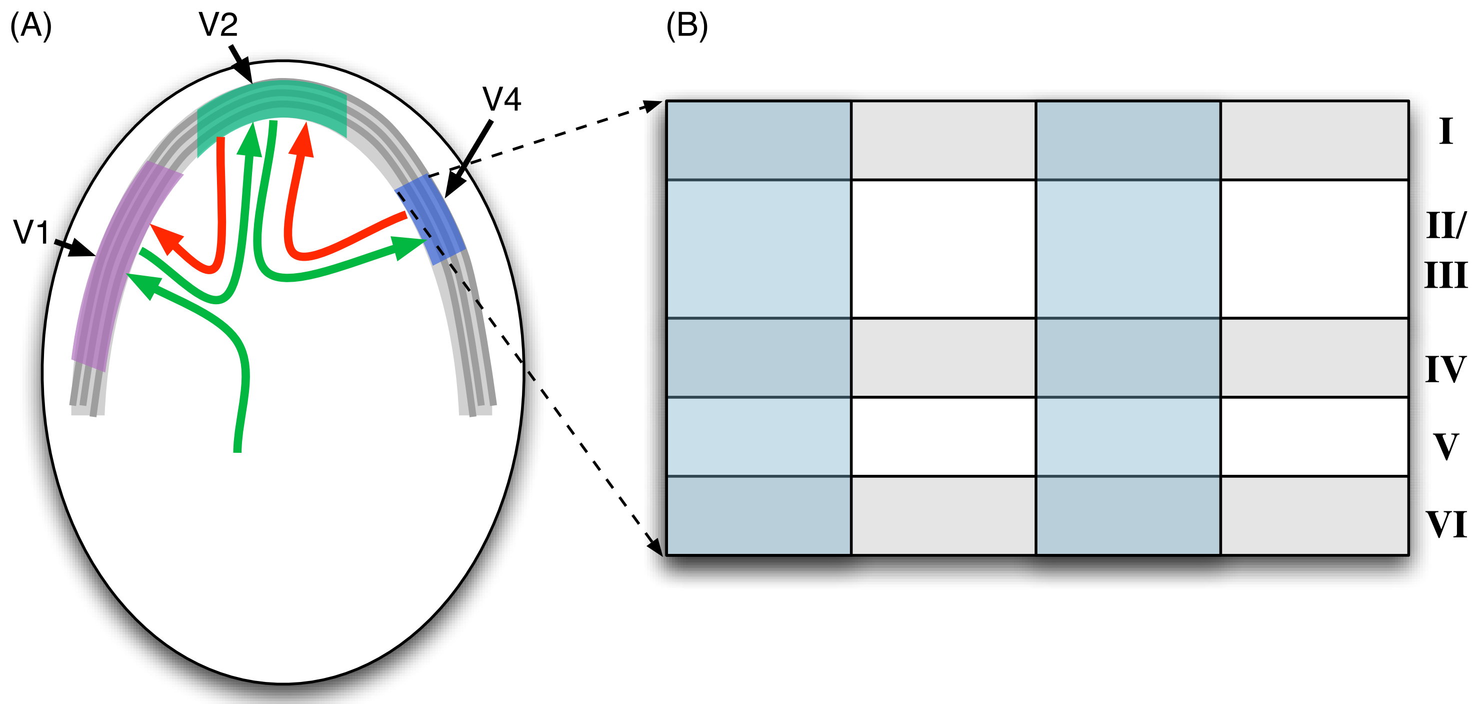
Numenta®

A Mathematical Model for Cortical Circuits

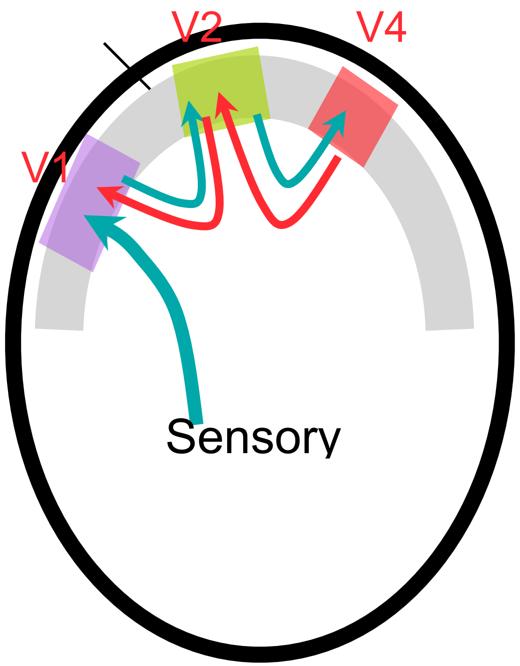
Dileep George

Neocortex Organization : 2 minute version

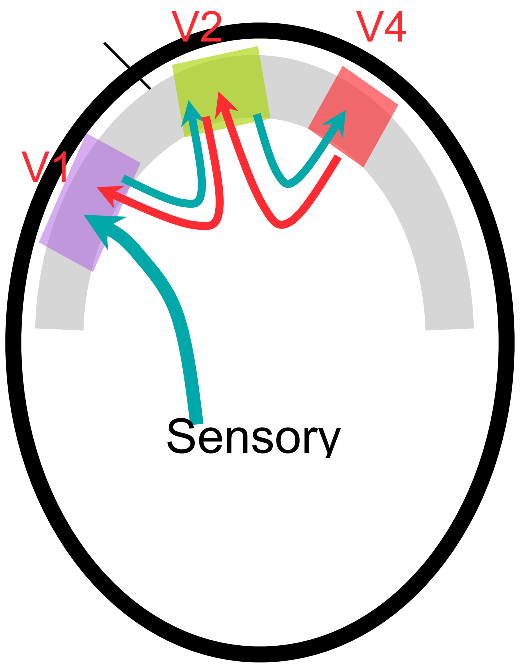


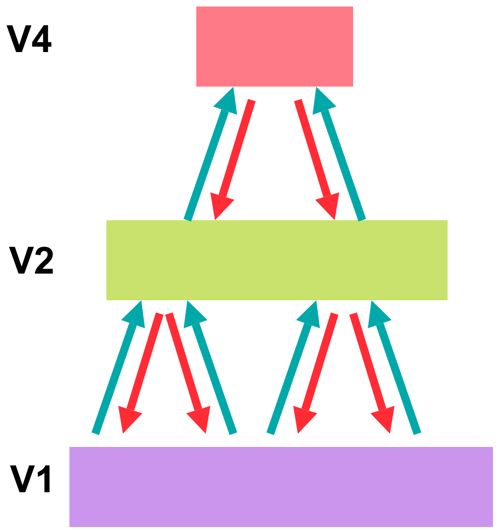
Cortical Hierarchy

Cortical Hierarchy

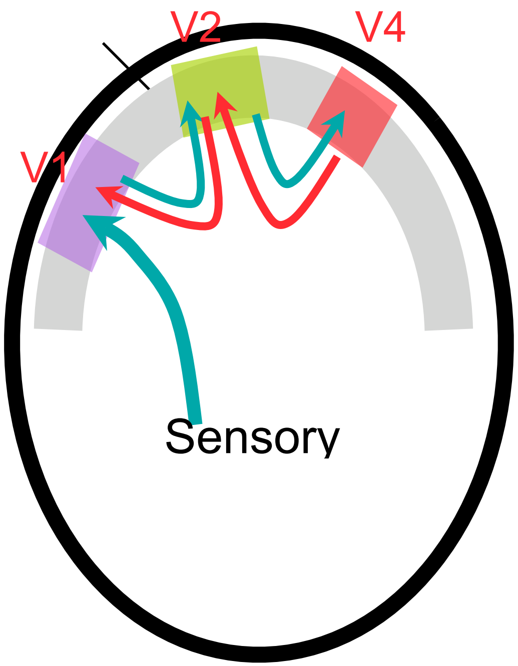


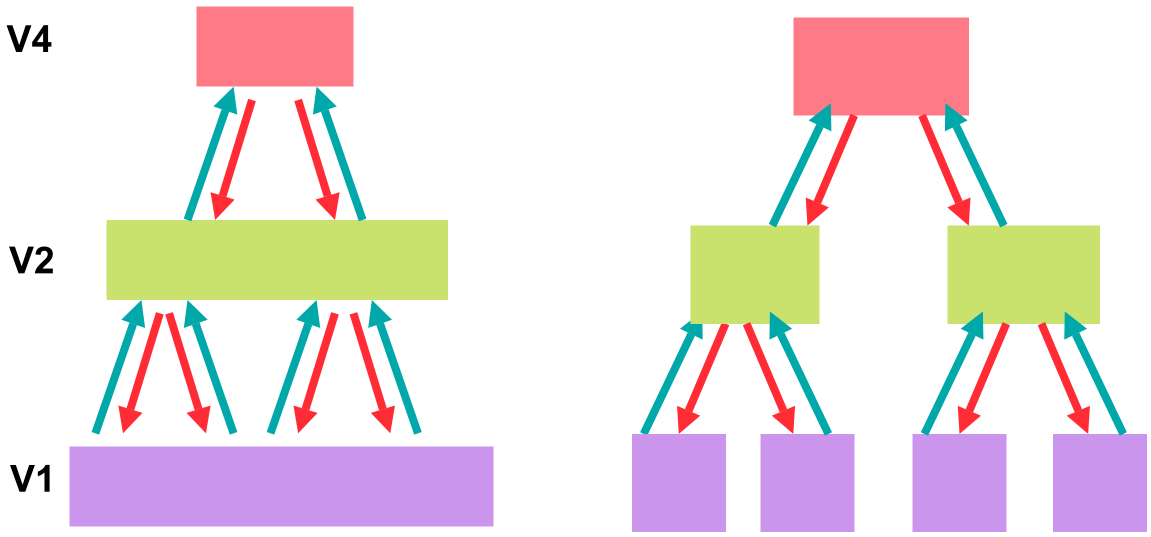
Cortical Hierarchy



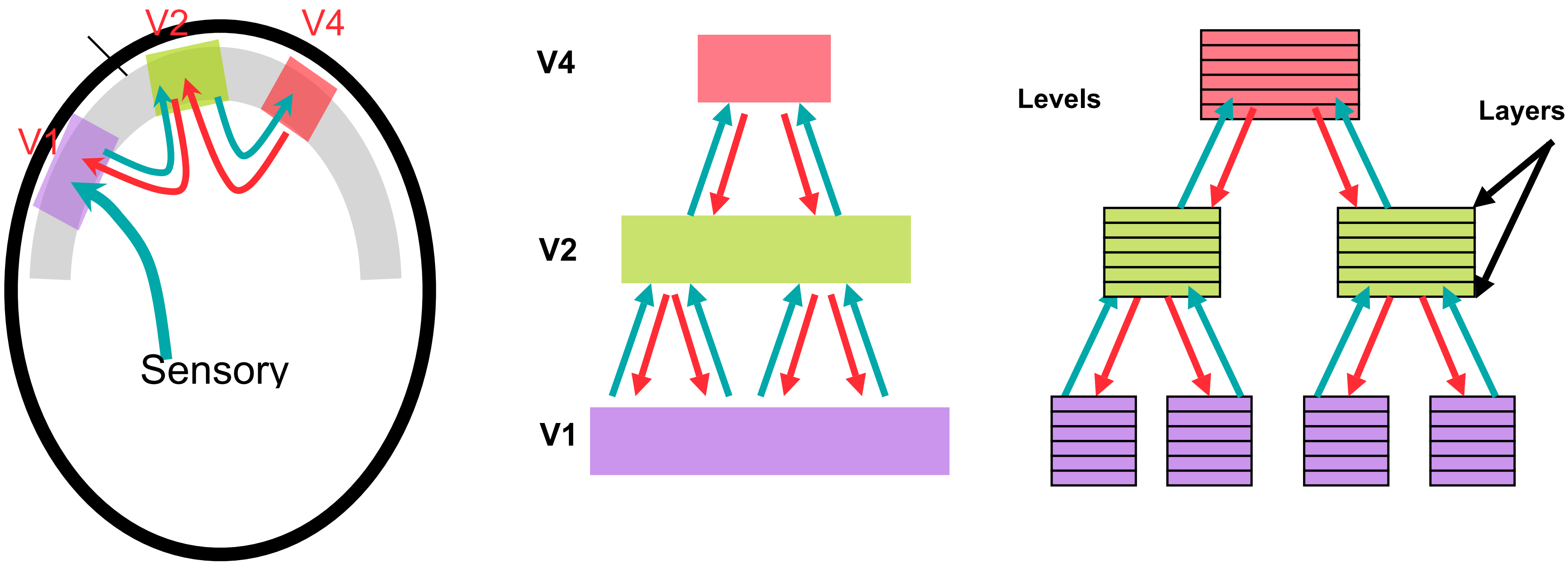


Cortical Hierarchy





Cortical Hierarchy



Hardware/software design challenges

Hardware/software design challenges

- **Adaptive computing**
  - Reconfigures connections.
  - Might need to allocate more resources during the initial stages of learning

Hardware/software design challenges

- **Adaptive computing**
  - Reconfigures connections.
  - Might need to allocate more resources during the initial stages of learning
- **Algorithms are still evolving**
  - Might be mature enough for some application areas
  - Inference algorithms might be more stable than learning algorithms

Hardware/software design challenges

- **Adaptive computing**
  - Reconfigures connections.
  - Might need to allocate more resources during the initial stages of learning
- **Algorithms are still evolving**
  - Might be mature enough for some application areas
  - Inference algorithms might be more stable than learning algorithms
- **Solution?**
  - Abstract and formalize
  - Parameterize the degrees of freedom

Aspiration

Aspiration

- Can we come to a set of mathematical equations that model cortical function?

Aspiration

- Can we come to a set of mathematical equations that model cortical function?
  - Something like Maxwell’s equations for electromagnetic waves

Aspiration

- Can we come to a set of mathematical equations that model cortical function?
  - Something like Maxwell's equations for electromagnetic waves
  - Assigns functions/equations for laminar circuits

Aspiration

- Can we come to a set of mathematical equations that model cortical function?
  - Something like Maxwell's equations for electromagnetic waves
  - Assigns functions/equations for laminar circuits

Aspiration

- Can we come to a set of mathematical equations that model cortical function?
  - Something like Maxwell's equations for electromagnetic waves
  - Assigns functions/equations for laminar circuits

- Audacious

Aspiration

- Can we come to a set of mathematical equations that model cortical function?
  - Something like Maxwell's equations for electromagnetic waves
  - Assigns functions/equations for laminar circuits
- Audacious
  - But this is the season of Hope and Change

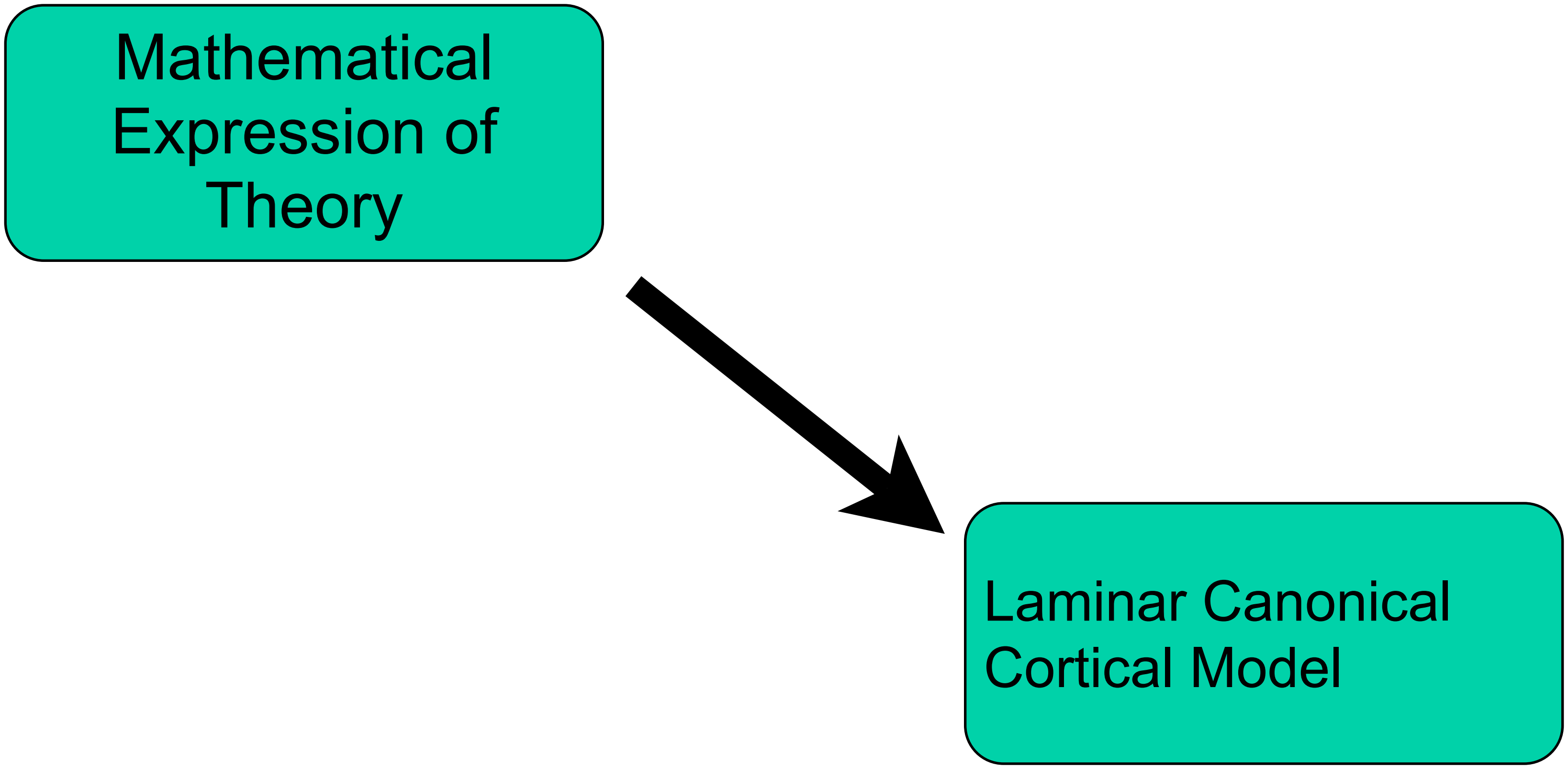
Aspiration

- Can we come to a set of mathematical equations that model cortical function?
  - Something like Maxwell's equations for electromagnetic waves
  - Assigns functions/equations for laminar circuits
- Audacious
  - But this is the season of Hope and Change
  - Yes we can!!

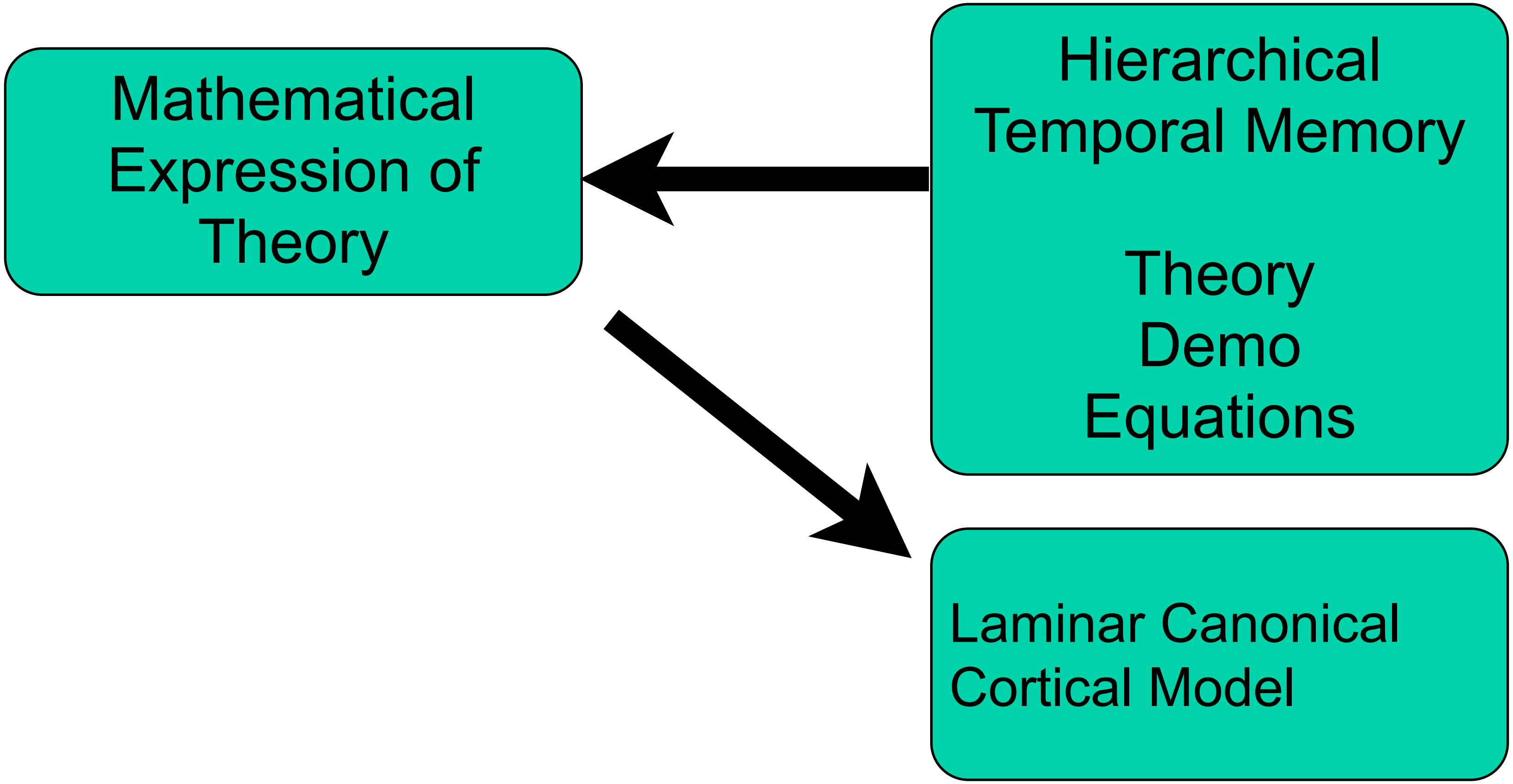
Agenda

Mathematical Expression of Theory

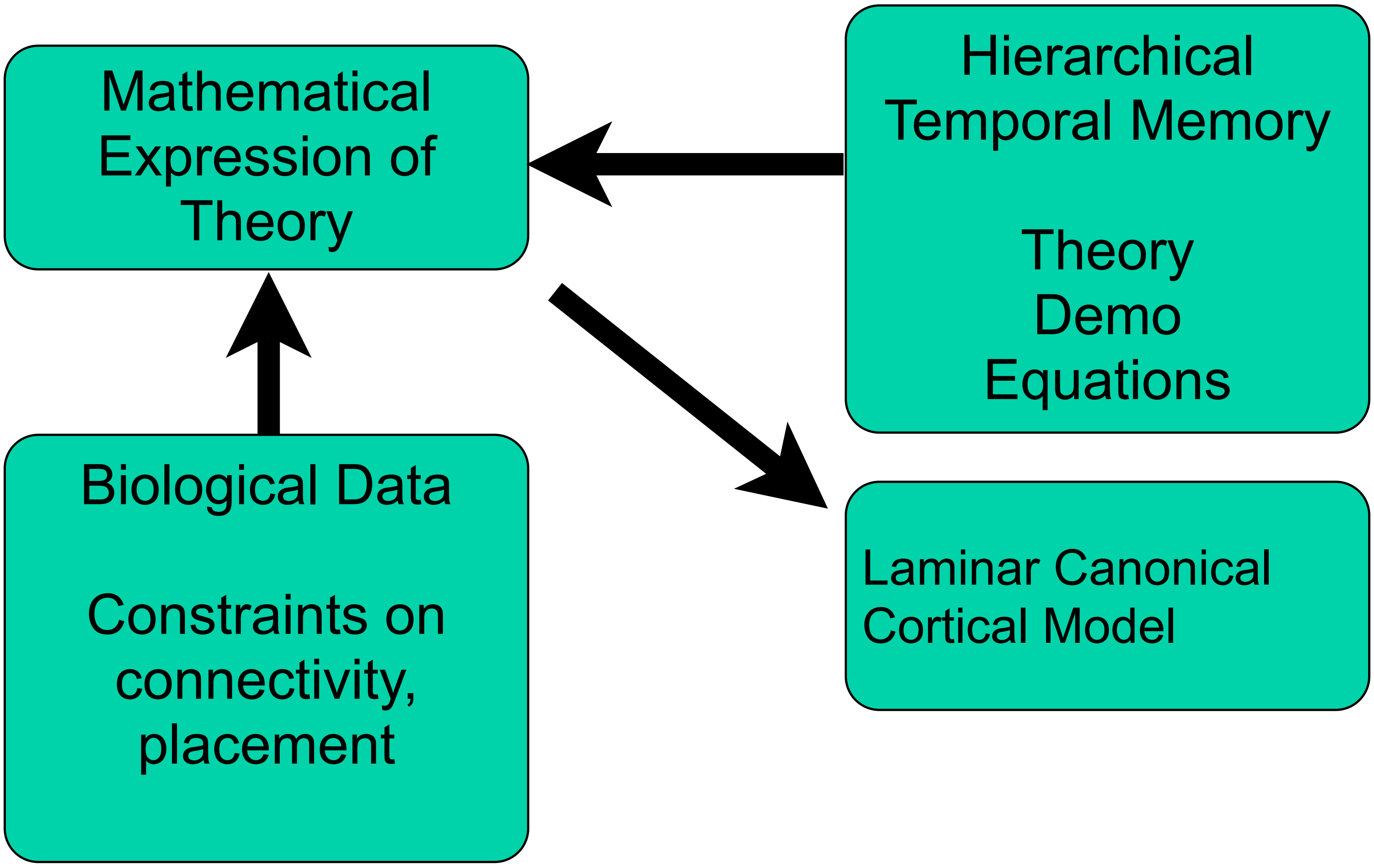
Agenda


Mathematical Expression of Theory
Laminar Canonical Cortical Model

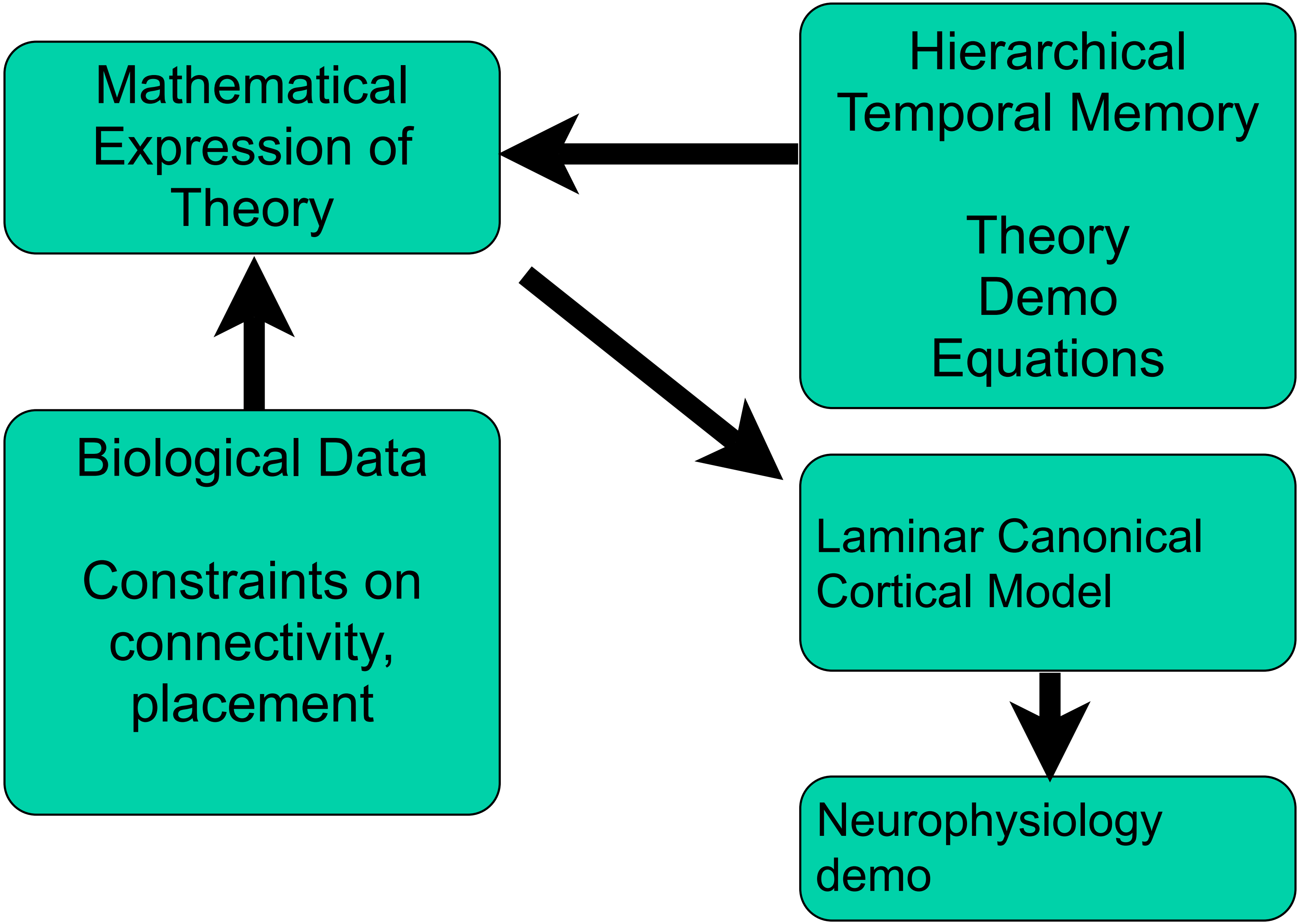
Agenda


Mathematical Expression of Theory
Hierarchical Temporal Memory
Theory
Demo
Equations
Laminar Canonical Cortical Model

Agenda


Mathematical Expression of Theory
Hierarchical Temporal Memory
Theory
Demo
Equations
Biological Data
Constraints on connectivity, placement
Laminar Canonical Cortical Model

Agenda


Mathematical Expression of Theory
Hierarchical Temporal Memory
Theory
Demo
Equations
Biological Data
Constraints on connectivity, placement
Laminar Canonical Cortical Model
Neurophysiology demo

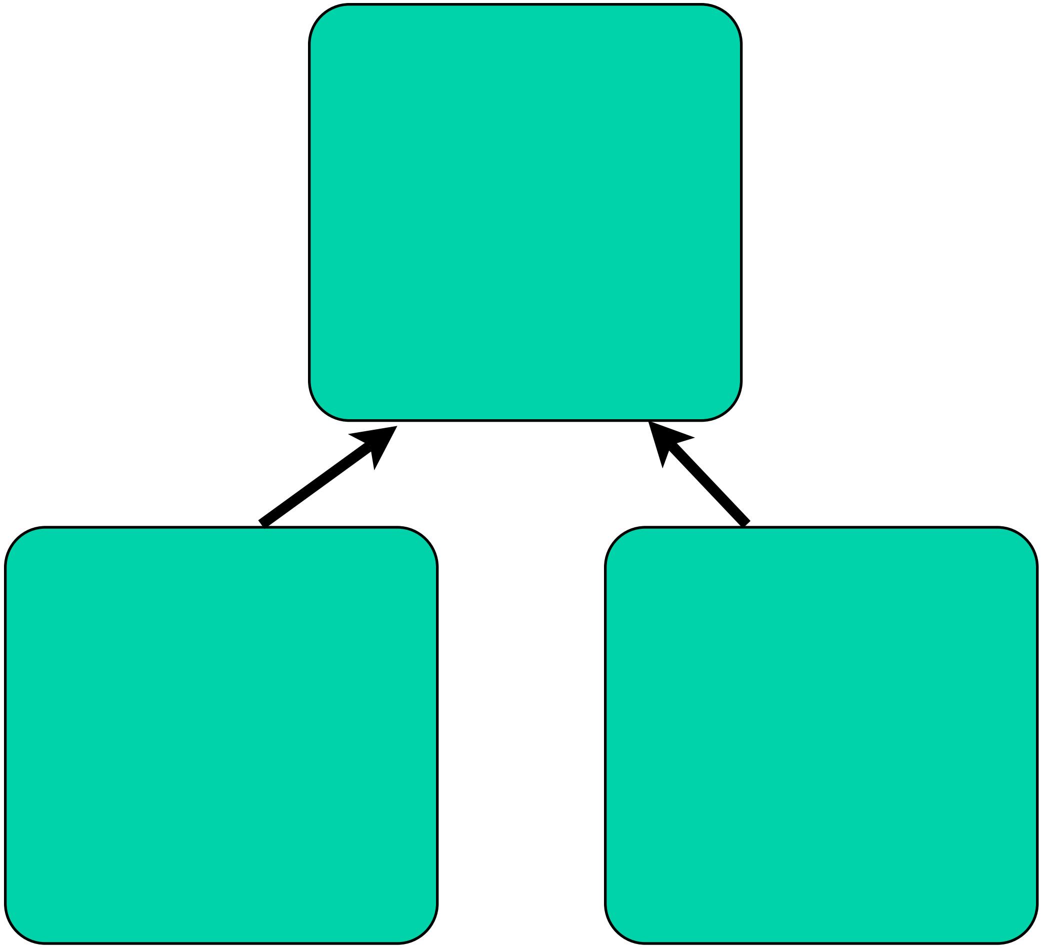
Hierarchical Temporal Memory

- Is a model for neocortical operation
  - Nodes organized in a hierarchy
  - Each node uses the same algorithm

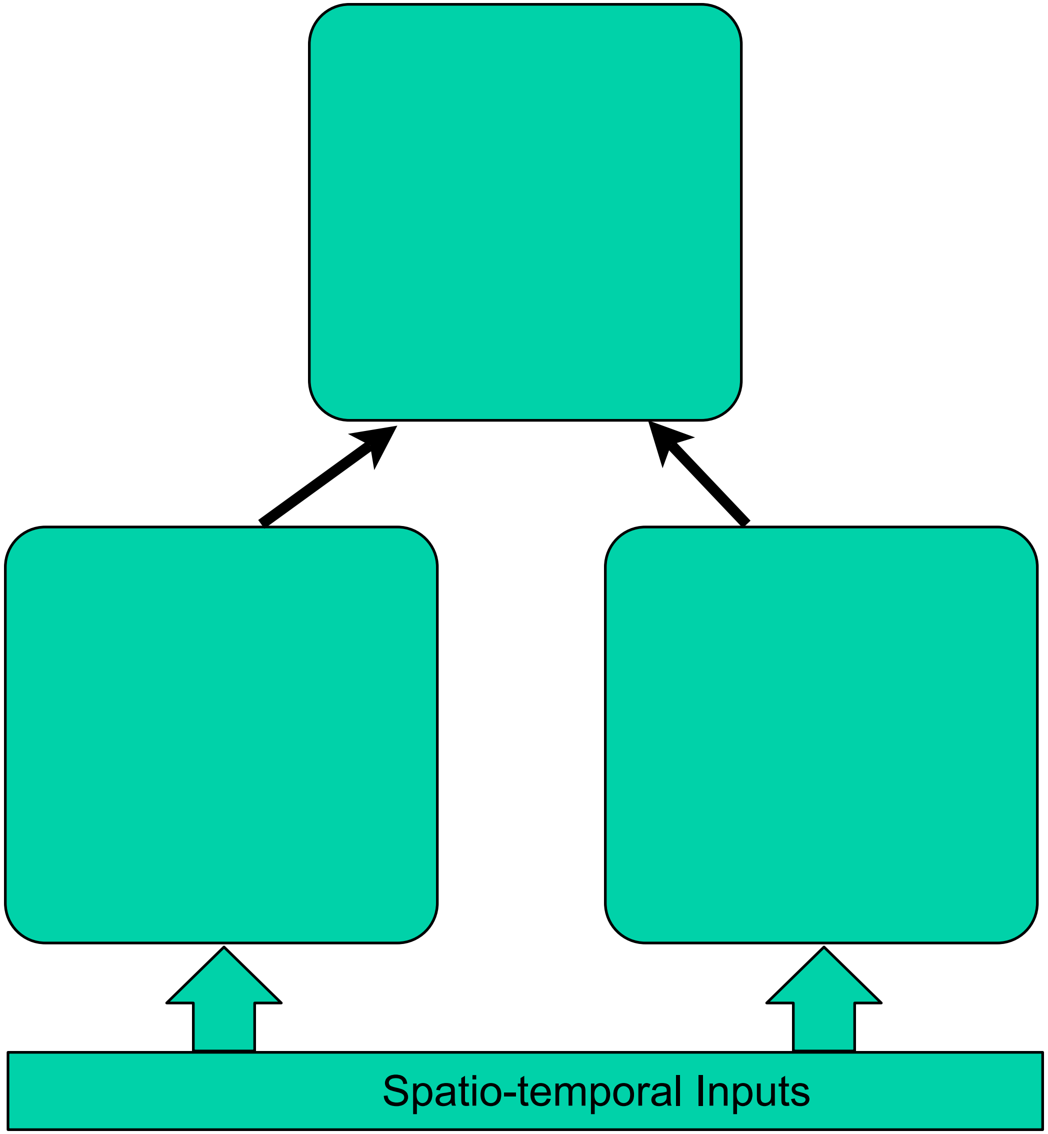
Features of HTM (Principles of Cortex-like computation)

- Hierarchy in space and time
- Feed-forward and feedback connections
- Common cortical algorithm
- Inference using Bayesian belief propagation
- Sparse Distributed Representations
- Prediction using temporal context
- Biologically accurate

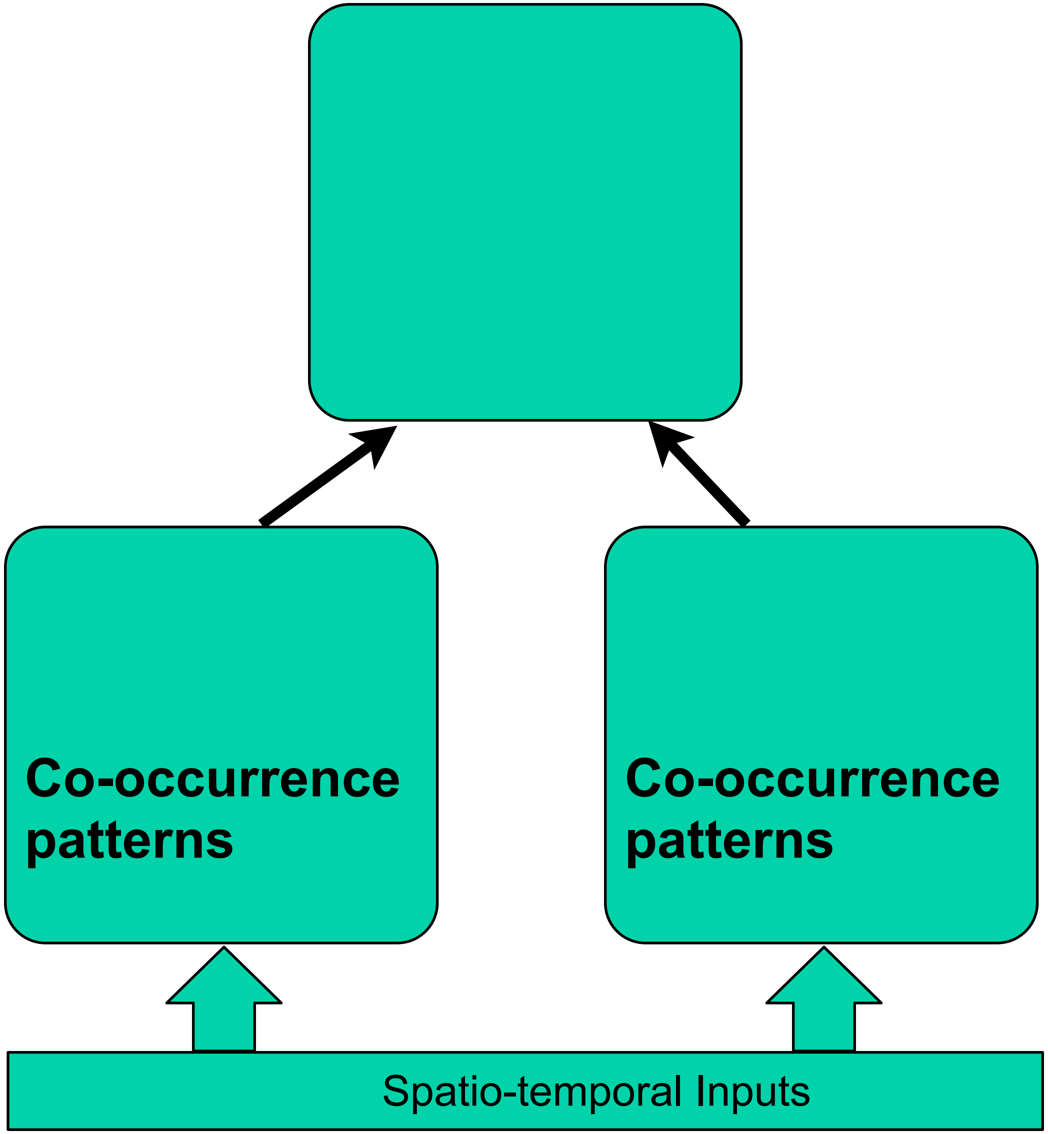
Hierarchical Temporal Memory



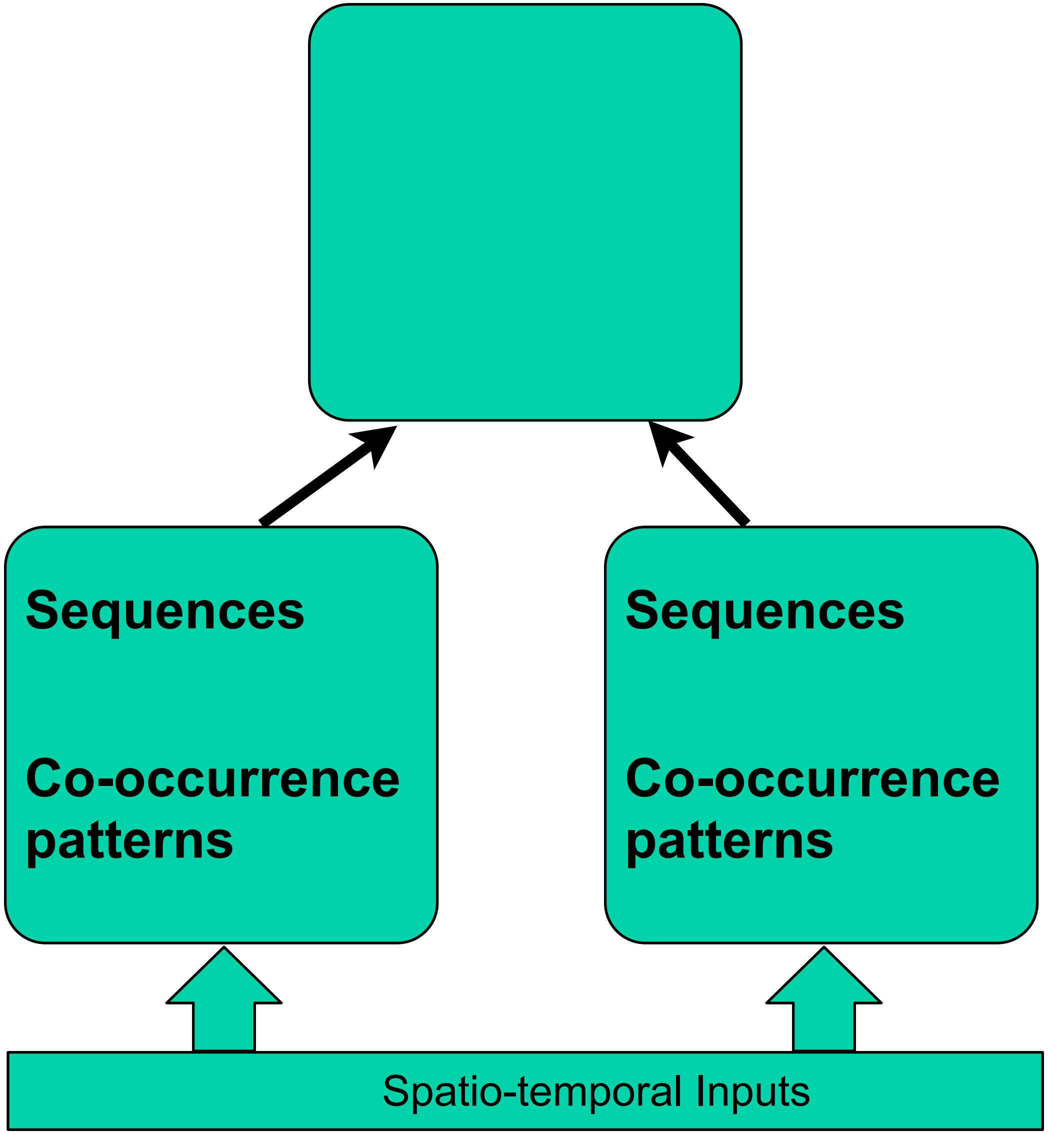
Hierarchical Temporal Memory



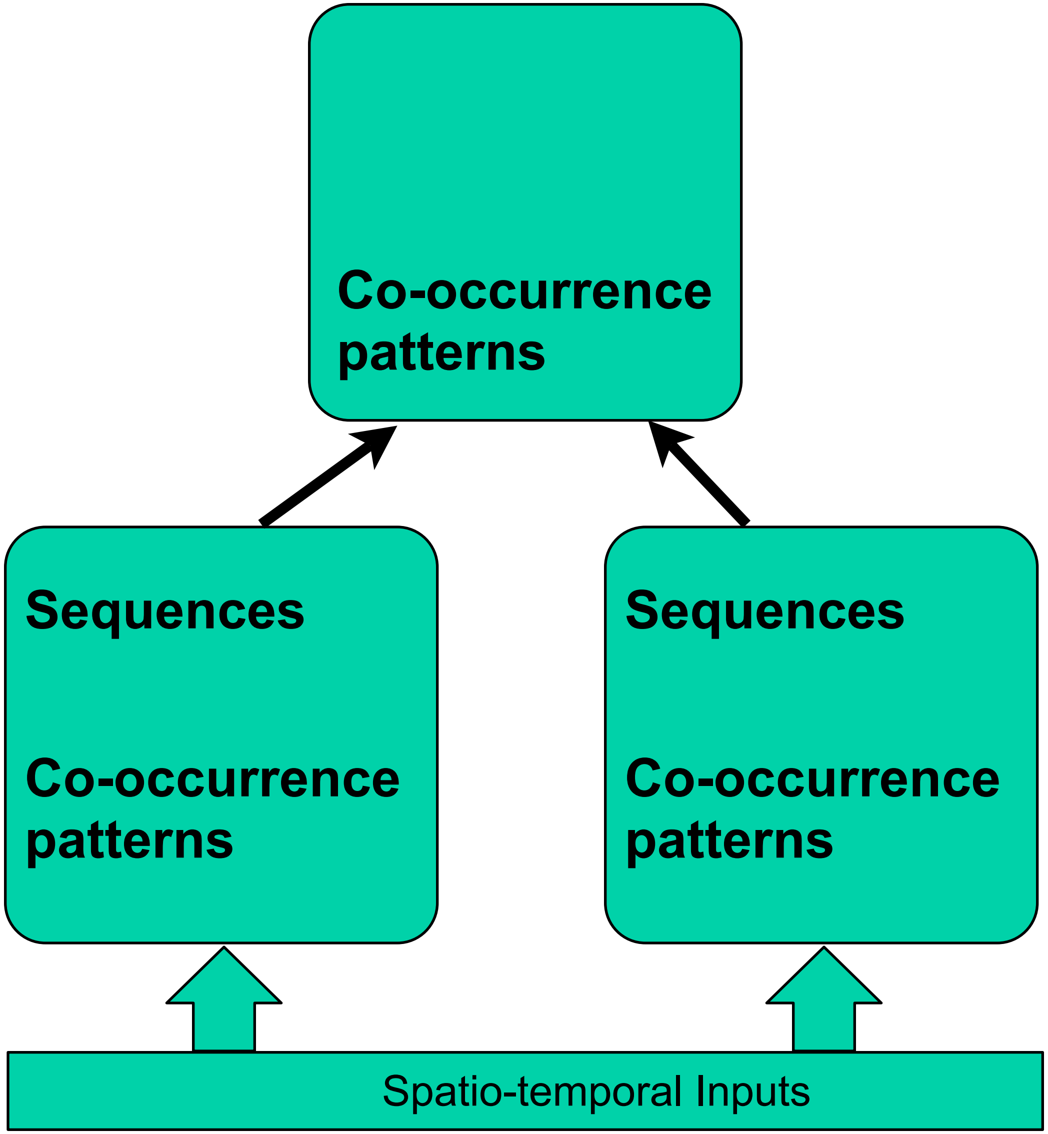
Hierarchical Temporal Memory



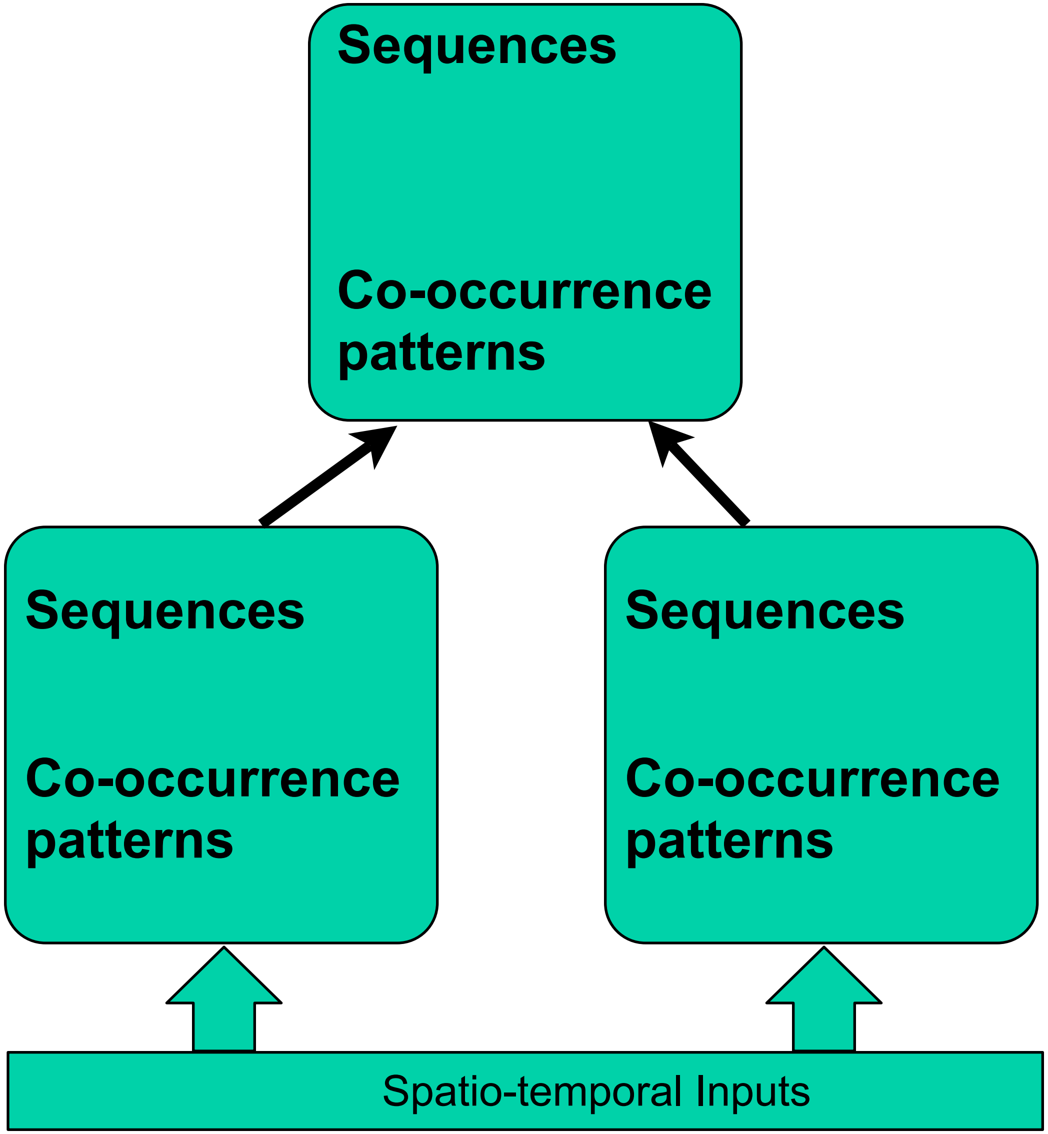
Hierarchical Temporal Memory



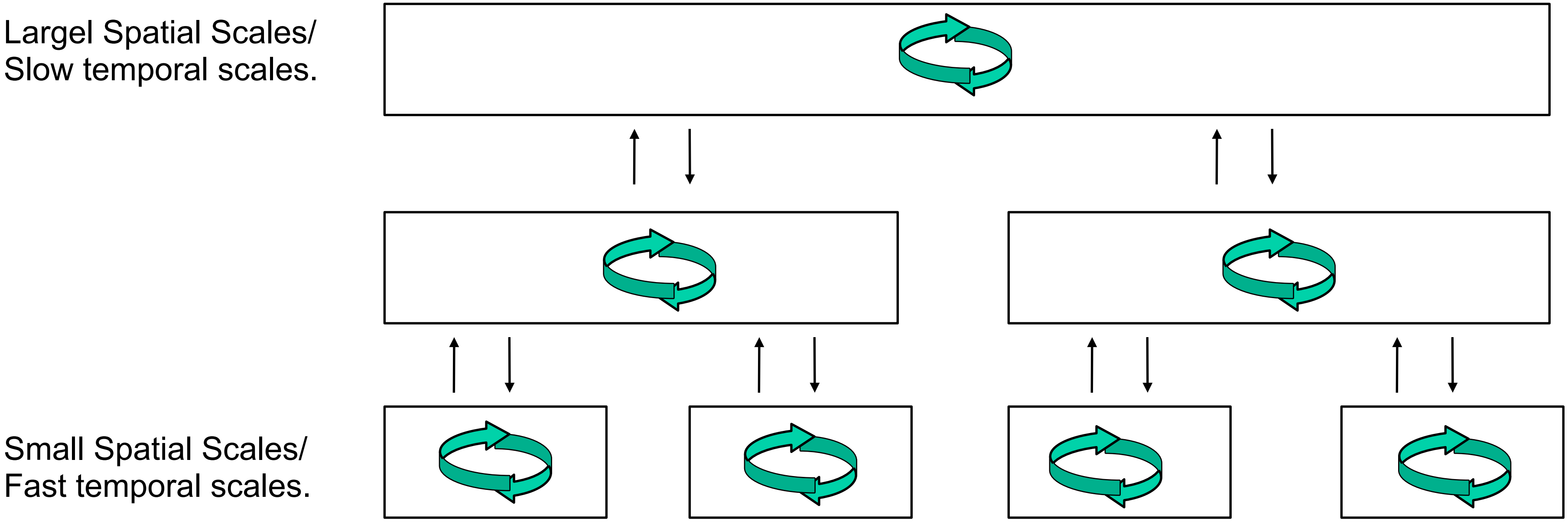
Hierarchical Temporal Memory



Hierarchical Temporal Memory

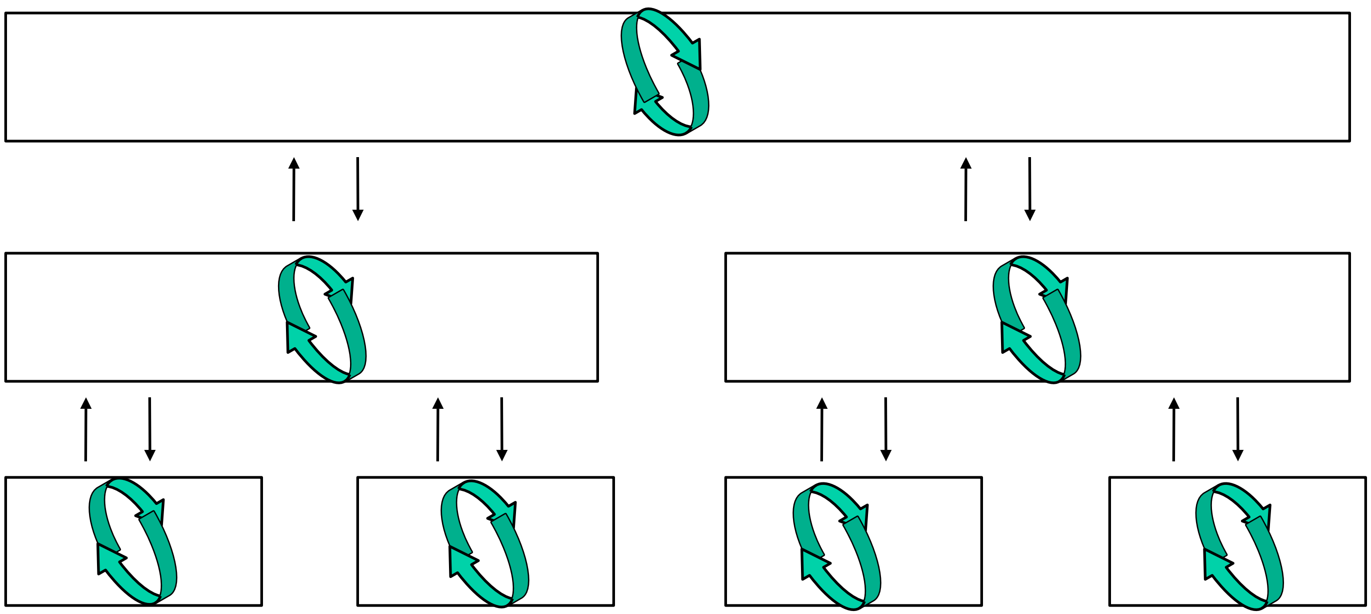


Spatial/temporal hierarchies in the world



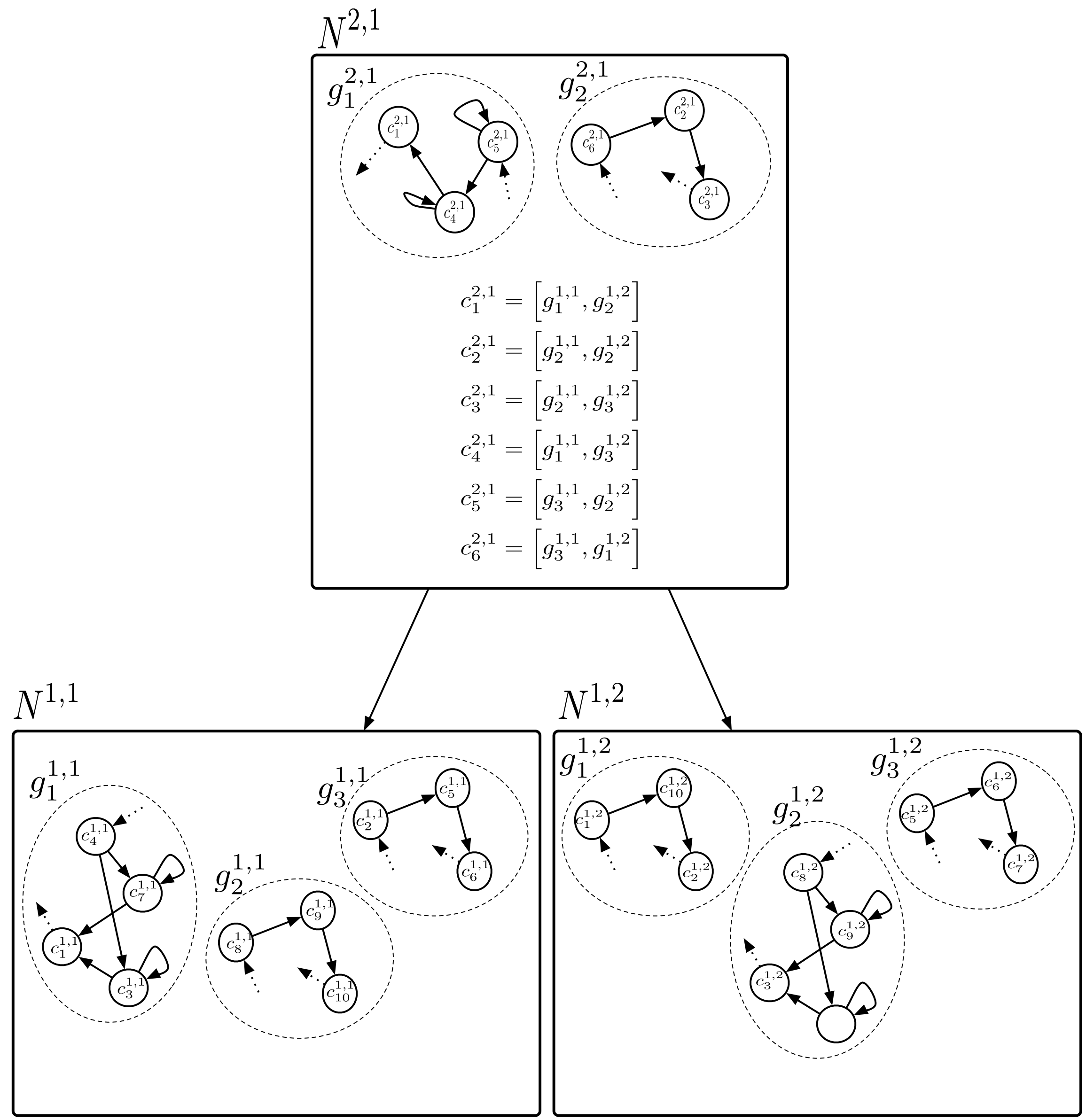
Spatial/temporal hierarchies in the world

Largel Spatial Scales/
Slow temporal scales.



Small Spatial Scales/
Fast temporal scales.

Generative Model : Hierarchical Temporal World

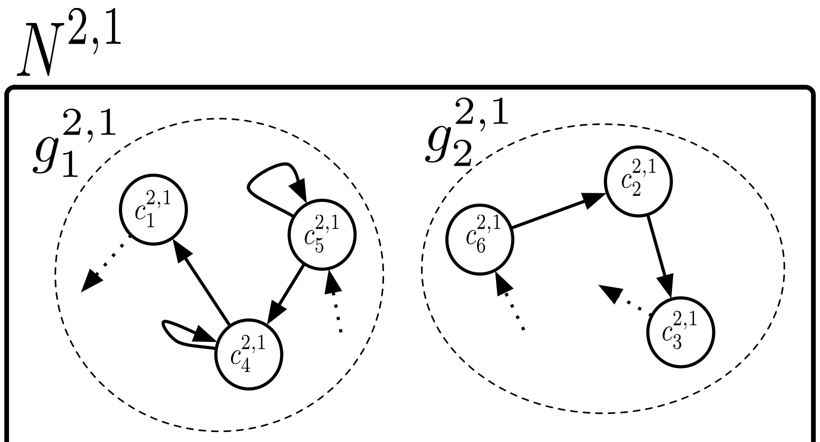


Learning in a node

- Level by level learning
- Each node
  - Stores co-occurrence patterns
  - Learns sequences

Learning in a node

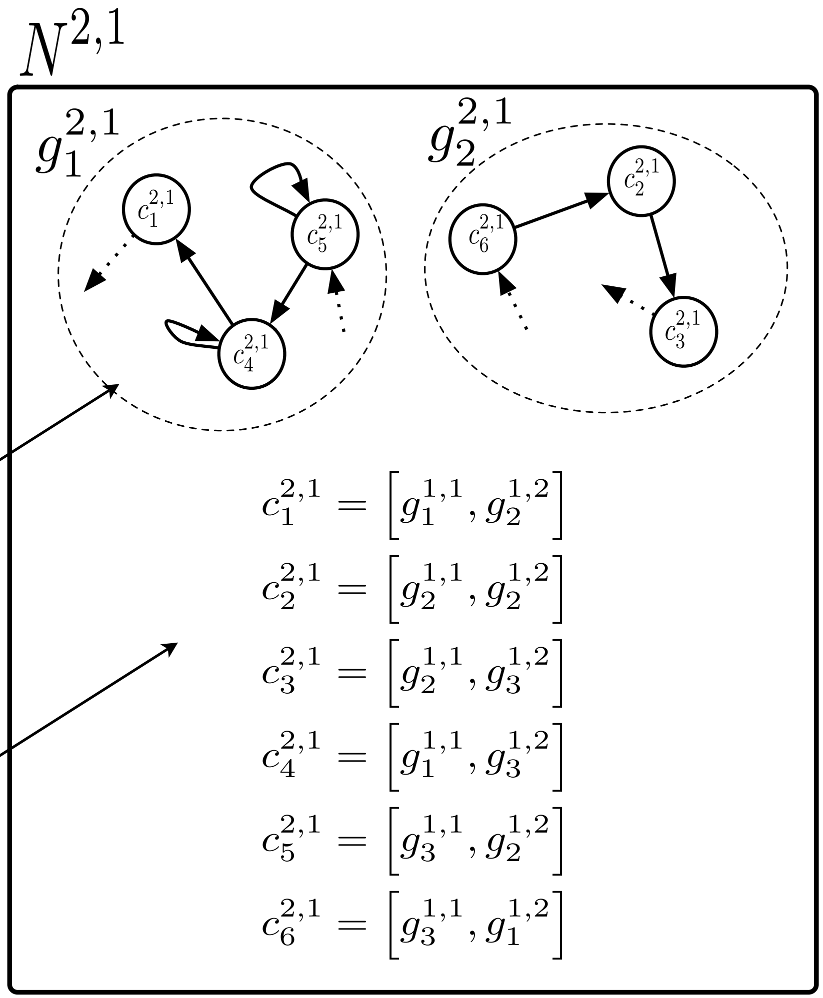
- Level by level learning
- Each node
  - Stores co-occurrence patterns
  - Learns sequences



$$c_1^{2,1} = \left[g_1^{1,1}, g_2^{1,2}\right]$$
$$c_2^{2,1} = \left[g_2^{1,1}, g_2^{1,2}\right]$$
$$c_3^{2,1} = \left[g_2^{1,1}, g_3^{1,2}\right]$$
$$c_4^{2,1} = \left[g_1^{1,1}, g_3^{1,2}\right]$$
$$c_5^{2,1} = \left[g_3^{1,1}, g_2^{1,2}\right]$$
$$c_6^{2,1} = \left[g_3^{1,1}, g_1^{1,2}\right]$$

Learning in a node

- Level by level learning
- Each node
  - Stores co-occurrence patterns
  - Learns sequences



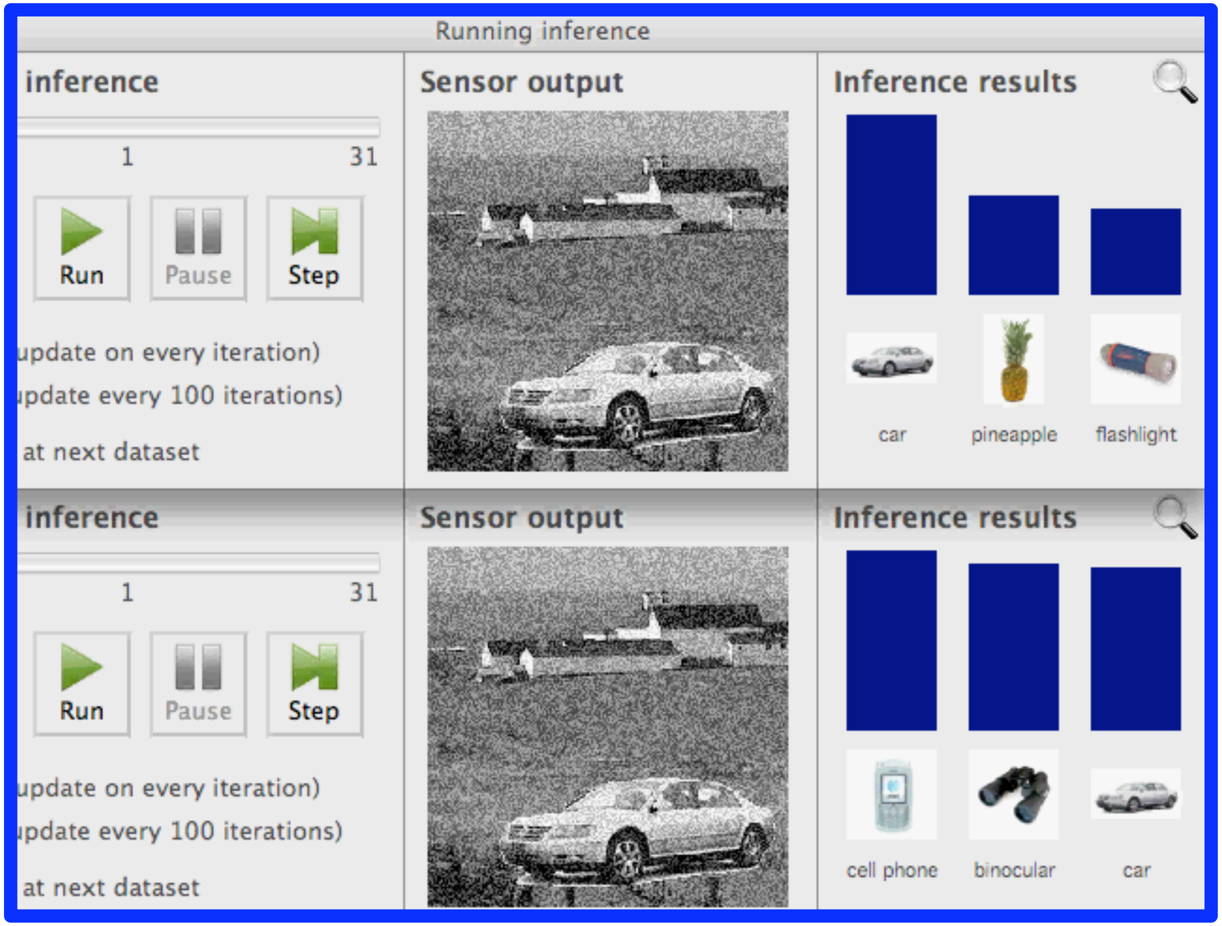
Markov chains (sequences)

Coincidence patterns

Demos

Video surveillance

Temporal Inference Demo (Preliminary Results)



← TBI

← non-TBI

What is happening behind the scenes?

Recognition/Prediction/Attention in HTMs

- Inference is done using Bayesian Belief Propagation on HTM hierarchy
  - *Probabilistic Reasoning in Intelligent Systems* by Judeal Pearl

Bayesian computation in the brain

Bayesian computation in the brain

- Hierarchical Bayesian inference is gaining acceptance as the framework for understanding cortical computation
    - Lee and Mumford 2003

Bayesian computation in the brain

- Hierarchical Bayesian inference is gaining acceptance as the framework for understanding cortical computation
    - Lee and Mumford 2003

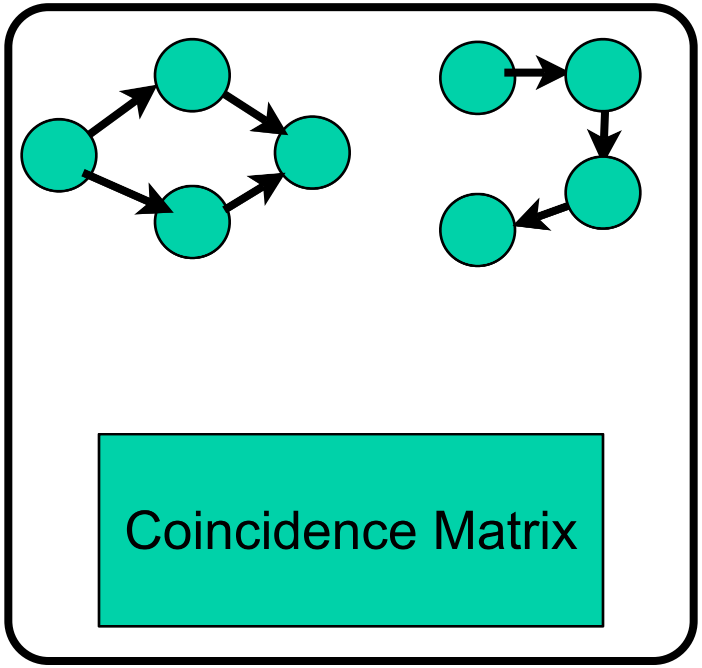
Bayesian computation in the brain

- Hierarchical Bayesian inference is gaining acceptance as the framework for understanding cortical computation
  - Lee and Mumford 2003
- "Bayesian framework is not yet a neural model. [Bayesian] framework currently helps explain the computations that underlie various brain functions, but not how the brain implements those computations"
  - Hegde & Felleman 2007

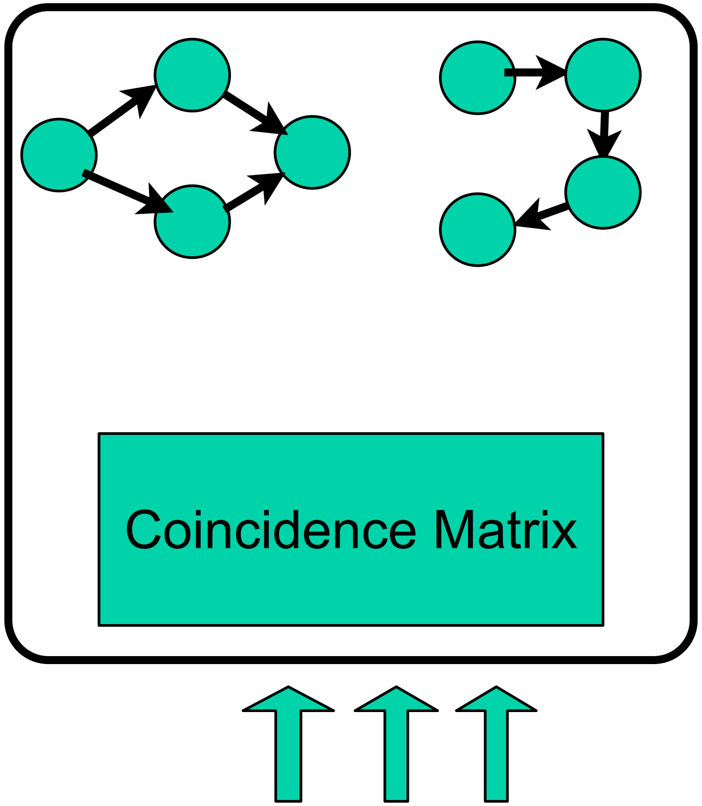
Our hypothesis

Bayesian Belief Propagation equations on HTMs correspond to the operations done by cortical microcircuits during inference.

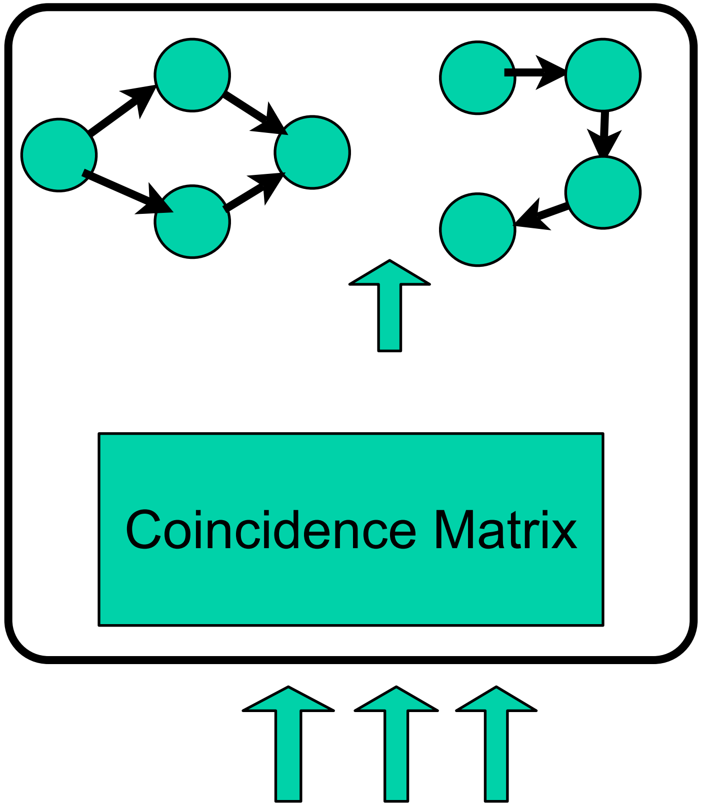
Inference in an HTM node



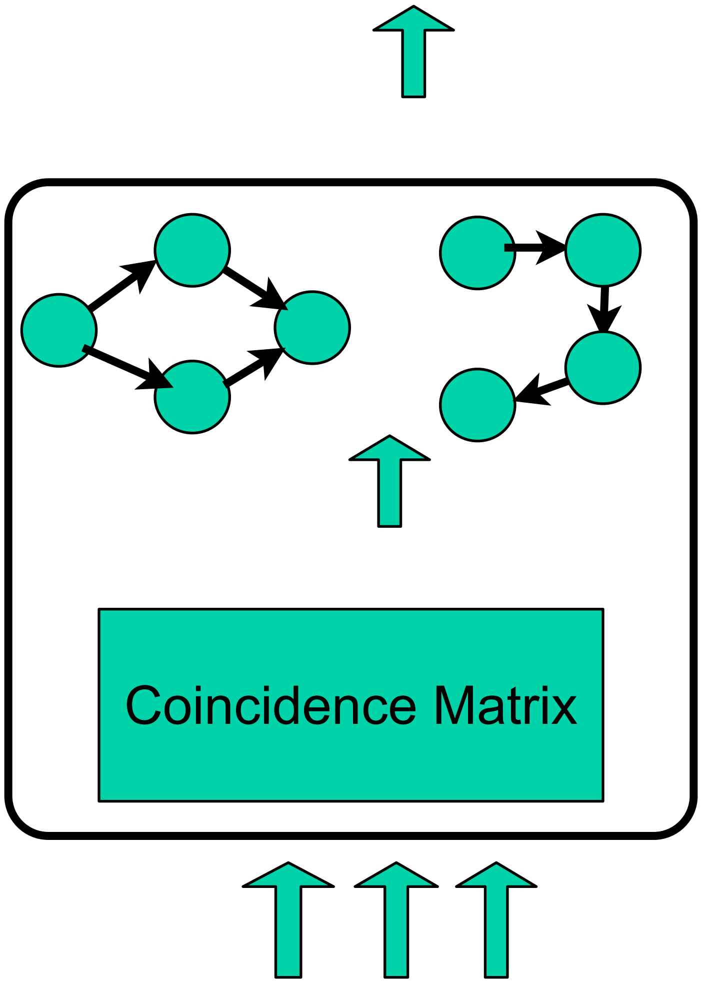
Inference in an HTM node



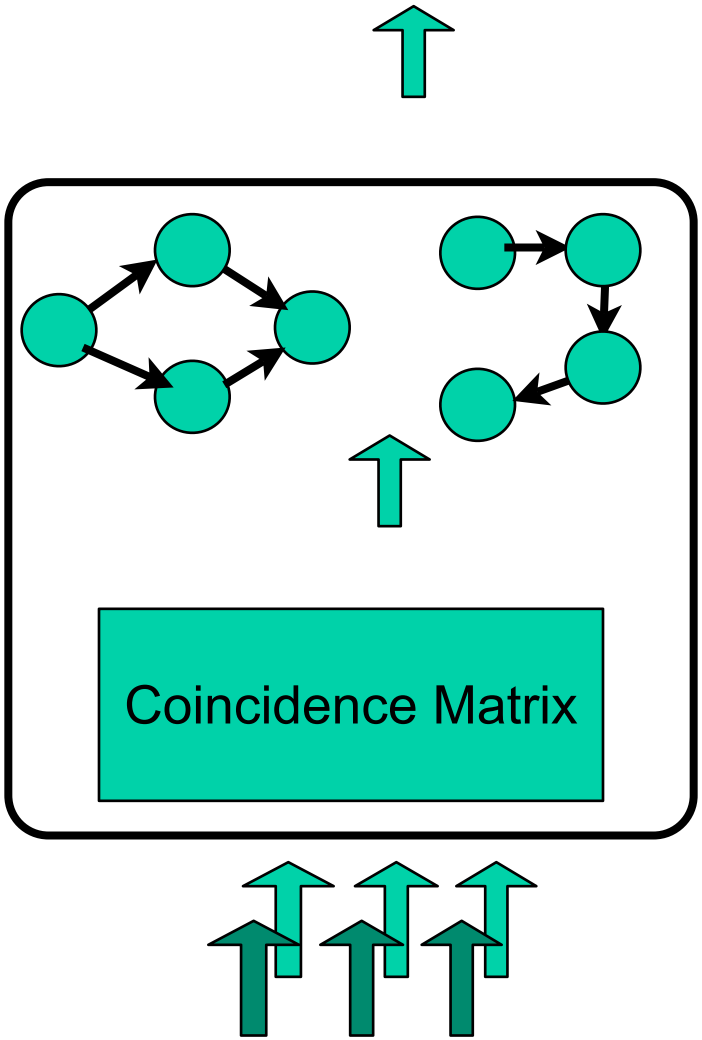
Inference in an HTM node



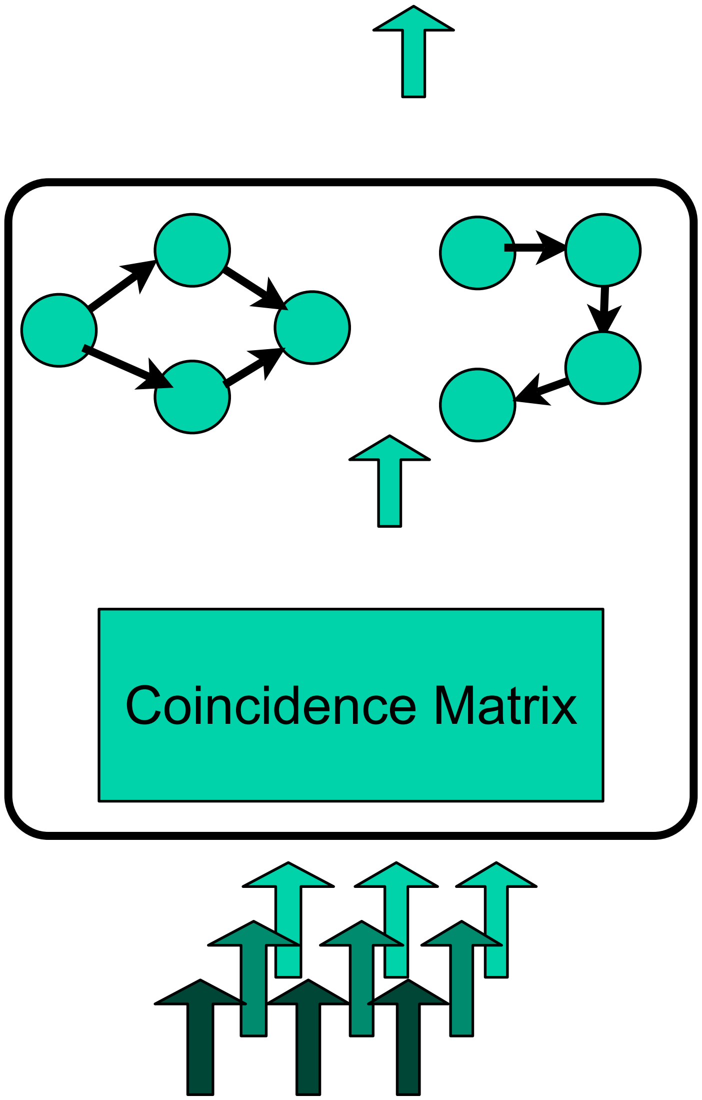
Inference in an HTM node



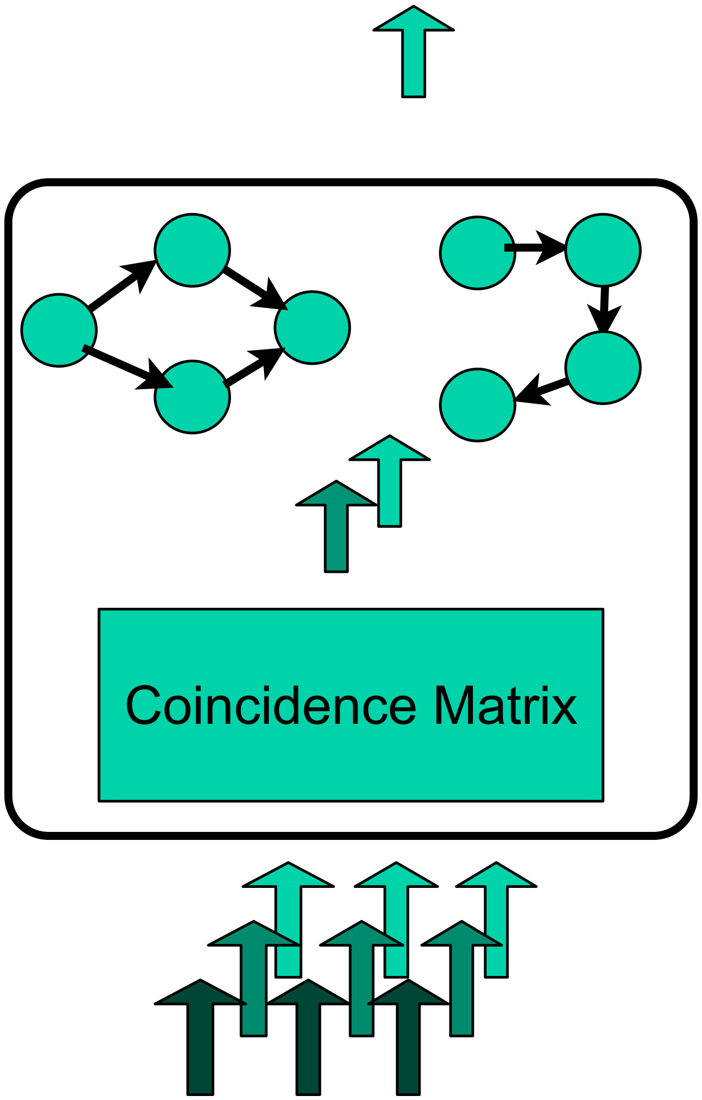
Inference in an HTM node



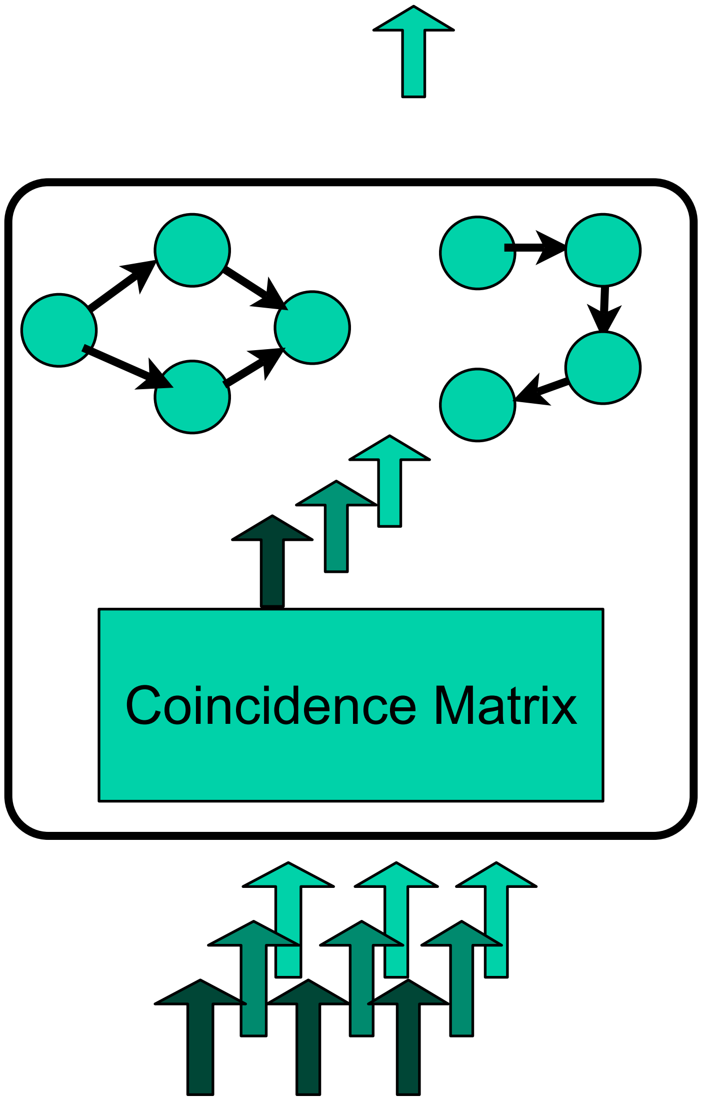
Inference in an HTM node



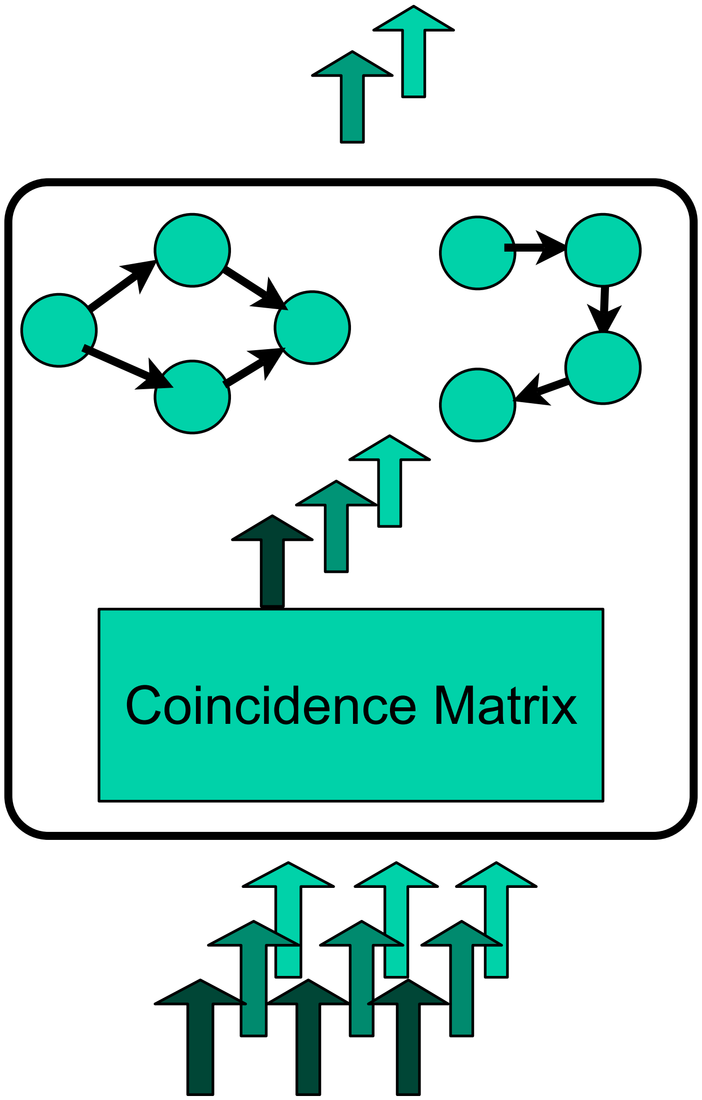
Inference in an HTM node



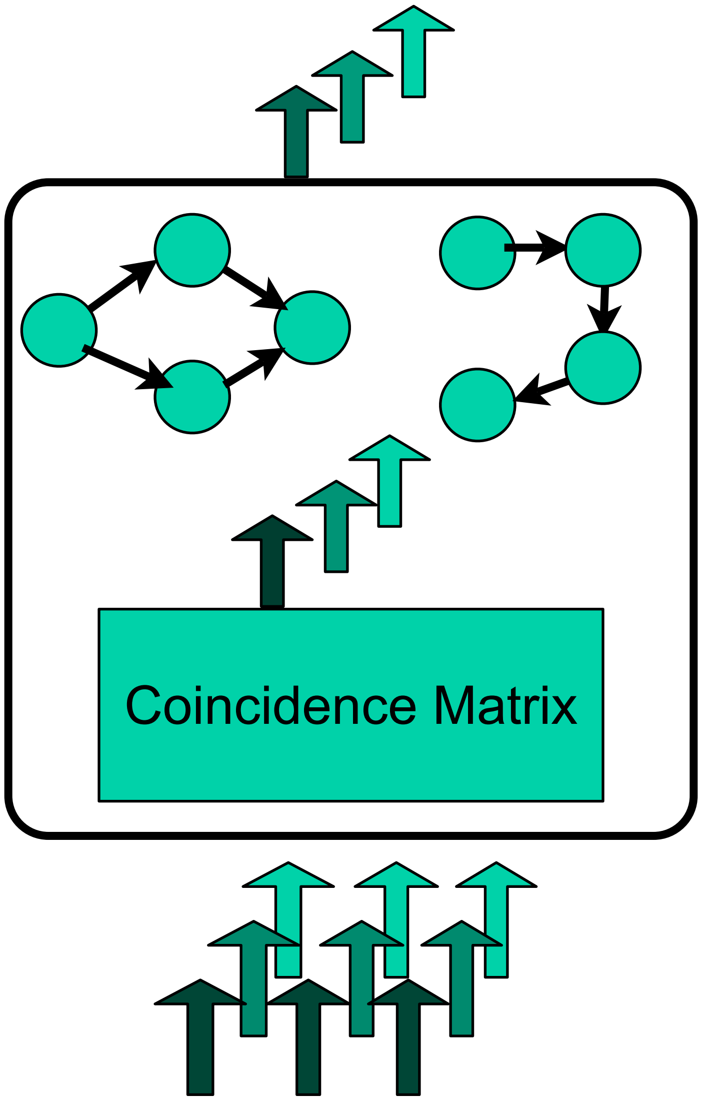
Inference in an HTM node



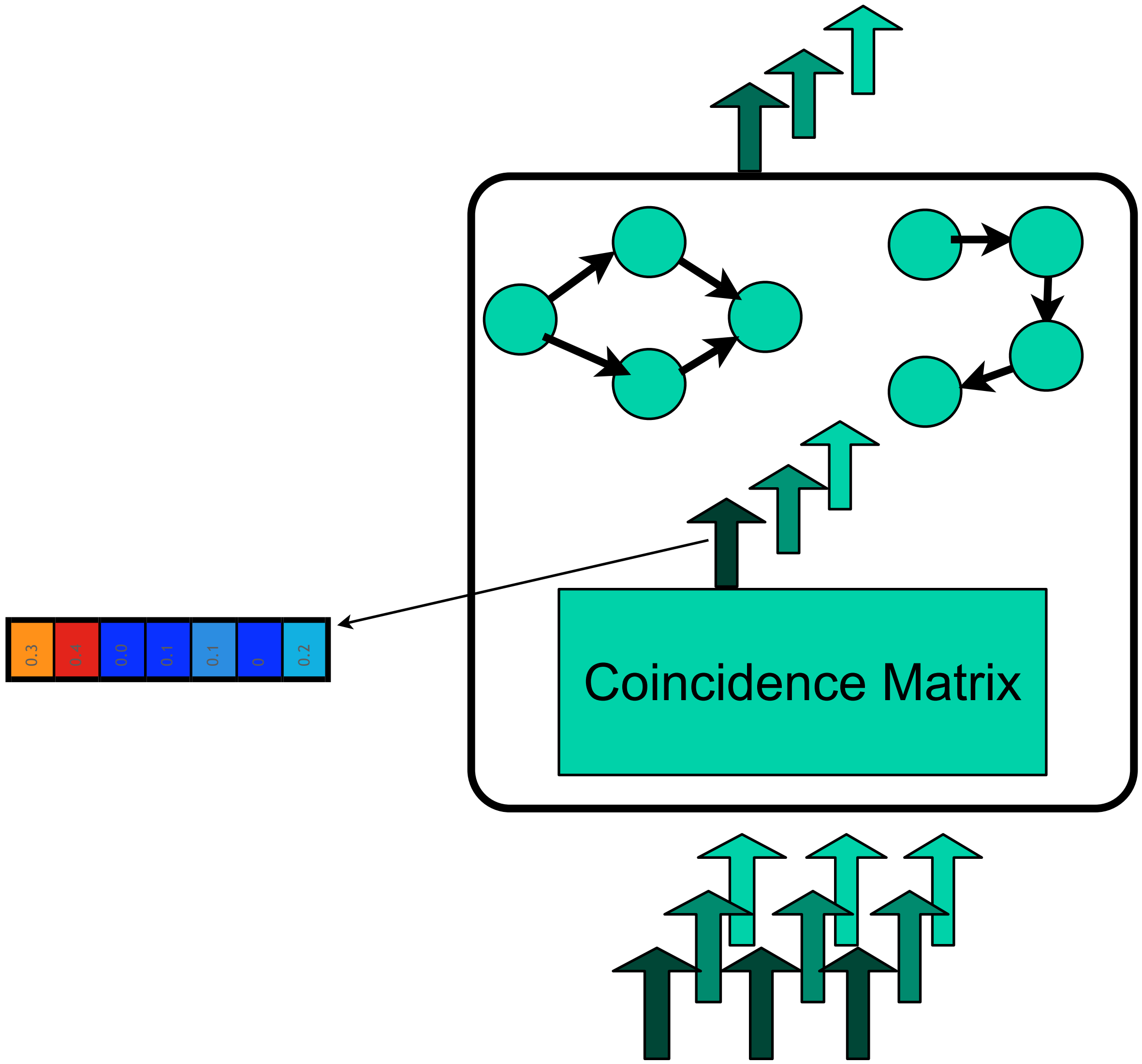
Inference in an HTM node



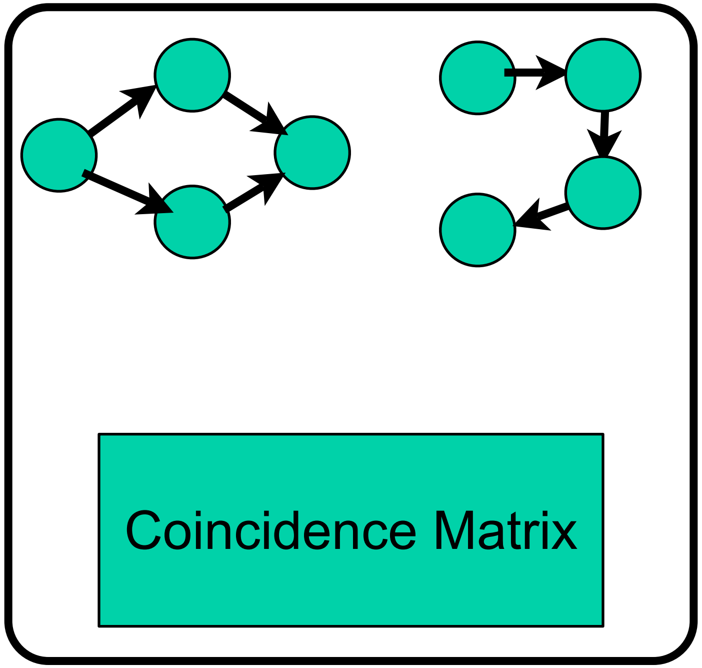
Inference in an HTM node



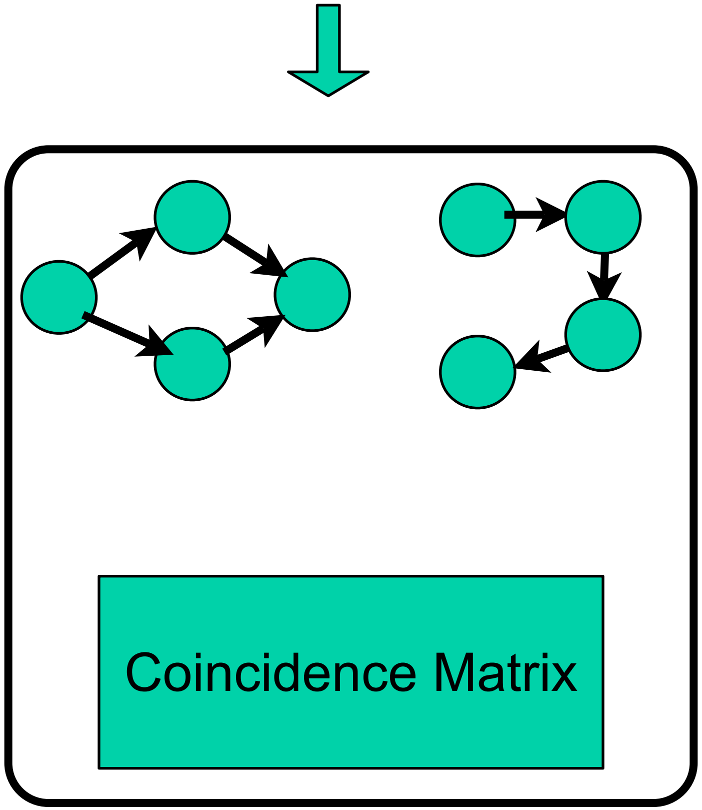
Inference in an HTM node



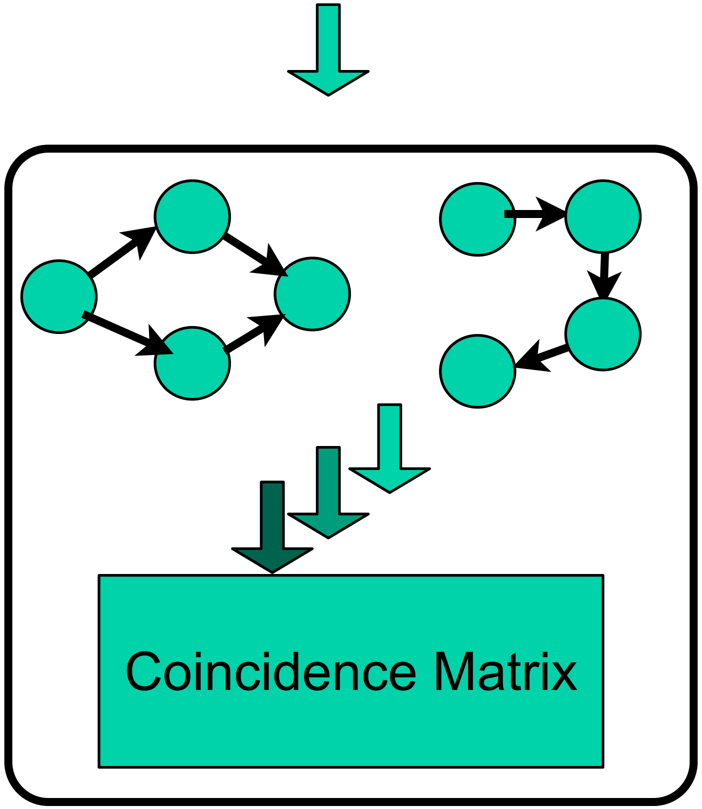
Inference in an HTM node



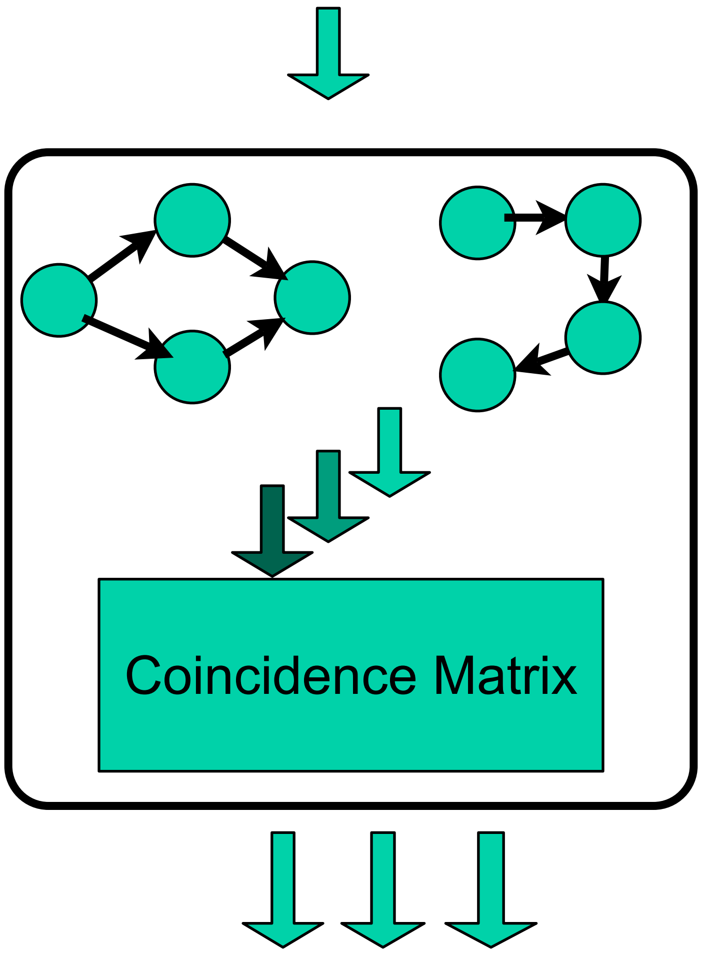
Inference in an HTM node



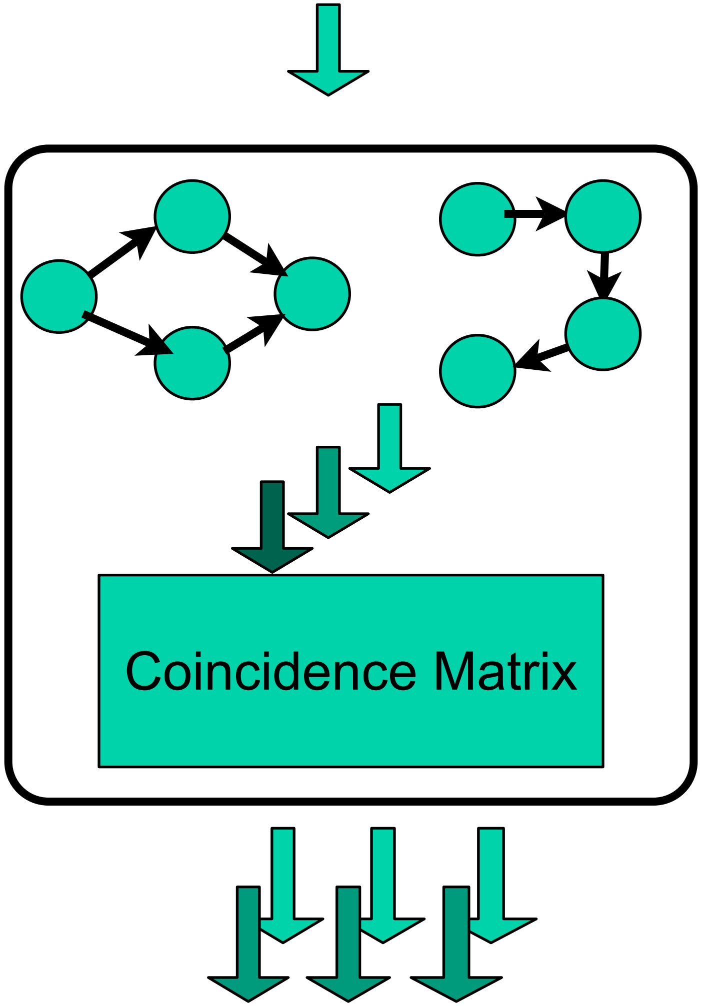
Inference in an HTM node



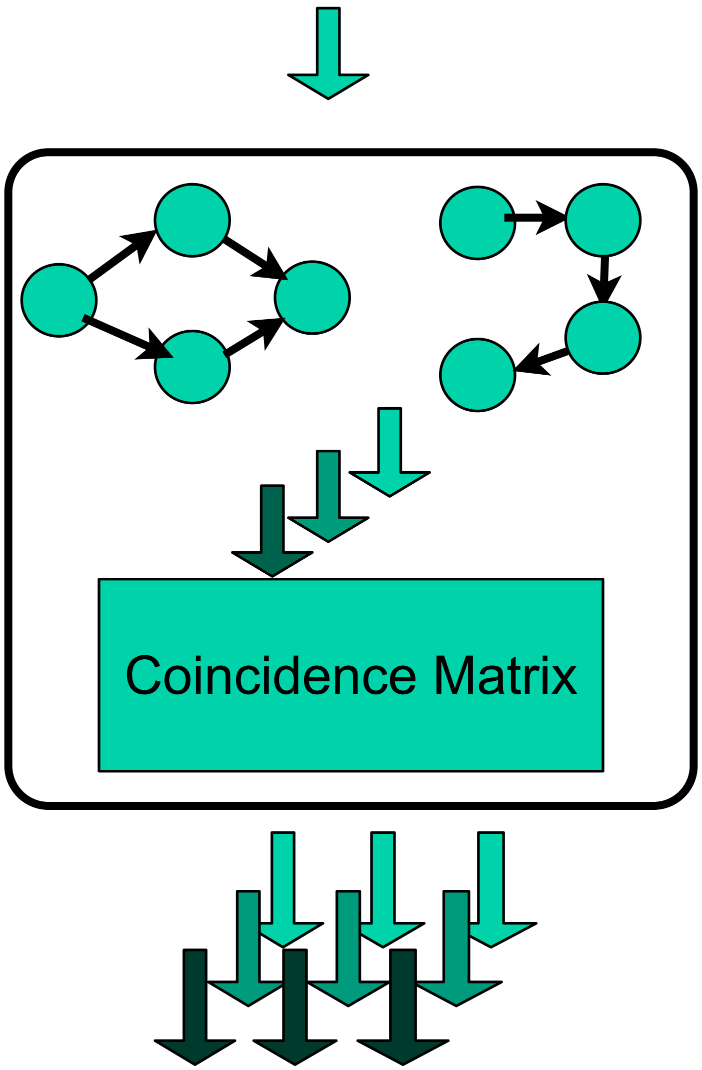
Inference in an HTM node



Inference in an HTM node

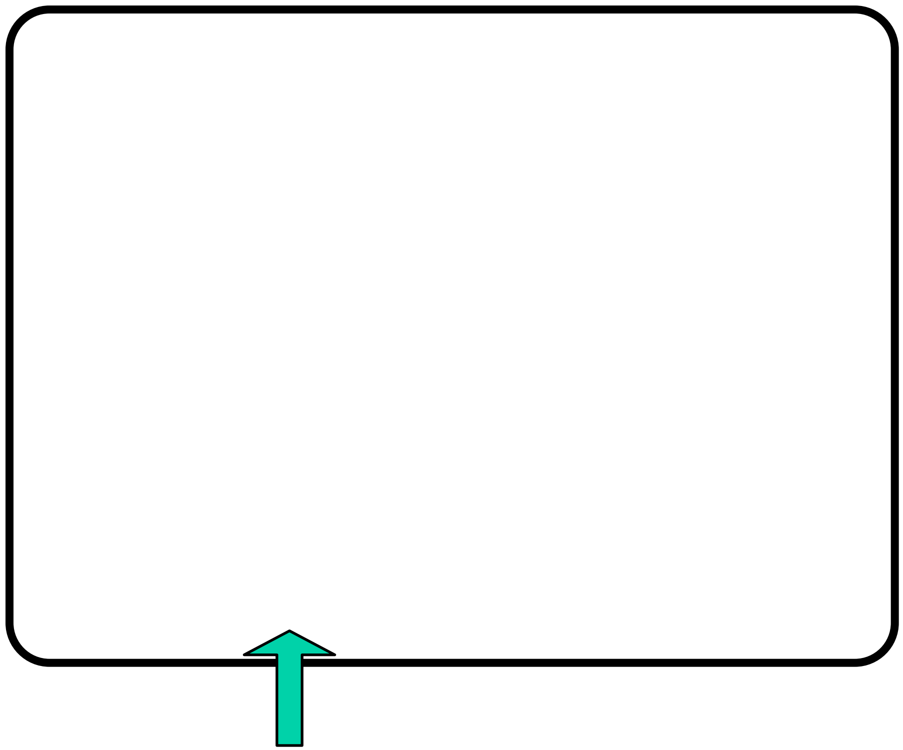


Inference in an HTM node

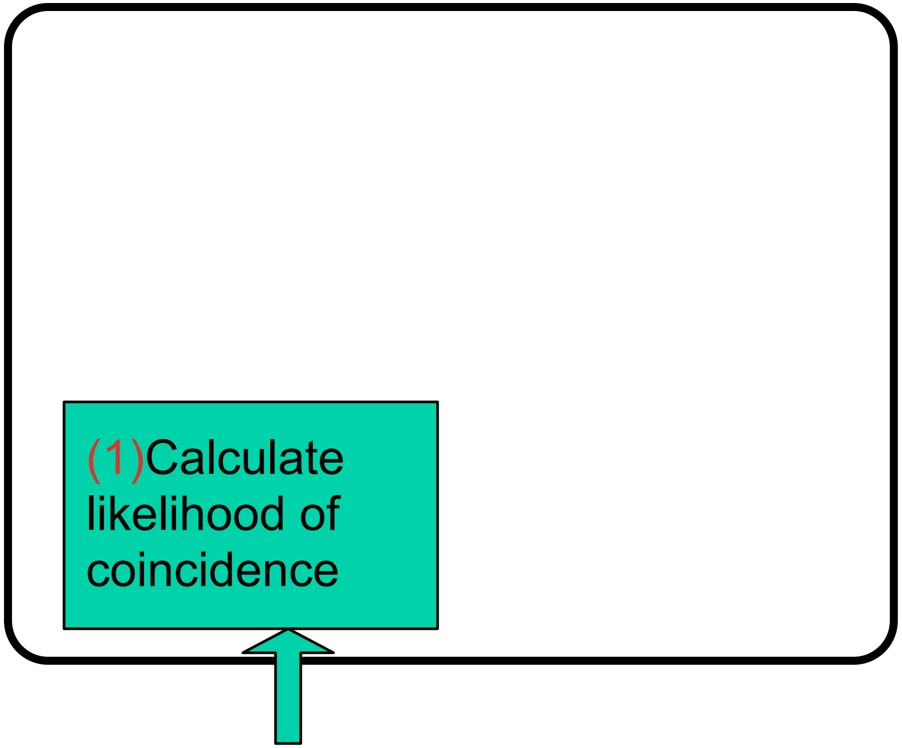


Inference in an HTM node

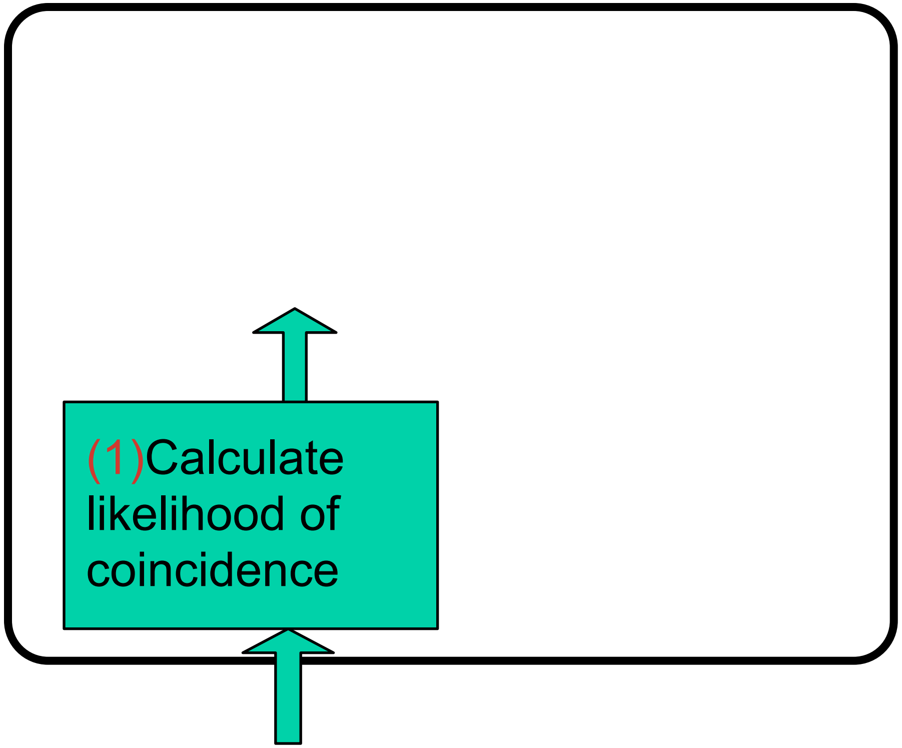
Inference in an HTM node



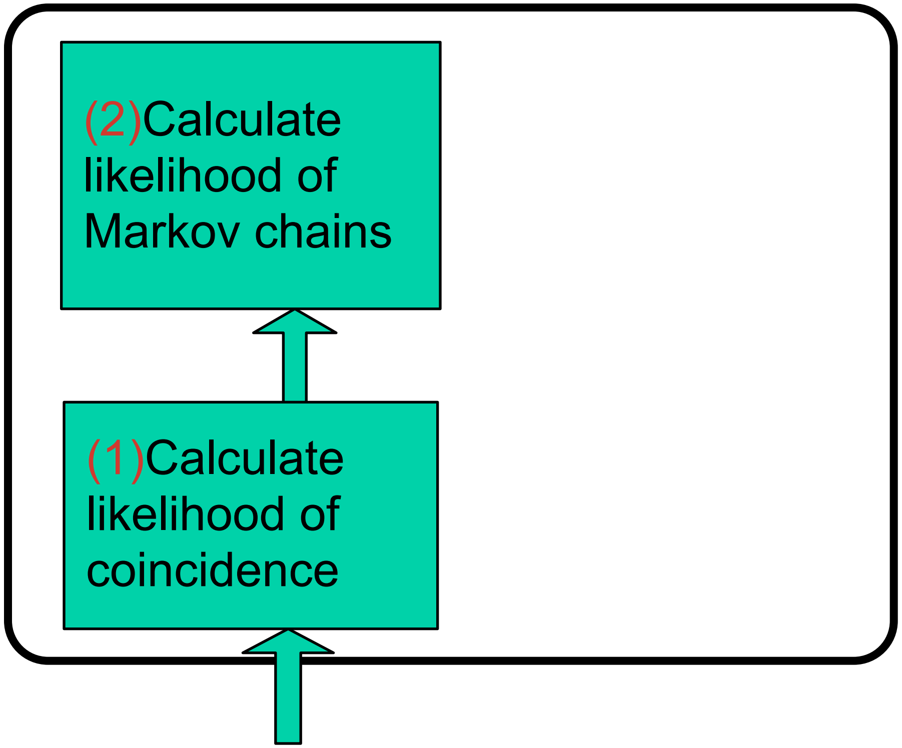
Inference in an HTM node



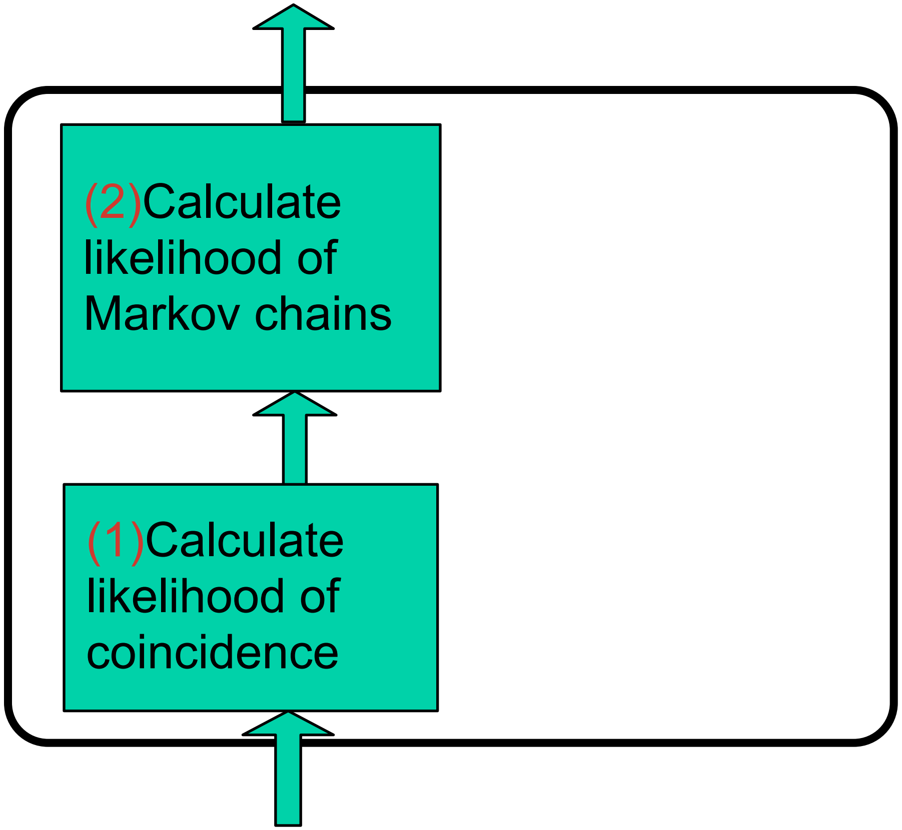
Inference in an HTM node



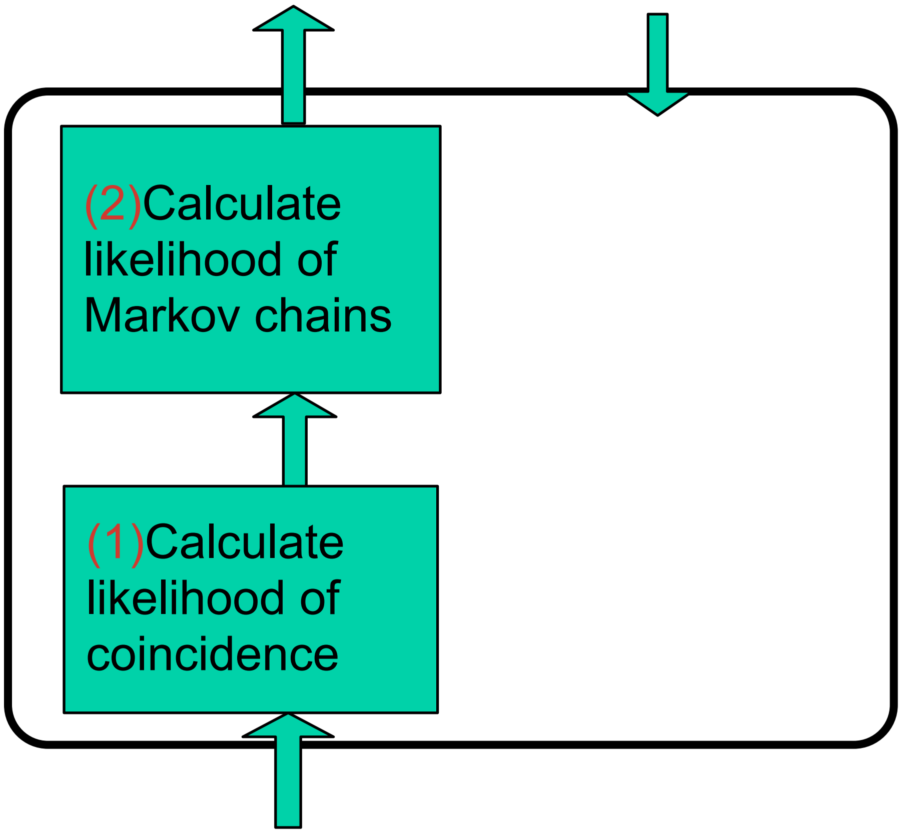
Inference in an HTM node



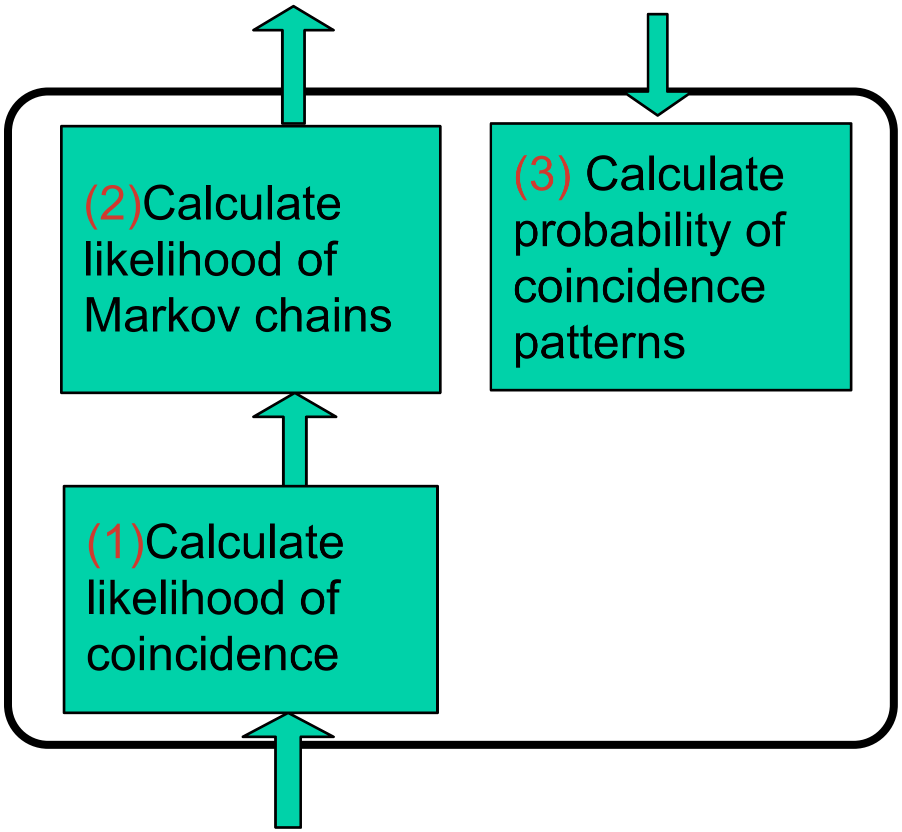
Inference in an HTM node



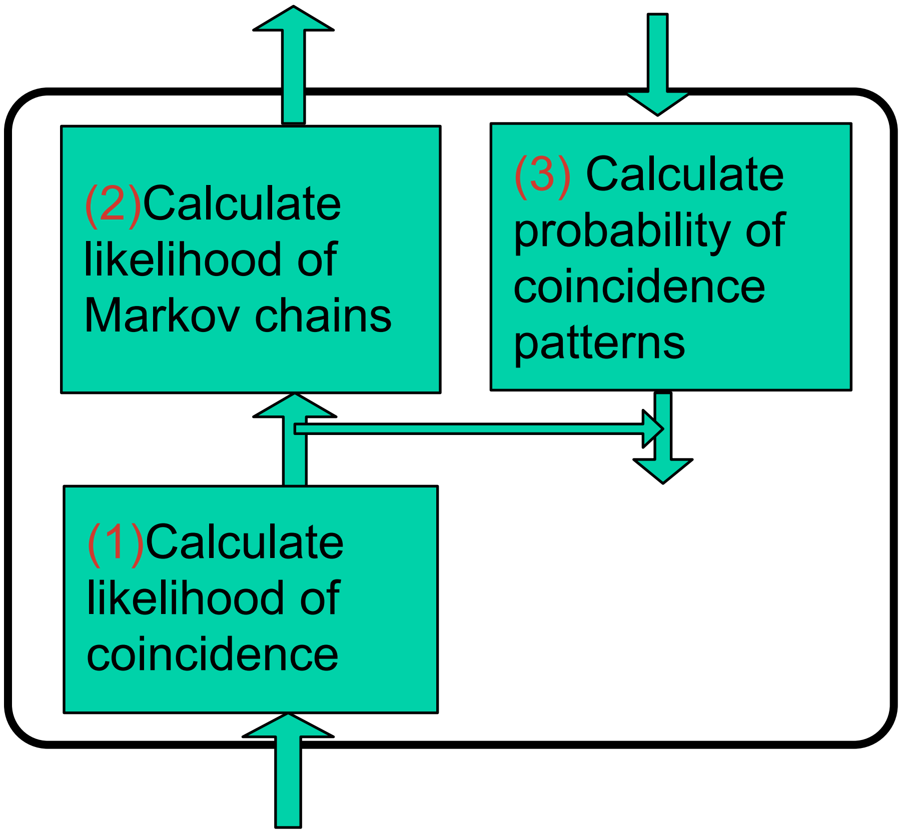
Inference in an HTM node



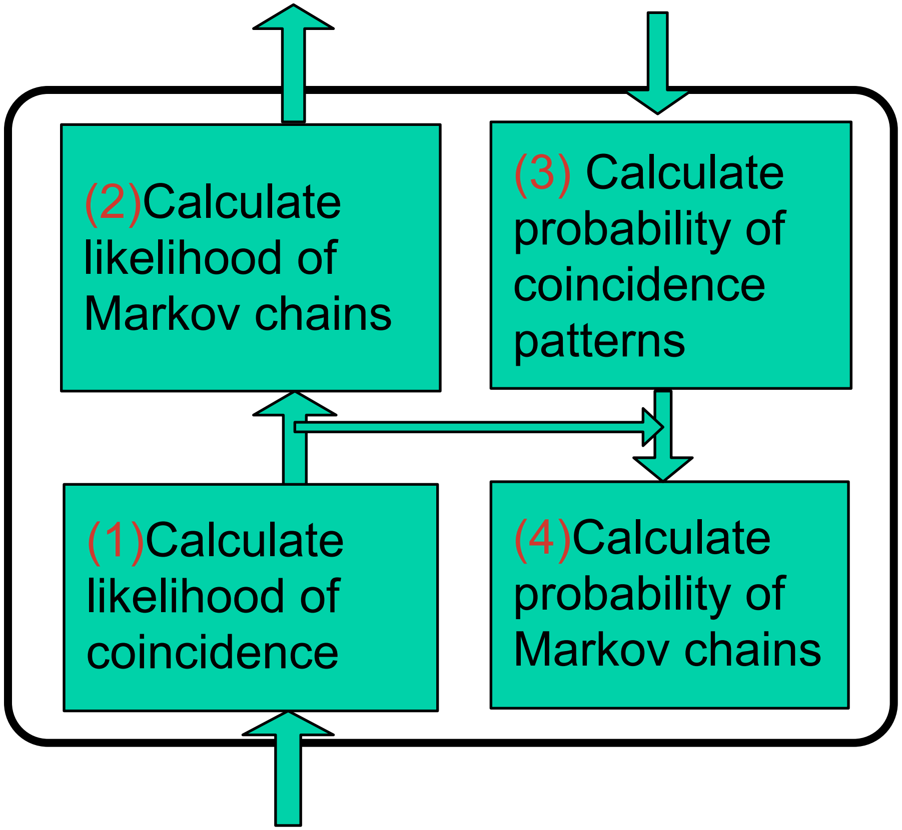
Inference in an HTM node



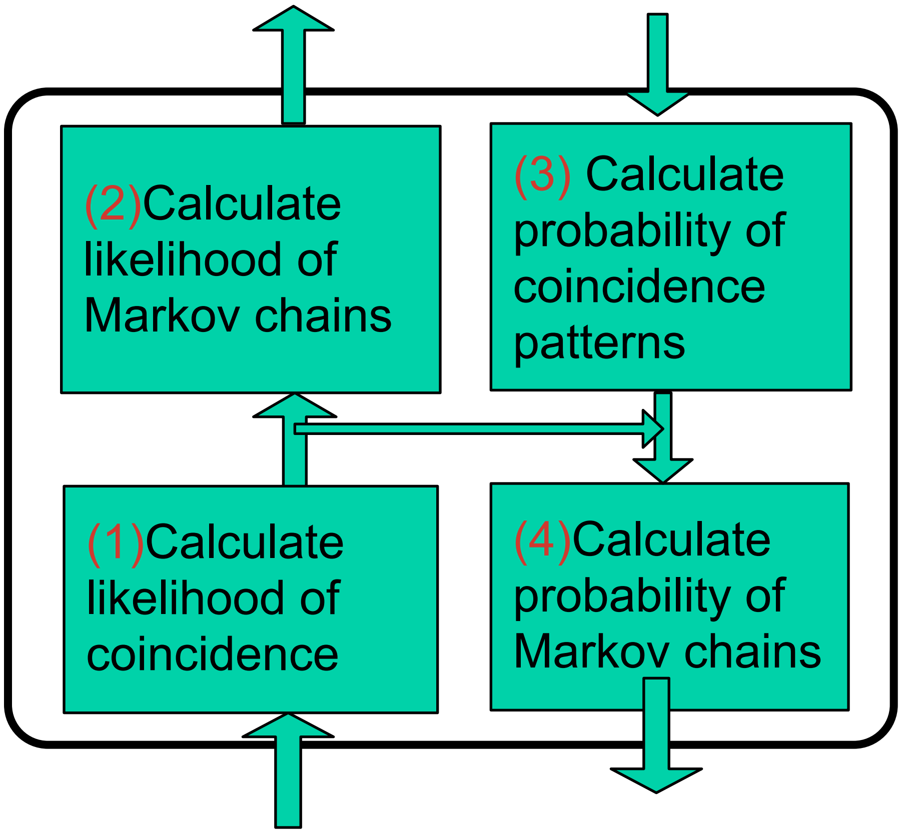
Inference in an HTM node



Inference in an HTM node



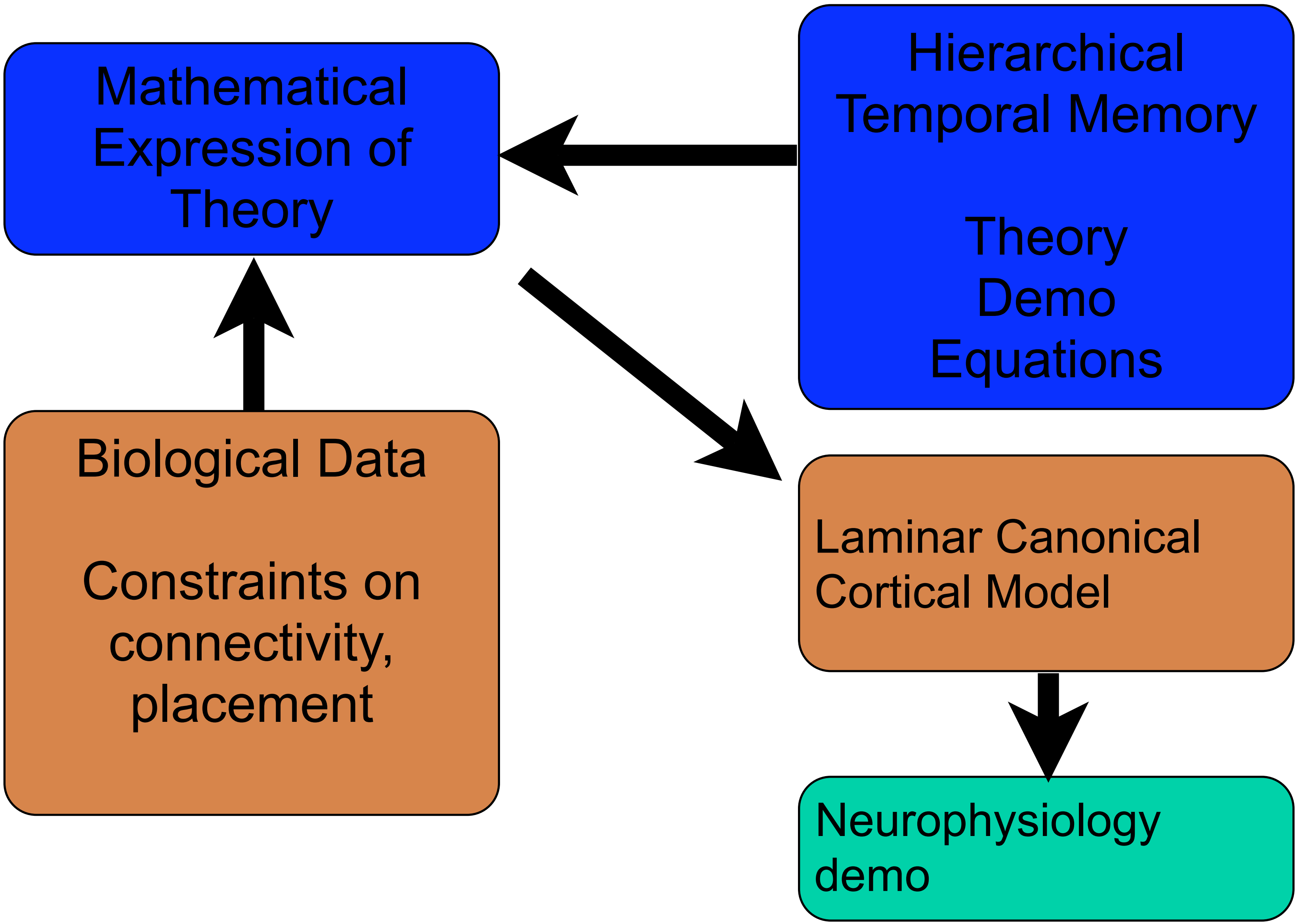
Inference in an HTM node



Belief propagation in HTMs : Equations

(1) Coincidence likelihood	$$y_t(i) = P(^-e_t \mid c_i(t)) \propto \prod_{j=1}^{M} \lambda_t^{m_j}(r_i^{m_j}) \quad (1)$$ where coincidence-pattern c_i is the co-occurrence of $r_i^{m_1}$'th Markov chain from child 1, $r_i^{m_2}$'th Markov chain from child 2, $\cdots$, and $r_i^{m_M}$'th Markov chain from child M.
(2) Markov chain likelihood	$$\lambda_t^k(g_r) = P(^-e_0^t \mid g_r(t)) \propto \sum_{c_i(t)\in C^k} \alpha_t(c_i, g_r) \quad (2)$$ $$\alpha_t(c_i, g_r) = P(^-e_t \mid c_i(t)) \sum_{c_j(t-1)\in C^k} P(c_i(t) \mid c_j(t-1), g_r)\alpha_{t-1}(c_j, g_r) \quad (3)$$ $$\alpha_0(c_i, g_r) = P(^-e_0 \mid c_i(t=0))P(c_i(t=0) \mid g_r) \quad (4)$$
(3) Coincidence Belief	$$Bel_t(c_i) \propto \sum_{g_r\in G^k} P(g_r \mid ^+e_0)\beta_t(c_i, g_r) \quad (5)$$ $$\beta_t(c_i, g_r) = P(^-e_t \mid c_i(t)) \sum_{c_j(t-1)\in C^k} P(c_i(t) \mid c_j(t-1), g_r)\beta_{t-1}(c_j, g_r) \quad (6)$$ $$\beta_0(c_i, g_r) = P(^-e_0 \mid c_i(t=0))P(c_i(t=0) \mid g_r, ^+e_0) \quad (7)$$
(4) Feedback messages	$$\pi^{child}(g_m) \propto \sum_{i\forall i} I(c_i)Bel(c_i) \quad (8)$$ where $$I(c_i) = \begin{cases} 1, & \text{if } g_m^{child} \text{ is a component of } c_i \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

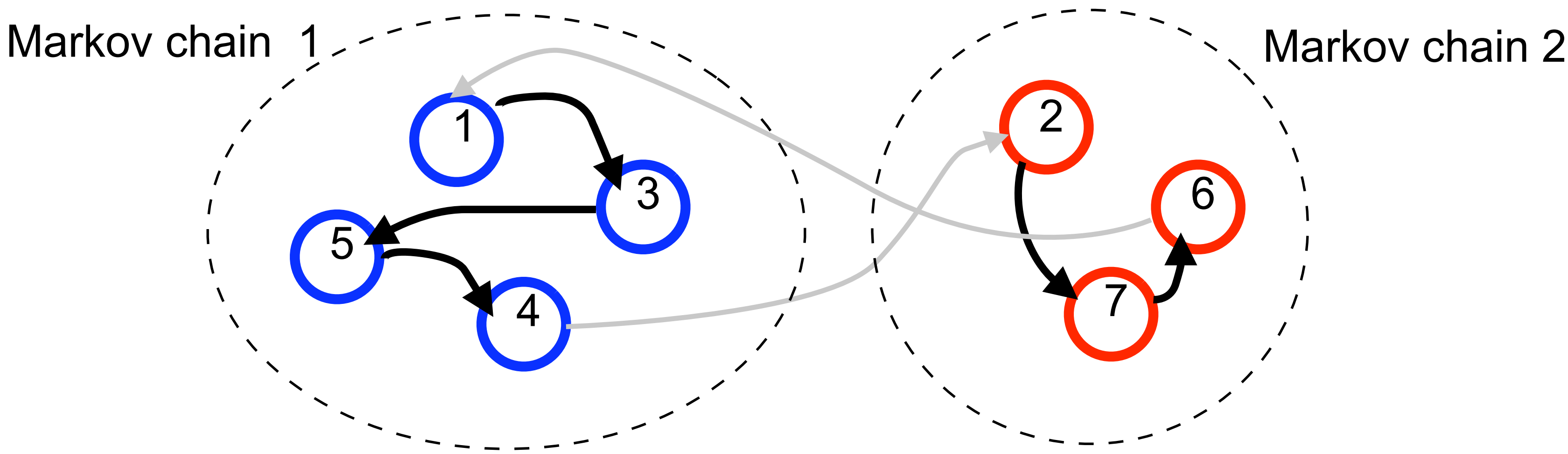
Agenda

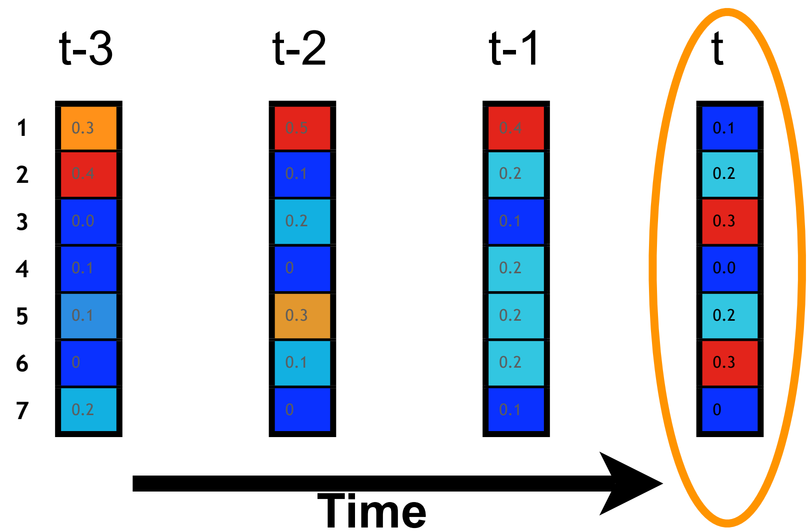

Mathematical Expression of Theory
Hierarchical Temporal Memory
Theory
Demo
Equations
Biological Data
Constraints on connectivity, placement
Laminar Canonical Cortical Model
Neurophysiology demo

Proposal:

A set of layer 2/3 neurons implement sequence inference using dynamic programming

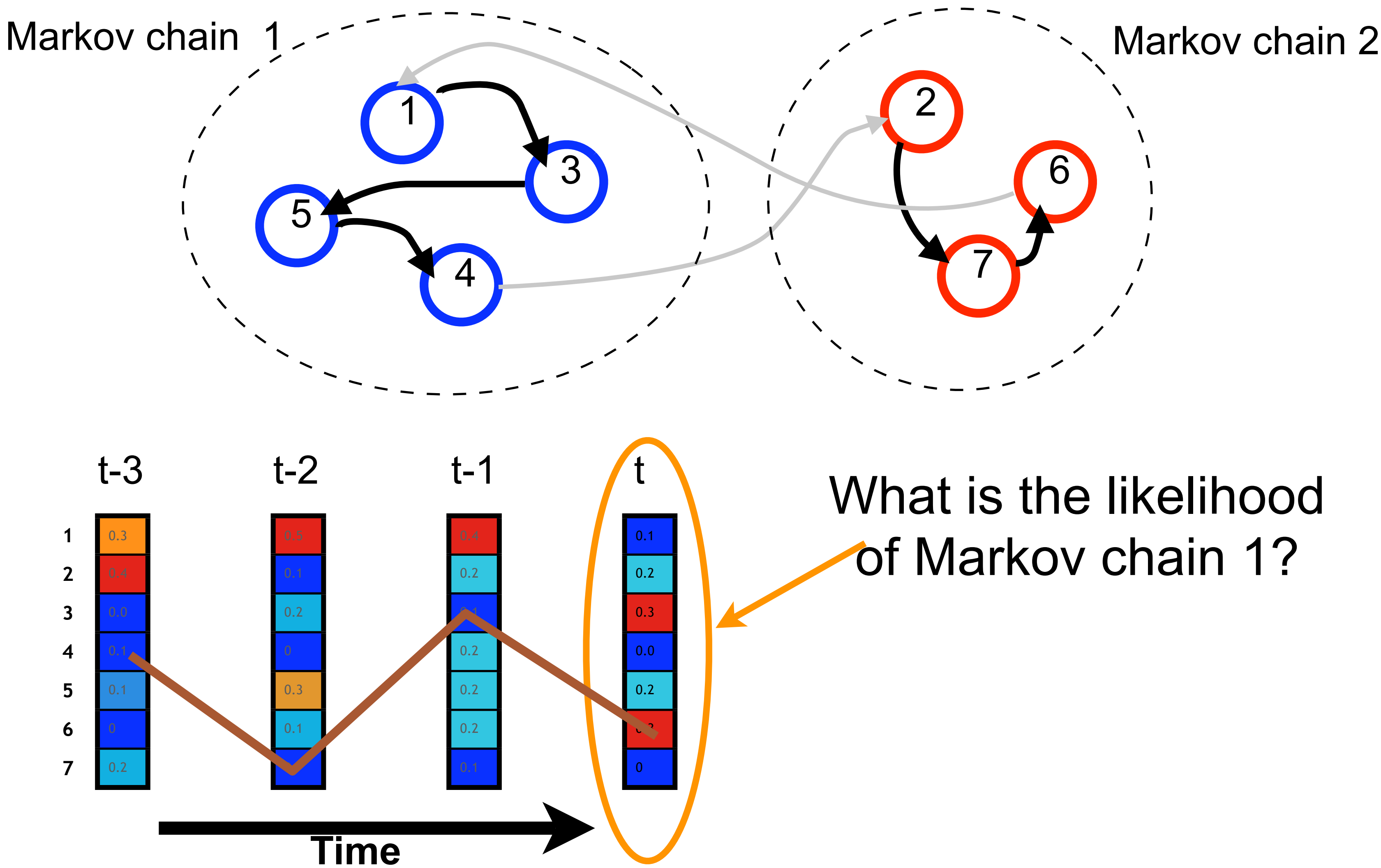
Dynamic programming for Markov chain inference



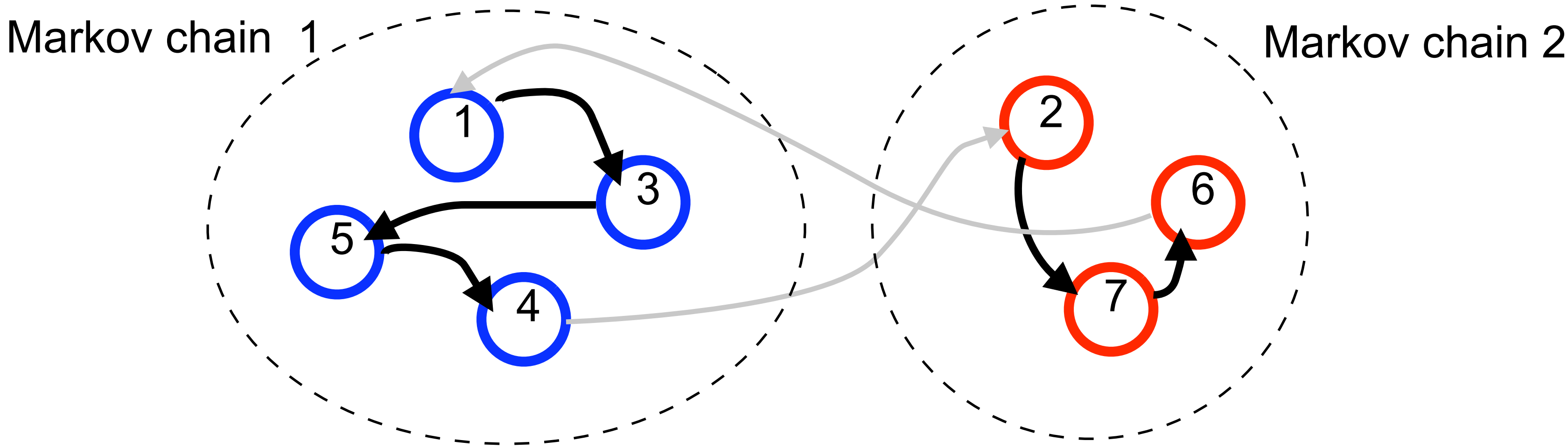


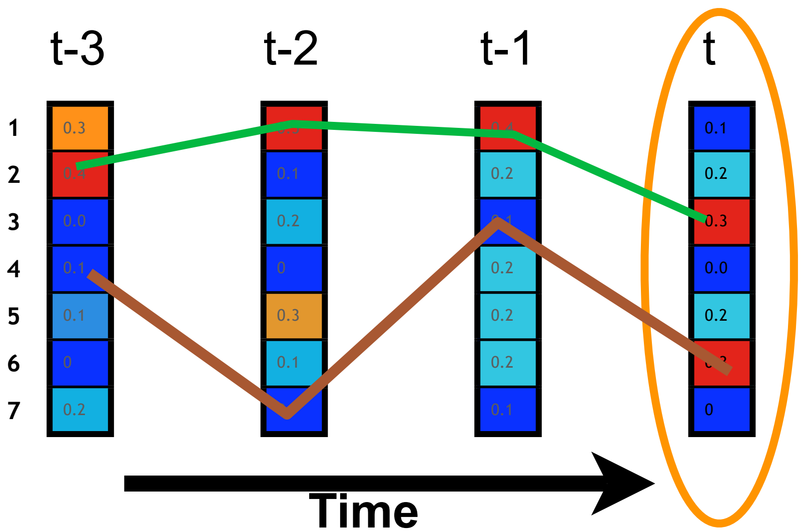
What is the likelihood of Markov chain 1?

Dynamic programming for Markov chain inference



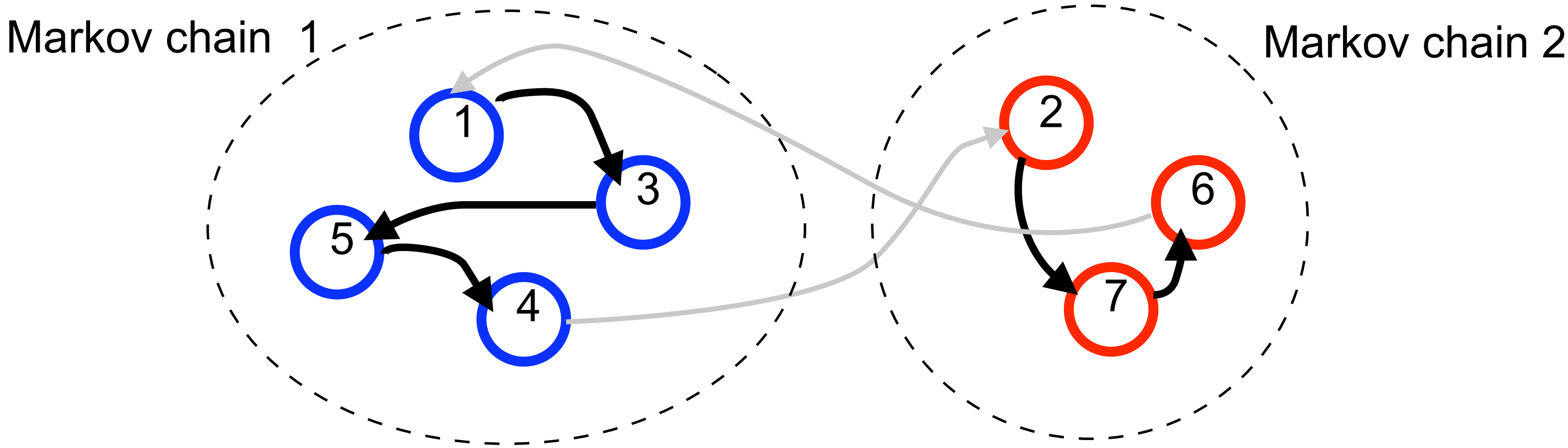
Dynamic programming for Markov chain inference

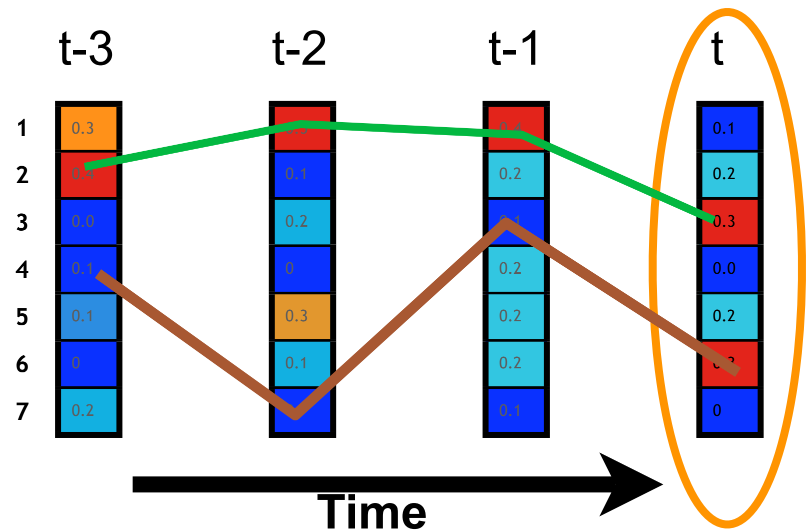




What is the likelihood of Markov chain 1?

Dynamic programming for Markov chain inference





What is the likelihood of Markov chain 1?

Dynamic programming calculates these likelihoods efficiently

Calculate the likelihood of Markov chains

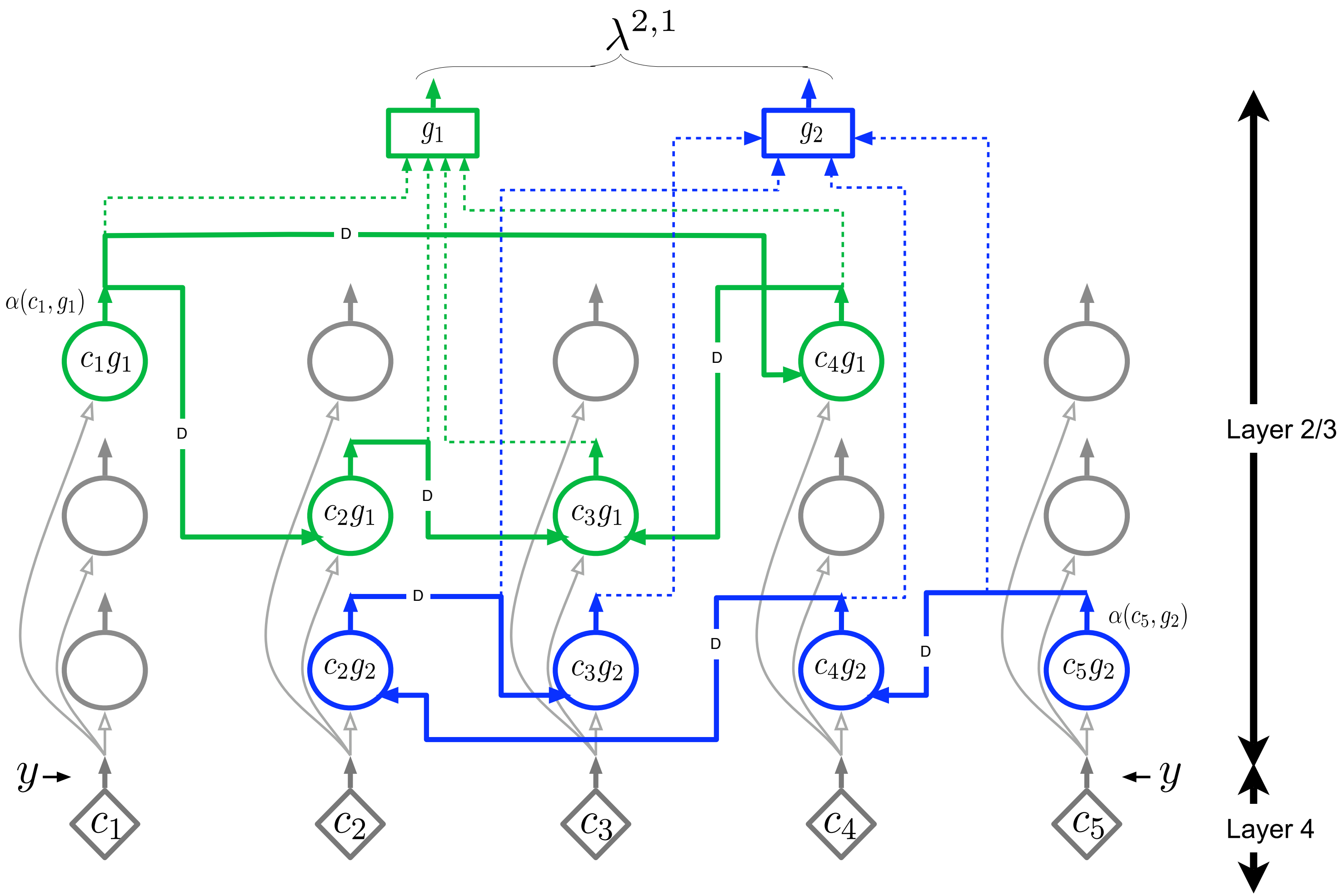
$$\alpha_t(c_i, g_r) = P(^{-}e_t | c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) | c_j(t-1), g_r) \alpha_{t-1}(c_j, g_r)$$

Calculate the likelihood of Markov chains

$$\alpha_t(c_i, g_r) = P(^{-}e_t | c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) | c_j(t-1), g_r) \alpha_{t-1}(c_j, g_r)$$

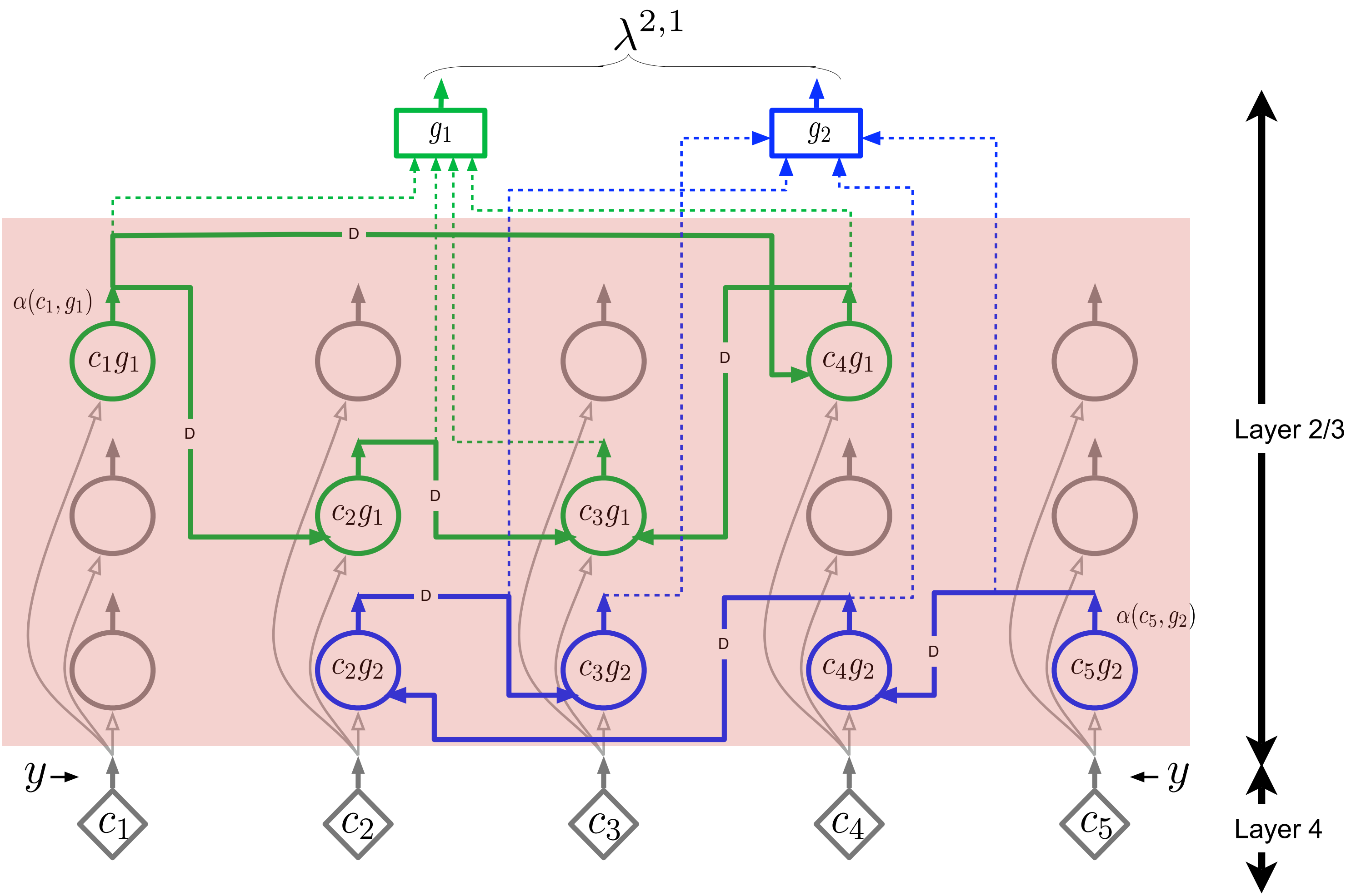
Calculate the likelihood of Markov chains

$$\alpha_t(c_i, g_r) = P(^{-}e_t | c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) | c_j(t-1), g_r) \alpha_{t-1}(c_j, g_r)$$



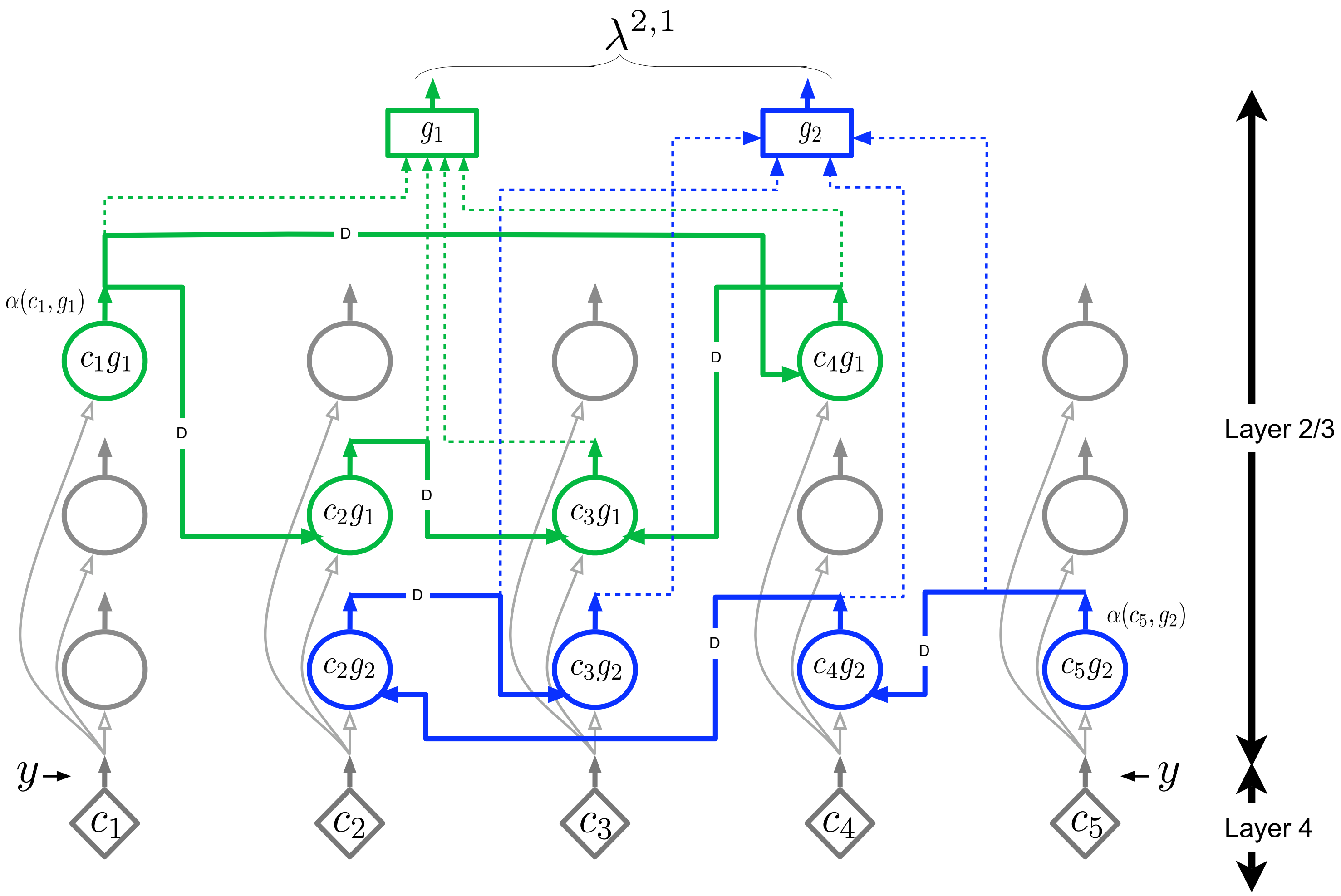
Calculate the likelihood of Markov chains

$$\alpha_t(c_i, g_r) = P(^{-}e_t | c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) | c_j(t-1), g_r) \alpha_{t-1}(c_j, g_r)$$



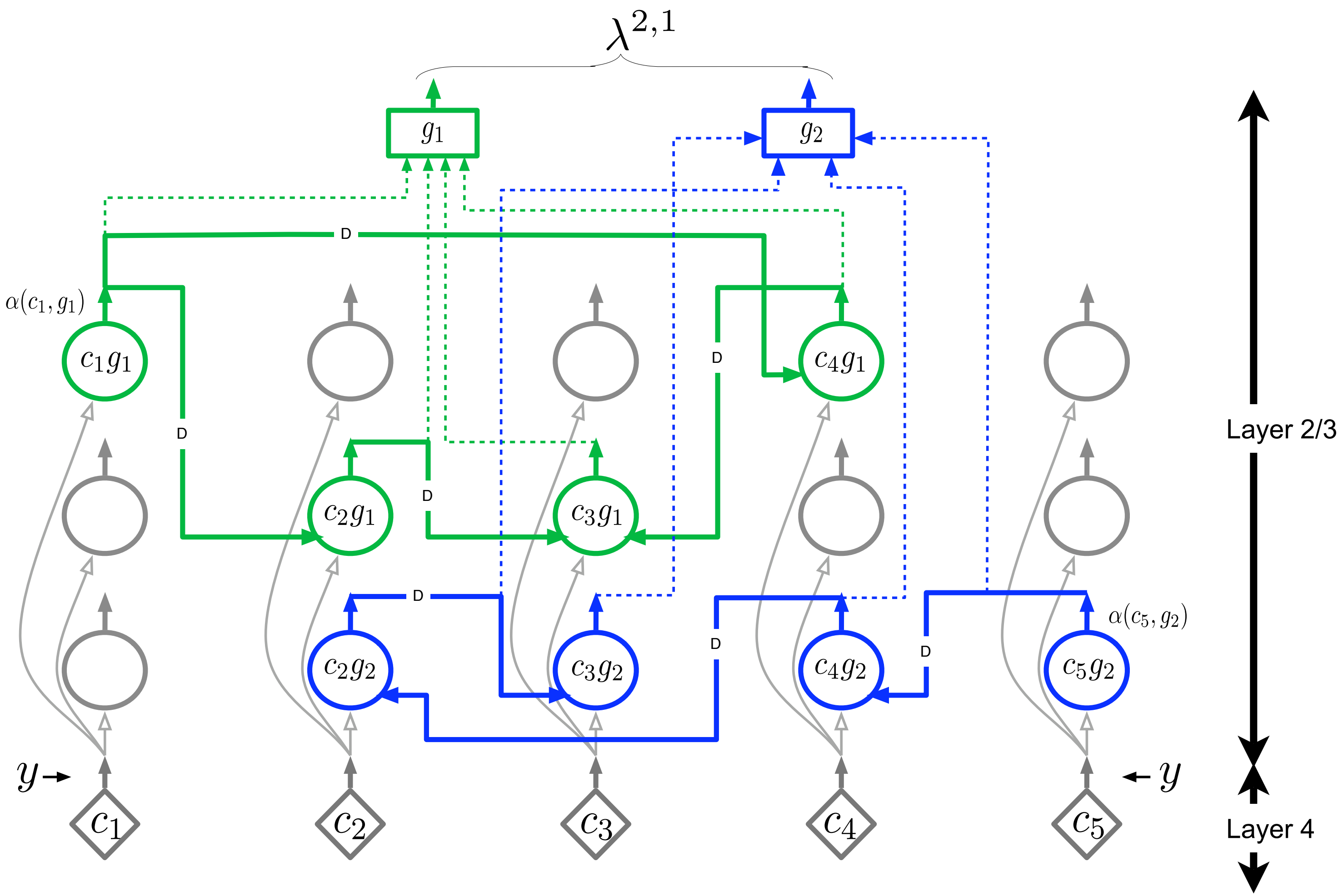
Calculate the likelihood of Markov chains

$$\alpha_t(c_i, g_r) = P(^{-}e_t|c_i(t)) \sum_{c_j(t-1)\in C^k} P(c_i(t)|c_j(t-1), g_r)\alpha_{t-1}(c_j, g_r)$$



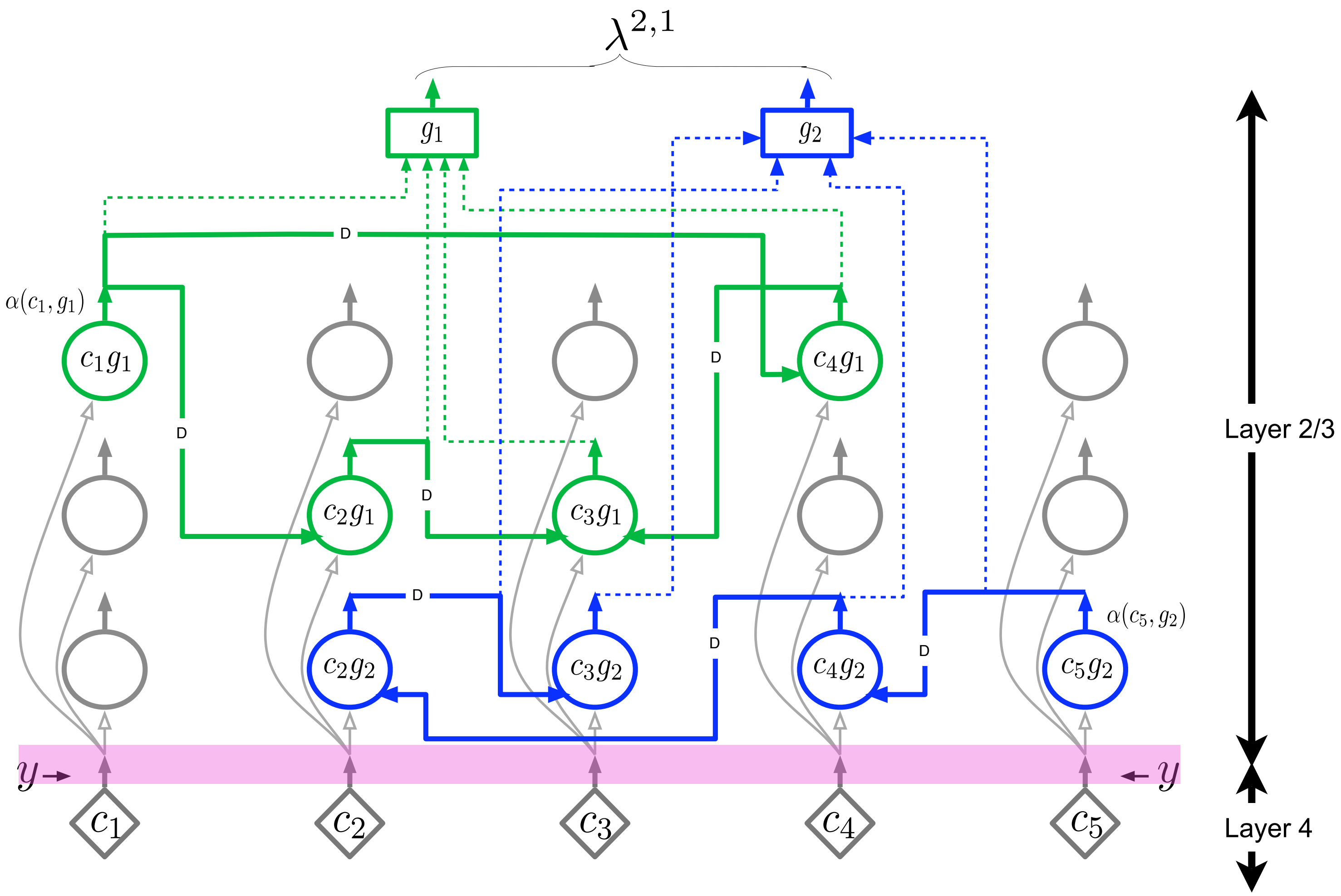
Calculate the likelihood of Markov chains

$$\alpha_t(c_i, g_r) = P(^- e_t | c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) | c_j(t-1), g_r) \alpha_{t-1}(c_j, g_r)$$

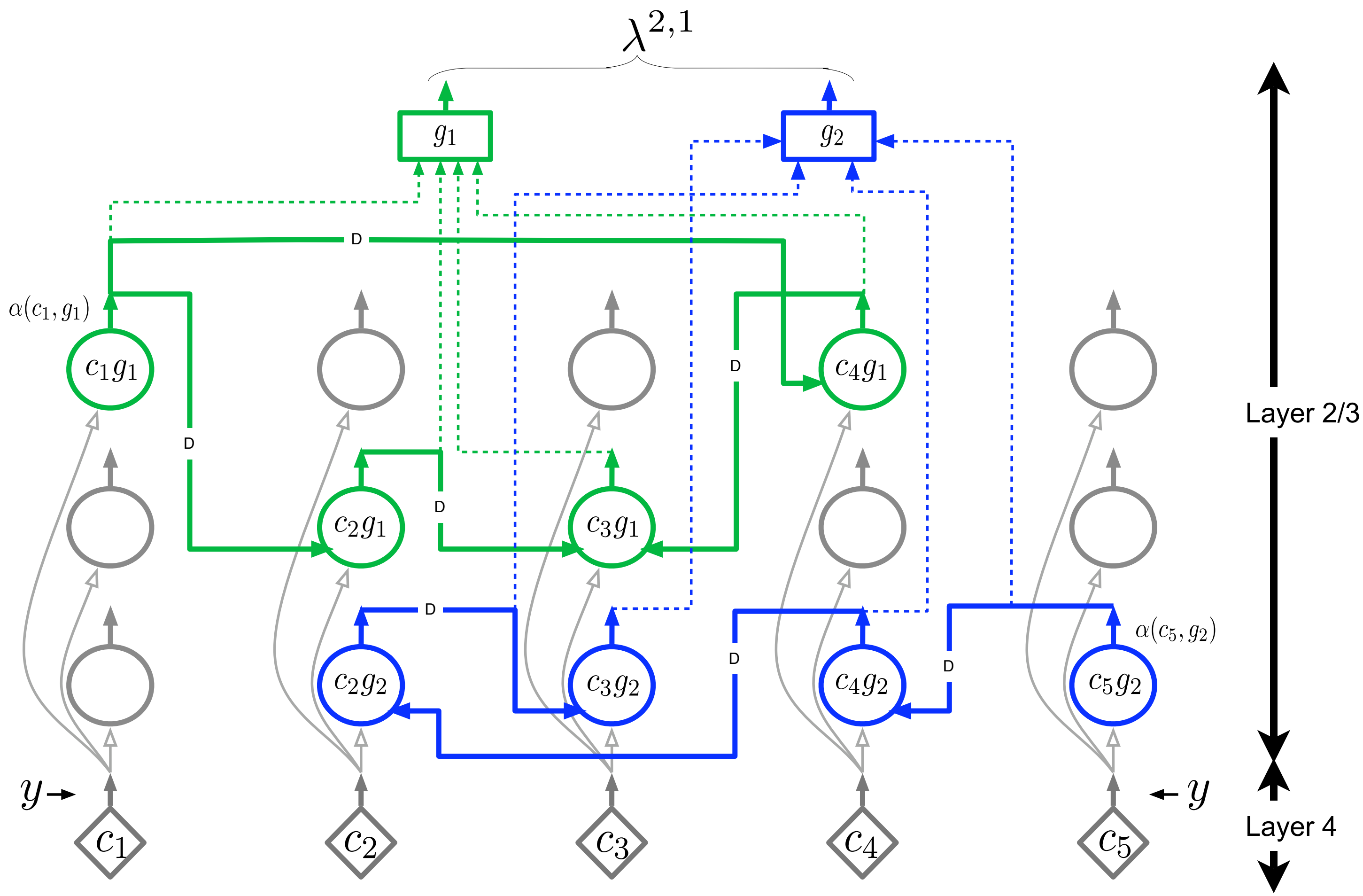


Calculate the likelihood of Markov chains

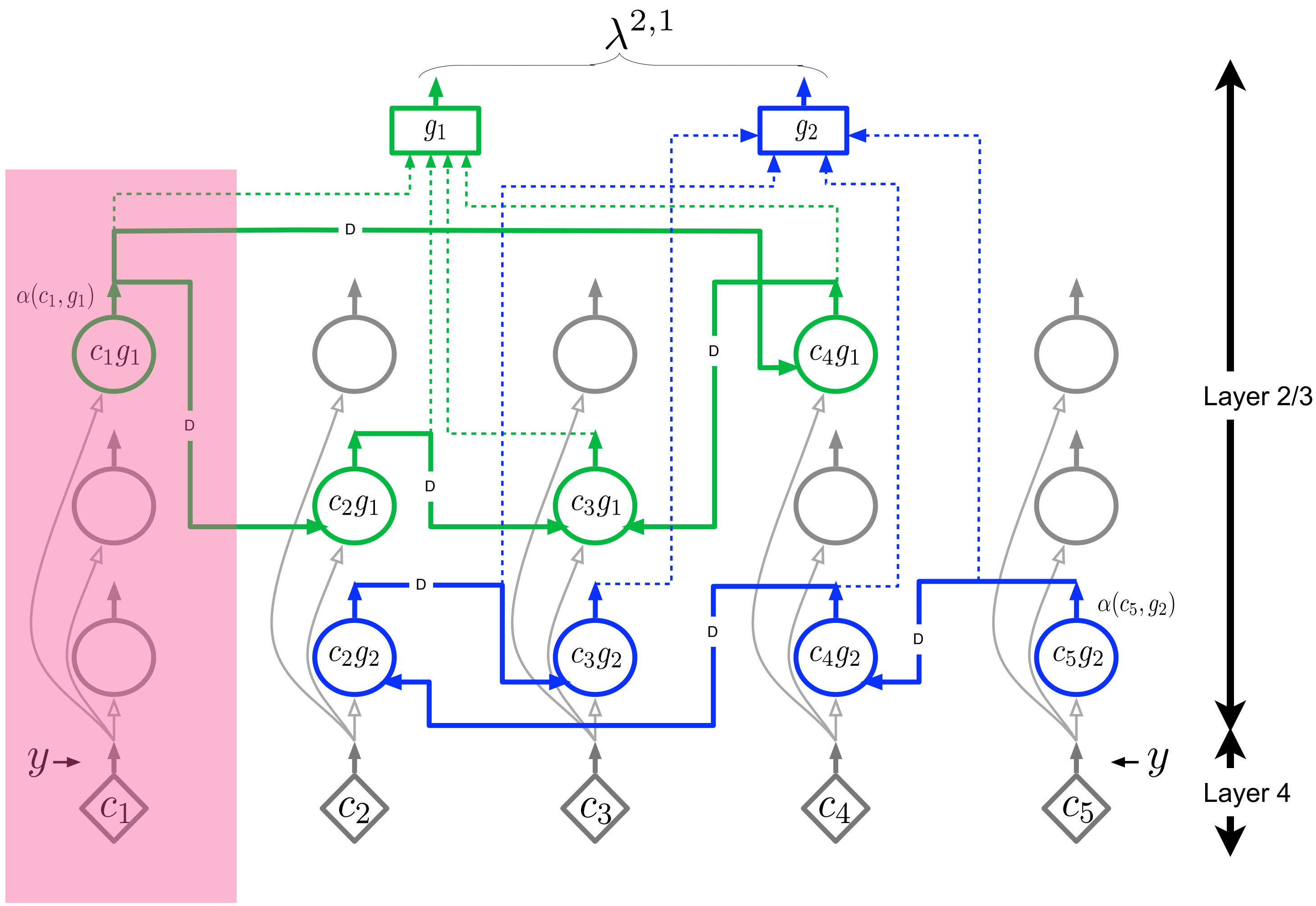
$$\alpha_t(c_i, g_r) = P(^{-}e_t | c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) | c_j(t-1), g_r) \alpha_{t-1}(c_j, g_r)$$



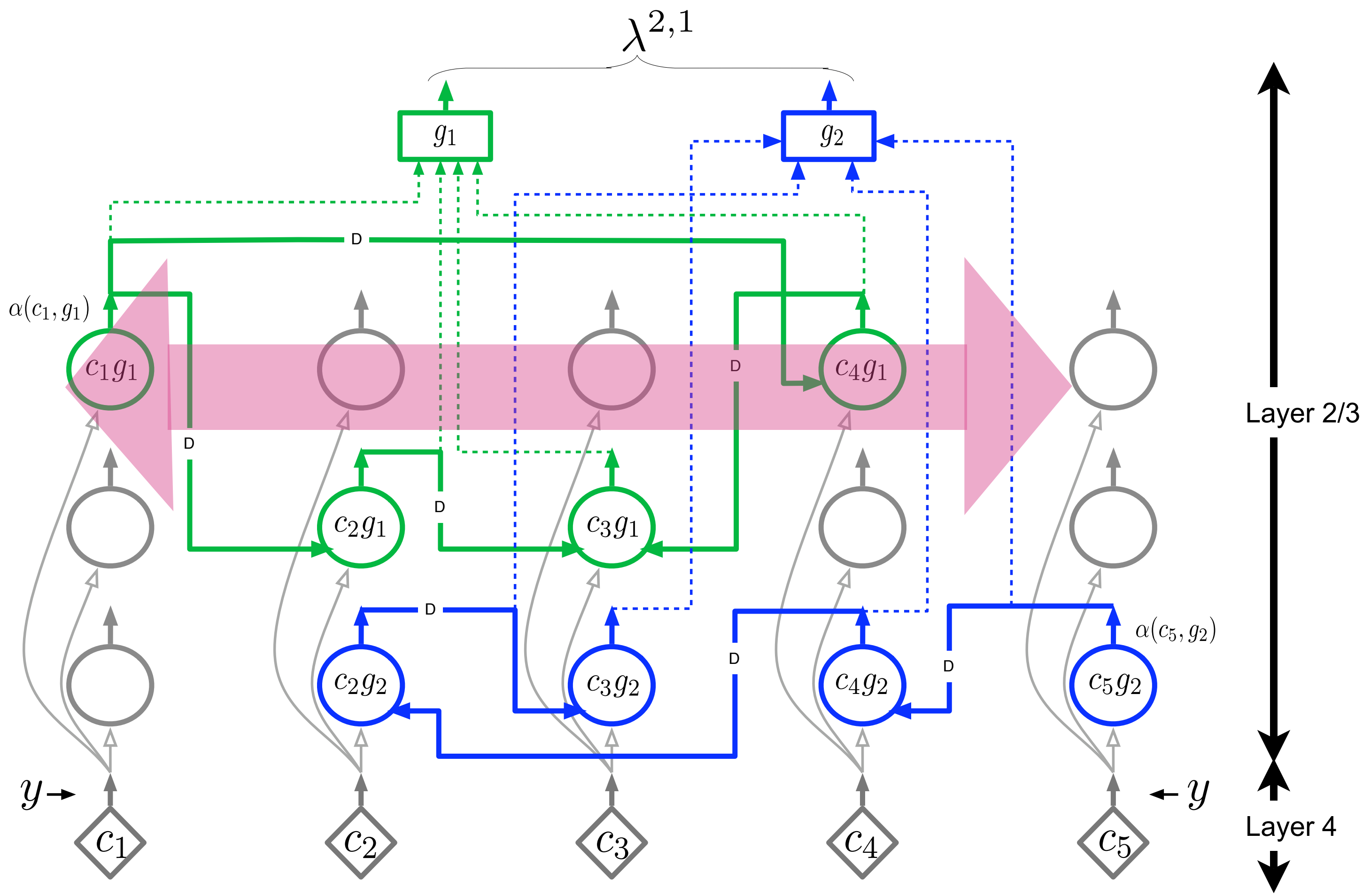
Layer 2/3 circuit (1)



Layer 2/3 circuit (1)



Layer 2/3 circuit (1)



Data : Layer 2/3

Data : Layer 2/3

- Major excitatory projection of layer 4 is to layers 2/3.
    - Thomson & Lamy 2007

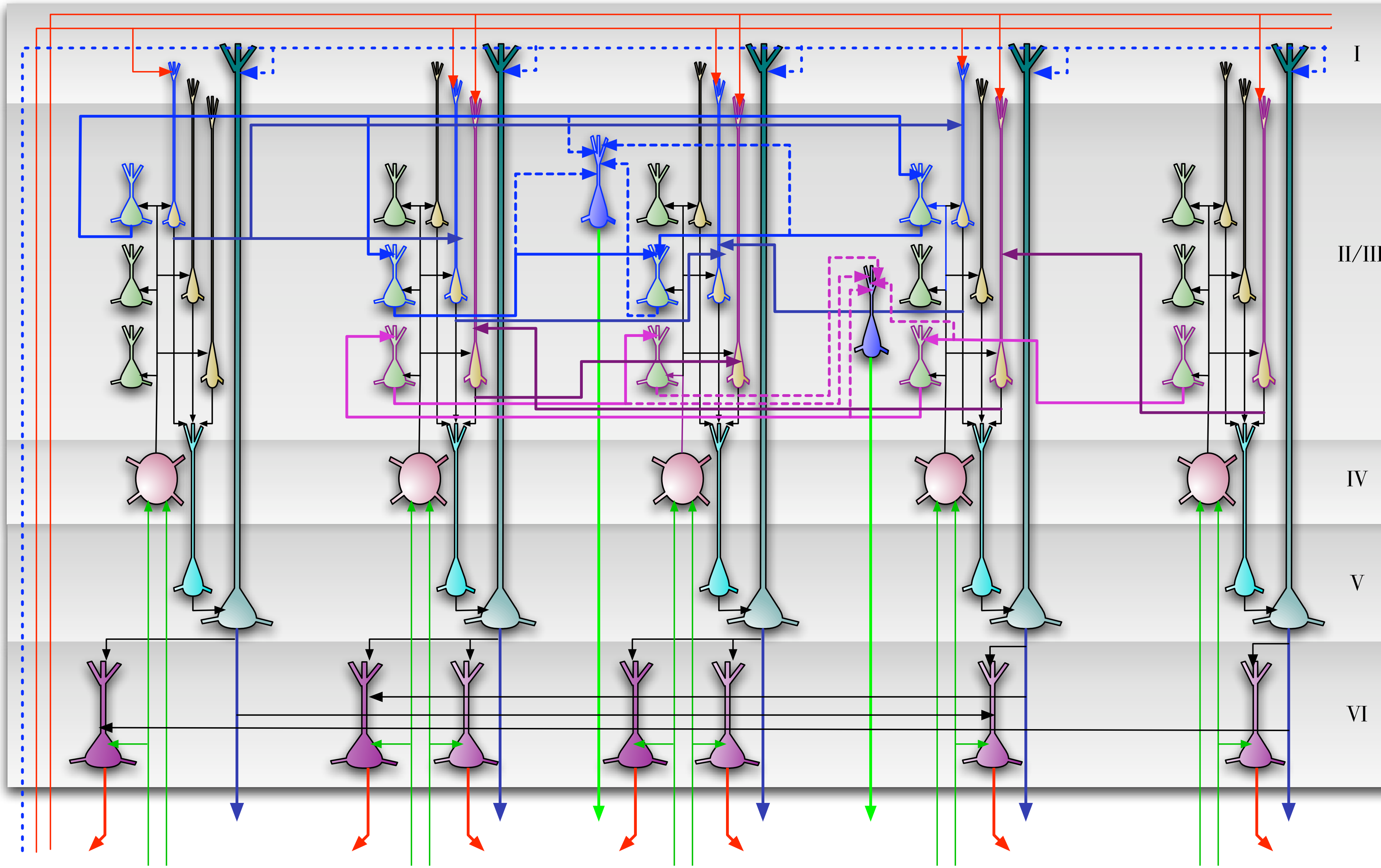
Data : Layer 2/3

- Major excitatory projection of layer 4 is to layers 2/3.
    - Thomson & Lamy 2007
- Layer 4 spiny neurons in the barrel cortex are characterized by vertically oriented intra-columnar axons that target Layer 2/3 pyramidal cells
    - Lubke & Feldmeyer 2007

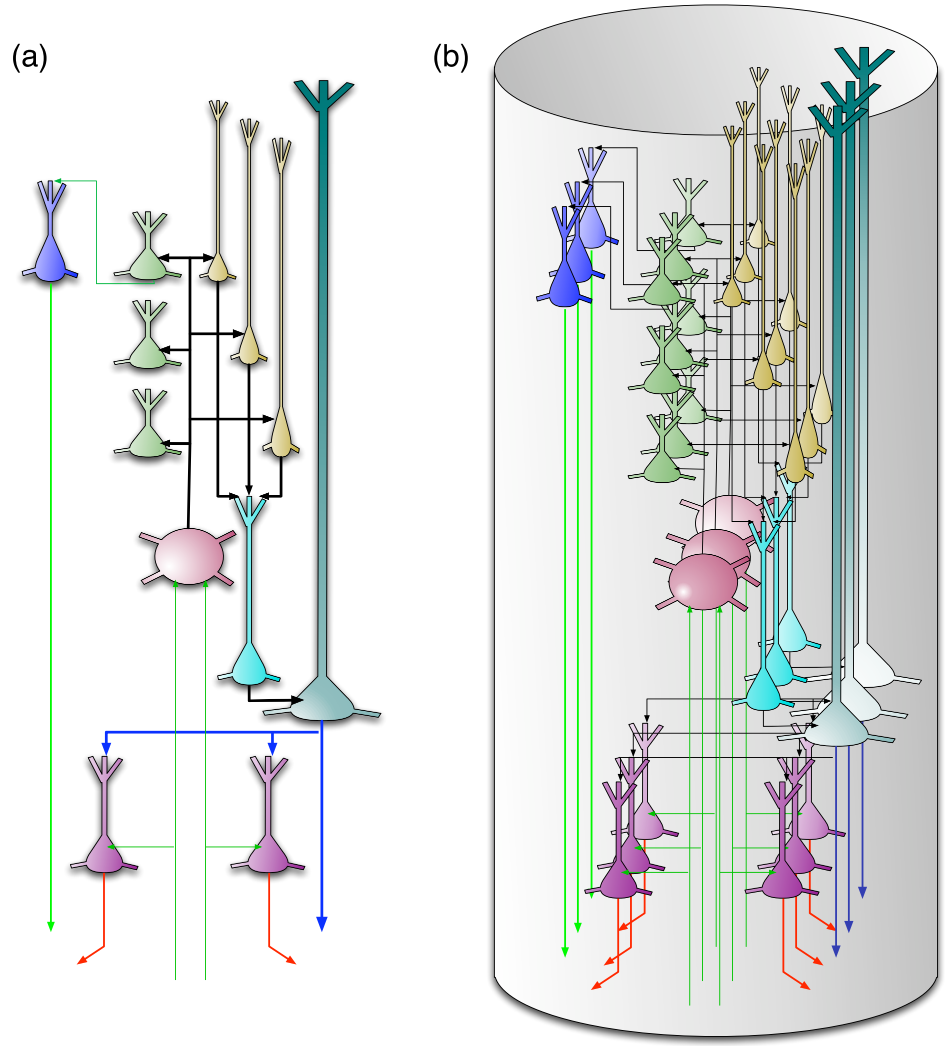
Data : Layer 2/3

- Major excitatory projection of layer 4 is to layers 2/3.
  - Thomson & Lamy 2007
- Layer 4 spiny neurons in the barrel cortex are characterized by vertically oriented intra-columnar axons that target Layer 2/3 pyramidal cells
  - Lubke & Feldmeyer 2007
- Cells in layer 2/3 are complex cells that prefer richer stimuli, such as motion in the preferred direction
  - Hirsch and Martinez 2006

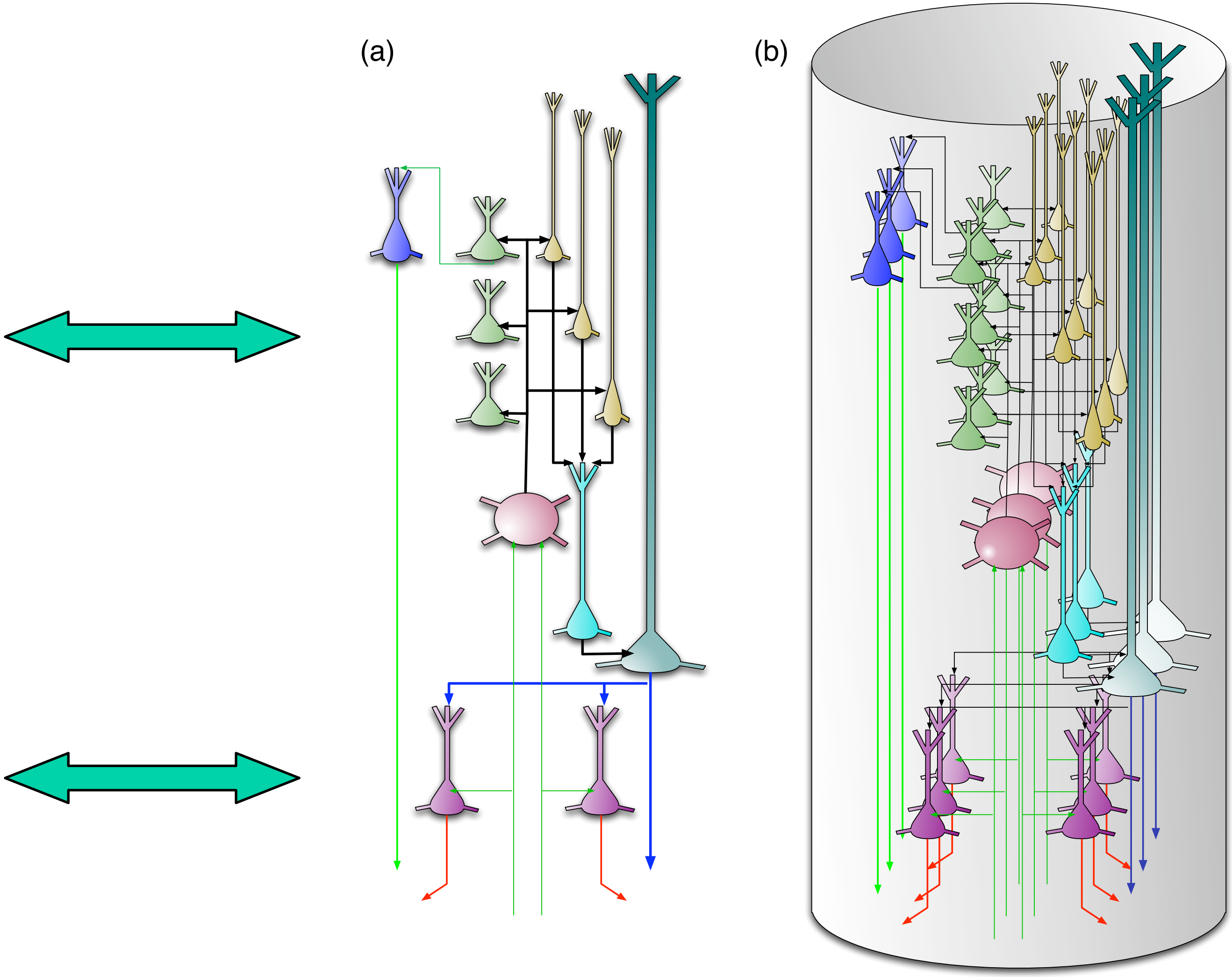
Cortical circuit from HTM belief propagation



Cortical Column



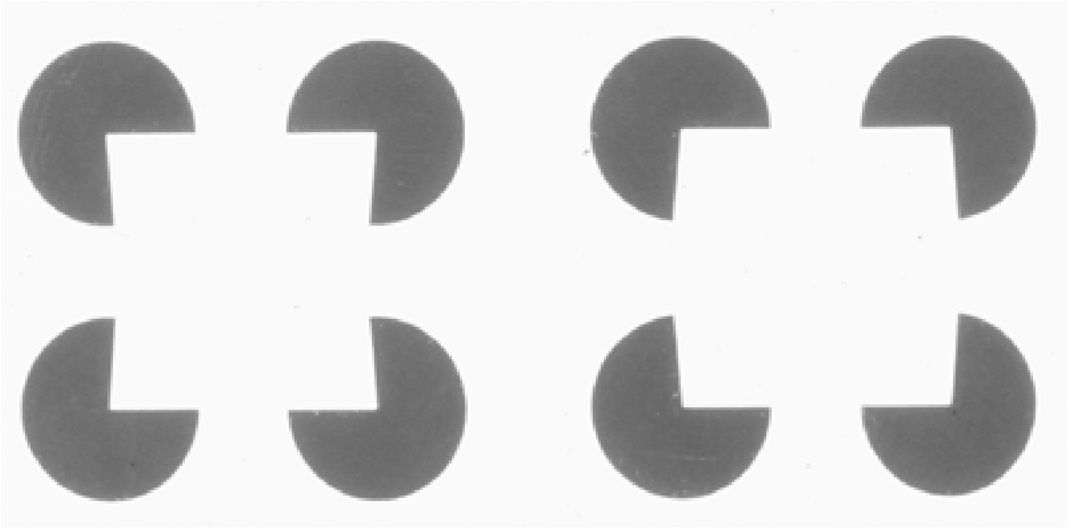
Cortical Column



Summary of laminar functions

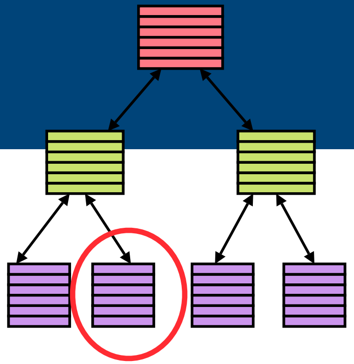
- Layer 4
  - Coincidence detection
- Layer 2/3
  - Feedforward inference on Markov chains (complex cells)
  - Feedback inference on Markov chains (complex cells)
- Layer 5
  - Assimilation (marginalization) of feedback information
  - Timing
- Layer 6
  - Computation of top-down messages from layer 5 outputs

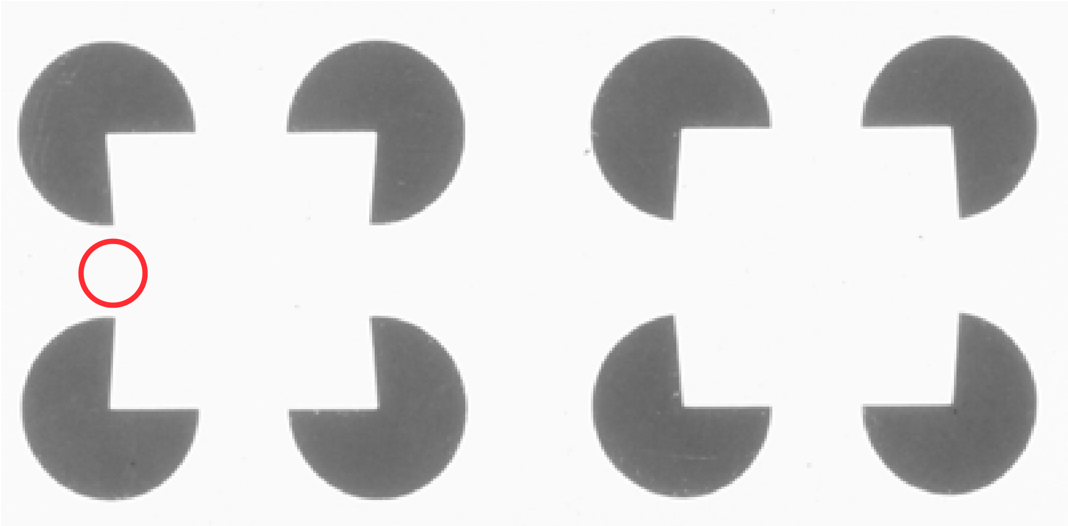
Illusory Contours

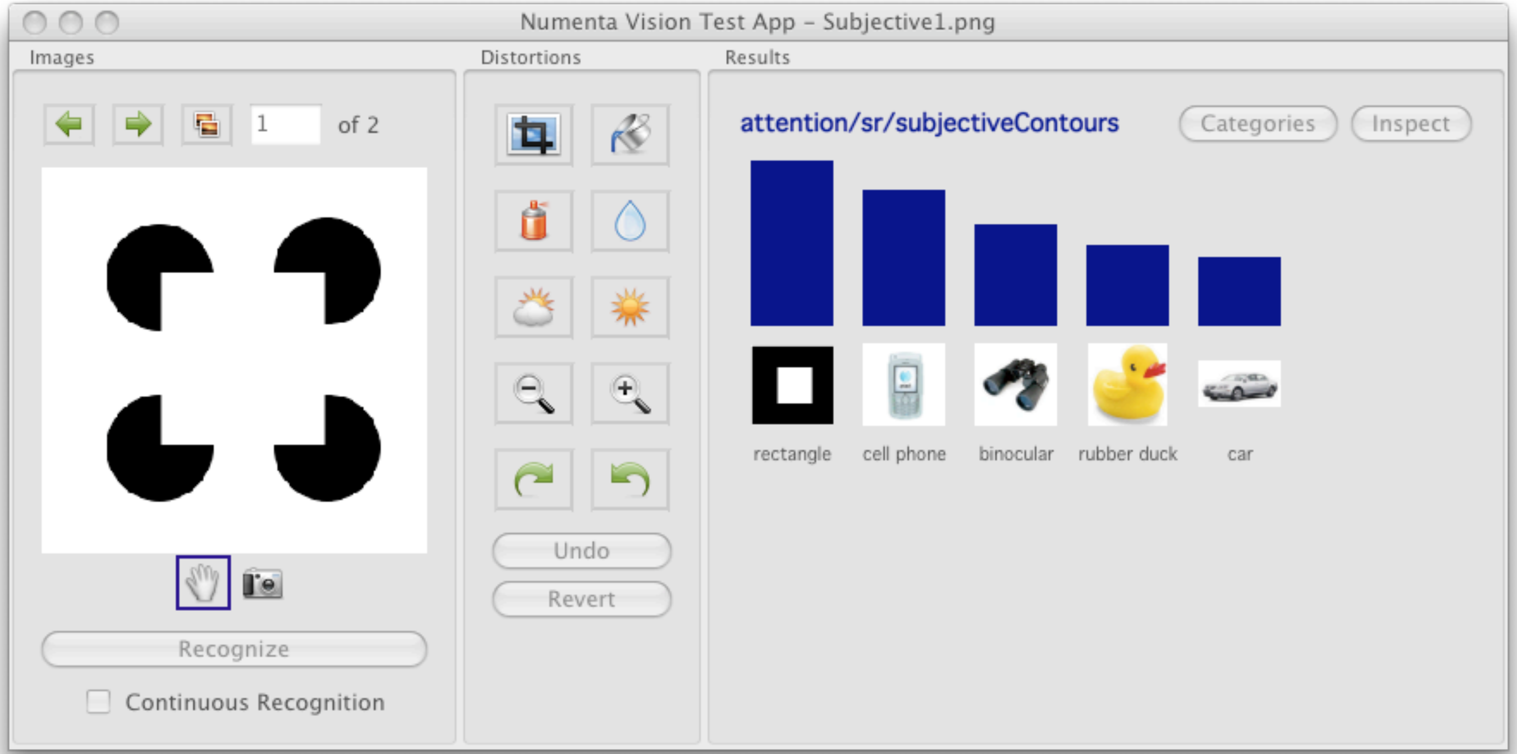


Lee and Mumford, J.Opt. Soc. America. July 2003

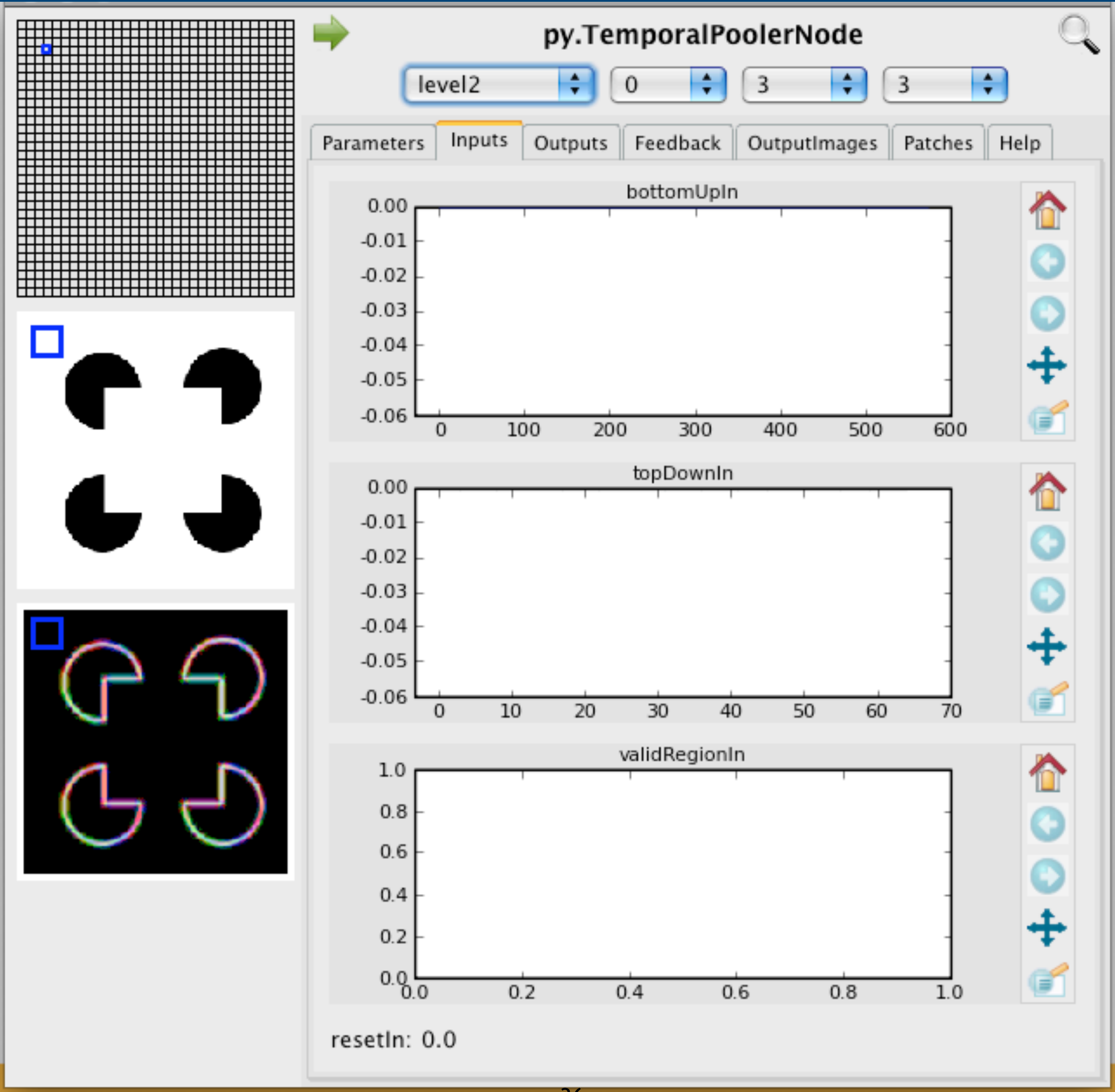
Illusory Contours



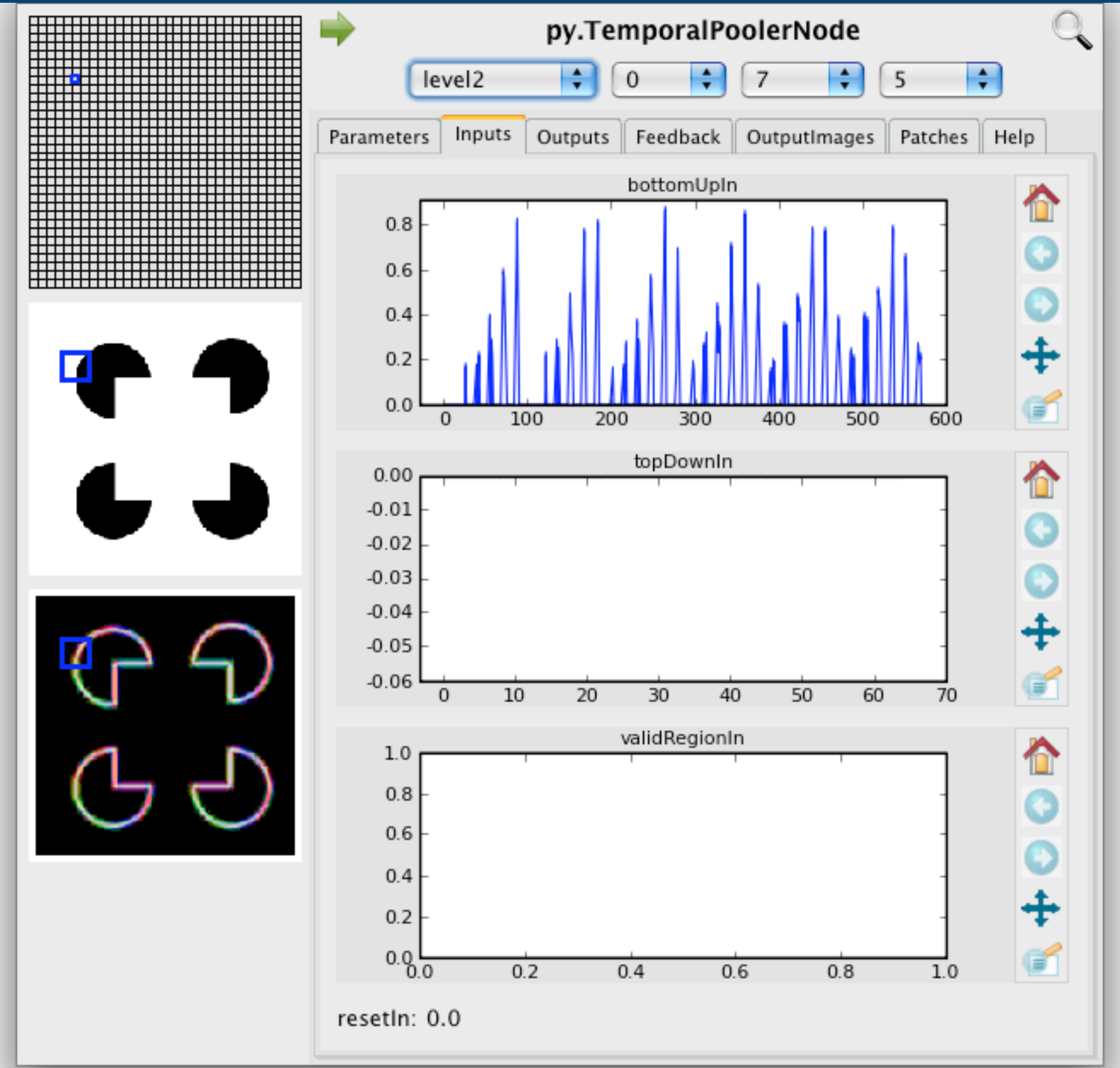



Numenta Vision Test App – Subjective1.png
Images
Distortions
Results
1
of 2
attention/sr/subjectiveContours
Categories
Inspect
rectangle
cell phone
binocular
rubber duck
car
Undo
Revert
Recognize
Continuous Recognition

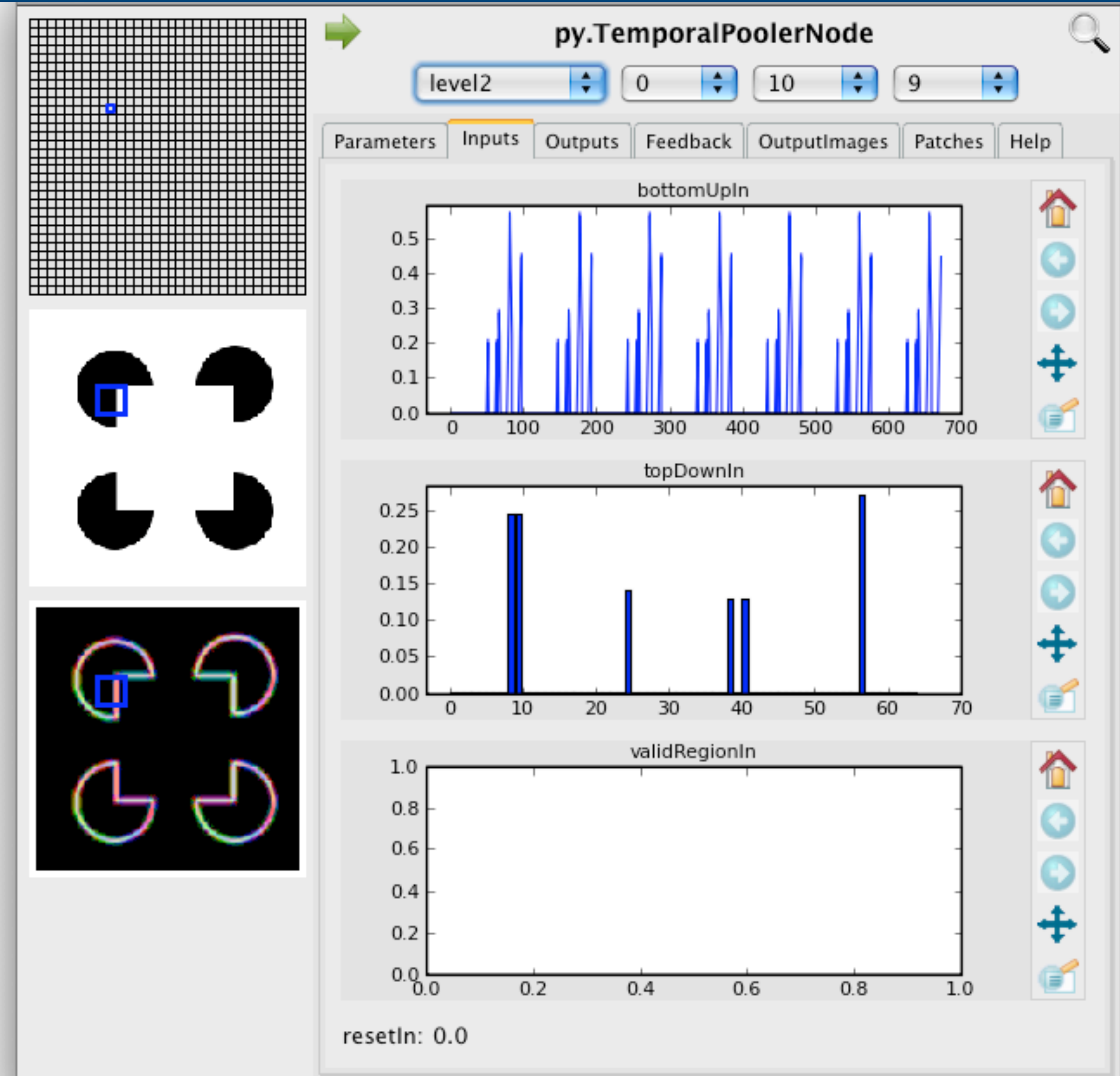
No contours


py.TemporalPoolerNode
level2
0
3
3
Parameters
Inputs
Outputs
Feedback
OutputImages
Patches
Help
bottomUpIn
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
0 100 200 300 400 500 600
topDownIn
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
-0.06
0 10 20 30 40 50 60 70
validRegionIn
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0
resetIn: 0.0

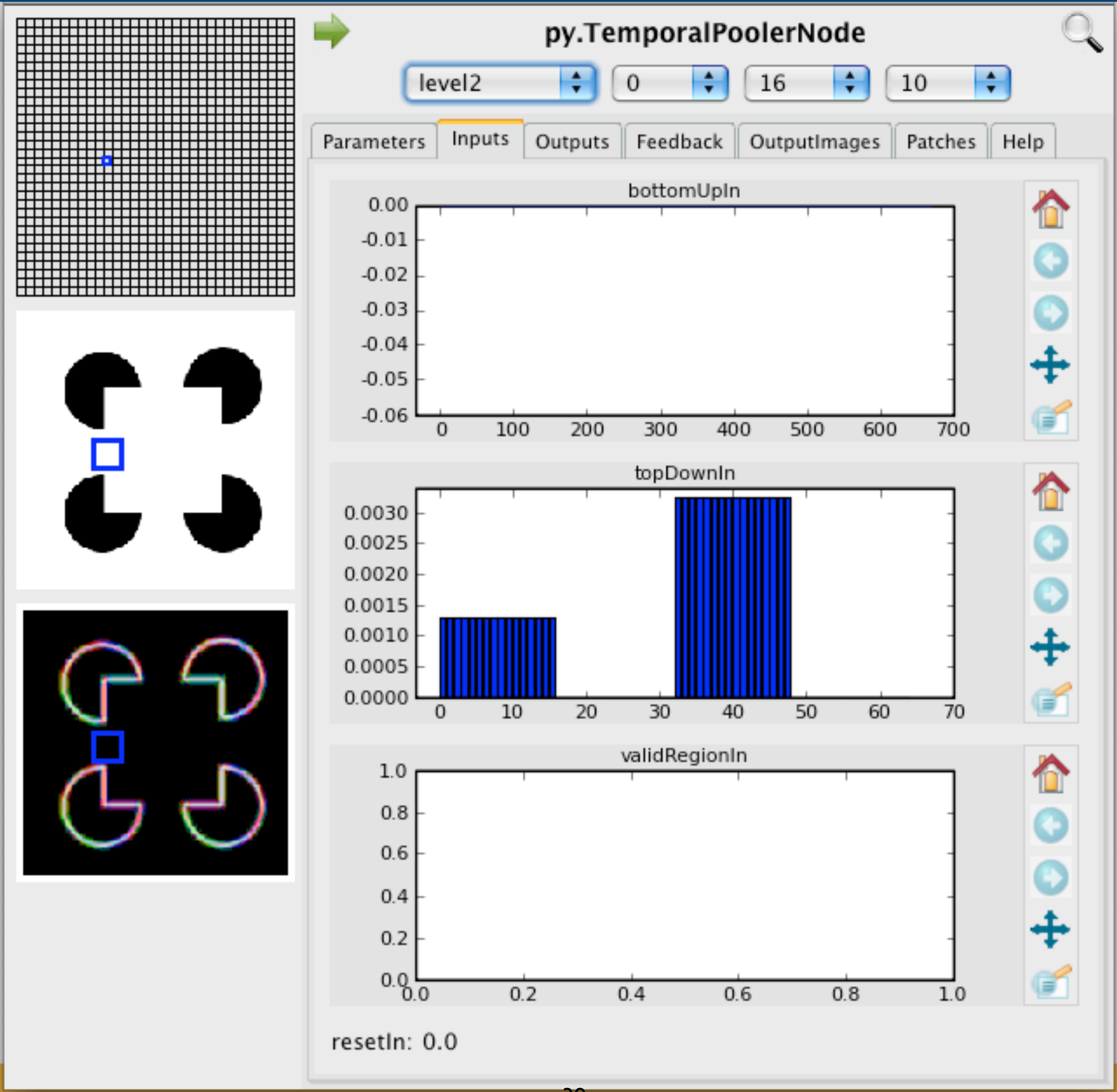
Bottom-up. No Top-down



Bottom-up And Top-Down



Subjective Contour



Computing platform prognosis

- Algorithms are mature enough in application areas like vision
- It is very likely that there will be embedded applications in near future
- The first embedded circuits are likely to be inference-only
  - Learning done offline using parallel software
- Time is right to start specing out a computational platform for HTM-like computation

To learn more

- Download Vision4 Demo from www.numenta.com
- Read On Intelligence
- Read white papers from Numenta website
- Read Dileep's PhD thesis *"How the brain might work...*
  - search for "Dileep thesis" on google

Thank You