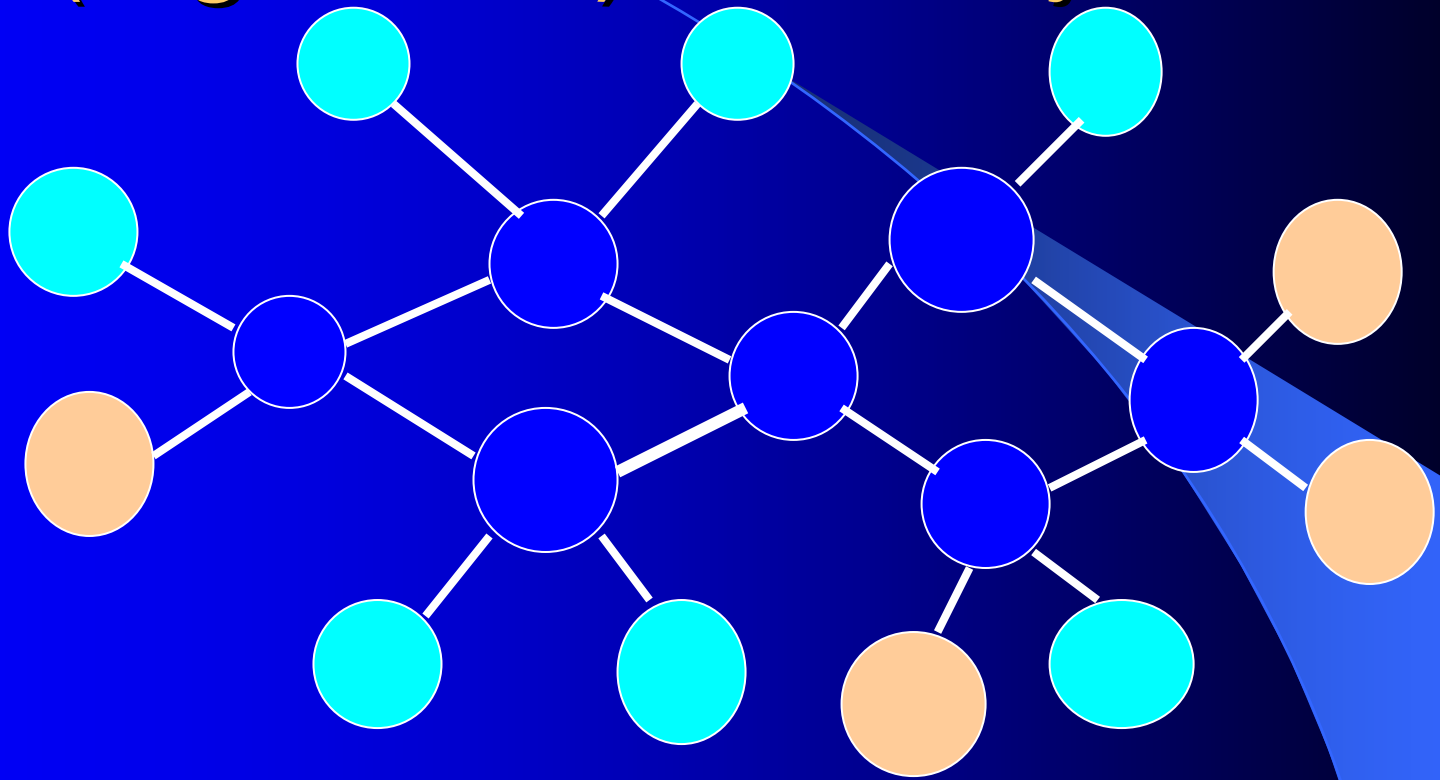


IDEA OF SRN: TWO TIME INDICES t vs.

n



Spatial Symmetry in the General Case (e.g. Grids): the Object Net



- 4 General Object Types (busbar, wire, G, L)
- Net should allow **arbitrary number** of the 4 objects
- How design ANN to input and output FIELDS -- variables like the SET of values for current ACROSS all objects?
- **Great preliminary success** (Fogel's Master Class Chess player; U. Mo. Power)
- **But how learn the objects and the symmetry transformations???? (Brain and images!!)**

From Neural Networks to the Intelligent Power Grid: What It Takes to Make Things Work

- What is an Intelligent Power Grid, and why do we need it?
- Why do we need neural networks?
- How can we make neural nets really work here, & in diagnostics/"prediction"/"control" in general?

Paul J. Werbos, pwerbos@nsf.gov

•“Government public domain”: These slides may be copied, posted, or distributed freely, so long as they are kept together, including this notice. But all views herein are personal, unofficial.

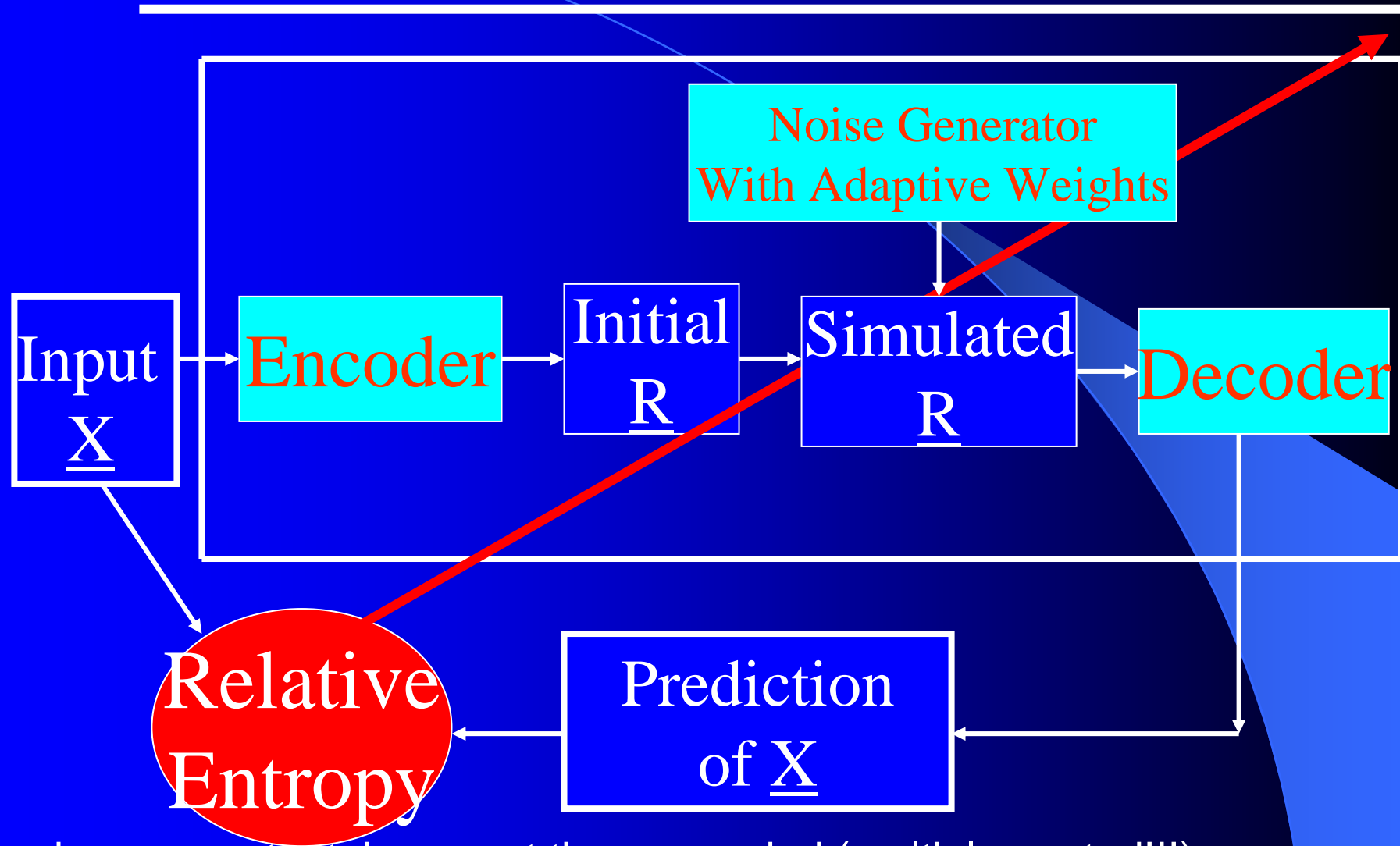
Quick Review of www.face-rec.org

- See Chellappa et al for review
- 3 best recognizers use ANNs, learning
- Wechsler, von der Malsburg: need to learn elastic symmetry transformations (e.g. curl up mouth), not just Euclidean
- Low-lying fruit: use CSRN or Object Net to learn elastic symmetry transformations, but how does brain do it? Foveal vision doesn't have Euclidean metric symmetry. (Though topology helps, connection learning.)

Fresh Look: Initial Approach to Brain-Like Symmetry Learning and Use

- First learn a family of vector maps \underline{f}_α such that:
 - $\Pr(\underline{f}_\alpha(\underline{x}(t+1)) | \underline{f}_\alpha(\underline{x}(t))) = \Pr(\underline{x}(t+1) | \underline{x}(t))$ for the same conditional probability distribution \Pr and all α .
- Exploit these symmetries via:
 - “Reverberatory generalization”: after observing or remembering the pair $\{\underline{x}(t+1), \underline{x}(t)\}$, also train on $\{\underline{f}_\alpha(\underline{x}(t+1)), \underline{f}_\alpha(\underline{x}(t))\}$.
 - “Multiplex gating”: after inputting $\underline{x}(t)$, pick α to map \underline{x} to $\underline{f}_\alpha(\underline{x}(t))$, and use that as input to a “universal” canonical prediction model. (e.g. Olshausen. Not the same as spontaneous or affective or salience gating.)
 - “Multimodular gating”: like multiplexed, but implement parallel (coordinated) copies of the canonical model to allow use on multiple objects in parallel at the same time.
- Human brains seem to exploit the first two (or second), but **how are the symmetry transforms learned? How far can a purely emergent kind of design get by learning?**

Solution Exists Off-the Shelf! (SEDP, HIC Chapter 13)

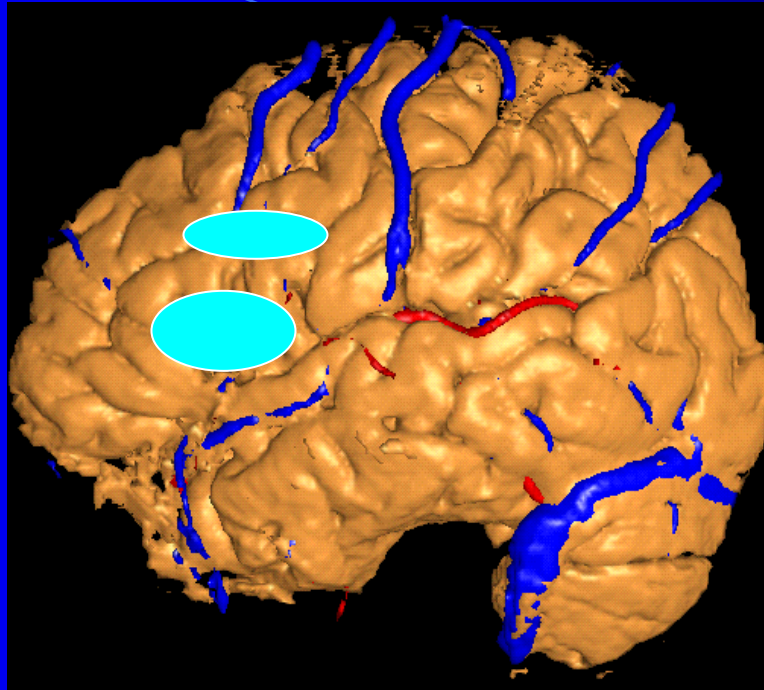


Learning symmetry takes most time; encoded (multiplex gated!!!) image allows fast learning of objects, faces, etc., as if brute force gating or transformation encoding!!

To get from SEDP to full Mammal Brain Like Spatial Complexity:

- Work to improve learning speed, robustness & generalization in SRN, TLRN, CSRN, Object Net, GDHP and SEDP – including memory-based learning as discussed often, & analysis of mathematical properties, toolkits, etc.
- Active control of saccade & efferent copy to encoder
- Test short-term object permanence (automatic), and augment long-term memory I/O interface for “object identity” and “world modeling.”

New Data on Complexity in the Brain



Petrides (IJCNN06) shows that dorsolateral (DL) and orbitofrontal (OF) prefrontal cortex – our “highest” brain centers – answer two basic questions:
OF: Where did I leave my car this time in the parking lot? (**space?**)
DL: What was I trying to do anyway? (**time?**)

- BUT: even bird brains (no neocortex) handle great spatial complexity & have big basal ganglia!!
- Hypothesis: SEDP fits pyramid cell geometry very well but is already be in old cortex (bird!)
- Neocortex (mouse) harnesses/alters stochastic mechanism in SEDP for creativity.
- OF strengthens object identity & world modeling & object-oriented action. (Test birds, lizards!)
- Temporal aggregation is by “re-entrant” mechanism, not explicit temporal hierarchy.

6 MegaChallenges for the 21st Century

- Key Challenges To Basic **Scientific Understanding**:
 - What is **Mind**? (how to build/understand intelligence)
 - Basic Science of Mind: Up To the Highest Kind of General Intelligence We see in the Smallest Mouse
 - Middle Sci. Mind: from Mouse to “Sapient,” the level of full use of symbolic reasoning integrated with meaning/empathy (human brain is “new”, still halfway there in its evolution)
 - Higher Sci. of Mind: Principles of higher levels in intelligent systems design, like quantum, multimodular, soul
 - How does the **Universe work**? (Quantum physics...)
 - What is **Life**? (e.g., quantitative systems biotechnology)
- Key Broader Challenges to **Humanity**:
 - **Sustainable growth** on earth
 - Global sustainable **energy/environment** & mid-term survival
 - “**yin sustainability**,” e.g. population, related women’s issues, peace
 - Cost-effective sustainable **space** settlement
 - Human potential -- growth/learning in brain, soul, integration (body)

