COMP9208
Artificial Intelligence and Society

Lecture 6: 12 September 2025

Prof. Mikhail Prokopenko
Dr. Sheryl Chang

Canvas: https://canvas.sydney.edu.au/courses/66452

Email: mikhail.prokopenko@sydney.edu.au
Office: J12 (Computer Science), level 3W, room 324

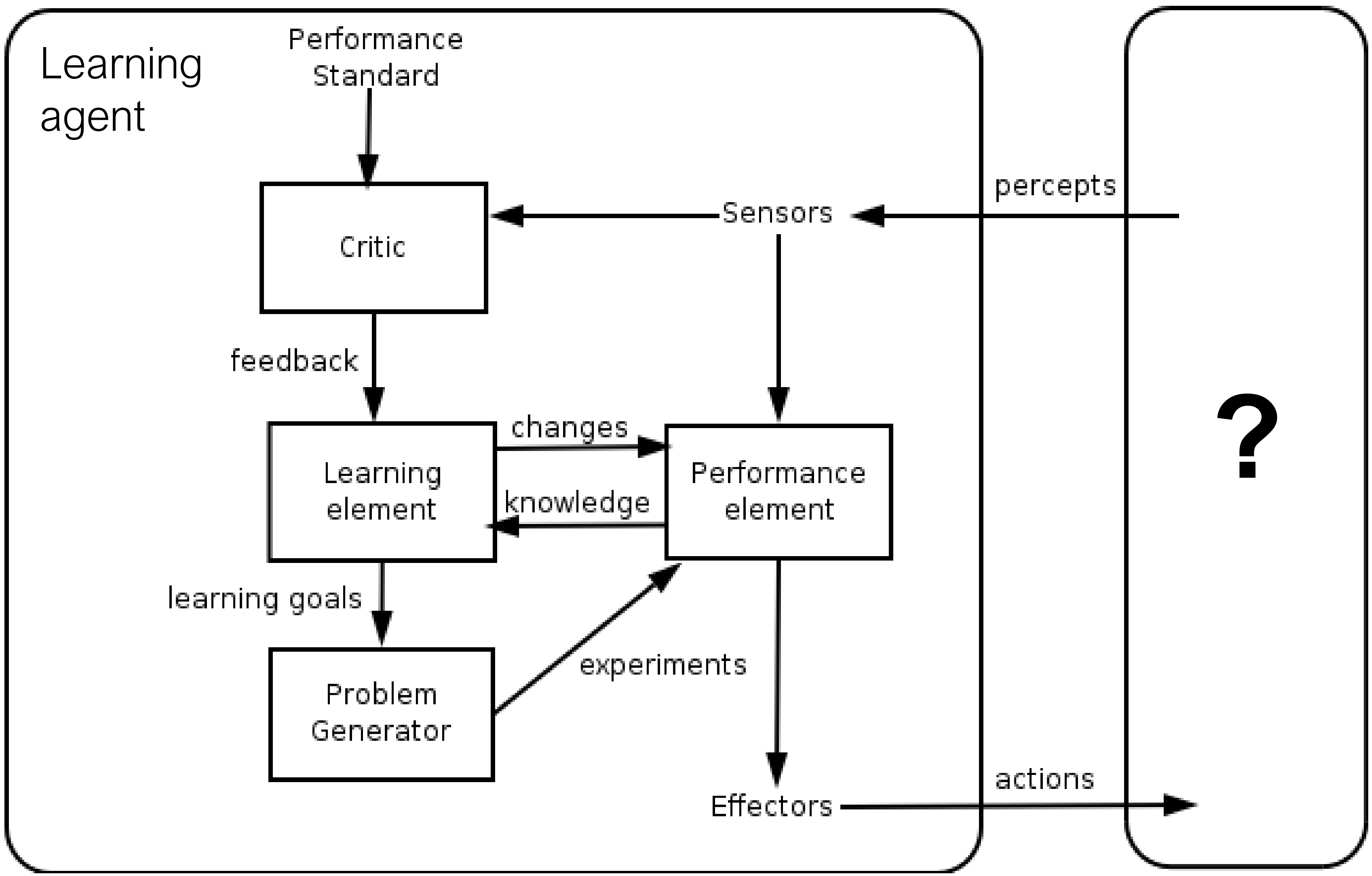
Date	Week: Lecture	Assignment
08/08	1: History of AI	
15/08	2: History of AI + Perception and Actuation	
22/08	3: History of AI + Representation and Reasoning	
29/08	4: Machine Learning	
05/09	5: Natural Interaction	
12/09	6: Distributed AI (swarm intelligence)	
19/09	7: From Good Old-Fashioned AI (GOFAI) to Artificial General Intelligence (AGI)	# 1: Individual assignment (article review)
26/09	8: Goal-driven AI: search & rescue vs lethal autonomous weapon systems (LAWS)	
03/10	**BREAK**	
10/10	9: Review: language models, search trees, situated behaviours, neural networks	
17/10	10: Robotic swarms: RoboCop vs RoboCup	
24/10	11: Group Presentations	# 2: Group assignment
31/10	12: Adversarial Machine Learning	
07/11	13: Limits of cognition	
17/11	14: STUVAC	# 3: Individual assignment (data analysis report)

Brief review of last week

?

Intelligent agents

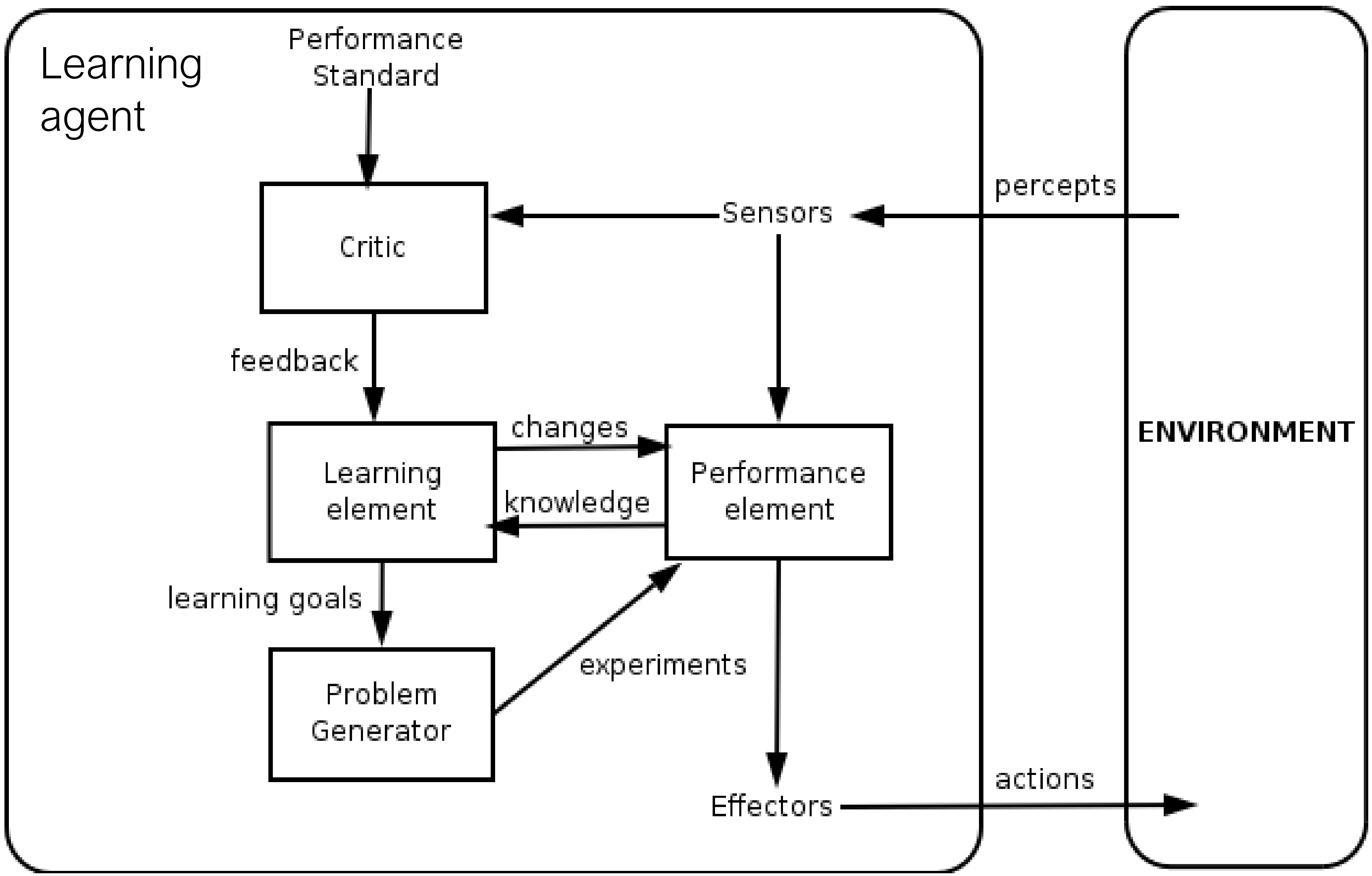
- an agent perceives its environment, processes its sensory information, takes actions autonomously in order to achieve goals (for example, by reacting or planning), and may improve its performance with learning or acquiring knowledge



https://en.wikipedia.org/wiki/Intelligent_agent

Intelligent agents

- an agent perceives its environment, processes its sensory information, takes actions autonomously in order to achieve goals (for example, by reacting or planning), and may improve its performance with learning or acquiring knowledge

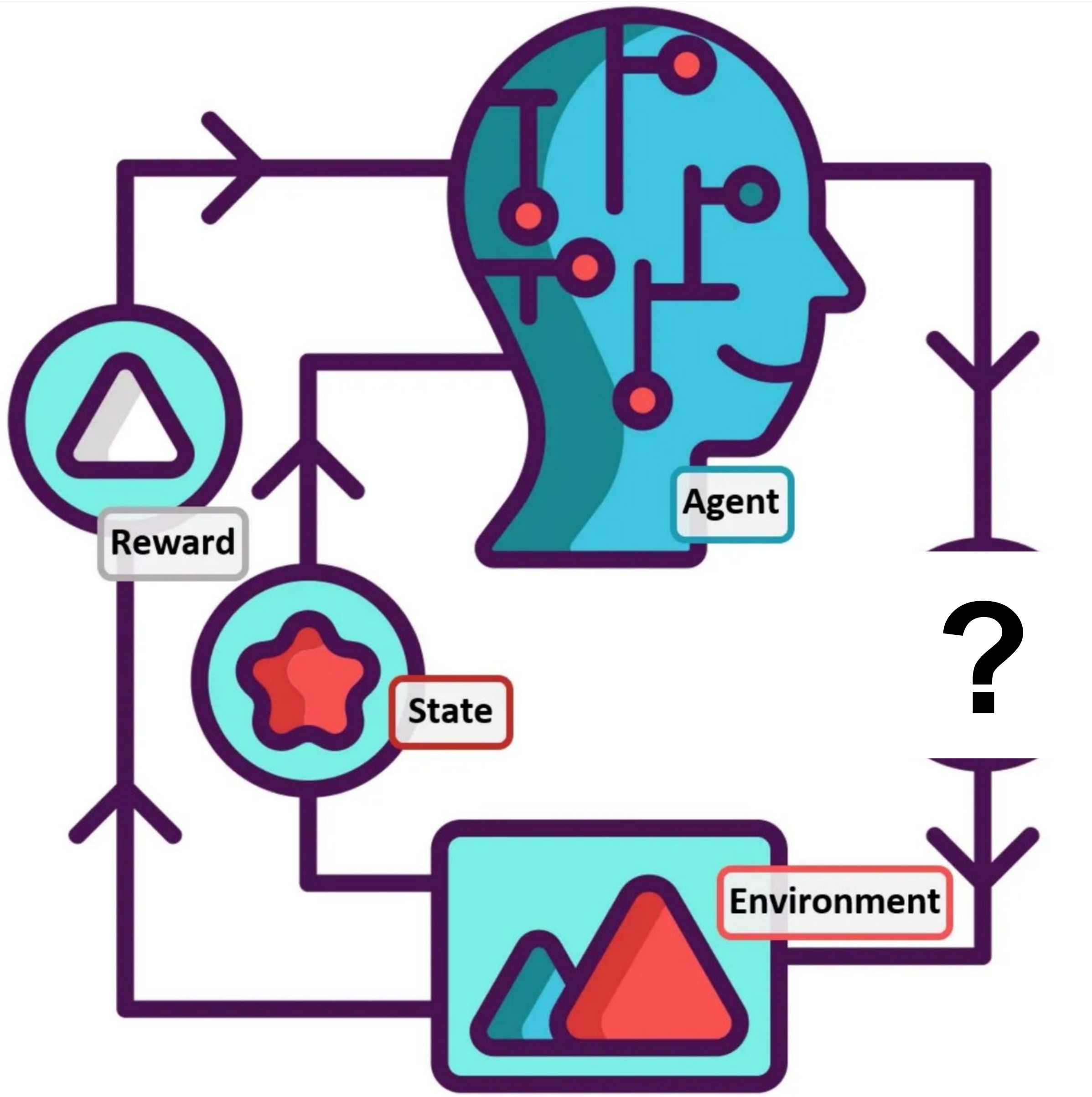


https://en.wikipedia.org/wiki/Intelligent_agent

Machine Learning

Reinforcement learning:

- agent learns to make decisions by doing actions in an environment to maximise cumulative reward
- the agent learns from the effects of its actions rather than from explicit instruction

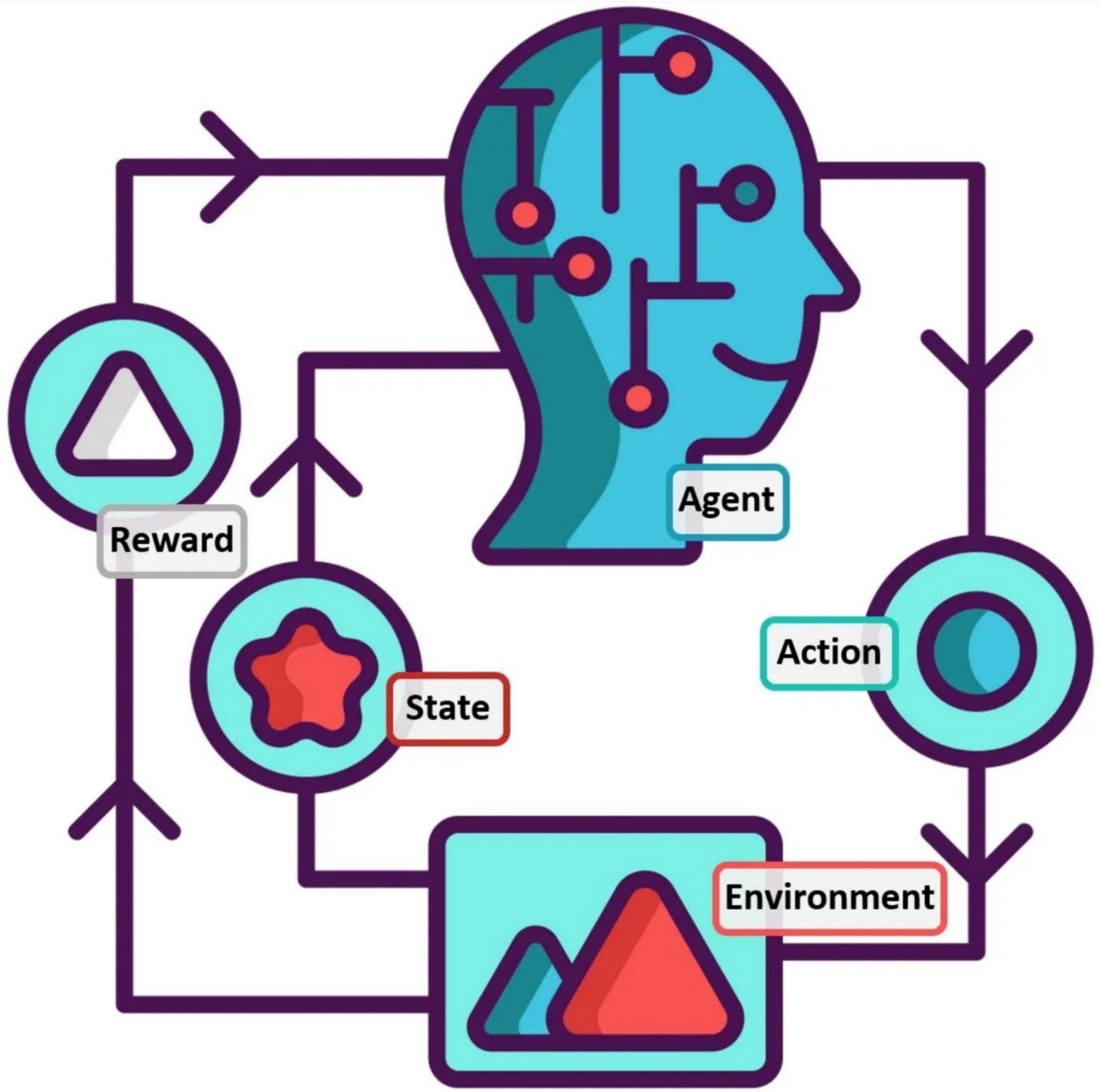


https://towardsdatascience.com/reinforcement-learning-fda8ff535bb6

Machine Learning

Reinforcement learning:

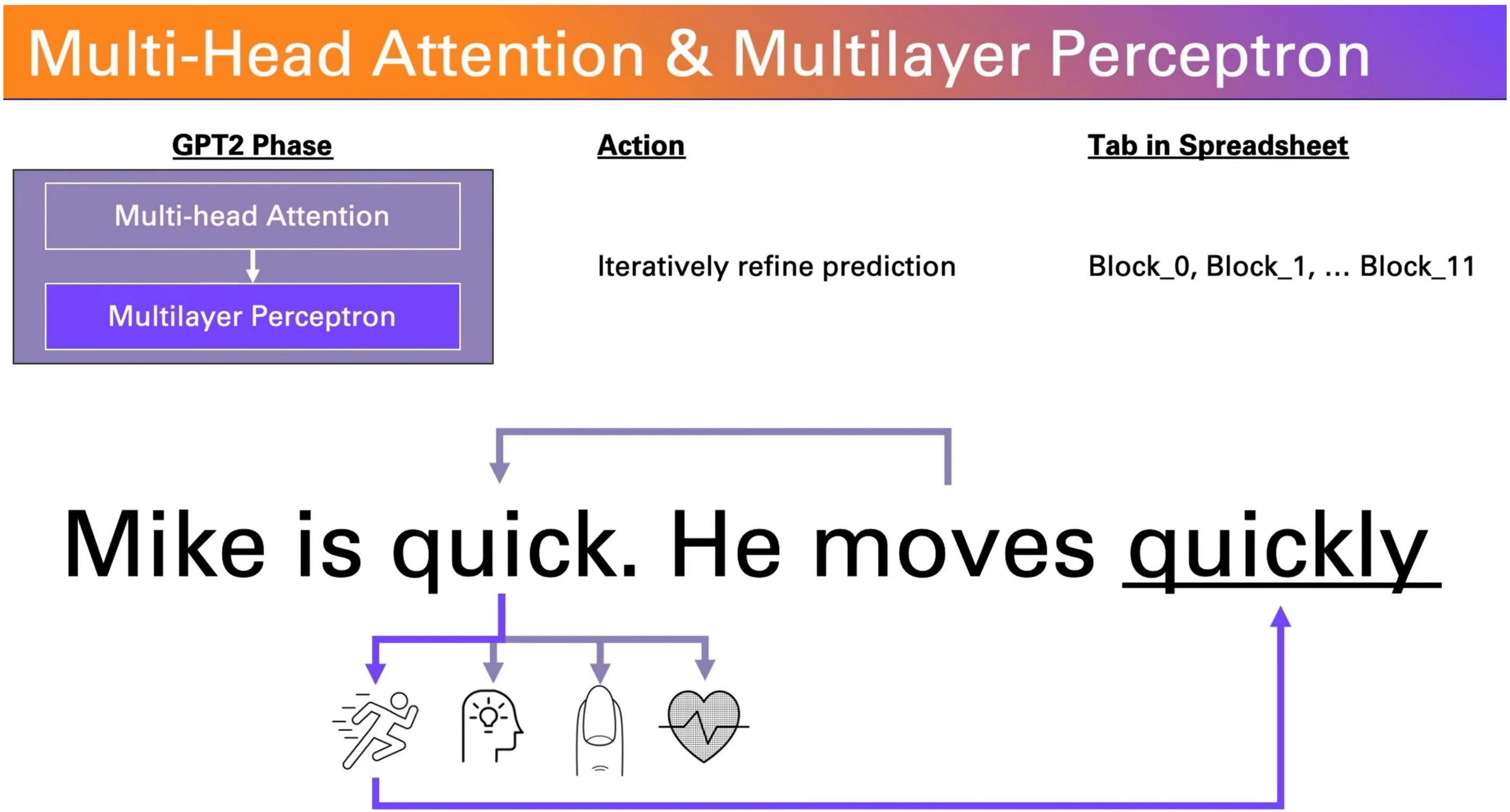
- agent learns to make decisions by doing actions in an environment to maximise cumulative reward
- the agent learns from the effects of its actions rather than from explicit instruction



https://towardsdatascience.com/reinforcement-learning-fda8ff535bb6

Natural Interaction

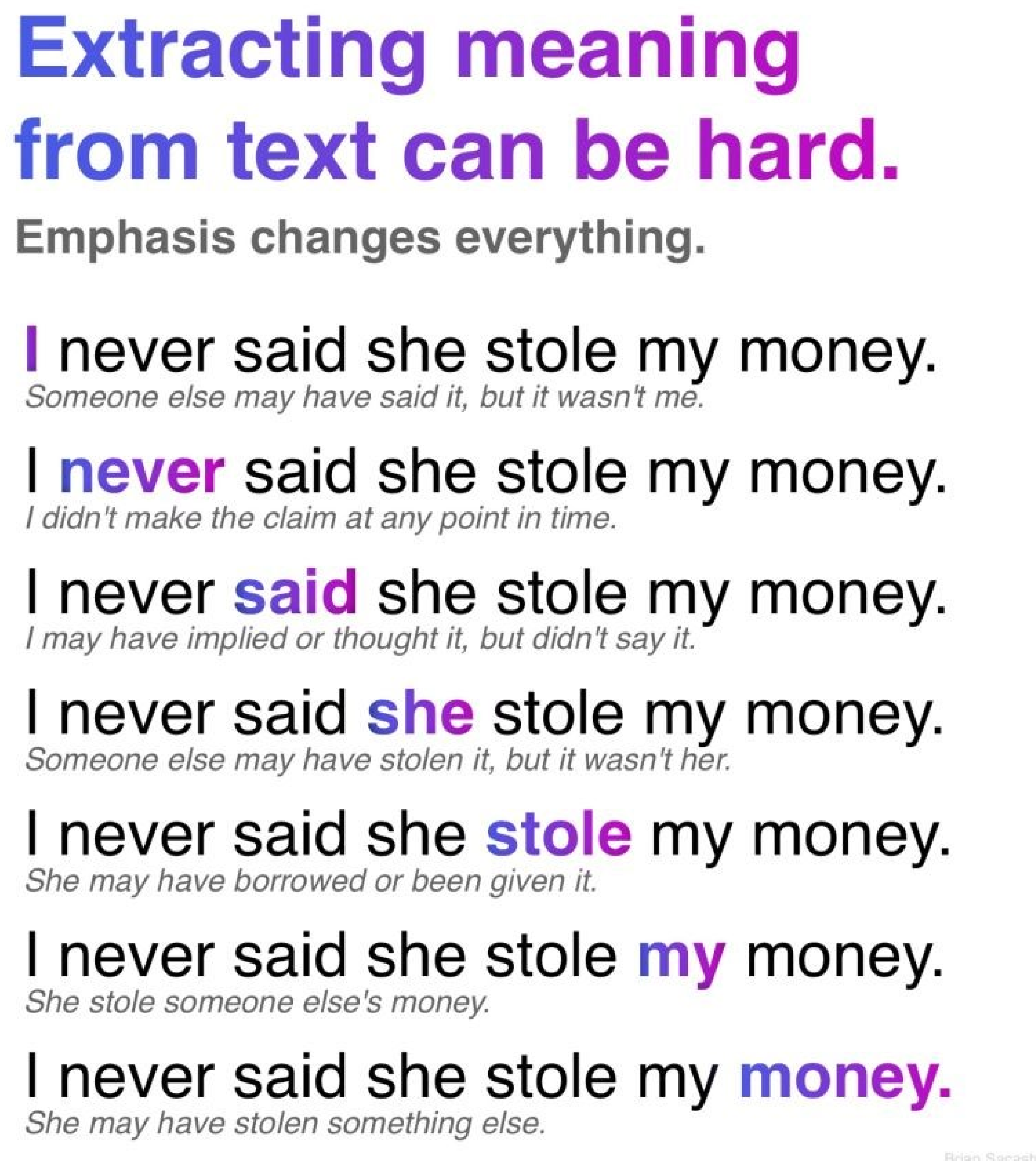
Natural Language Processing (NLP): the ability of AI systems to understand, interpret, and generate human language



https://spreadsheets-are-all-you-need.ai/index.html#watch-the-lessons

Natural Interaction

Natural Language Processing (NLP): the ability of AI systems to understand, interpret, and generate human language



Source: Brian Sacash (Internet)

Natural Interaction

Ethical concerns:

- misuse of technology
  -
  -

?

- autonomy and manipulation
  - virtual assistants or chatbots designed to persuade users to make purchases or share personal information
  - AI systems used in political campaigns to influence voter behavior
- psychological impact
  - users developing emotional attachments to AI companions or virtual assistants
  - reduced face-to-face interactions due to reliance on AI communication tools
- accessibility and inclusion
  - speech recognition systems may struggle with non-standard speech patterns or disabilities
  - interfaces may not be accessible to visually or hearing-impaired users

ChatGPT 4o

Natural Interaction

Ethical concerns:

- misuse of technology
    - deepfake technology used to create realistic but fake videos or audio recordings
    - social engineering attacks using AI to mimic trusted voices or personas
- autonomy and manipulation
    - virtual assistants or chatbots designed to persuade users to make purchases or share personal information
    - AI systems used in political campaigns to influence voter behavior
- psychological impact
    - users developing emotional attachments to AI companions or virtual assistants
    - reduced face-to-face interactions due to reliance on AI communication tools
- accessibility and inclusion
    - speech recognition systems may struggle with non-standard speech patterns or disabilities
    - interfaces may not be accessible to visually or hearing-impaired users

ChatGPT 4o

Homework

- check the information sources provided on different slides (at least some)
- “A low-code way to learn AI. Learn how AI works from a real LLM implemented entirely in Excel”:
  https://spreadsheets-are-all-you-need.ai/index.html#watch-the-lessons
- Please try to watch the first few video lessons before the next lecture, and prepare to give a few brief comments about the video(s):
  - what was most striking about the AI technology in the video(s)?
  - what societal impact did the AI make?
  - what were the main AI limitations, biases and risks?

?

Questions

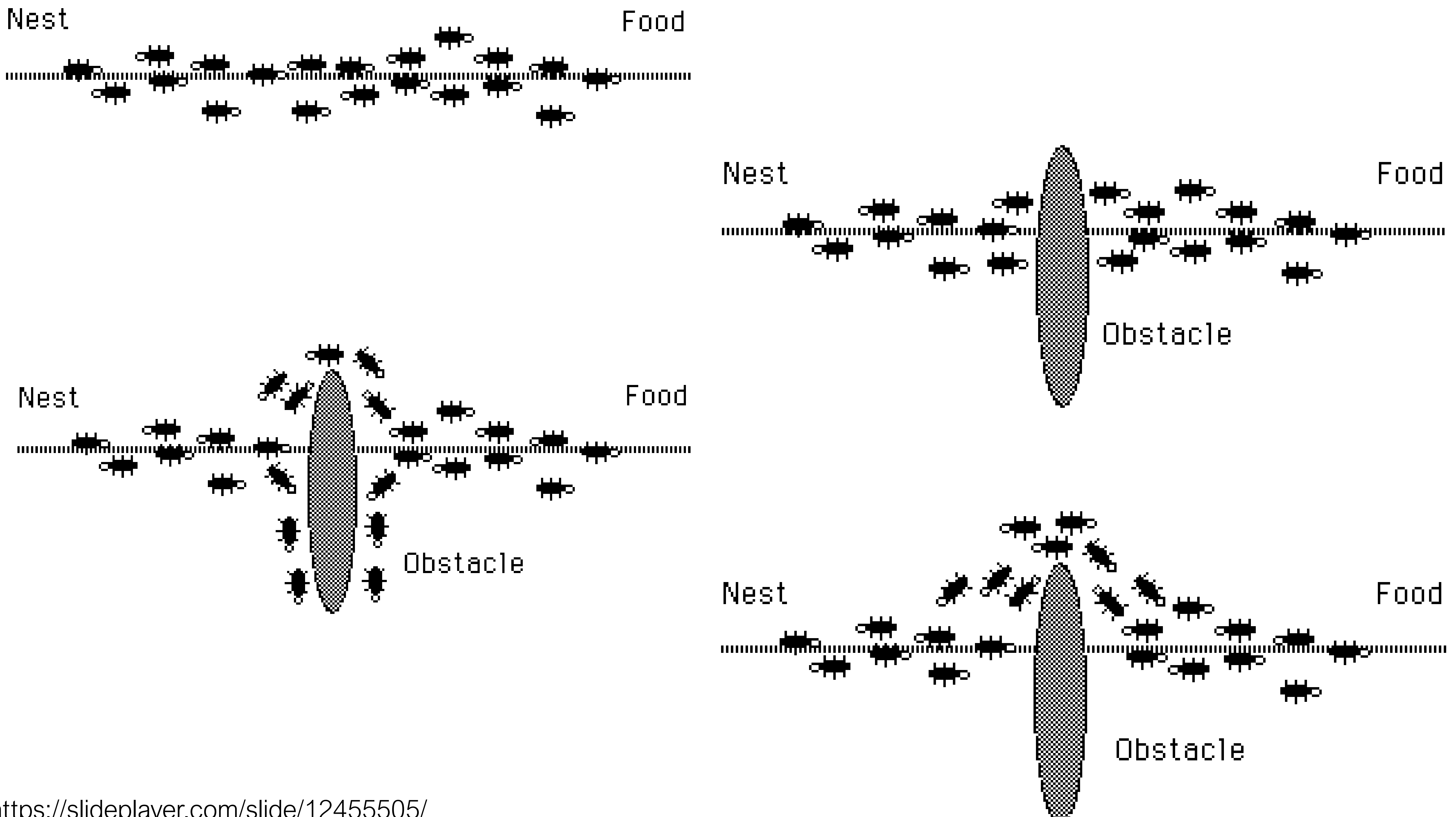
?

Distributed AI

- multi-agent systems (MAS)
- distributed machine learning (DML)
- distributed constraint satisfaction problems (DCSP)
- distributed reasoning
- swarm intelligence:
  - Ant Colony Optimization (ACO): optimisation algorithms inspired by the foraging behavior of ants
  - Particle Swarm Optimization (PSO): optimisation algorithms based on the social behavior of birds flocking or fish schooling
  - Artificial Immune Systems (AIS): anomaly detection, pattern recognition, and optimisation algorithms inspired by the human immune system's ability to recognize and eliminate pathogens
- evolutionary algorithms:
  - genetic algorithms
  - evolutionary strategies
  - genetic programming

Distributed AI

Swarm intelligence: Ant Colony Optimisation



https://slideplayer.com/slide/12455505/

International Space Station: a long-duration mission

Contributors
...with a tribute to A. J. Farmer and D. C. Price

W.H. Prosser[1*], S. G. Allison [1], S.E. Woodard[1], R. A. Wincheski[1], E.G. Cooper[1], D.C. Price[2], M. Hedley[3], M. Prokopenko[3], D. A. Scott[2], A. Tessler[1] and J. L. Spangler[4]

[1]NASA Langley Research Center, MS 231, Hampton, VA 23681, USA
[2]CSIRO Industrial Physics, PO Box 218, Lindfield, NSW 2070, Australia
[3]CSIRO ICT Centre, PO Box 76, Epping, NSW 1710, Australia
[4]Lockheed Martin Aeronautics Co., NASA Langley Research Center, Hampton, VA 23681, USA

D. A. Scott[1], A. Batten[1], G. C. Edwards[1], A. J. Farmer[1], M. Hedley [2], N. Hoschke[1], P. Isaacs[1], M. Johnson [2], A. Murdoch[1], C. Lewis[1], D. C. Price[1], M. Prokopenko [2], P. Valencia[2] and P. Wang [2]

[1] CSIRO Industrial Physics
Lindfield, NSW Australia

[2] CSIRO Information and Communications Technologies (ICT) Centre, Marsfield, NSW Australia

Typical shuttle launch

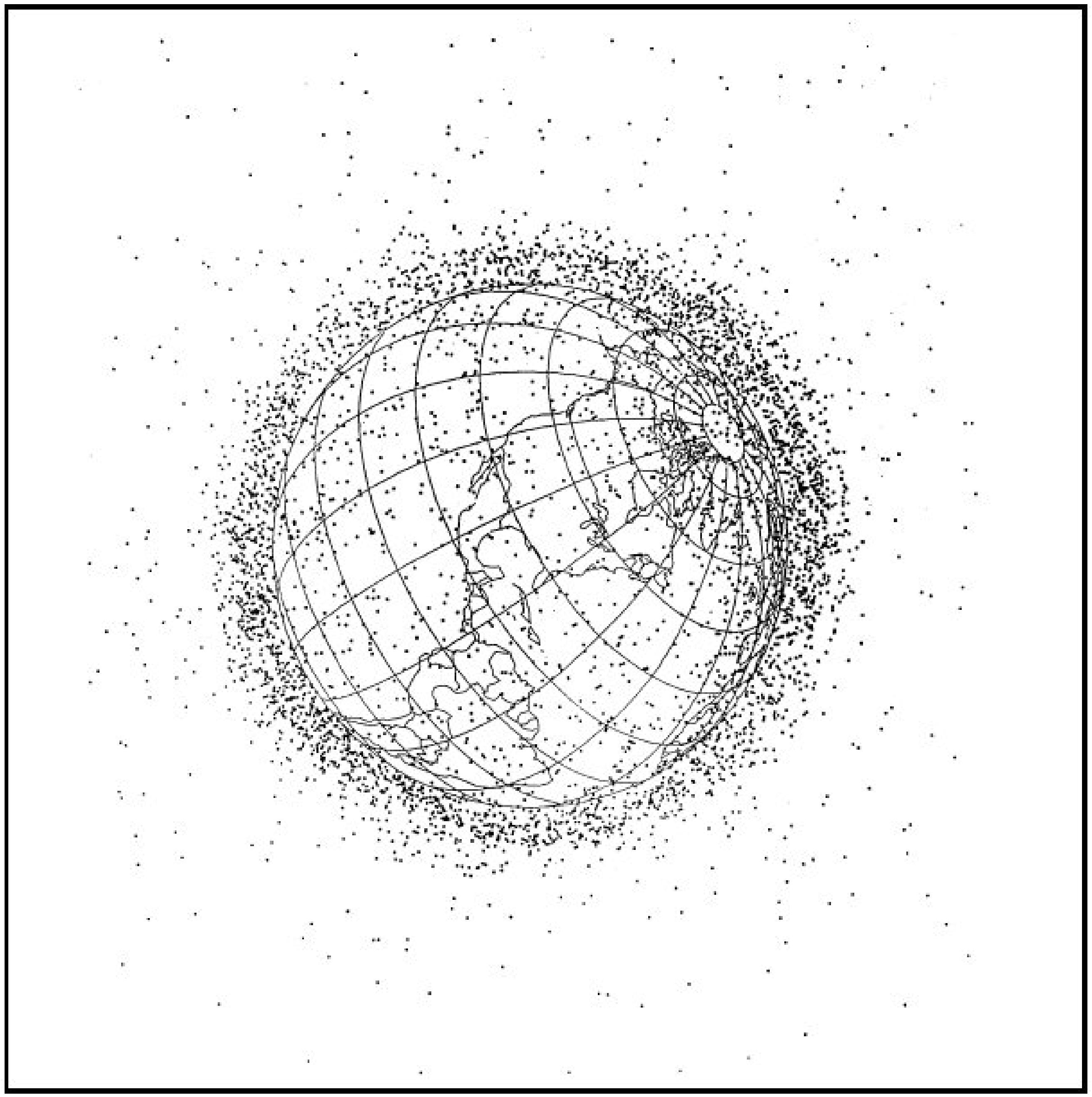
Towards self-repairing “ageless” aerospace vehicles (AAV)

ISS005E21472

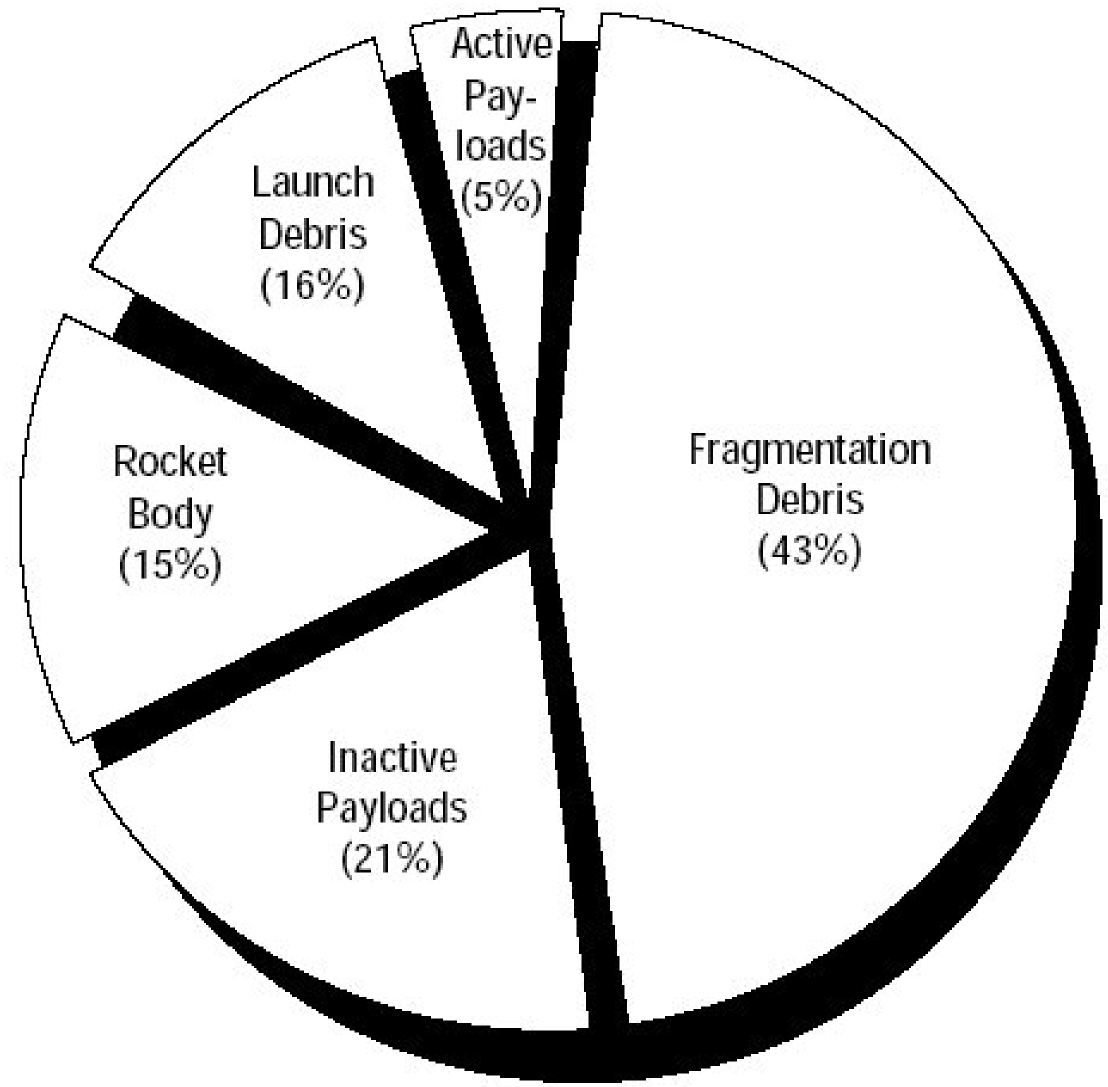
Langley Research Center, NASA

CSIRO Australia

Orbital debris



The Earth debris environment in July 1987



Sources of the catalogued debris population

Structural Health Monitoring and Management (SHMM)

- **General**
  - damage detection → self-repair
  - integrated with other subsystems:

    navigation, control, propulsion, etc.

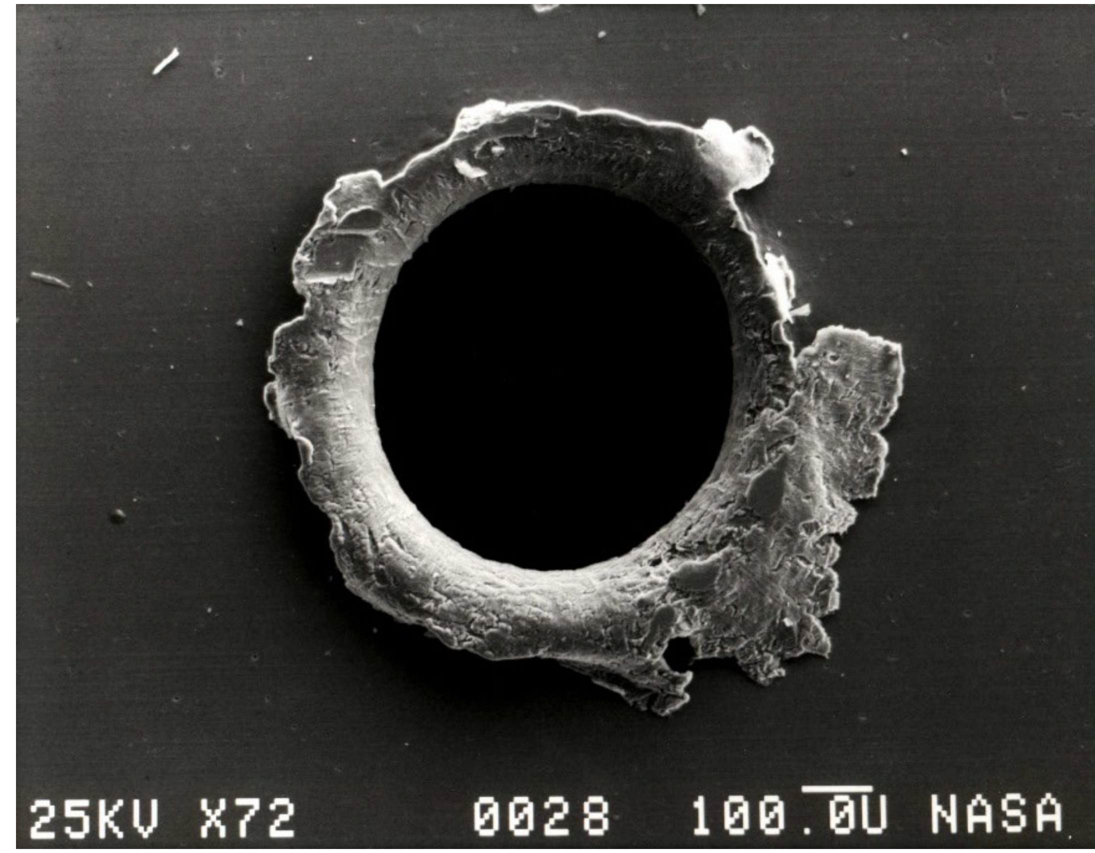
- **Specific**
  - multiple impacts

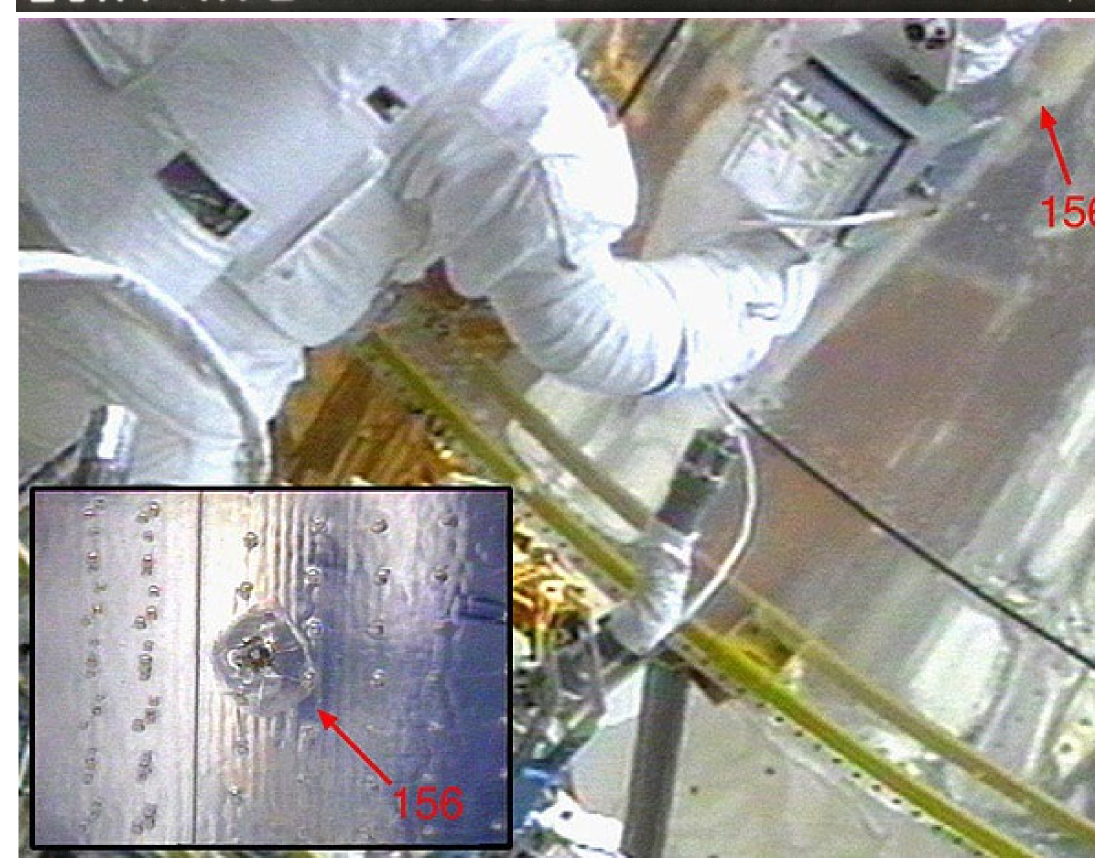
    micro-meteoroids, particles, …
  - damage severity:

    non-destructive → major damage
  - distributed in time:

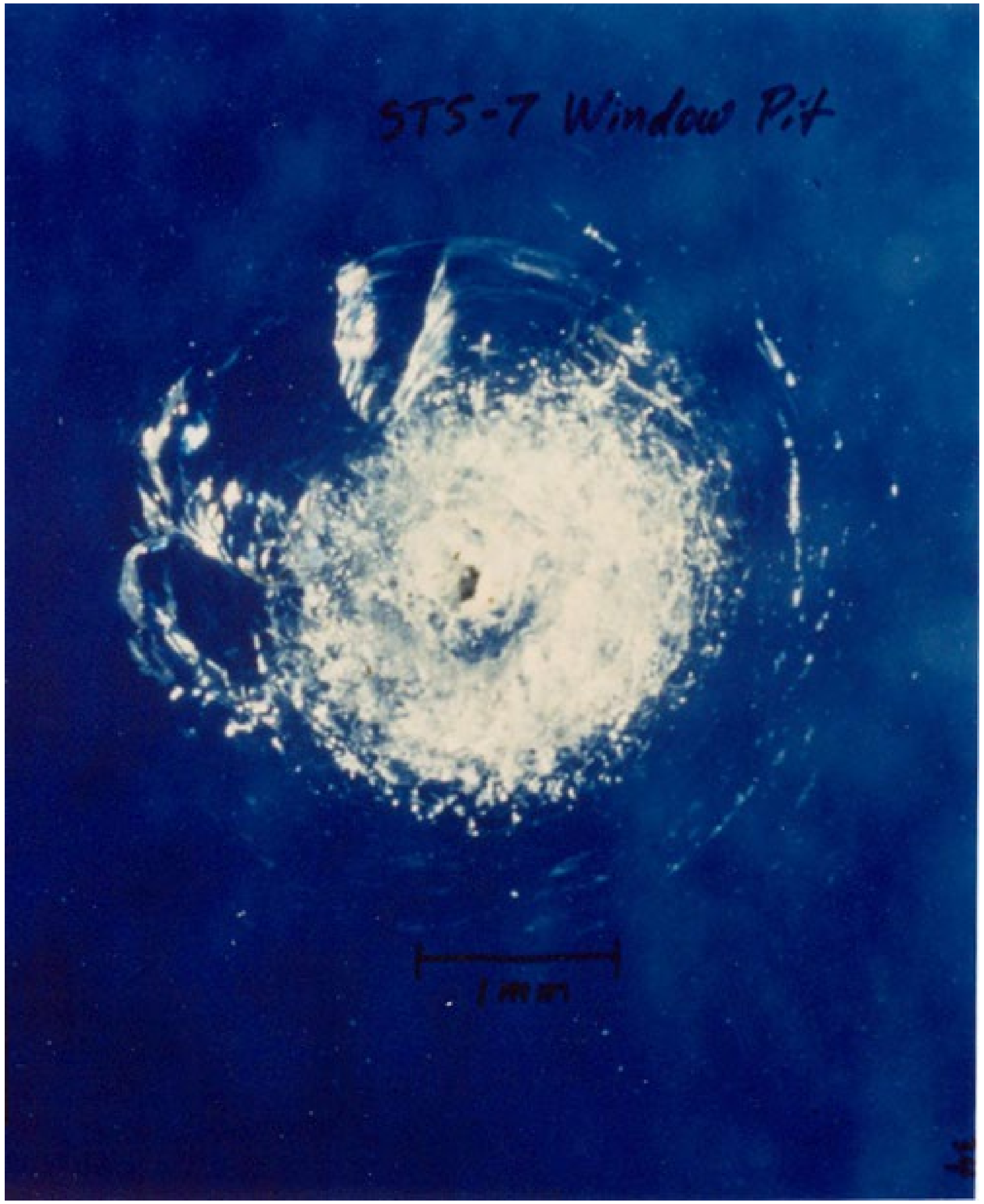
    nanoseconds → slow transients
  - electronic failures

    connectivity losses, errors, …





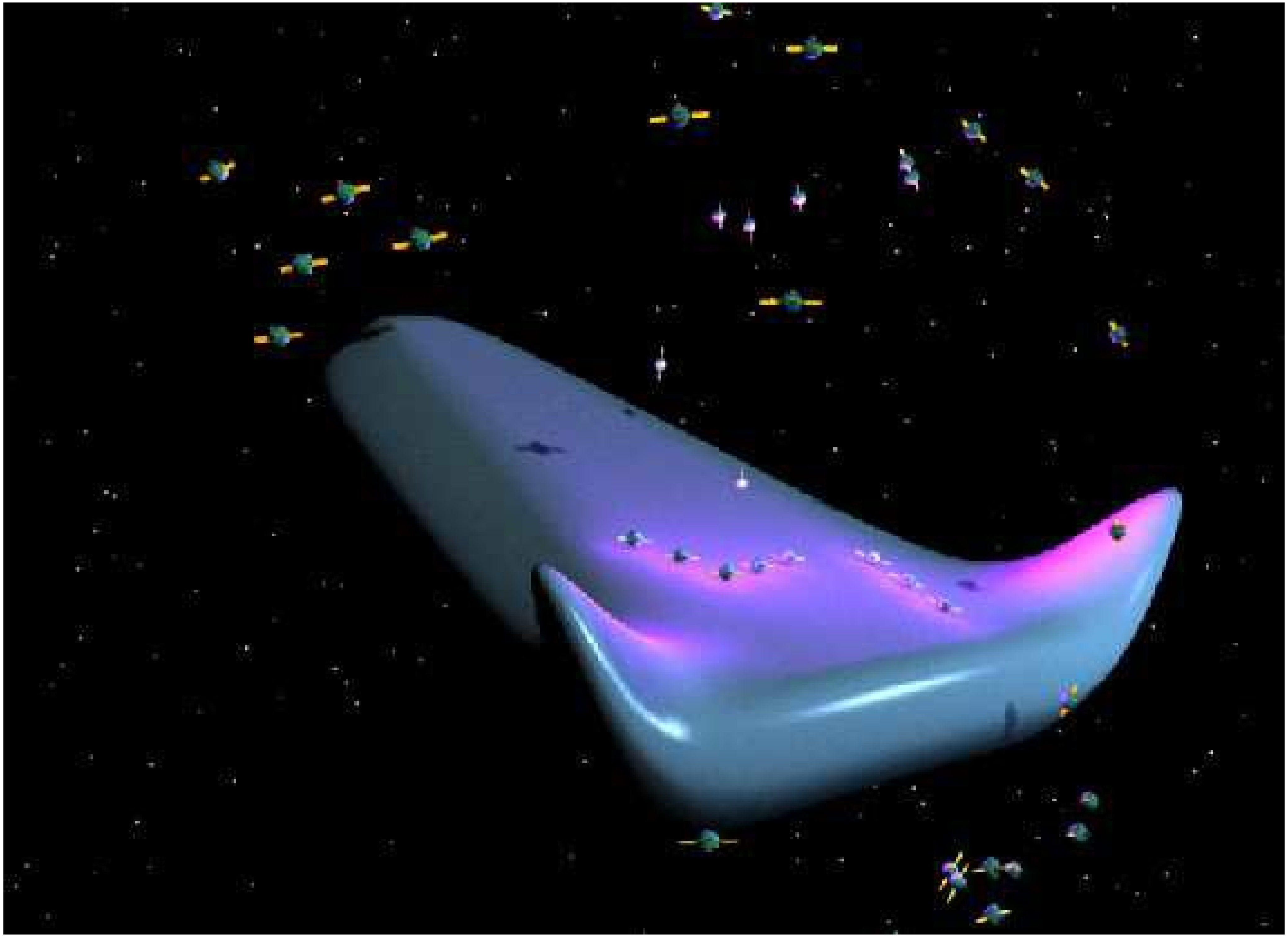
State of the practice (circa 2002)



Window pit from orbital debris
on STS-007

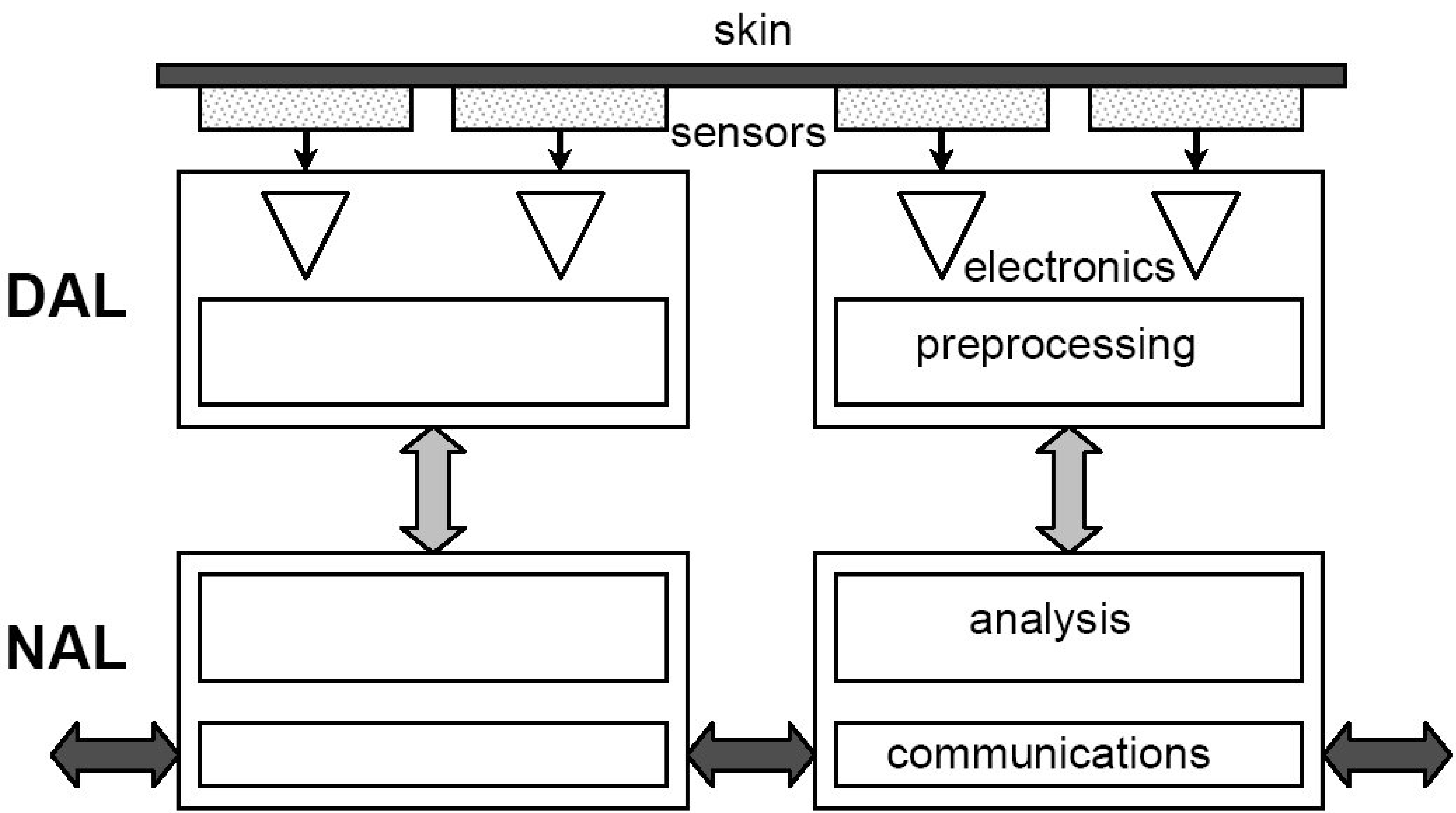
Space Shuttle window being inspected for
orbital debris impacts

NASA early vision

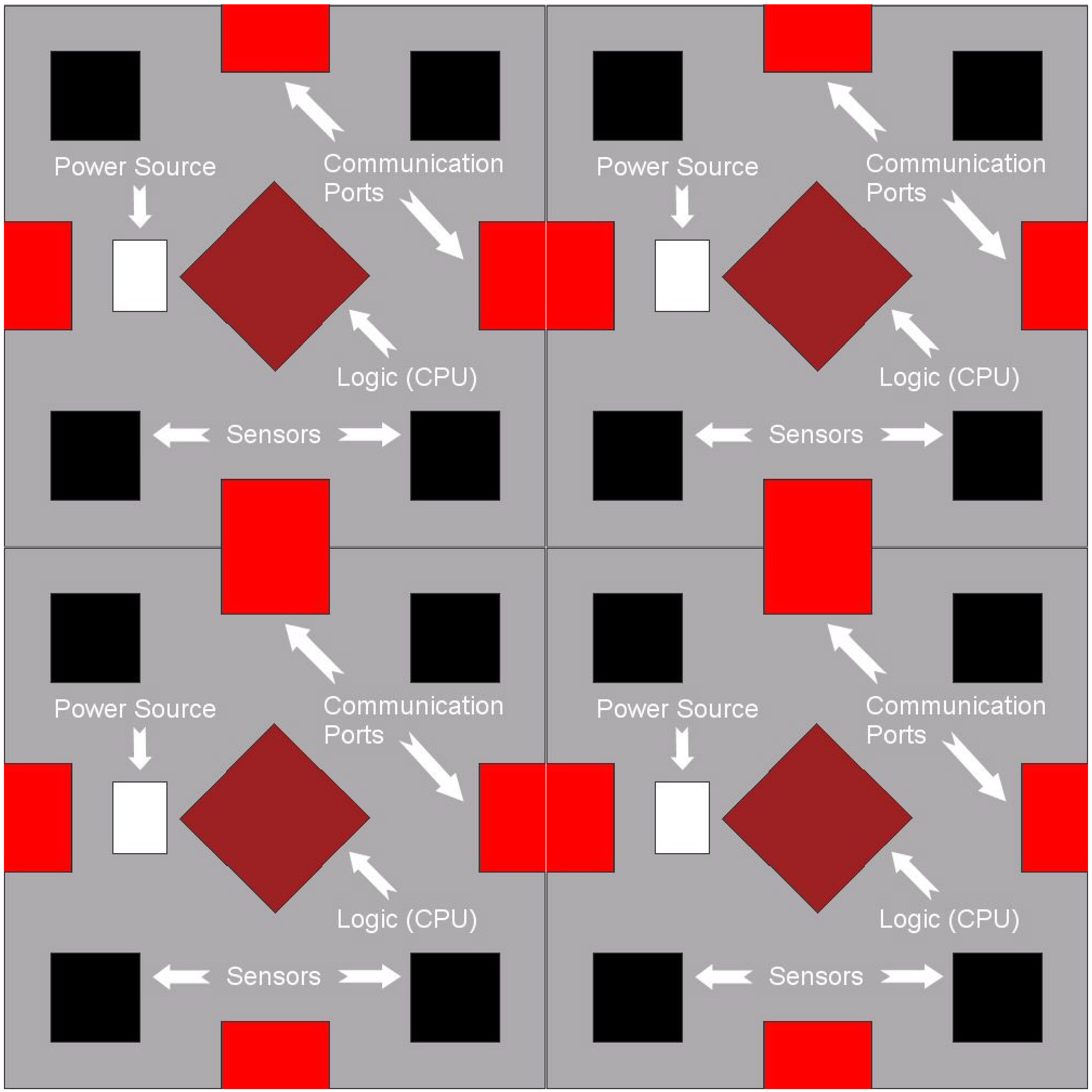


Swarm robots in a loop with embedded SHMM networks

Conceptual layer structure



High-level design: sensor & communication network



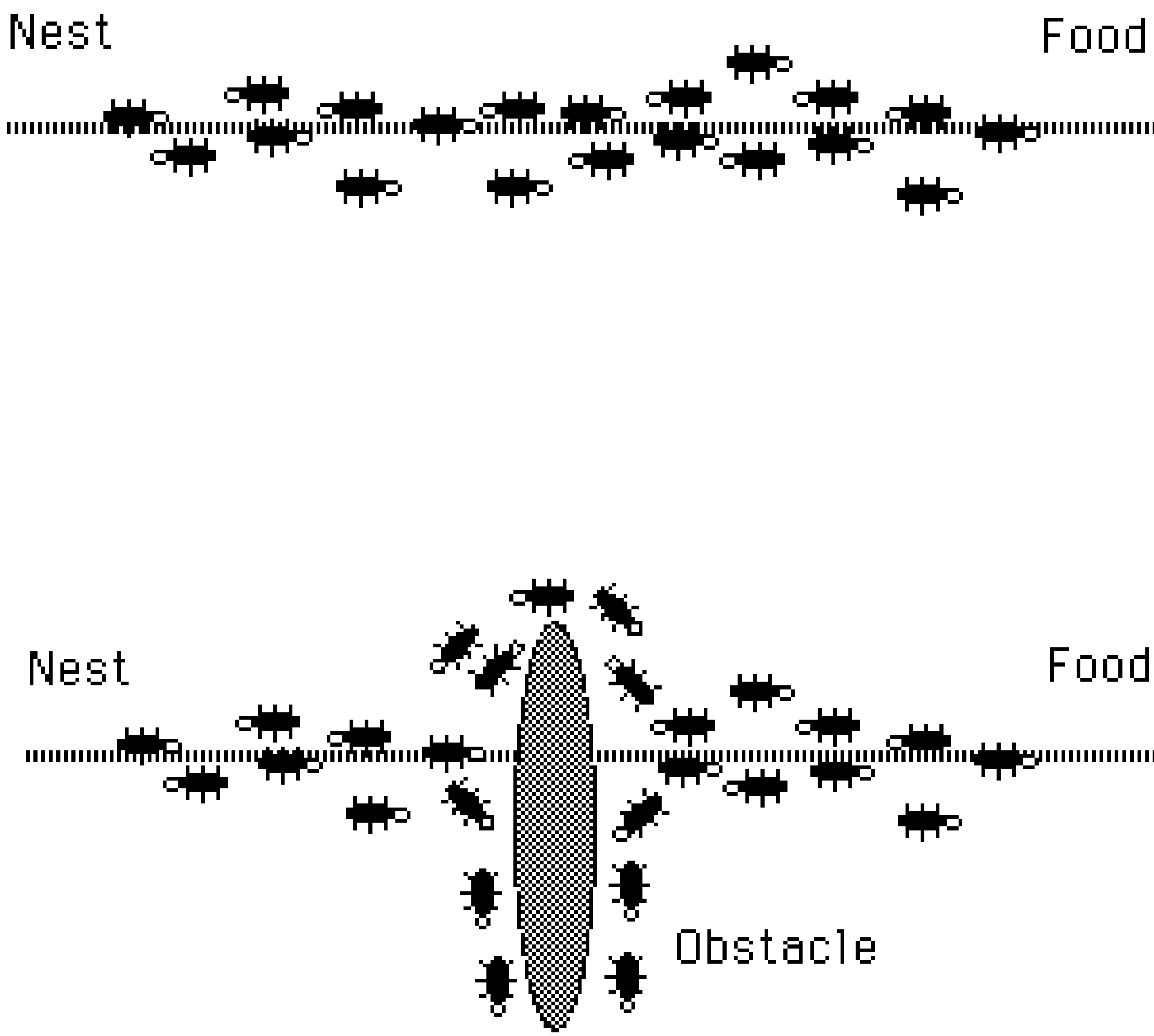
- distributed processing units
- locally connected
- embedded sensors
- variable sensor density
- easily replaceable cells

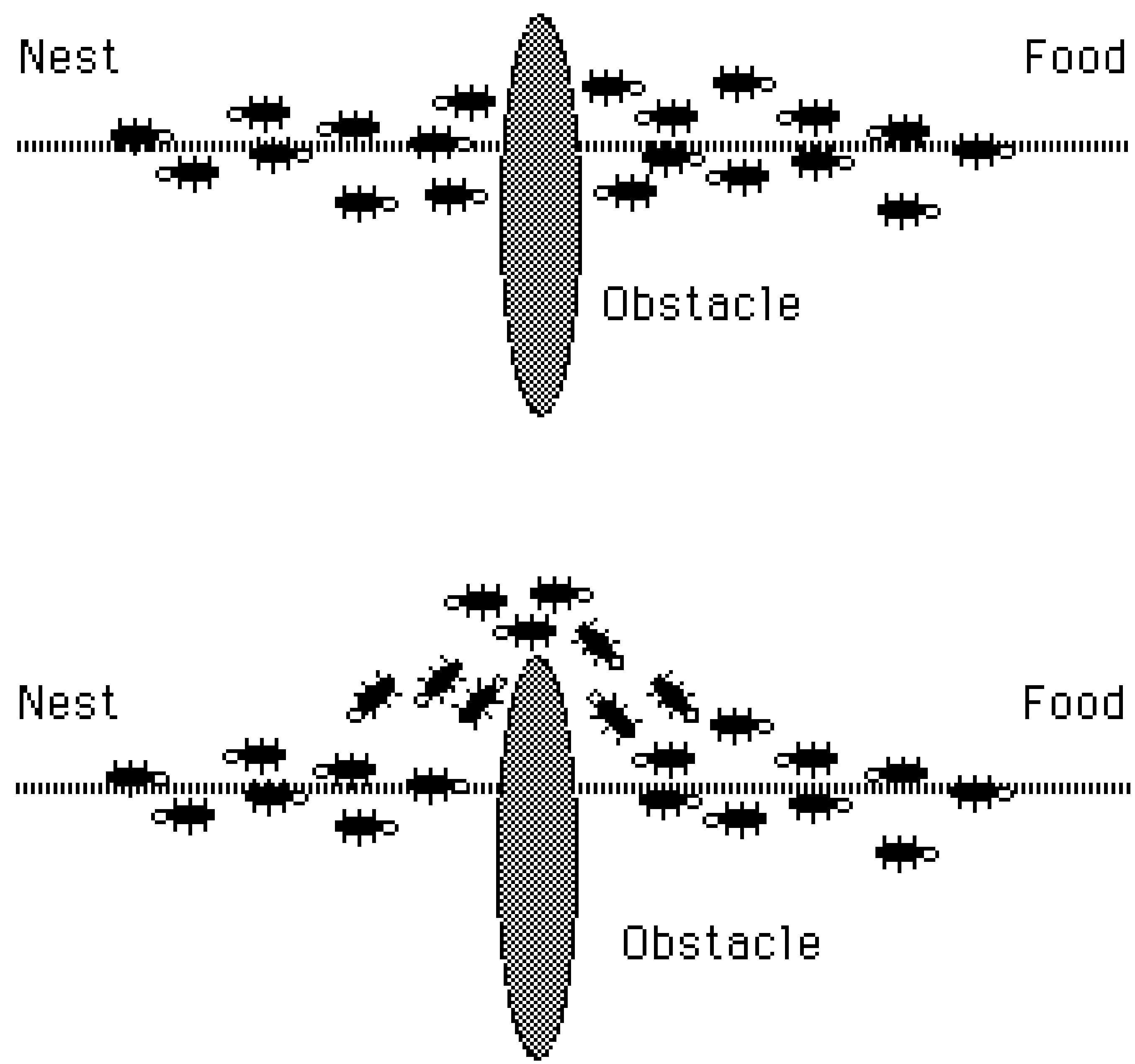
Break

?

Distributed AI

Swarm intelligence: Ant Colony Optimisation





https://slideplayer.com/slide/12455505/

Self-organising impact networks: connecting impact points

Self-organising impact networks: stable path

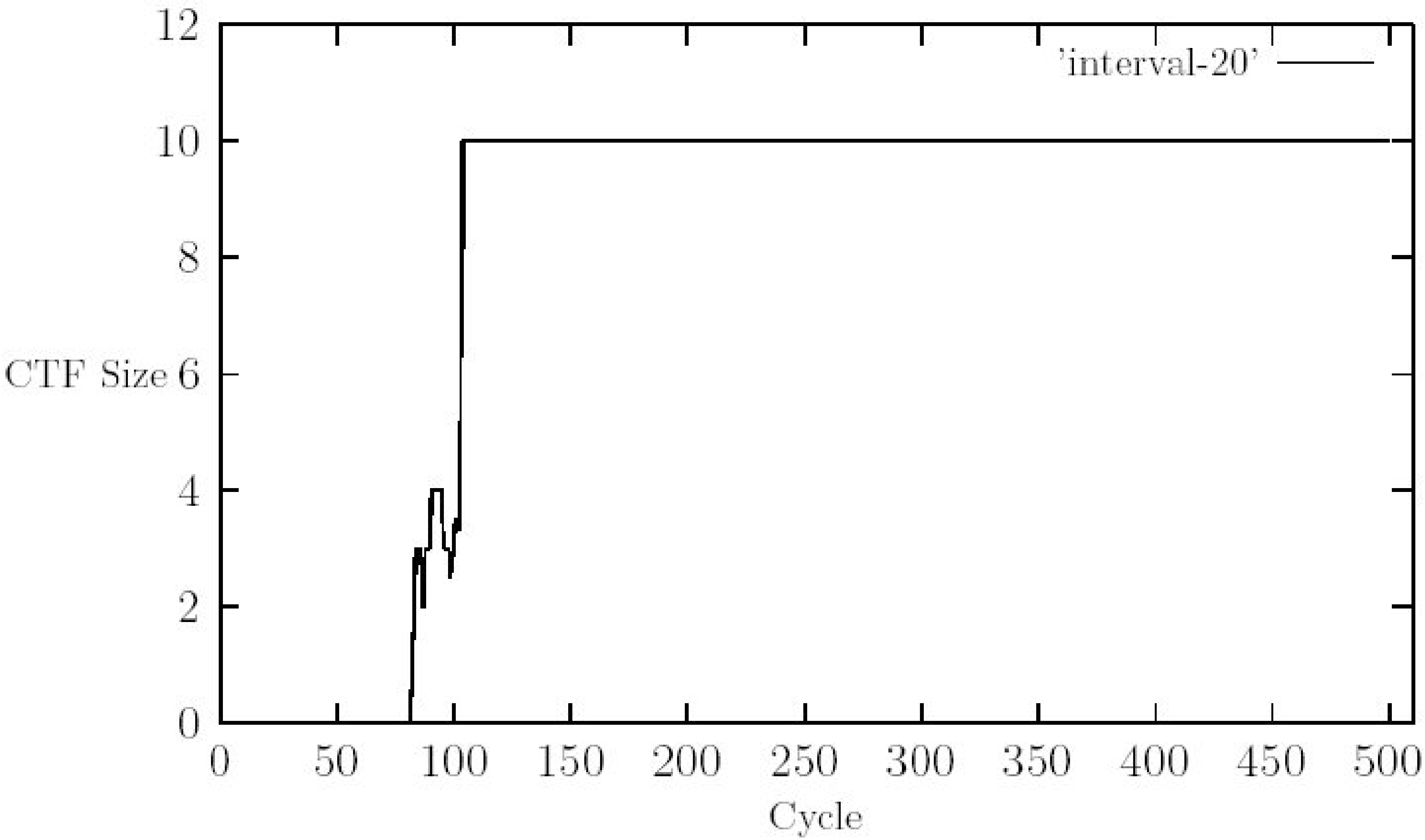


Fig. 7. Average size $H(T)$. Robust trails with $T = 20$.

Self-organising impact networks: unstable path

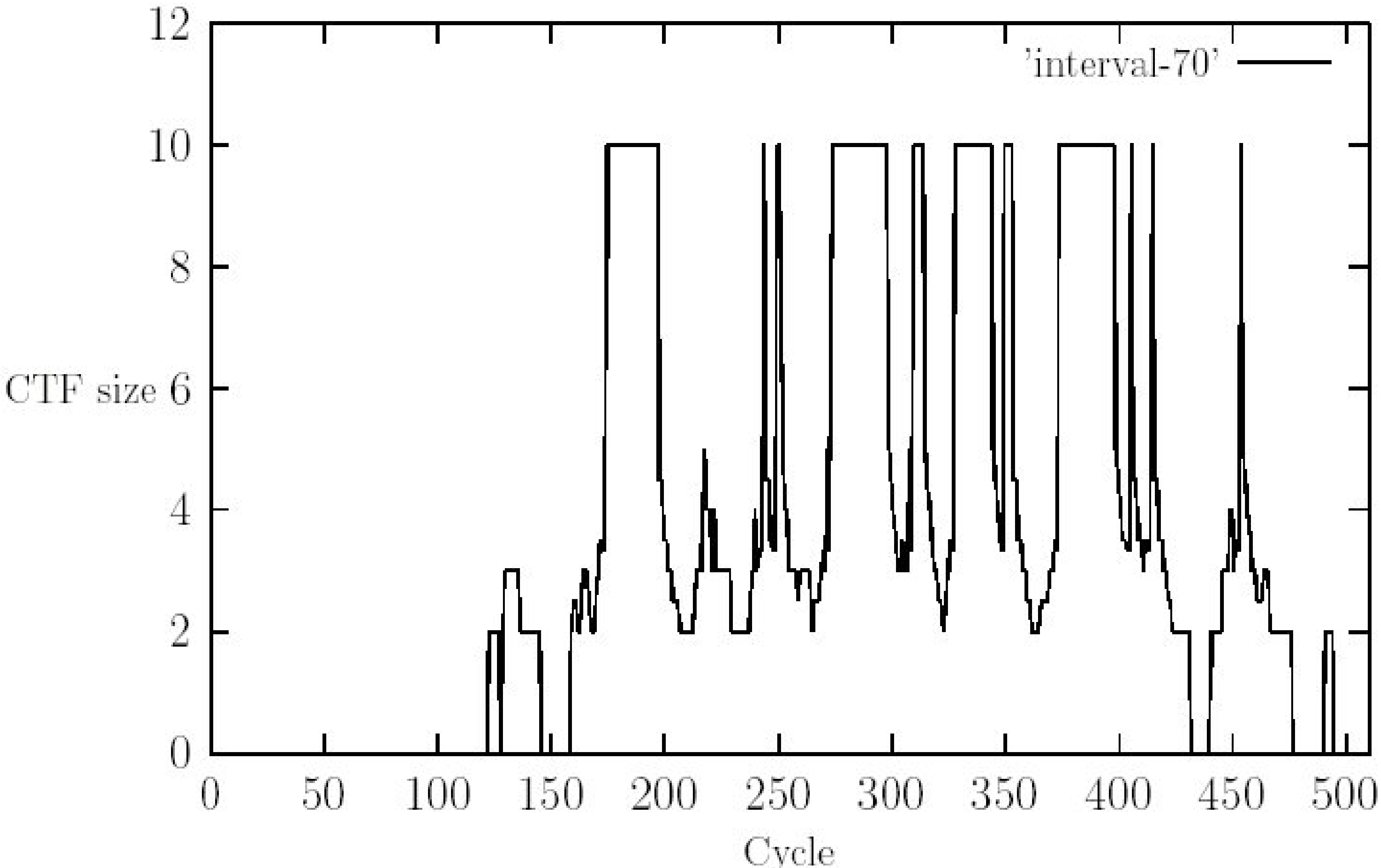
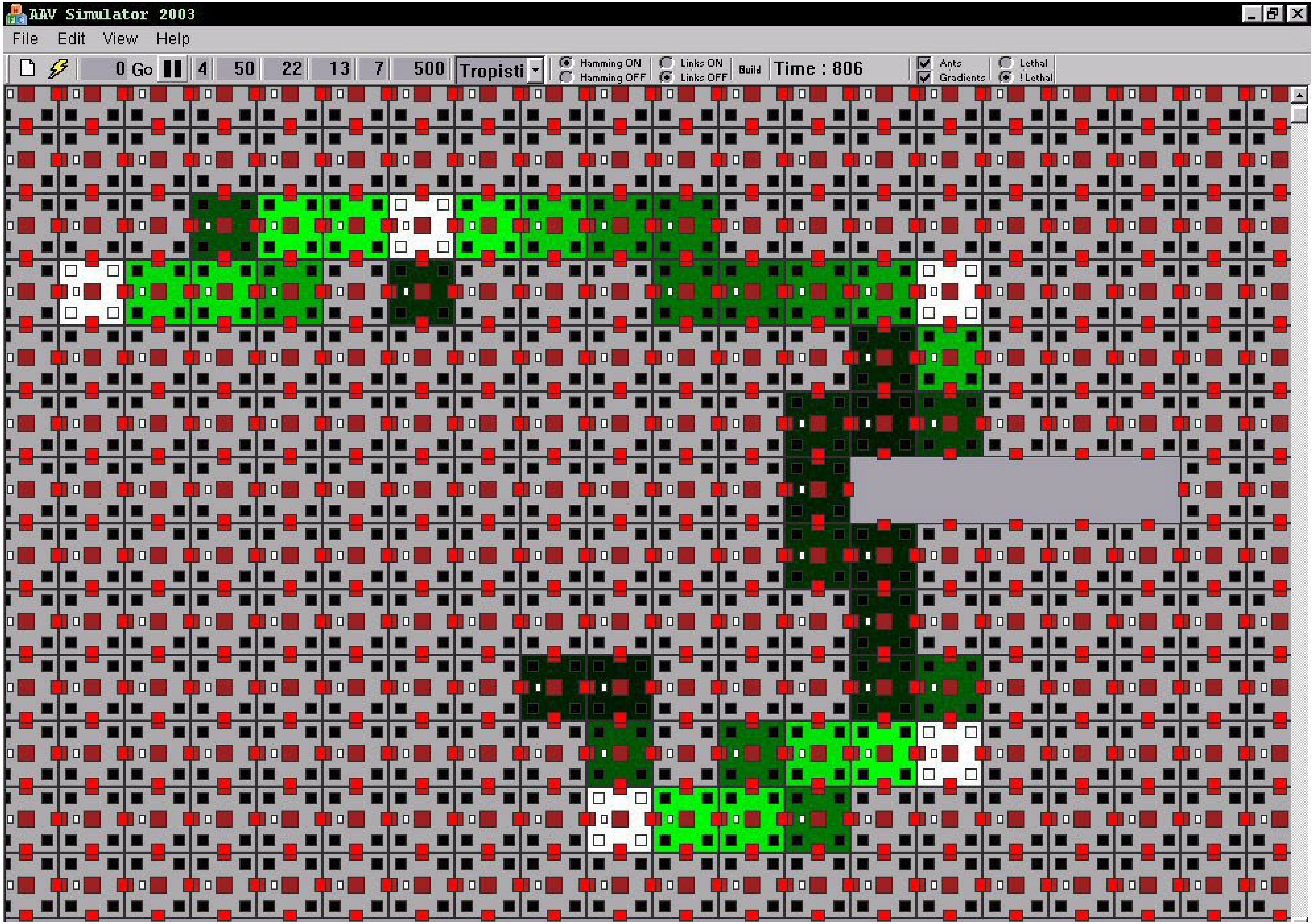


Fig. 8. Average size $H(T)$. Unstable trails with $T = 70$.

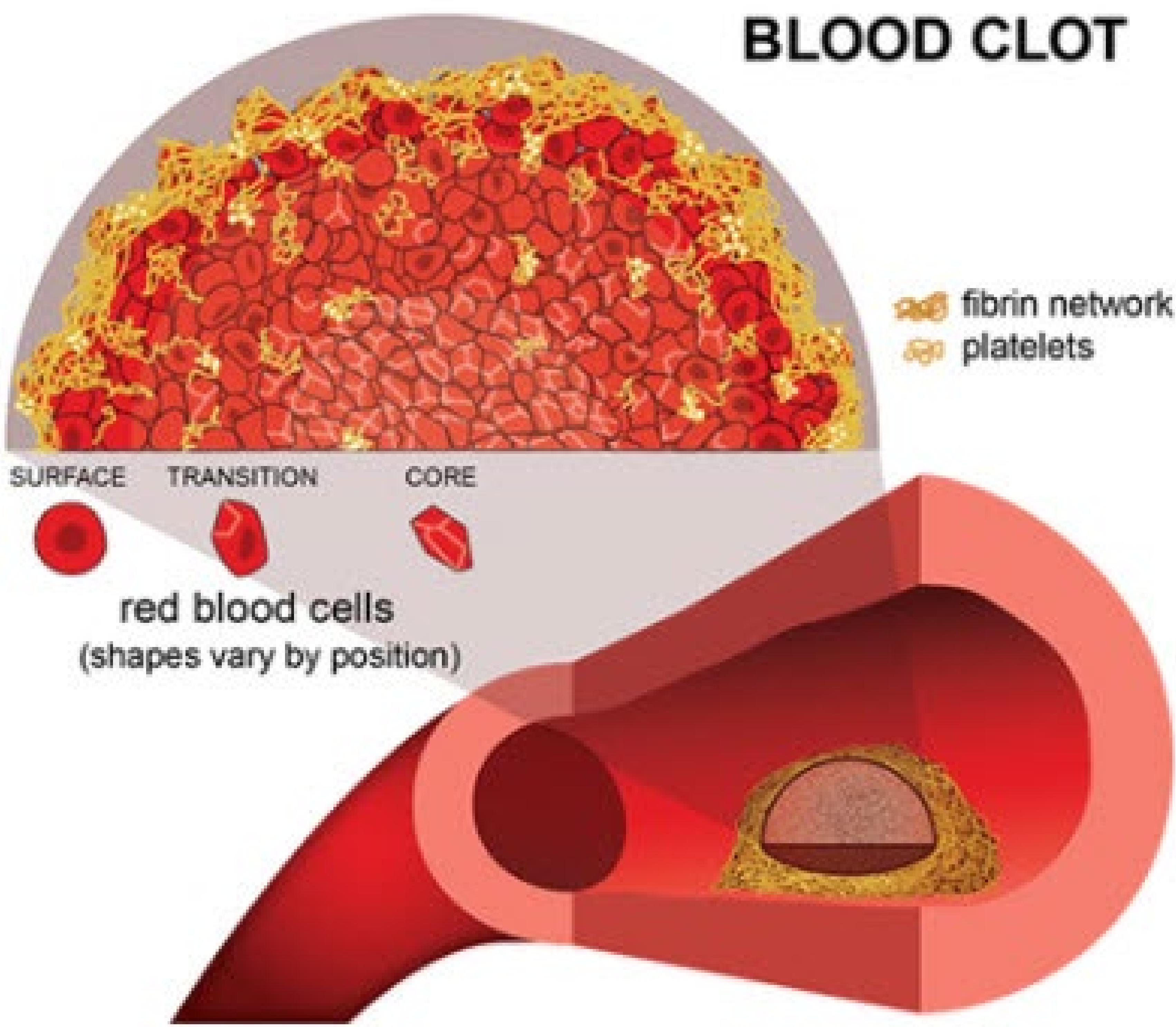
Self-organising impact networks: connecting impact points



Questions

?

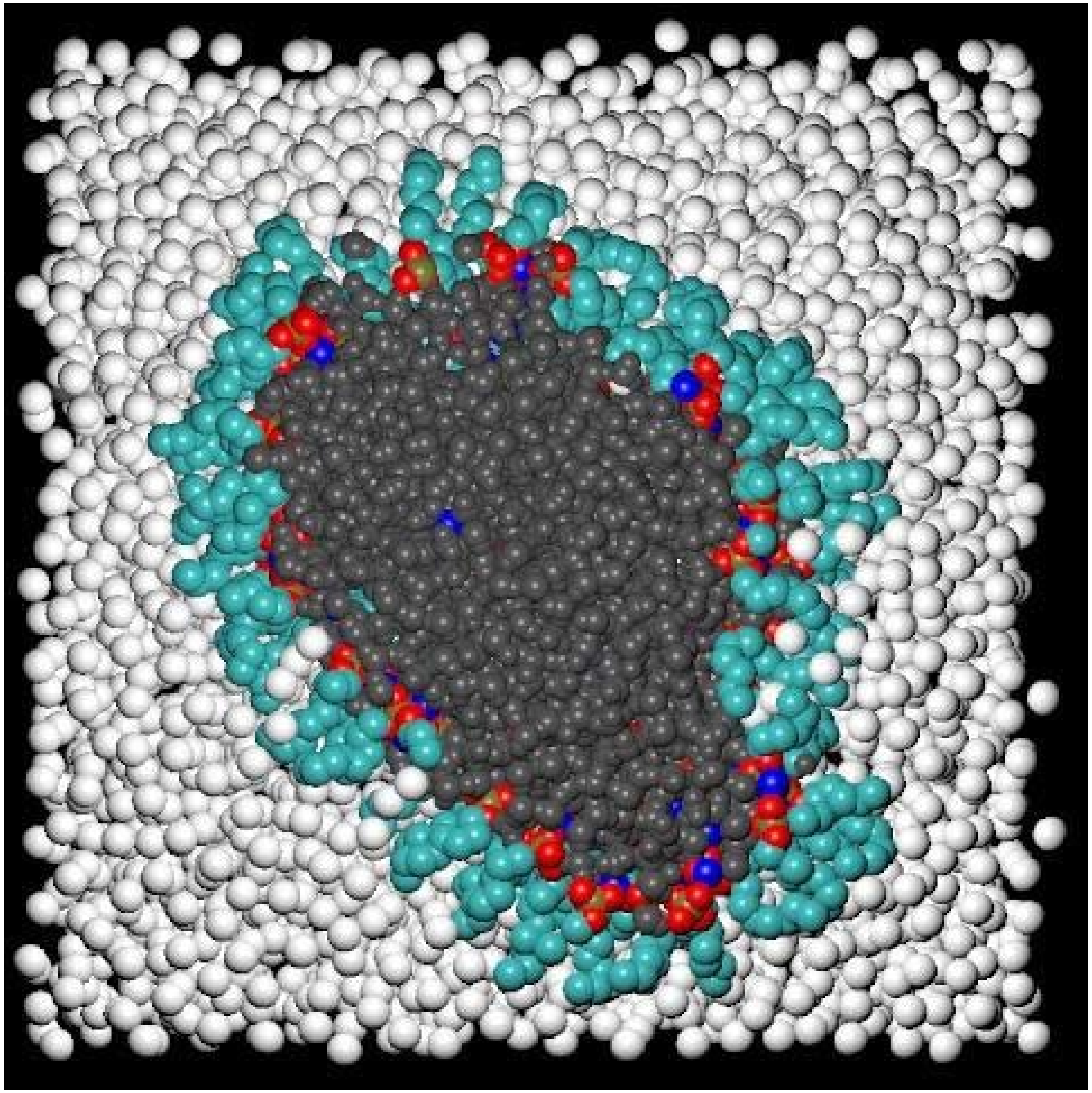
Damage containment: wound healing (biological analogy)



https://atlanticcoastrehab.blog/scientists-optically-clear-blood-clots-study-structure-pathogenesis

Damage containment: reverse micelles (biological analogy)

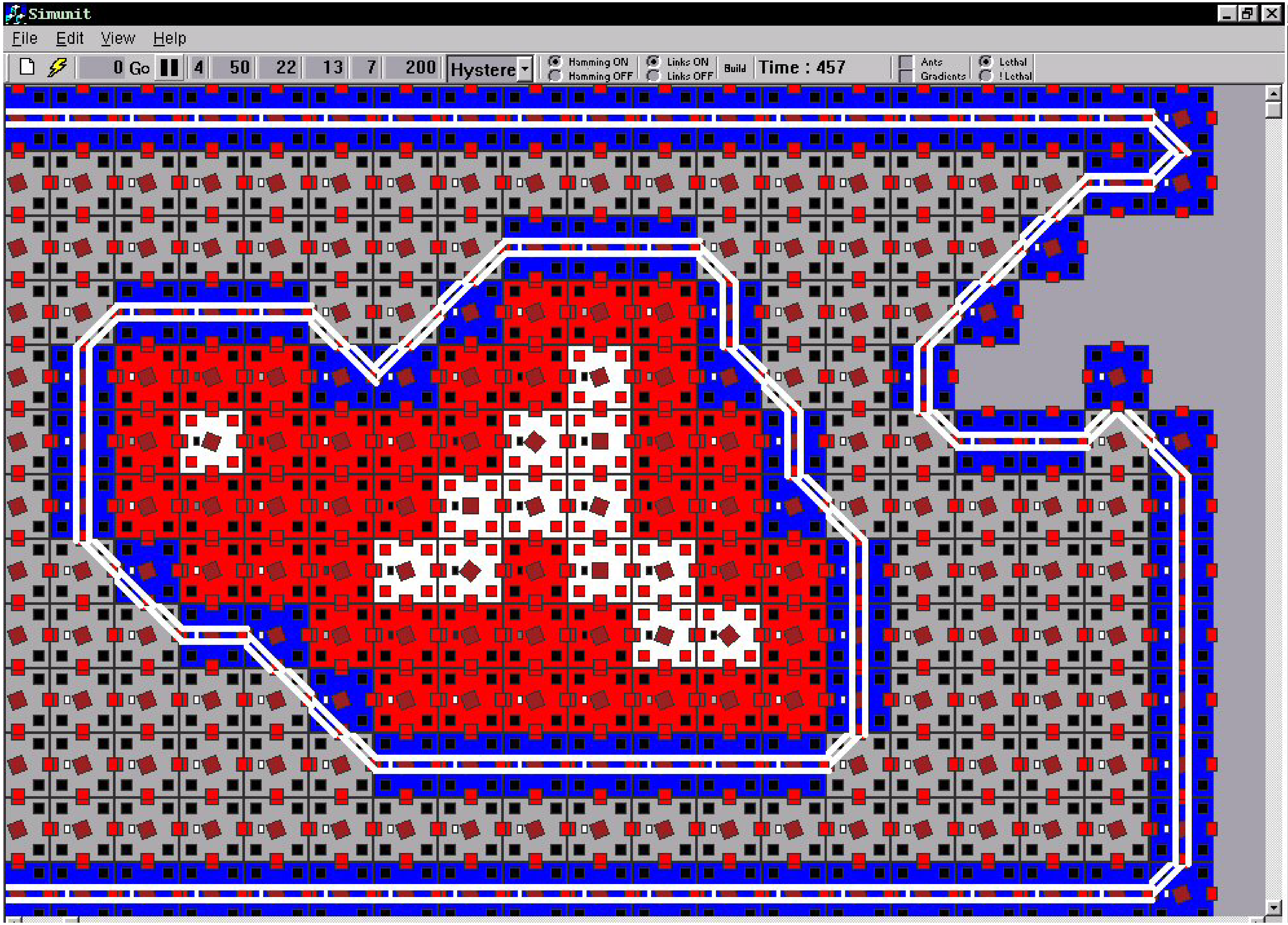
stable assemblies of molecules organised around a core of water (water droplets)



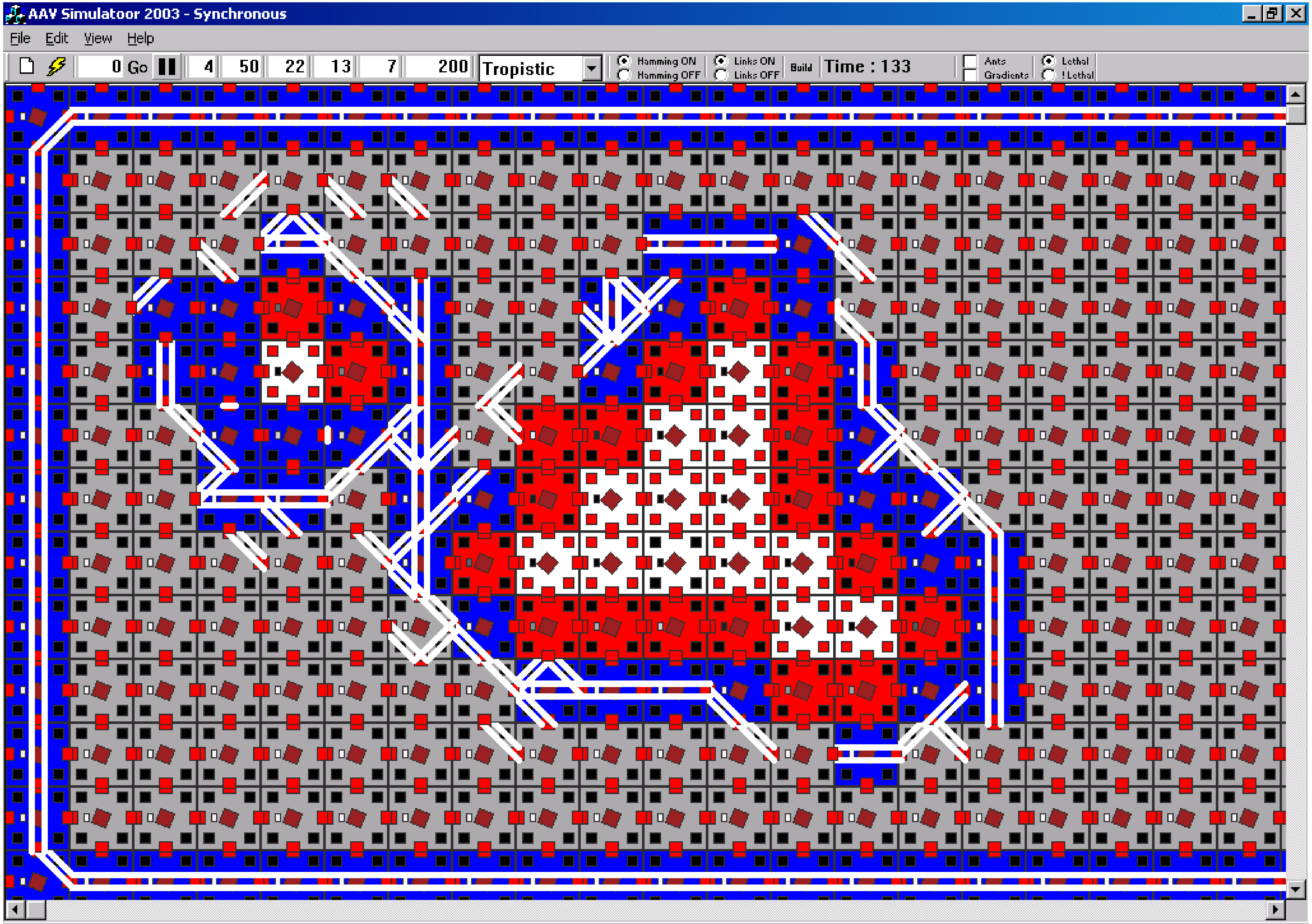
Berkowitz Group: Molecular Dynamics, The University of North Carolina at Chapel Hill

Self-organising impact boundaries: damage containment

Self-organising stable impact boundaries



Self-organising unstable impact boundaries



Self-organising impact boundaries: cell-to-cell connection rules

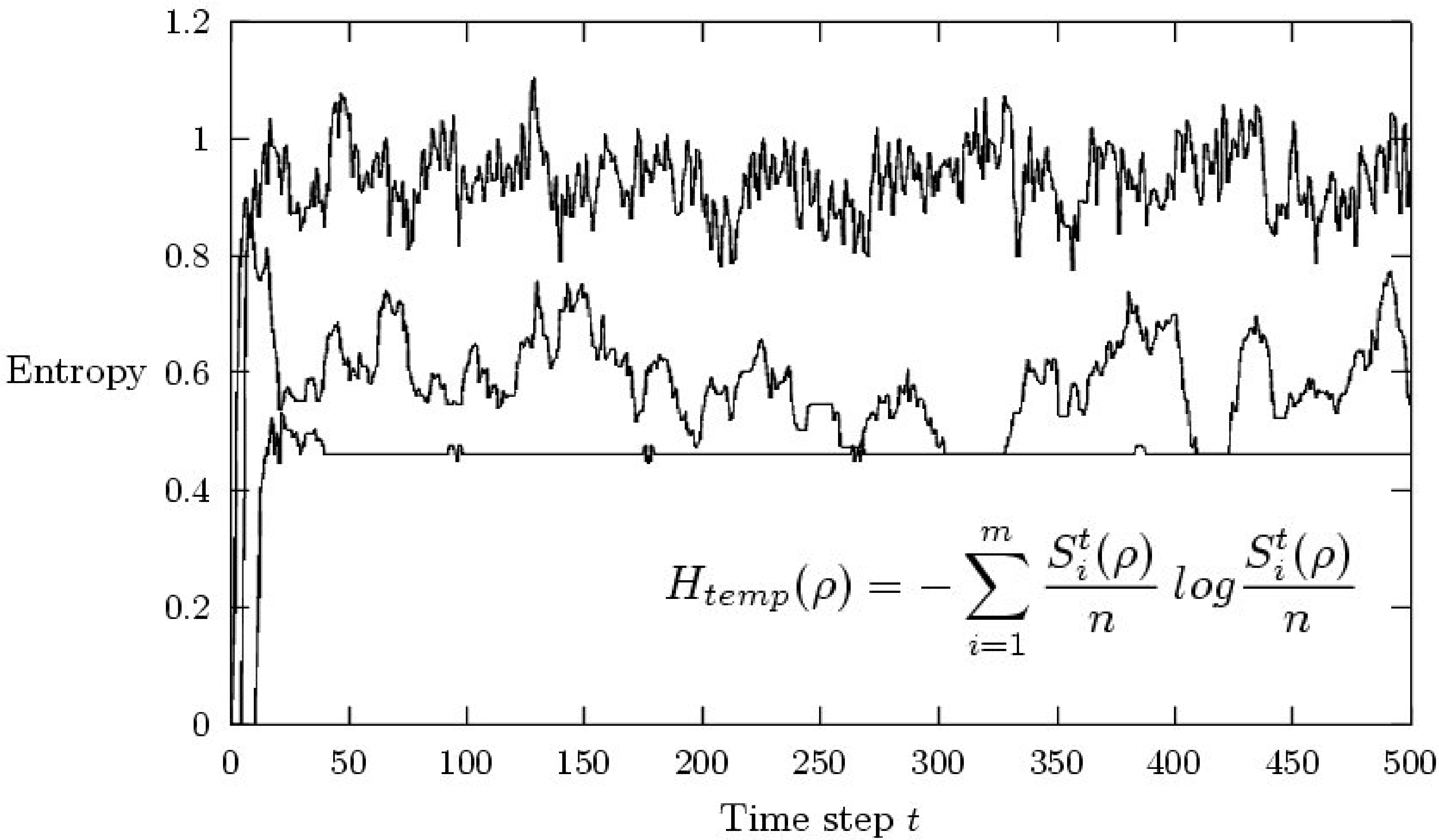


Fig. 3. Temporal entropy $H_{temp}(\rho)$ for $\rho = 0$ (top), $\rho = 4$ (middle), and $\rho = 10$ (bottom).

Self-organising impact boundaries: standard deviation of rules

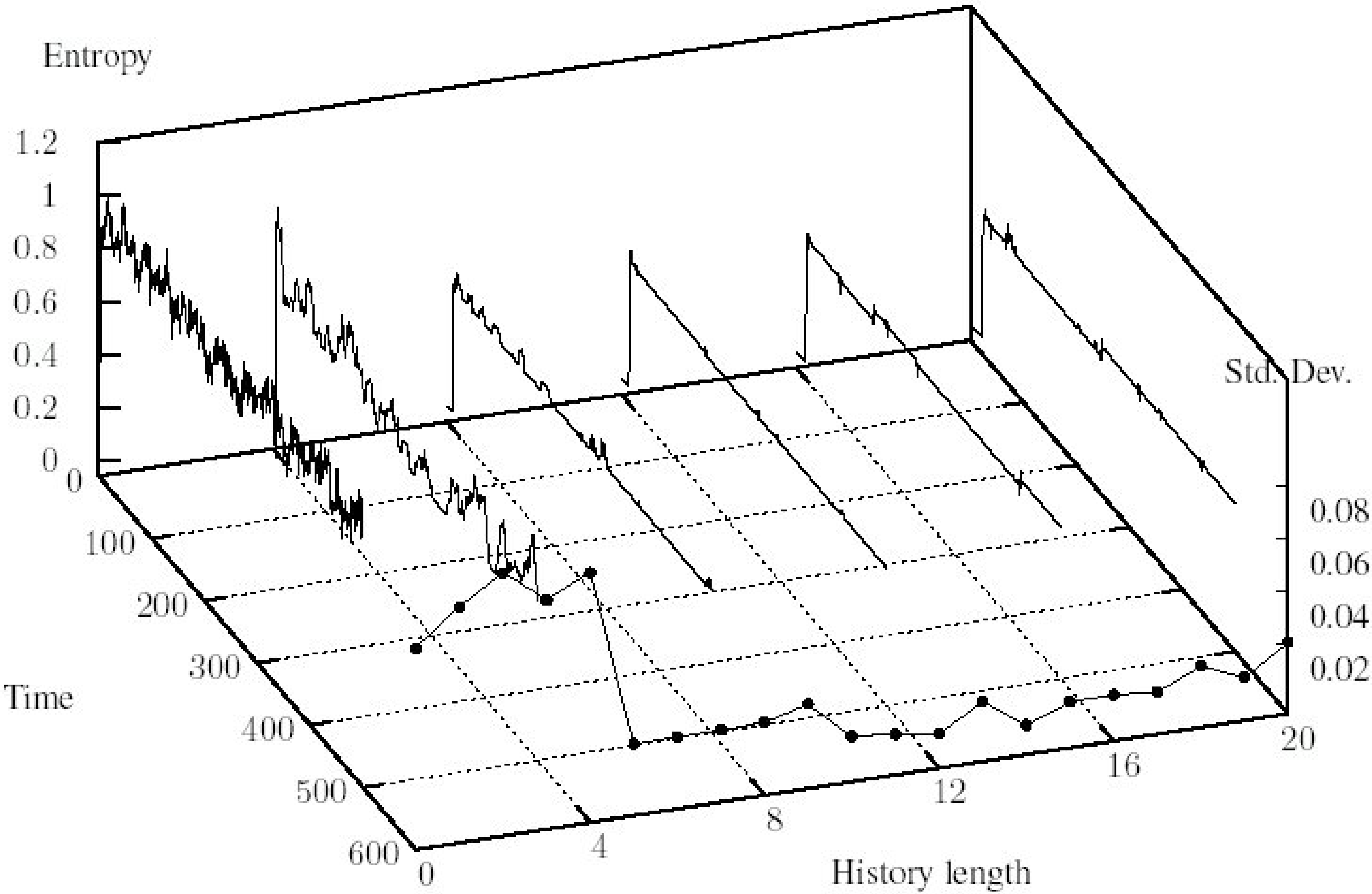


Figure 6: Background (left z-axis): temporal entropy $H_{temp}(\mu)$ for $\mu = 0$ (left), $\mu = 4$ (strongly fluctuating), and $\mu = 8$, $\mu = 12$, $\mu = 16$, $\mu = 20$ (right). Foreground (right z-axis): standard deviation $\sigma_{temp}(\mu)$ of temporal entropy $H_{temp}(\mu)$.

Self-organising impact boundaries: length (size) of boundary

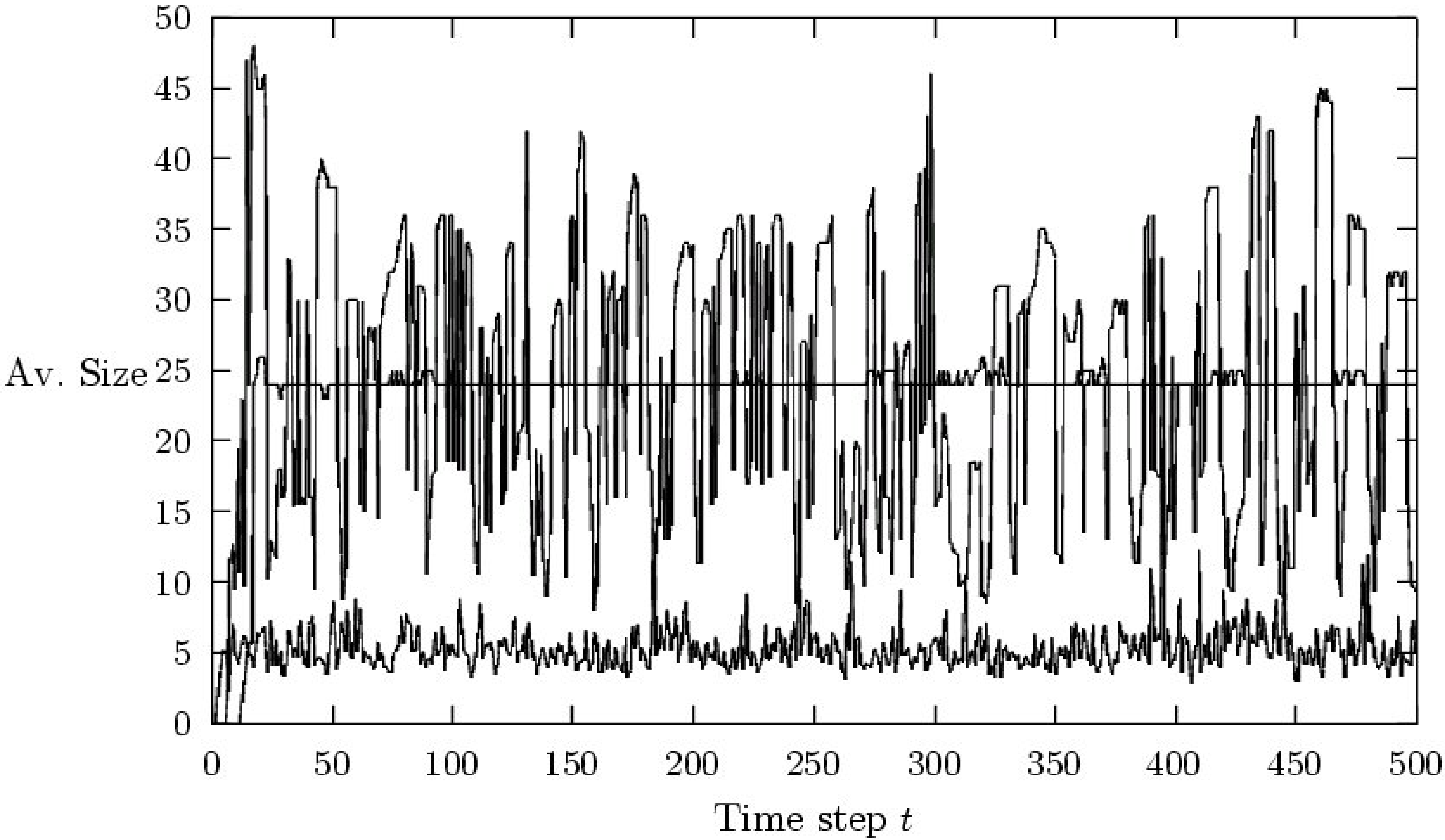


Fig. 5. Average size $H_{sp}(\rho)$ of CBF's, for $\rho = 0$ (bottom), $\rho = 4$ (strongly fluctuating), and $\rho = 10$ (almost straight line).

Self-organising impact boundaries: length (size) of boundary

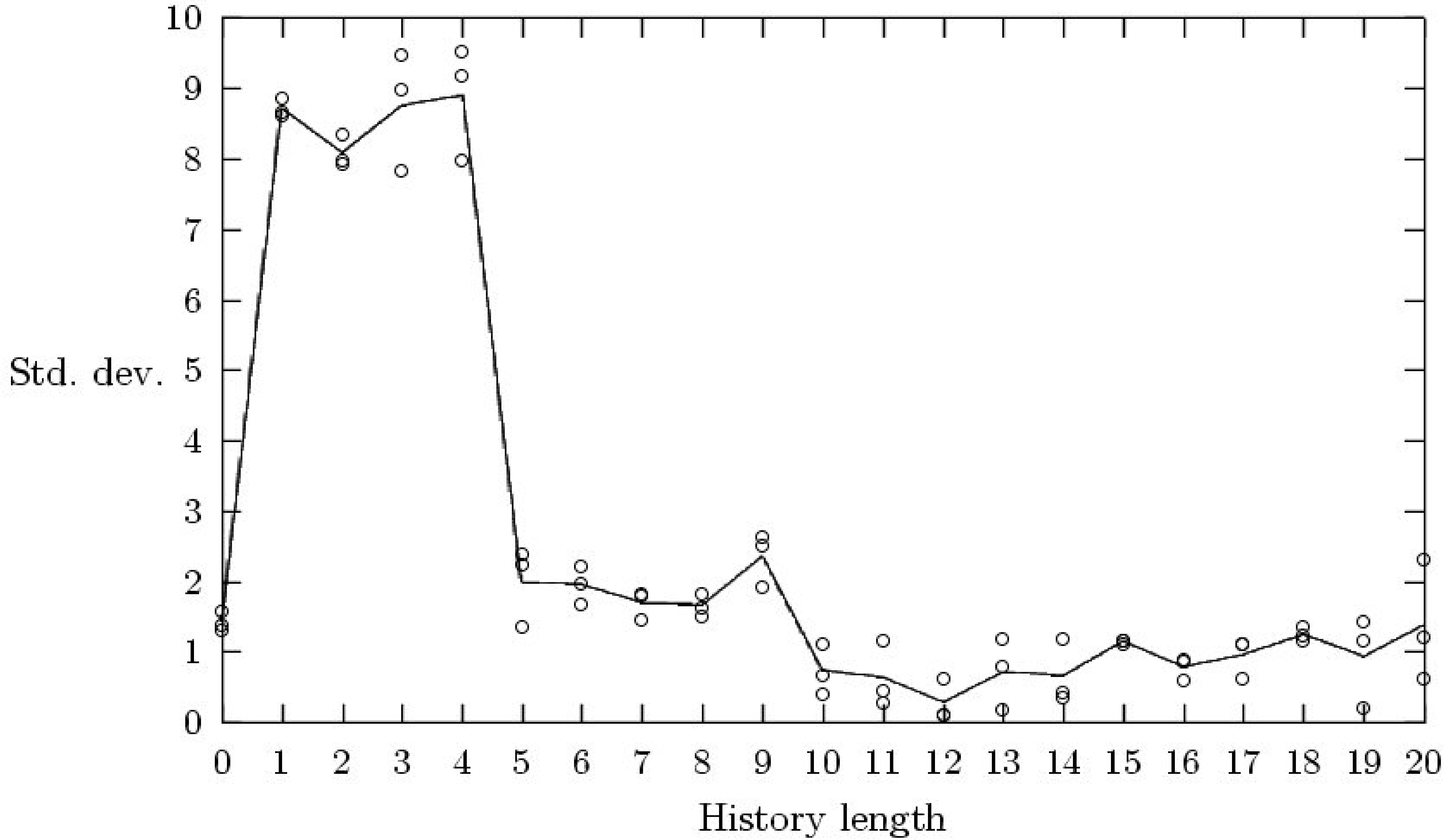


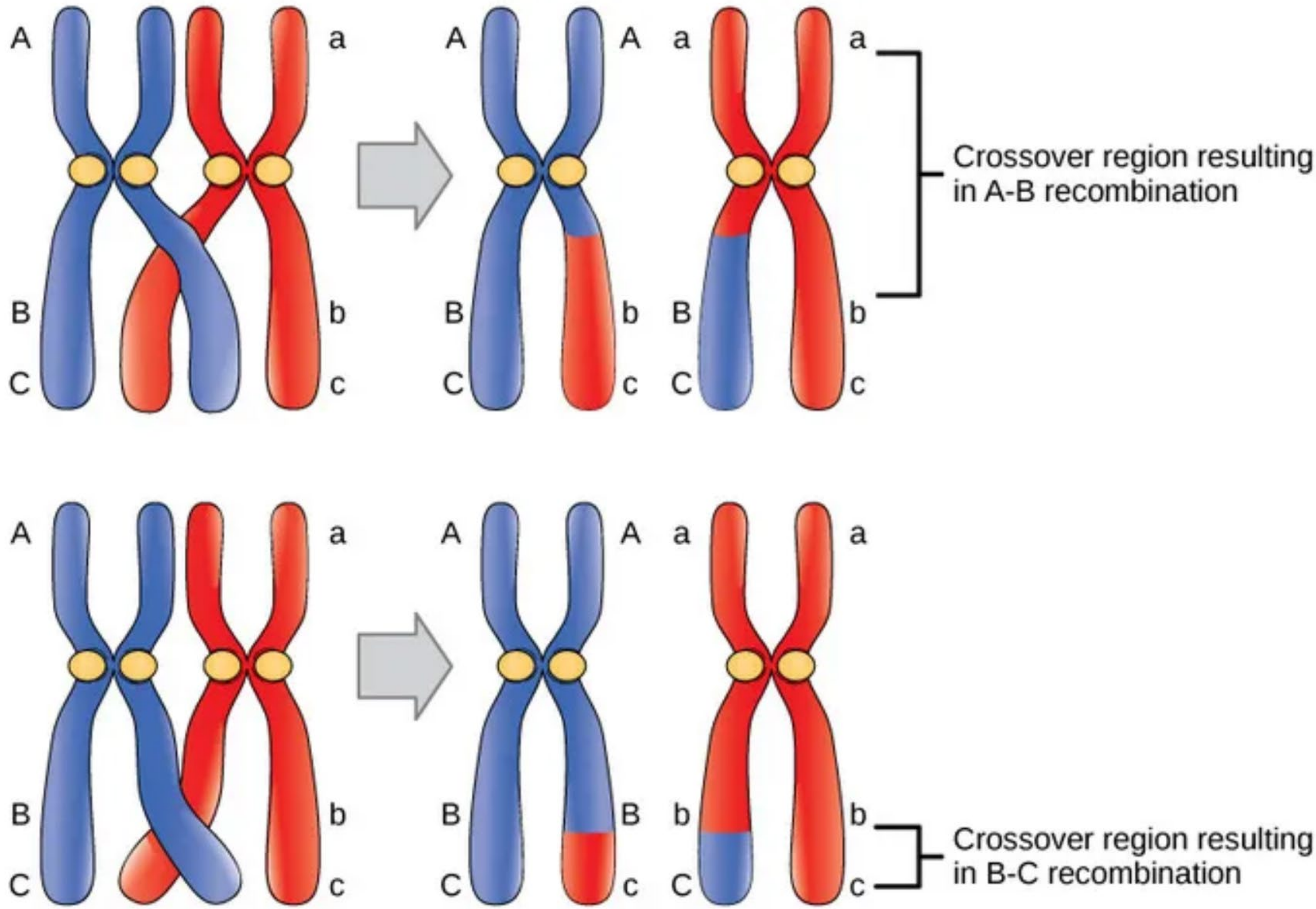
Fig. 6. Standard deviation $\sigma_{sp}(\rho)$ of average size $H_{sp}(\rho)$.

Questions

?

Distributed AI

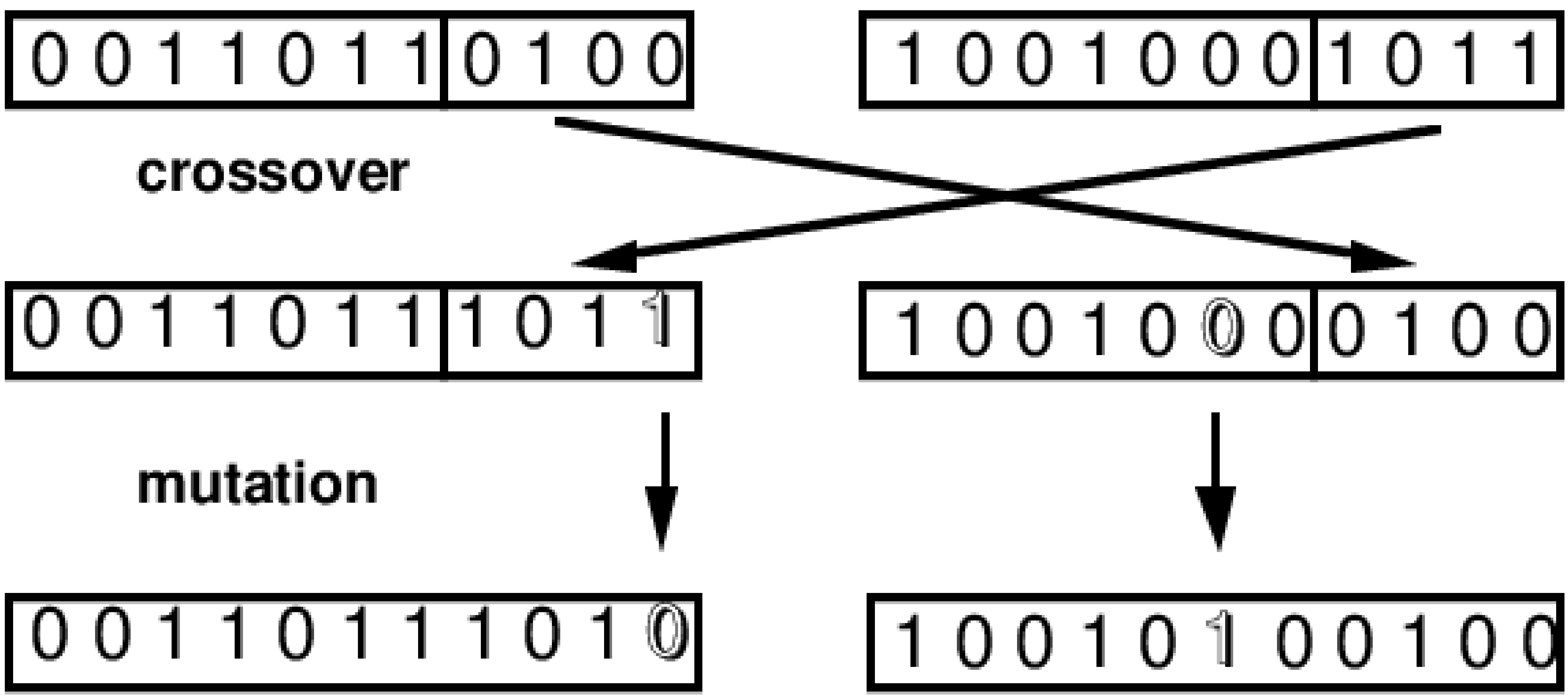
Genetic algorithms



https://jackwestin.com/resources/mcat-content/analytic-methods/gene-mapping-crossover-frequencies

Distributed AI

Genetic algorithms

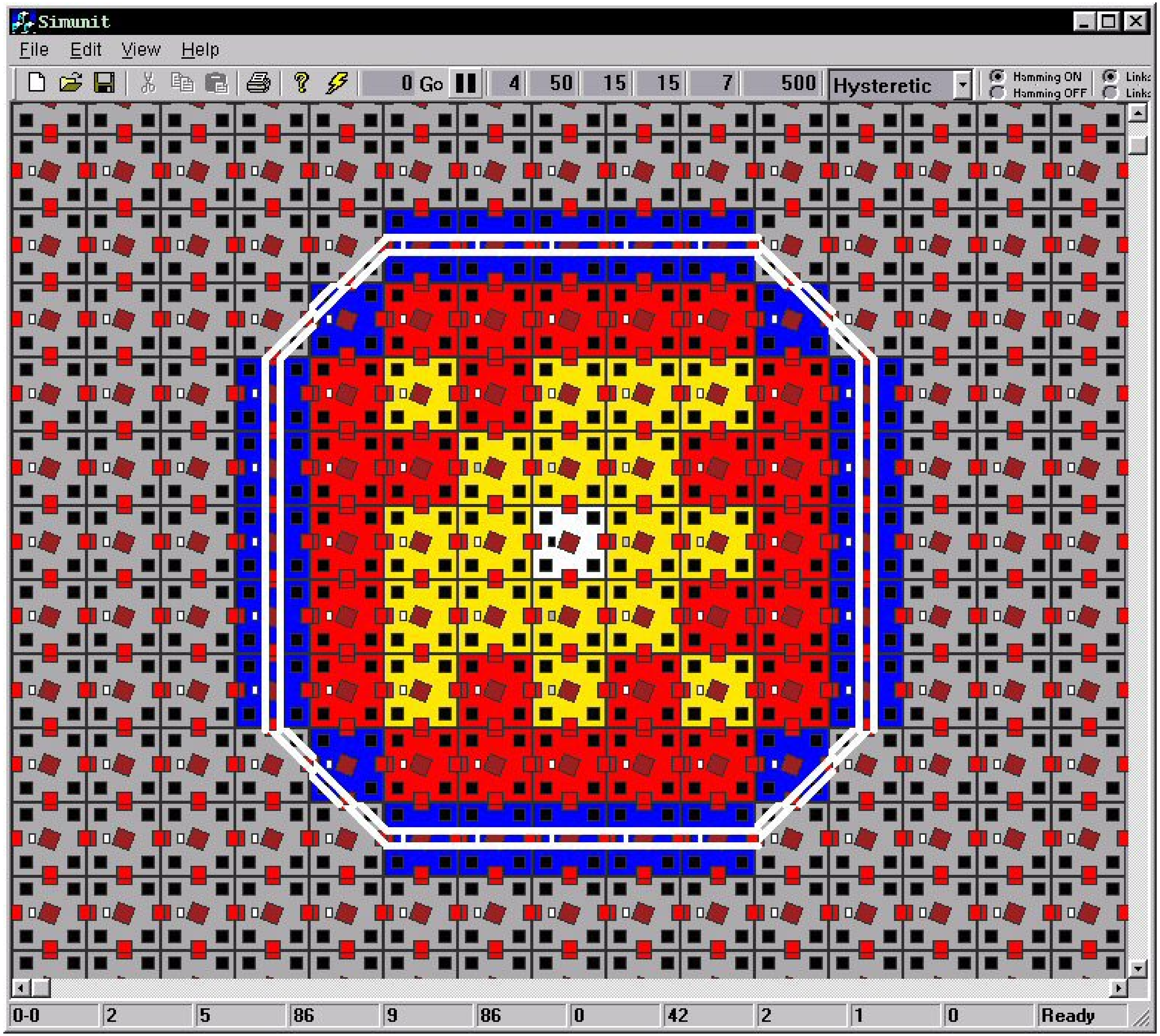


https://www.researchgate.net/publication/2263212_Graph_Coloring_Algorithms_for_Assignment_Problems_in_Radio_Networks

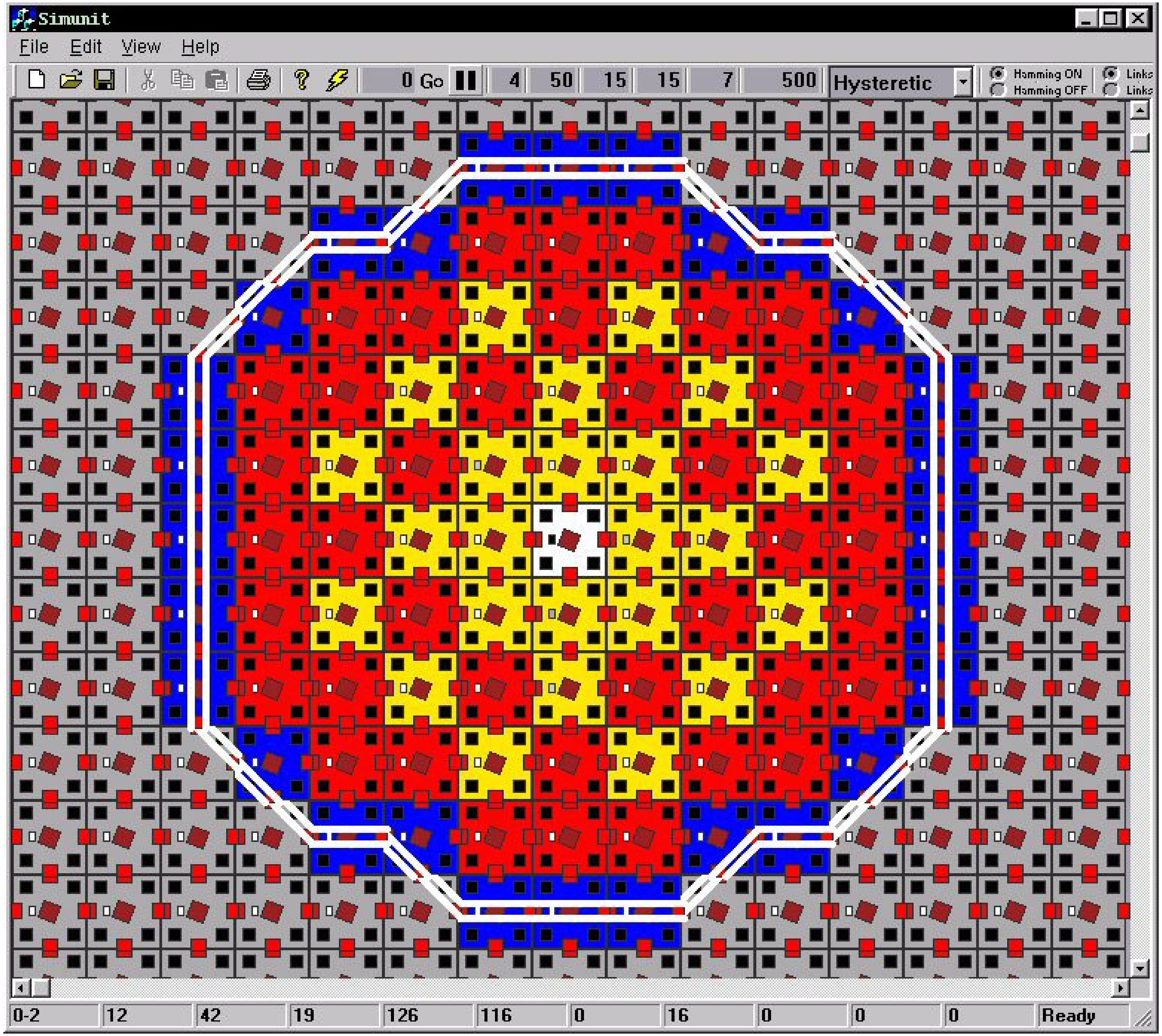
Self-organising impact boundaries: objective (fitness) function

$$f(\beta) = \begin{cases} M & \text{if} \quad \overline{H_{sp}} \leq 16; \\ \frac{1}{2}\left(4.0\ \sigma_{sp}^2 + 10^5\ \sigma_{temp}^2\right) + & \\ \qquad + \rho + \tau + \pi + \beta\ \overline{H_{sp}} & \text{if} \quad \overline{H_{sp}} > 16 \end{cases}$$

“Octagons”: small boundaries & long memory



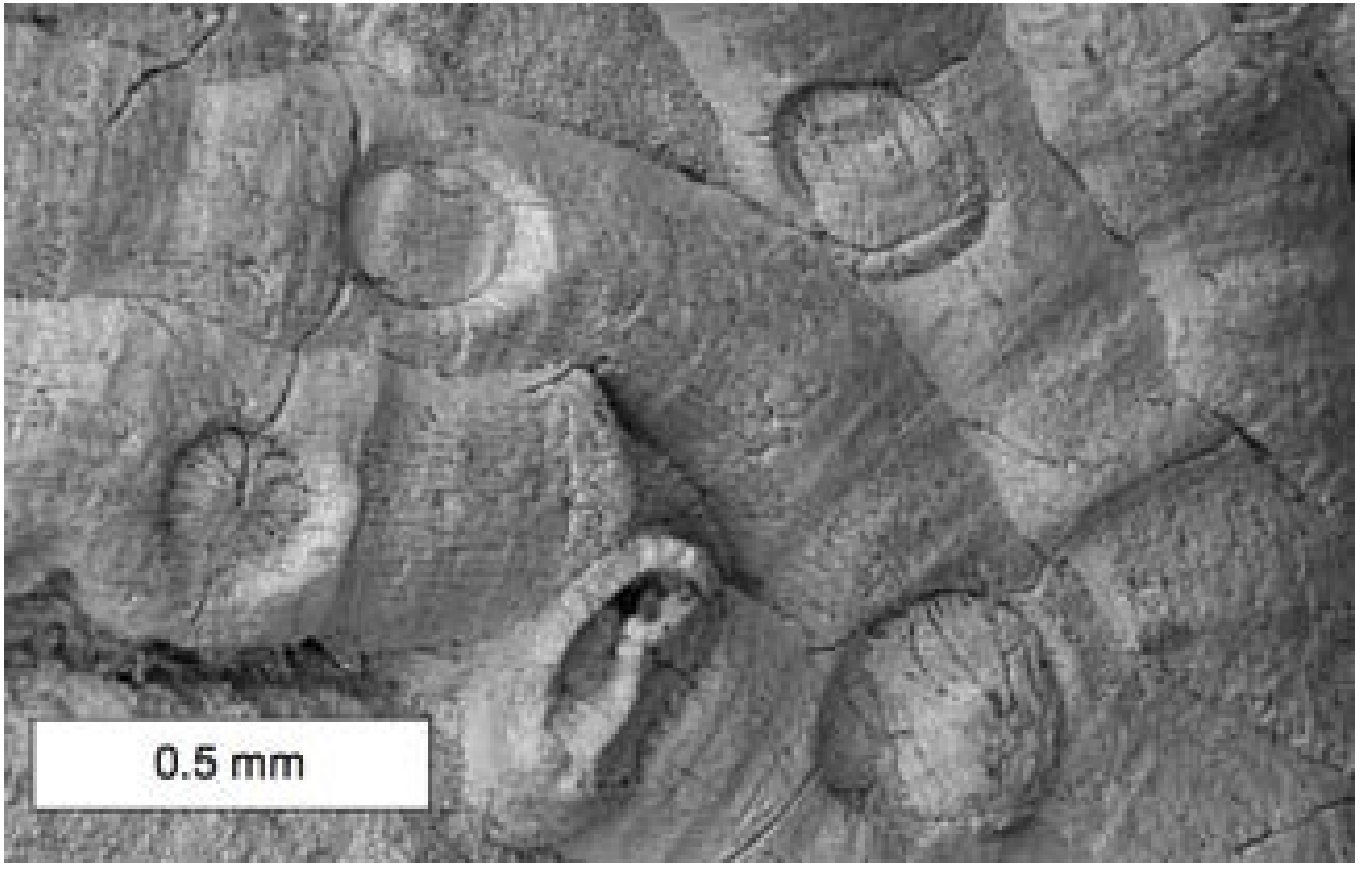
“Morphs”: large boundaries & short memory



Questions

?

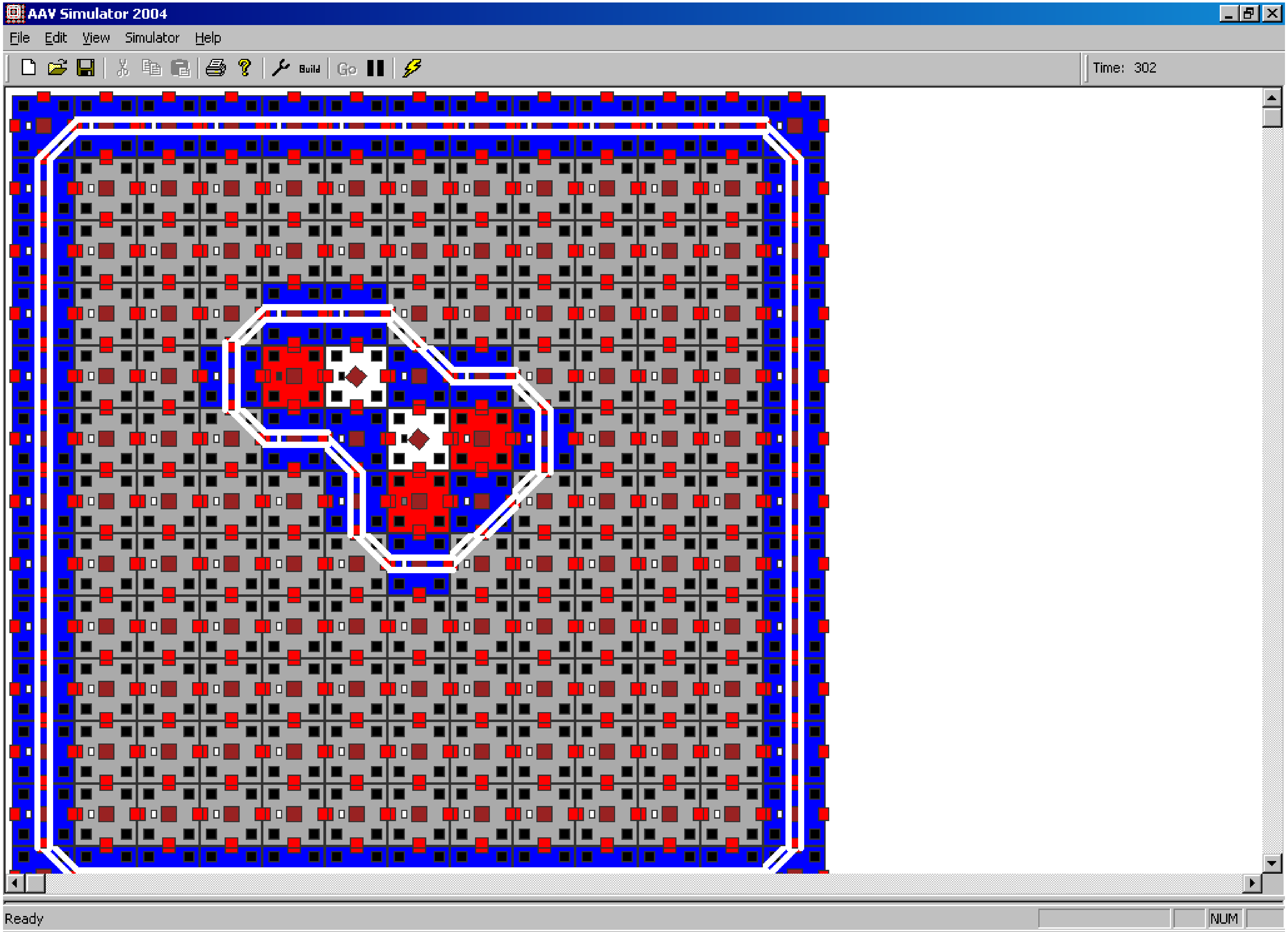
Self-repair: hederellid tubes (biological analogy)



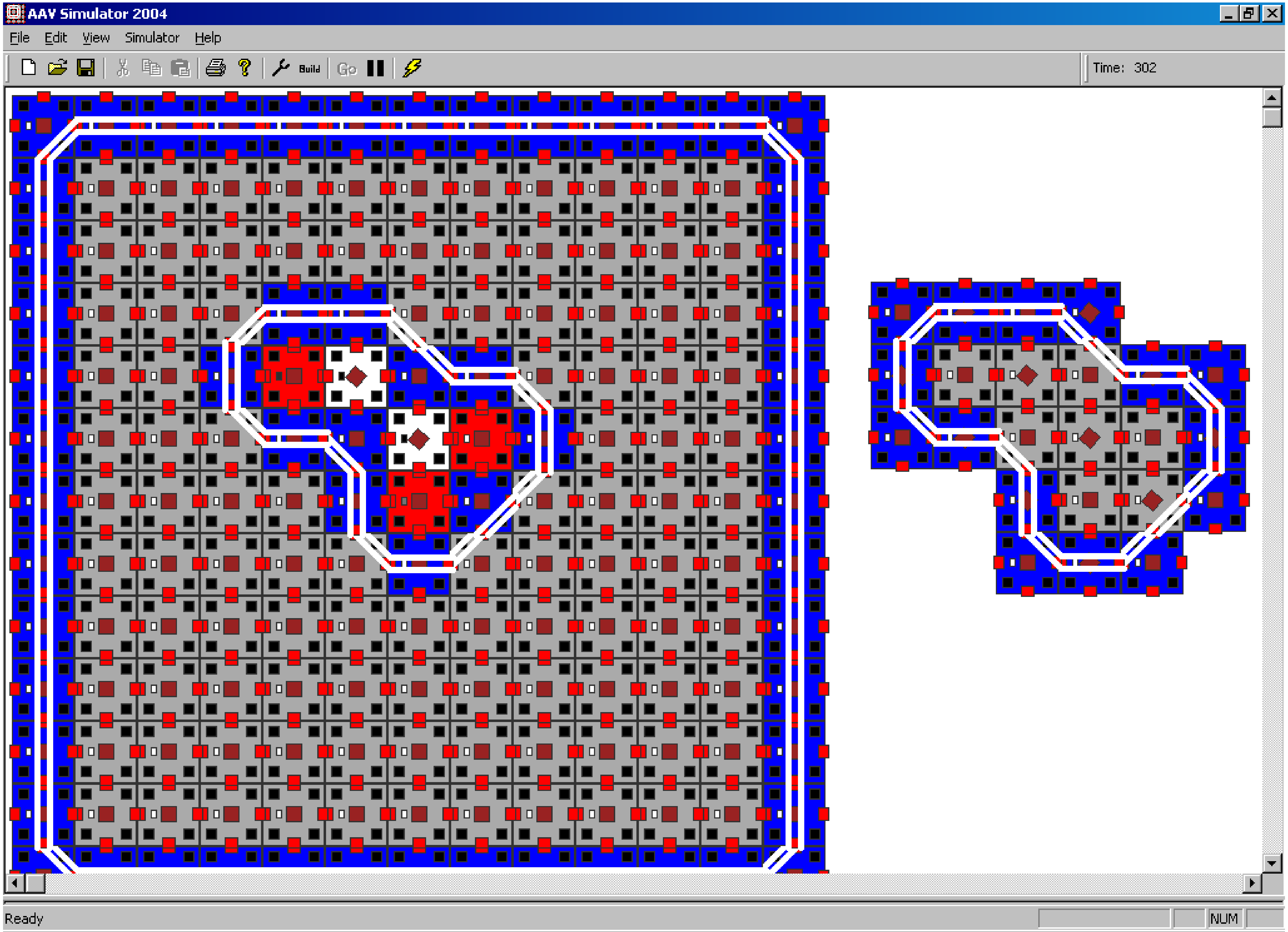
- each hole (drilled by a predator) is repaired with a patch
- skeletal patches are secreted by internal tissues
- damaged sections are closed off with skeletal plugs

Mark A. Wilson (Wooster, Ohio) and Paul D. Taylor (The Natural History Museum, London)
Geology; July 2006; v. 34; no. 7; p. 565-568

Shape replication: self-repair



Shape replication: self-repairable encoding



Shape replication: self-repairable encoding

Questions

?

Benefits of bio-inspired approaches

- efficiency: more flexible structures
  - robustness to multiple concurrent cell failures
  - scalability
- safety: damage detected at an early stage
  - continuous monitoring, evaluation and diagnosis of damage
  - longer-term prognosis for remediation responses
- cost: reduced maintenance down-time
  - self-monitoring
  - self-maintenance
  - self-repair

Distributed AI

Ethical concerns:

- bias and fairness
  - bias in data collected from different sources can lead to skewed decision-making in applications like hiring, lending, or law enforcement.
  - disparities in data representation from different demographics can result in biased models and unfair treatment
- privacy and data security
  - sensitive (personal or commercial-interest) data may be transmitted and stored across multiple devices and locations, increasing the risk of data breaches
  - adversarial attacks on distributed nodes can compromise the integrity of the entire system
  - data poisoning attacks where malicious data is introduced to corrupt the AI system
- accountability and transparency
  - difficulty in tracing decisions back to specific data sources or algorithms in distributed environments
  - hard to ensure transparency and explainability of AI decisions across multiple agents and nodes

ChatGPT 4o

Distributed AI

Ethical concerns:

- ethical use and misuse
  - use of distributed AI for mass surveillance without proper oversight or consent
  - deployment of AI systems that manipulate user behavior or spread misinformation
- environmental impact
  - large-scale distributed AI systems may consume vast amounts of electricity, contributing to carbon emissions
  - use of high-performance computing infrastructure that requires substantial energy resources
- social and economic disparities
  - disparities in access to distributed AI infrastructure and capabilities between developed and developing regions
  - potential job displacement and economic disruption caused by the deployment of distributed AI systems

ChatGPT 4o

Distributed AI

Mitigation of ethical concerns:

- bias and fairness
  - ensuring diverse and representative data collection practices
  - regularly auditing AI systems for bias and implementing fairness-aware algorithms
- privacy and data security
  - implementing strong encryption methods for data transmission and storage
  - ensuring compliance with privacy laws and regulations
  - using differential privacy to protect individual data while allowing aggregated data analysis
  - implementing robust security protocols and regular security audits
  - using secure multi-party computation and federated learning to enhance security
  - developing and deploying methods to detect and mitigate adversarial attacks and data poisoning
- accountability and transparency
  - developing clear frameworks for accountability that specify responsibilities during the AI lifecycle
  - implementing explainable AI (XAI) techniques making the decision-making process more transparent
  - maintaining comprehensive logs of data sources, algorithms, and decision processes

ChatGPT 4o

Distributed AI

Mitigation of ethical concerns:

- ethical use and misuse
  - establishing and enforcing ethical guidelines and regulations for the use of distributed AI
- environmental impact
  - developing energy-efficient algorithms and hardware for distributed AI
  - promoting the use of renewable energy sources to power AI infrastructure
  - conducting environmental impact assessments for large-scale AI projects
- social and economic disparities
  - promoting inclusive access to AI technologies and infrastructure
  - implementing policies and programs to support workforce transition and reskilling
  - engaging in ethical impact assessments to understand and mitigate the social consequences of AI deployment

ChatGPT 4o

Homework

- check the information sources provided on different slides (at least some)

- check some publications describing the AAV project:
  - M. Prokopenko, P. Wang, D.C. Price, P. Valencia, M. Foreman, A.J. Farmer. Self-organizing hierarchies in sensor and communication networks, *Artificial Life*, 11(4), 407-426, 2005.
  - M. Prokopenko, P. Wang, M. Foreman, P. Valencia, D.C. Price, G.T. Poulton. On connectivity of reconfigurable impact networks in Ageless Aerospace Vehicles, *Journal of Robotics and Autonomous Systems*, 53(1), 36-58, 2005.
  - P. Wang, M. Prokopenko. Evolvable recovery membranes in self-monitoring aerospace vehicles. In *Proceedings of the 8th International Conference on Simulation of Adaptive Behaviour (SAB-2004)*, Los Angeles, USA, July 2004.
  - M. Prokopenko, P. Wang. On self-referential shape replication in robust aerospace vehicles. In *Proceedings of the 9th International Conference on the Simulation and Synthesis of Living Systems (ALIFE9)*, Boston, USA, 2004.

- prepare to give a few brief comments about the project:
  - ➢ what was most striking about the AI technology in the video(s)?
  - ➢ what societal impact did the AI make?
  - ➢ what were the main AI limitations, biases and risks?

?

https://www.prokopenko.net/publications-aav.html