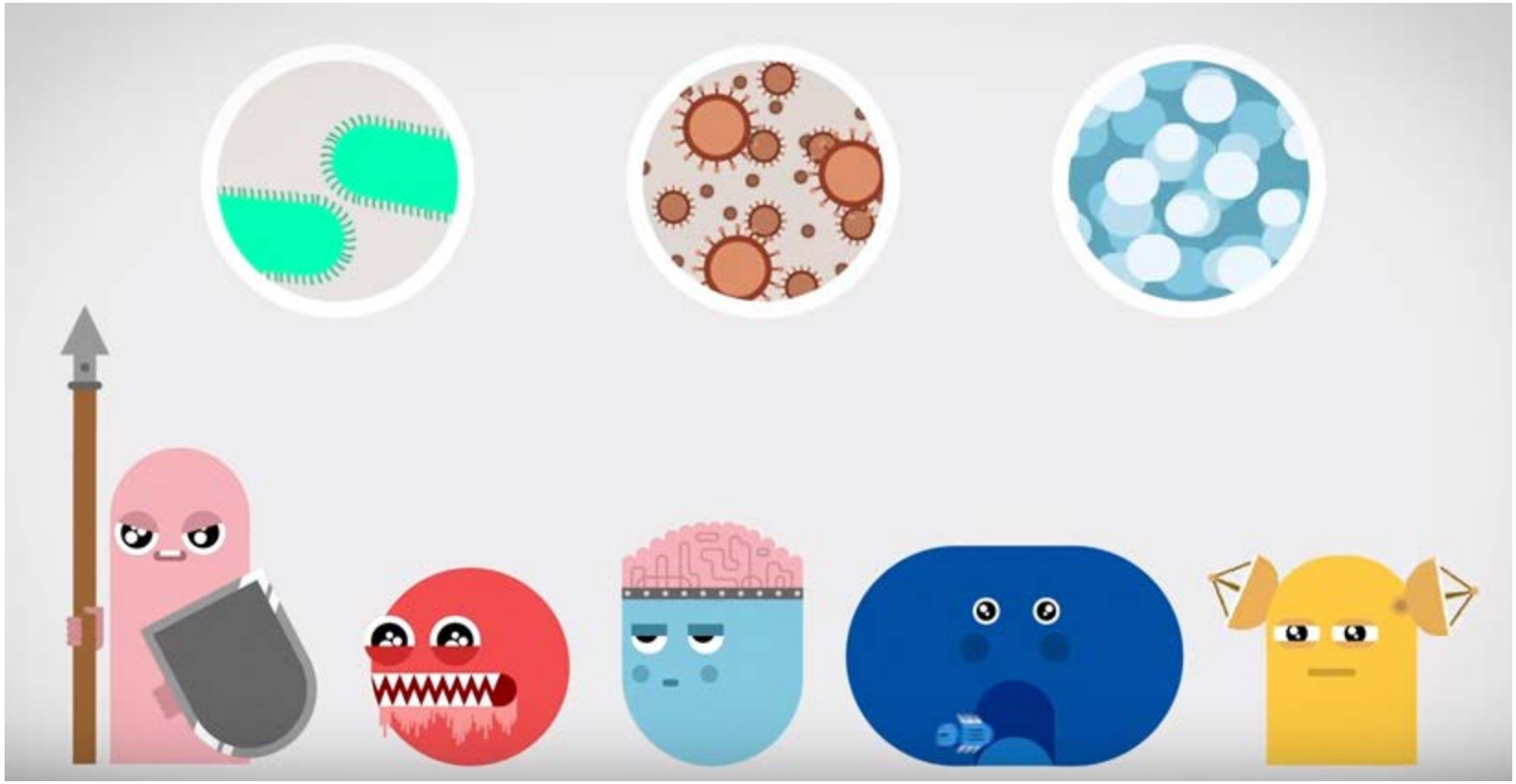
Architecture of an Artificial Immune System

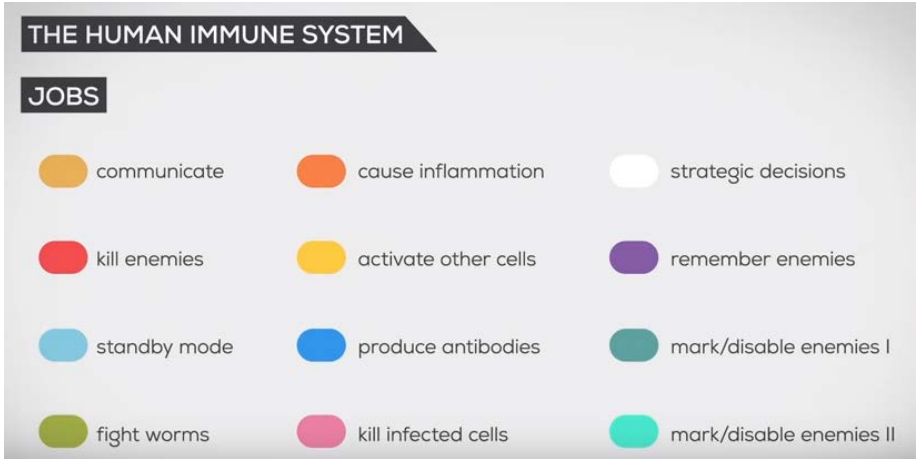
Steven A. Hofmeyr, S. Forrest

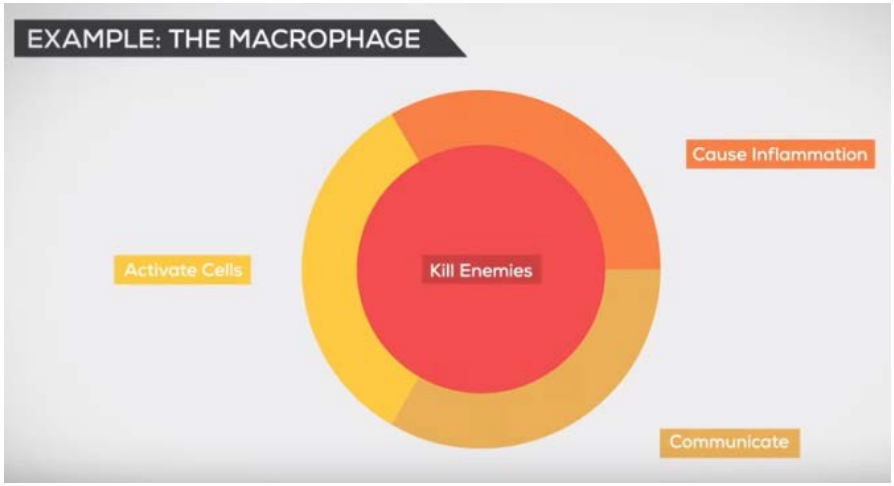
Presentation by
Ranjani Srinivasan
(ranjanis@andrew.cmu.edu)

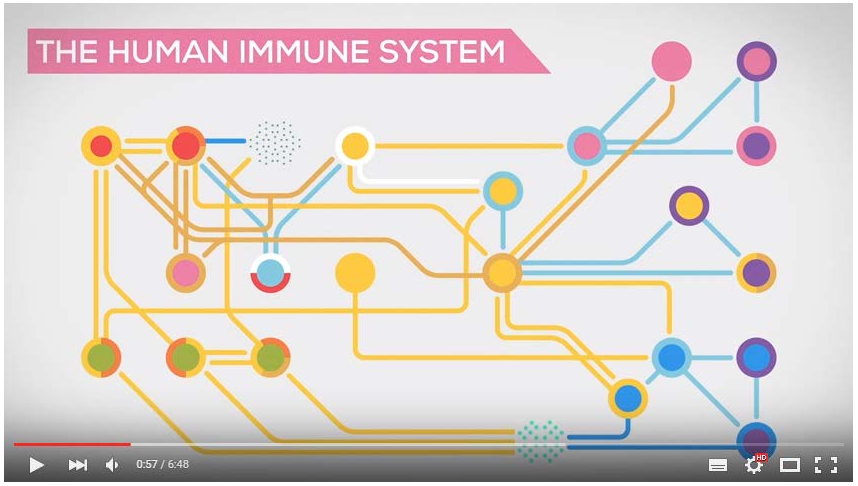
The Fellowship of the Immune System- The Defenders of the Realm



The Fellowship of the Immune System- The Defenders of the Realm







Courtesy: Kurzgesagt – In a Nutshell, YouTube Channel. The Immune System explained

Properties of the Immune System (IS)

- Diverse, Distributed, Error Tolerant, Dynamic and Self-monitoring

- Robustness
- Adaptable – Recognize and respond to new infections, retain memory
- Autonomous – No outside control, difficult to impose outside control or inside centralized control.

(Exceptions exist)

Basic Components of Immune Response

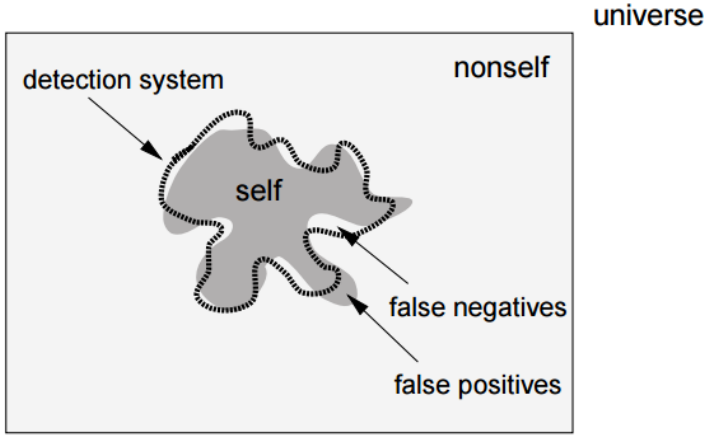
- Localized Interactions – chemical bonding
- Dynamic system of circulation
- Decentralized; no hierarchical organization
- Self-from-Nonself
- Improve to Harmful Nonself-from-EverythingElse
- Two sub-problems
  - Detection
  - Elimination

ARTificial Immune System

- ARTIS – Incorporates (most) properties of IS
- Independent of problem domain
- Situating in a domain can reduce unnecessary features or tailor features to problem

Problem definition

- Protein Chains – binary strings length l.
- Disjoint (assumption) subsets of Universe U, S and N
- Discrimination or Classification Task
- Errors – False Positives and False Negatives



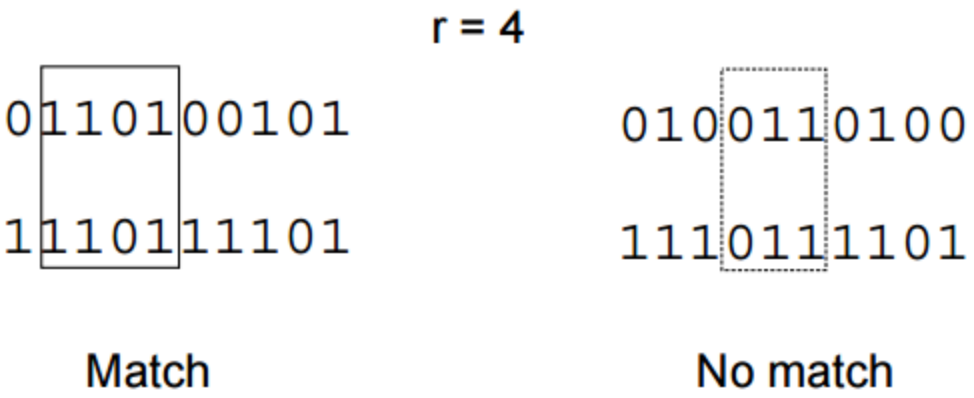
Detectors

- Modeled after one class – Lymphocytes
- Combines properties of B-cells, T-cells, and antibodies
- Distributed environment modeled as graph

$$G = (V,E)$$

- Affinity to epitopes (region on pathogen) – approximate string matching
- *r-contiguous bits* (More biologically consistent)

Detectors

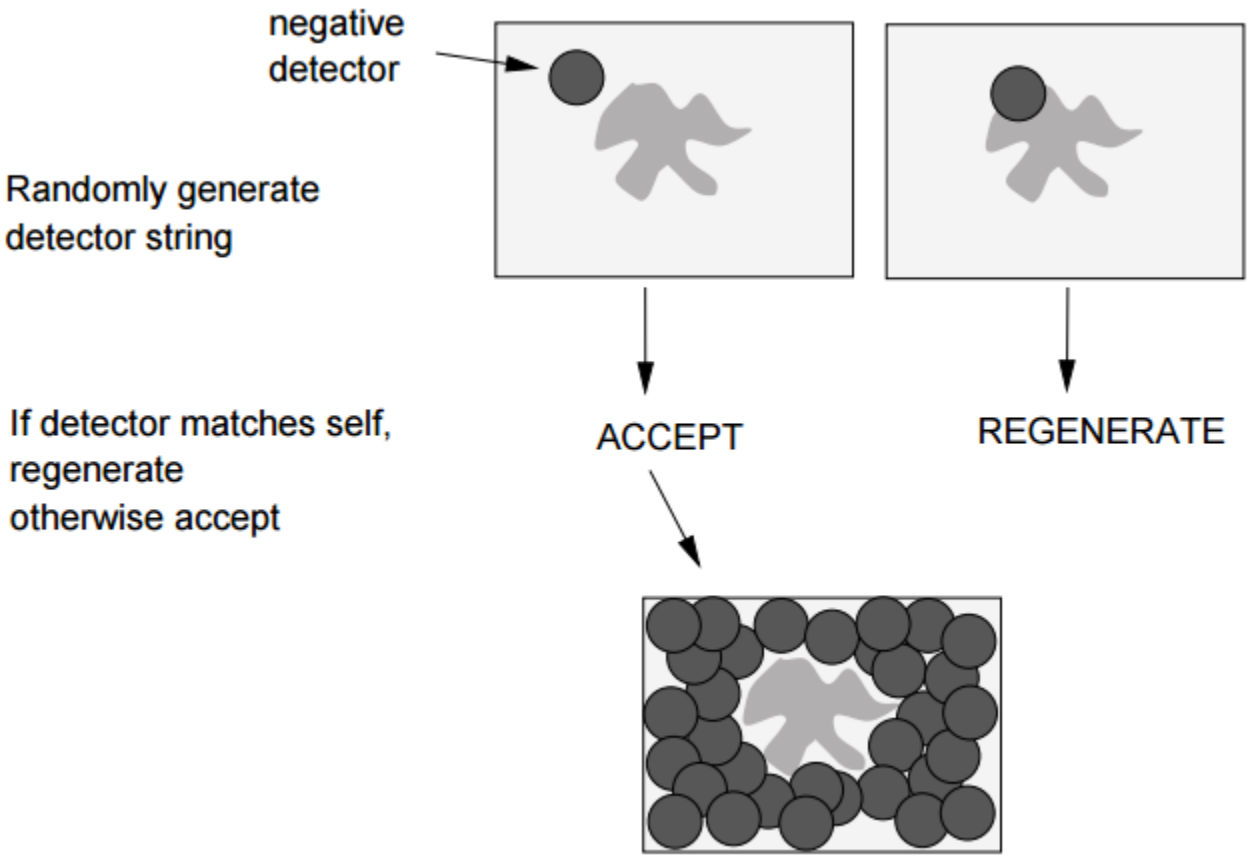


- Activation of lymphocyte – when binding receptors exceeds threshold
- Modeling Activation Threshold – Match atleast τ strings in given time. Decay match count (ϒ). Once activated, reset count to zero.

Training the Detection System

- Negative Selection Algorithm
- Tolerization – in Thymus. Training set of self
- Assumption: *Self occurs frequently compared to non-self*
- ARTIS – Distributed Tolerization

Negative Selection Algorithm



Memory

- Rapid and efficient secondary response
- Associative
- Activated lymphocytes clone; Retain memory cells
- Multiple detectors at node in competition. Closest match – winner. Spread to neighboring nodes
- Memory detectors have lower activation thresholds => rapid response

Sensitivity

- Cytokines (chemicals) – signal to nearby IS cells
- Detection node D_i , local sensitivity ω_i,
- Threshold of detectors at D_i is (τ - ω_i,)
- Matches at i go up, sensitivity is increased by 1
- Temporal horizon with decay rate Υ_w

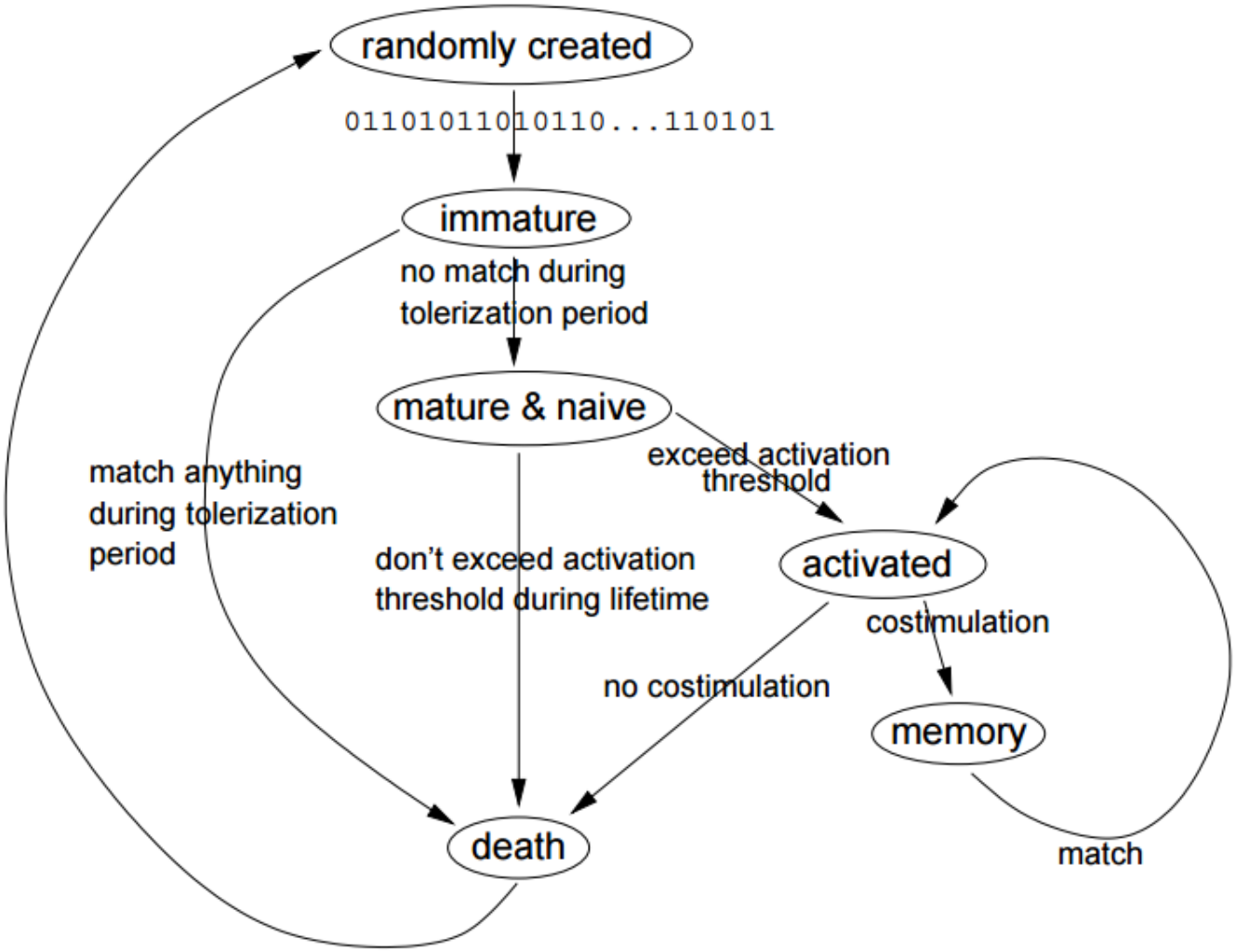
Co-stimulation

- T-cells require second signal of “damage”
- Model crude approximation of co-stimulation: human operator
- Co-stimulation delay T

Lifecycle of Detector

- Lympocytes short-lived. Dynamic population
- Model: p_{death} for mature detectors
- Exception: Memory Detectors. Die only if no co-stimulation
- Problem?
- Limit fraction of memory detectors m_d
- LRU (least Recently used) => Least useful (Is this a valid assumption?)

Lifecycle of Detector

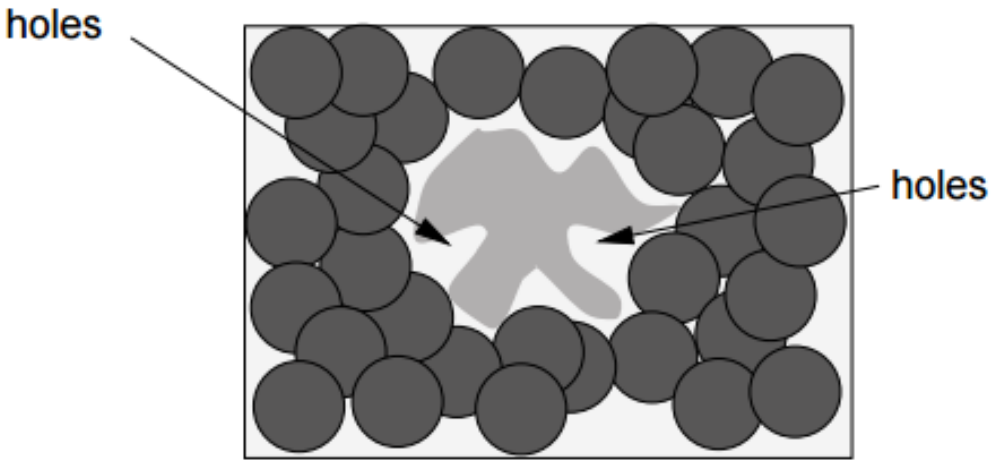


Representation

- Population level diversity – MHC.

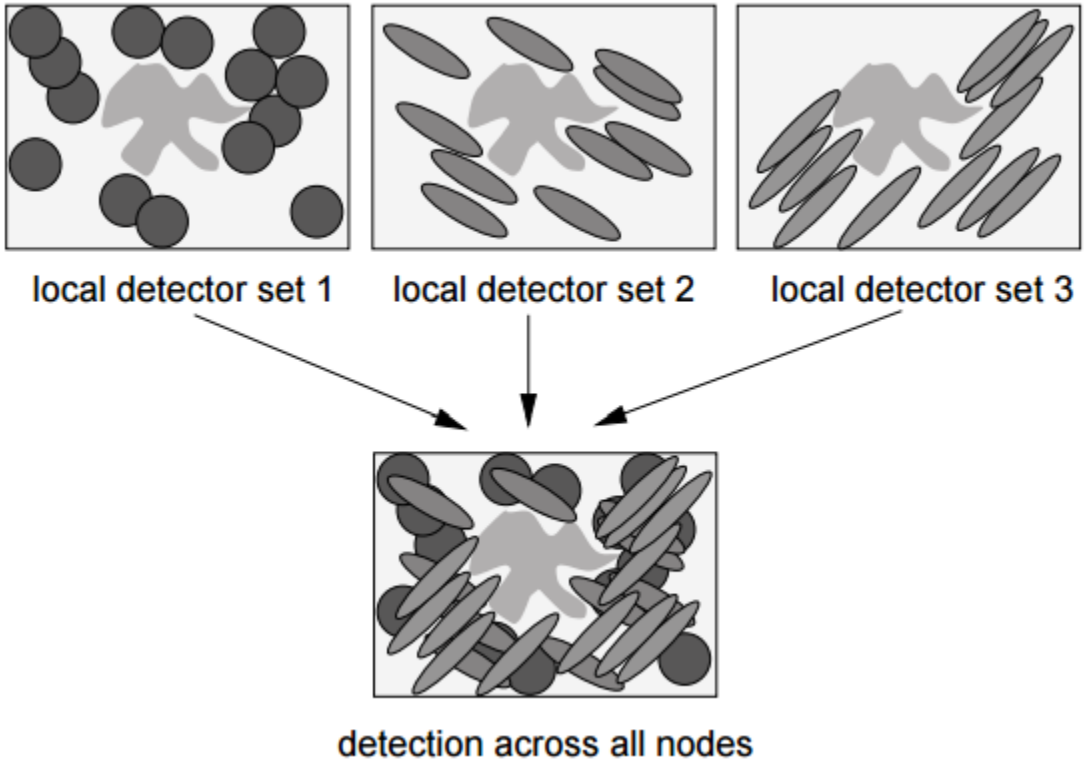
Holes – occur if every match has a self counterpart;

No valid detectors can be generated



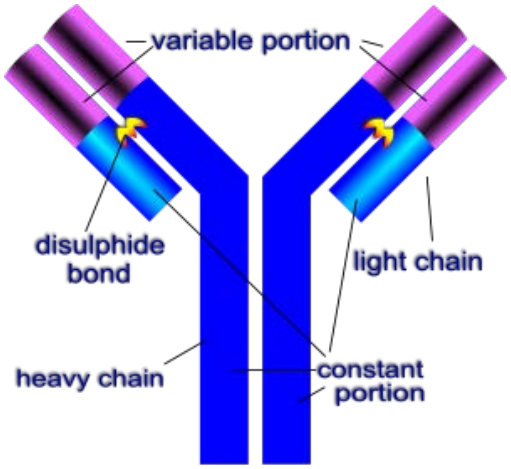
Representation

- Each node with different representation. Modify all incoming detectors



Response

- Effector selection – many kinds
- B-cells – antibodies; Variable and constant regions
- Isotype Switching
- Implementation: Augment detector with effector choice



Summary

What makes a system suitable for ARTIS?

- Pattern classification and Response
- Distributed architecture, scalable to arbitrary number of nodes
- Require detection of novel anomalous patterns
- Dynamic but normal behavior changes slowly
- Robust solution with no central control

Criticism

- Activation Threshold reduces false positive.
  But introduces paths of attack
  - Infrequent anomalous connections
  - Attack has fewer connections than threshold
- Assumption of infrequently occurring non-self
- LRU model assumption