COMP9208
Artificial Intelligence and Society

Lecture 7: 19 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

?

Distributed AI

- multi-agent systems (MAS)
- distributed machine learning (DML)
- distributed constraint satisfaction problems (DCSP)
- distributed reasoning
- swarm intelligence:

?

- distributed evolutionary algorithms:
  - genetic algorithms
  - evolutionary strategies
  - genetic programming

Distributed AI

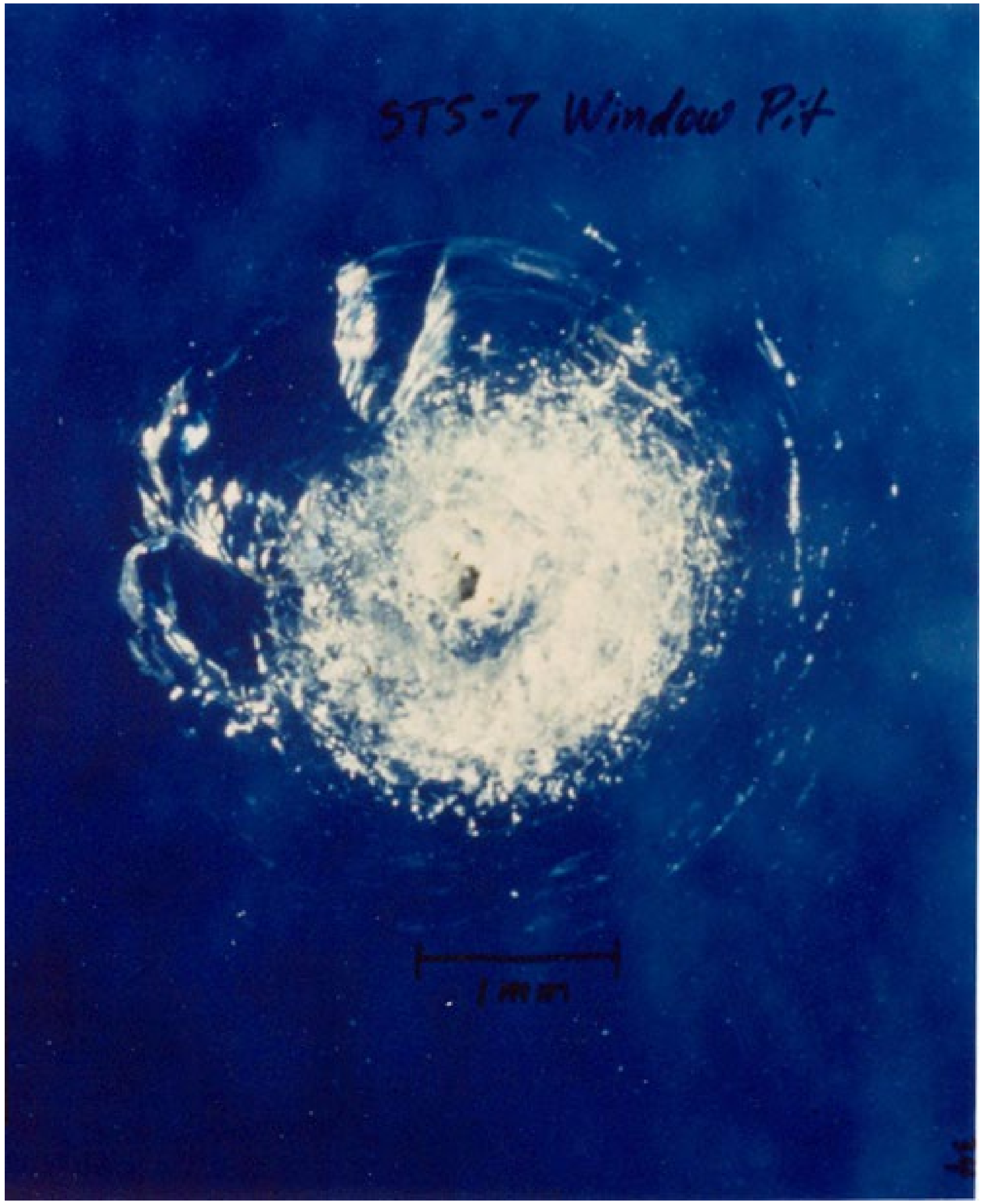
- multi-agent systems (MAS)
- distributed machine learning (DML)
- distributed constraint satisfaction problems (DCSP)
- distributed reasoning
- swarm intelligence:
  - Ant Colony Optimization (ACO): optimisation algorithms inspired by the foraging behavior of ants
  - Particle Swarm Optimization (PSO): 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
- distributed evolutionary algorithms:
  - genetic algorithms
  - evolutionary strategies
  - genetic programming

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

Langley Research Center, NASA

CSIRO Australia

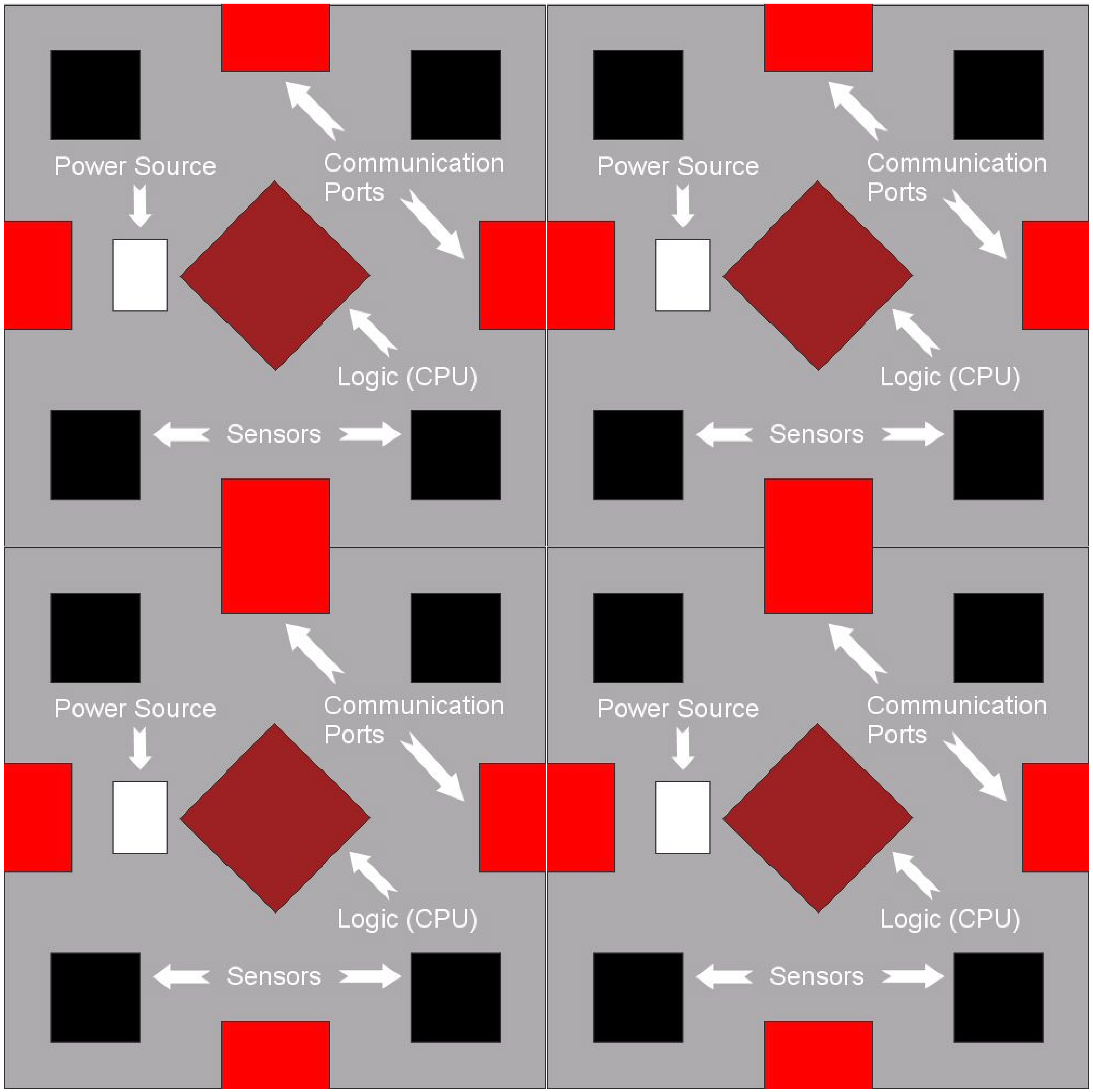
State of the practice (circa 2002)



Window pit from orbital debris on STS-007

Space Shuttle window being inspected for orbital debris impacts

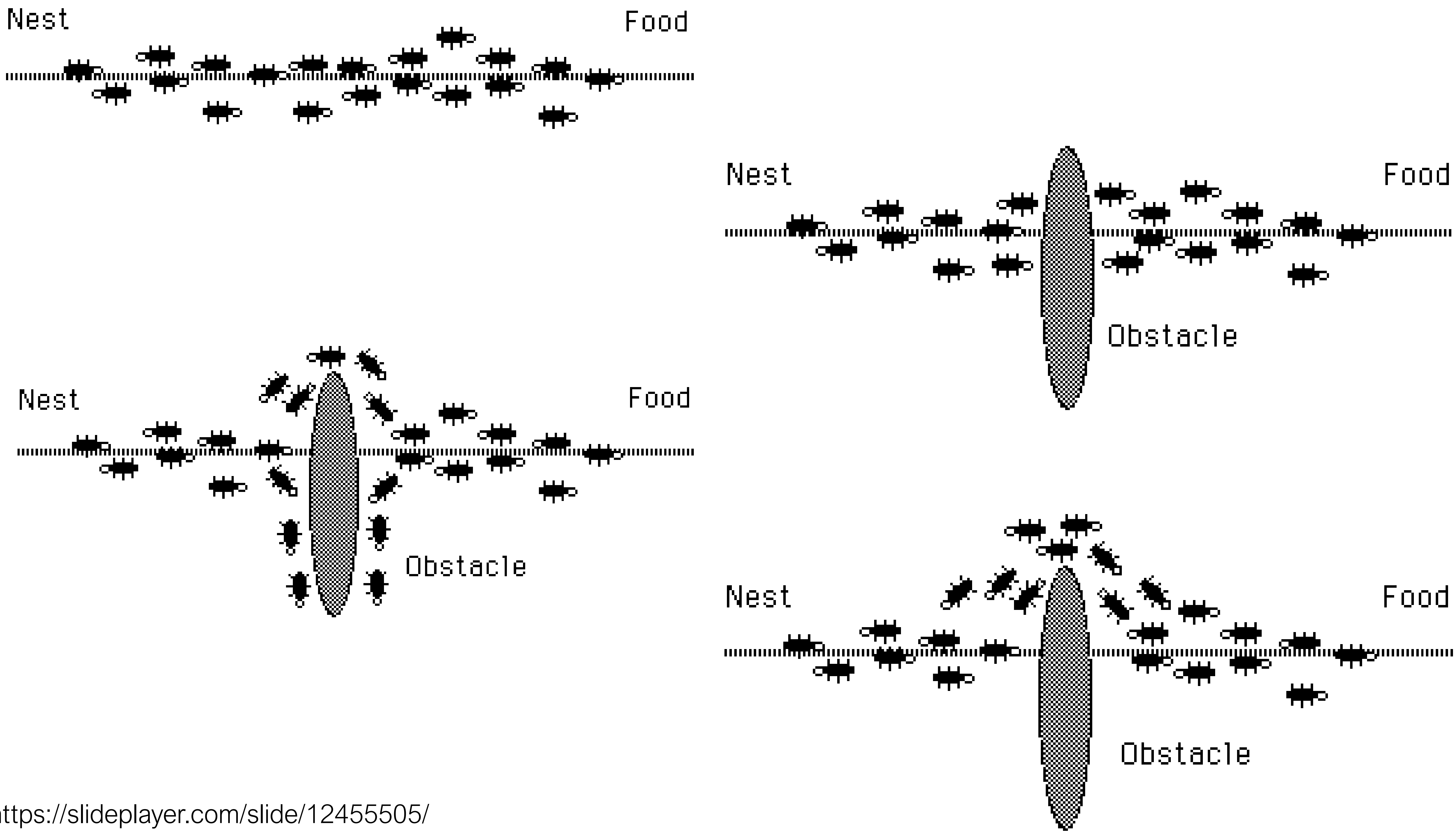
High-level design: sensor & communication network



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

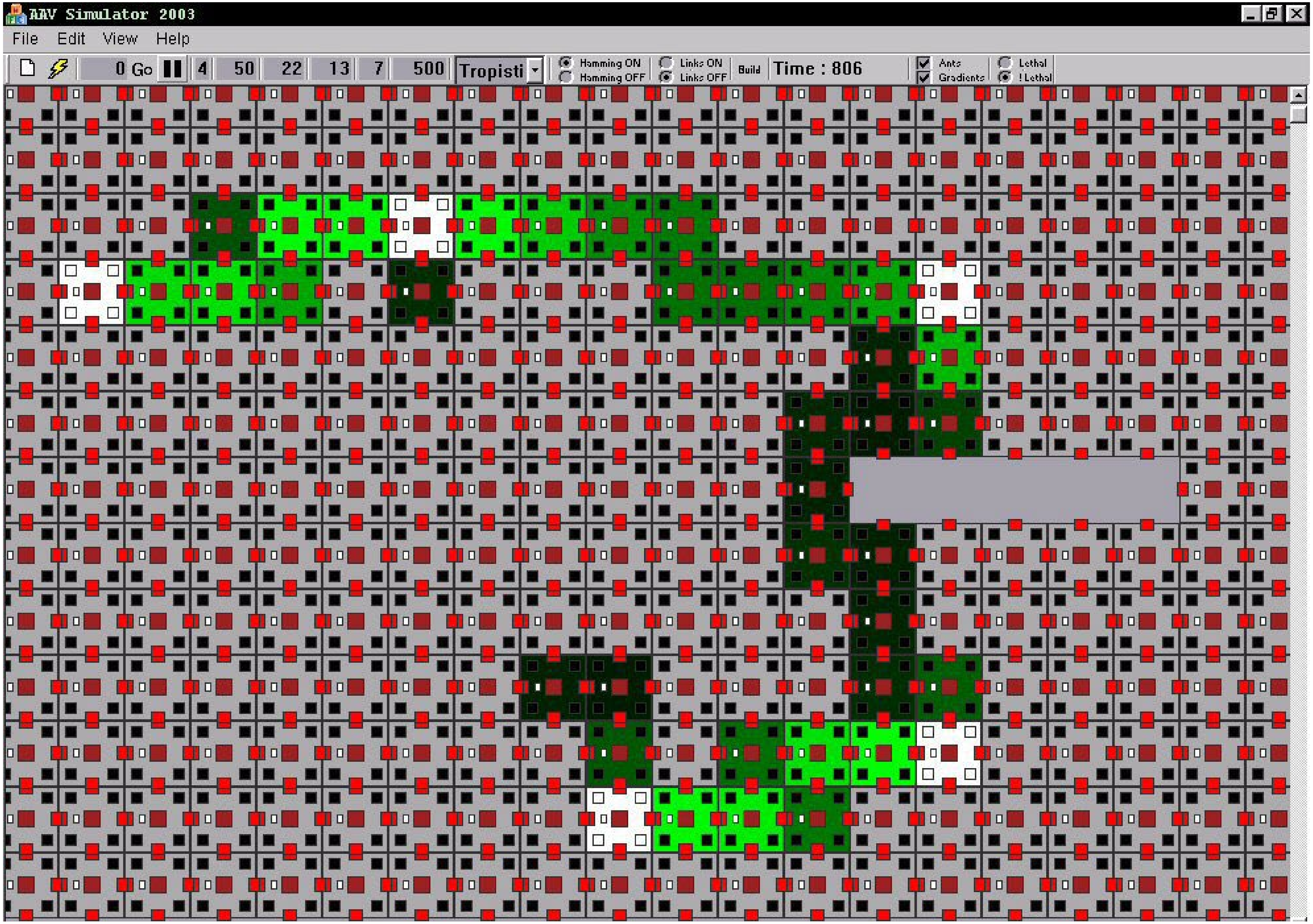
Distributed AI

Swarm intelligence: Ant Colony Optimisation

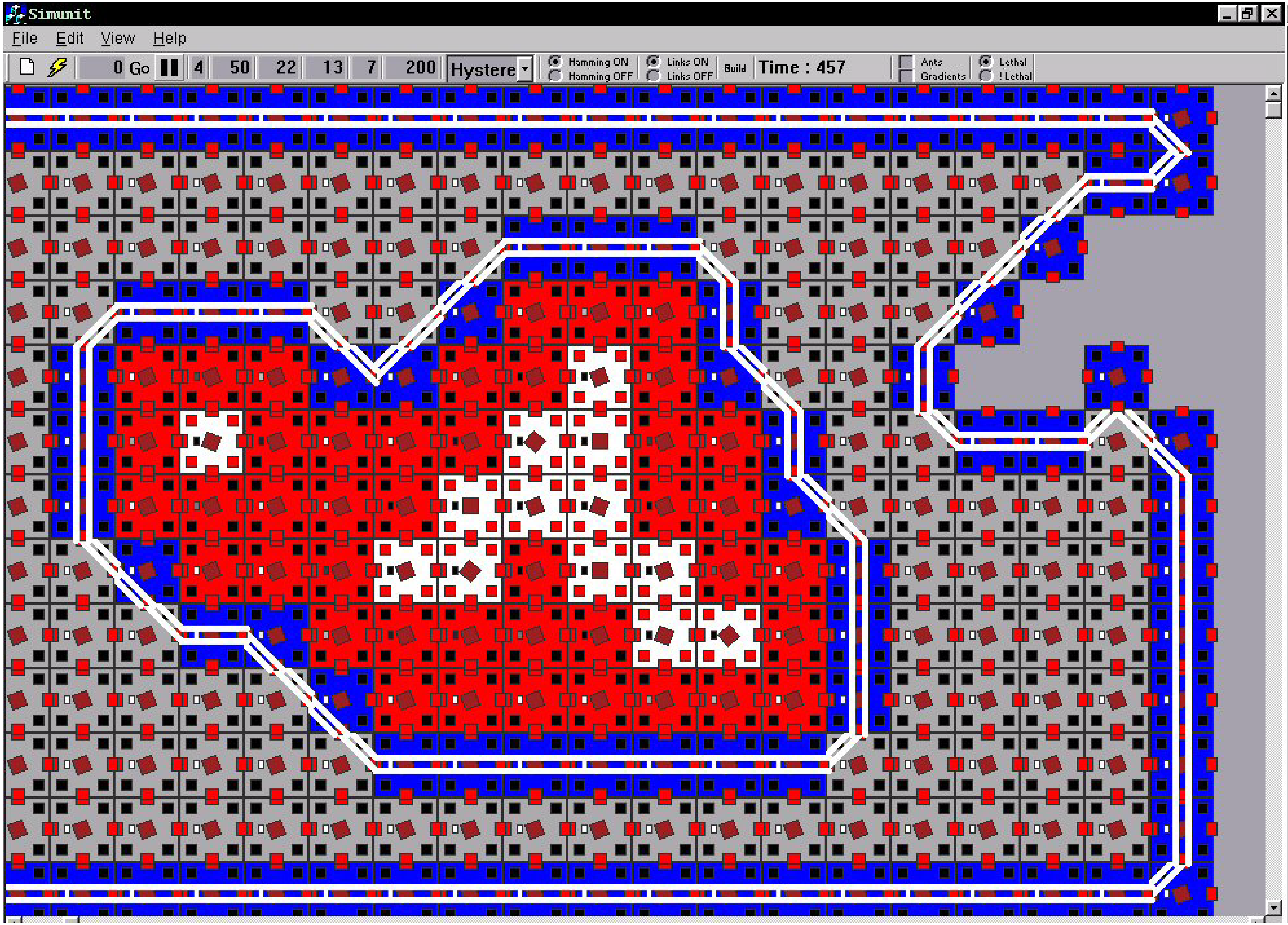


https://slideplayer.com/slide/12455505/

Self-organising impact networks: connecting impact points

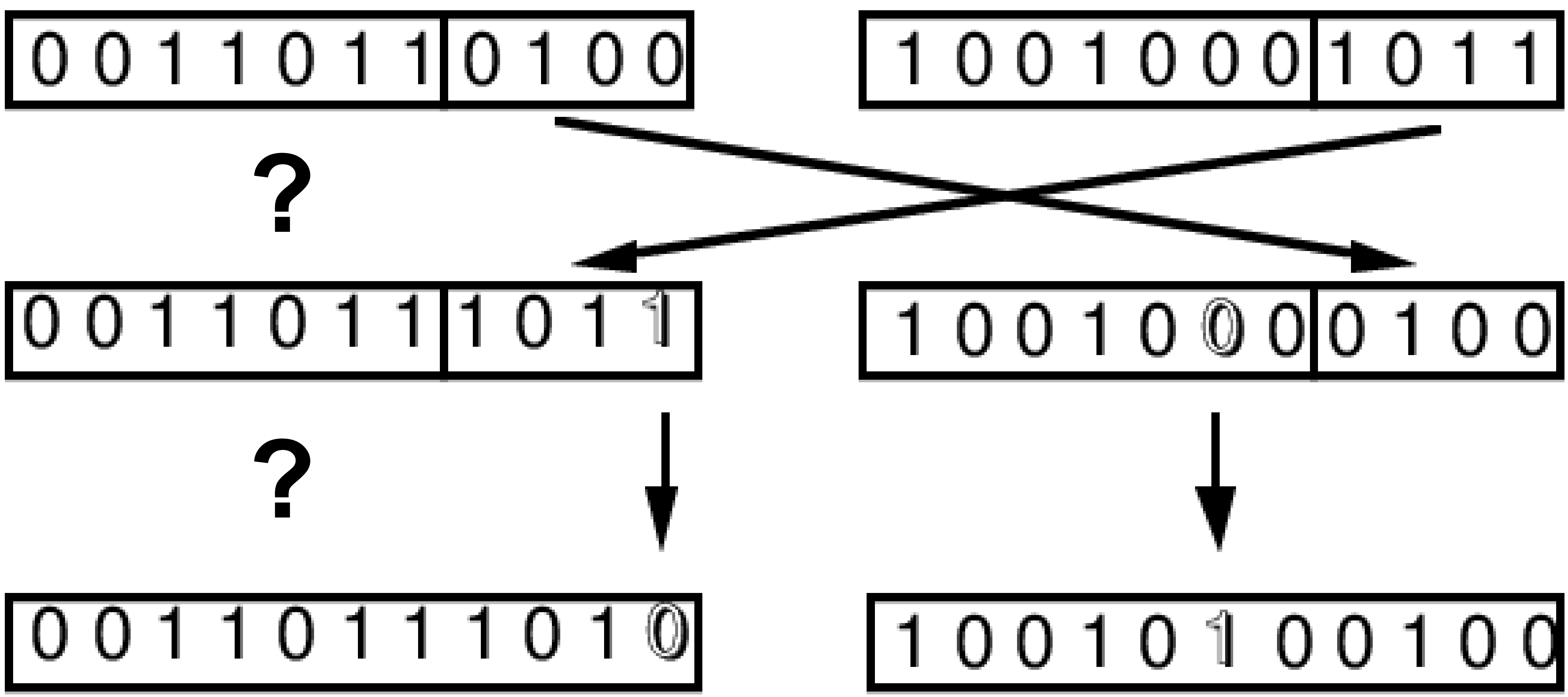


Self-organising stable impact boundaries



Distributed AI

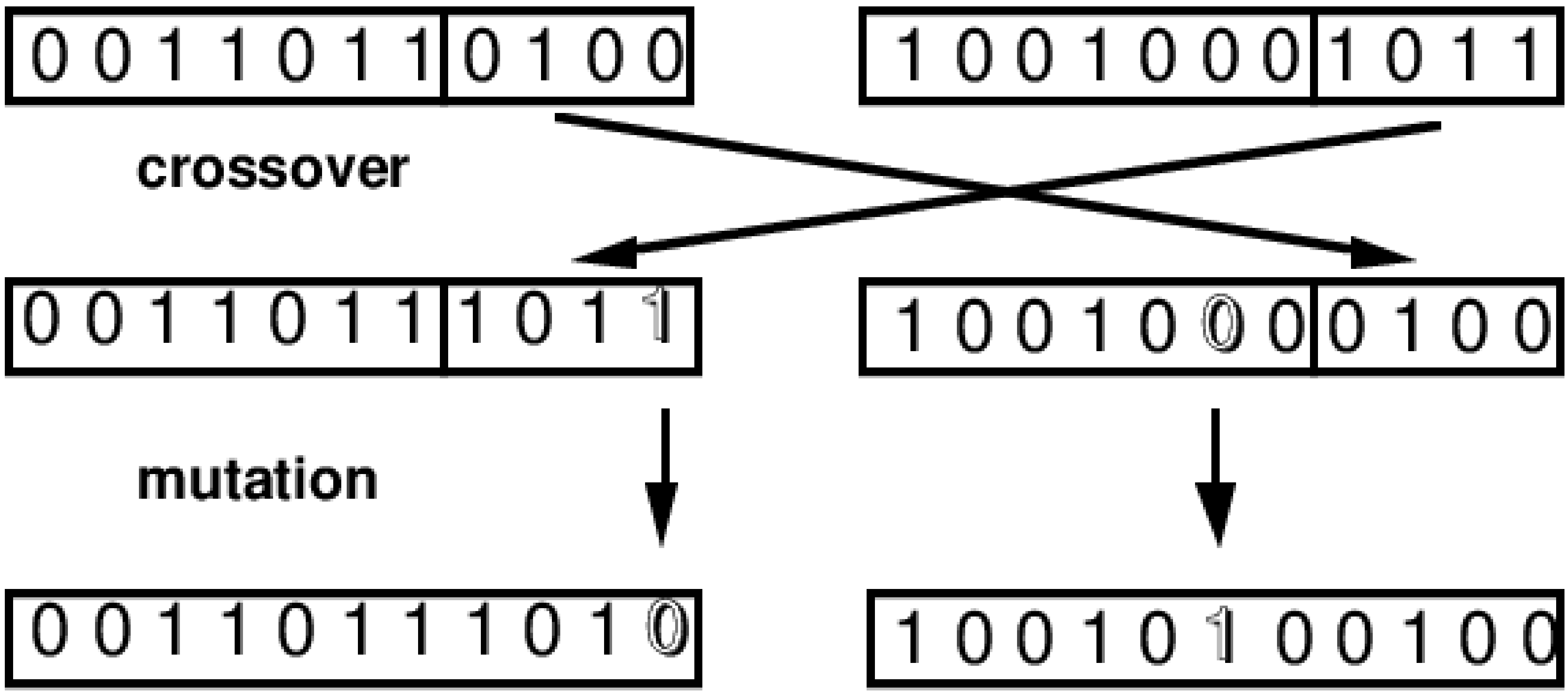
Genetic algorithms



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

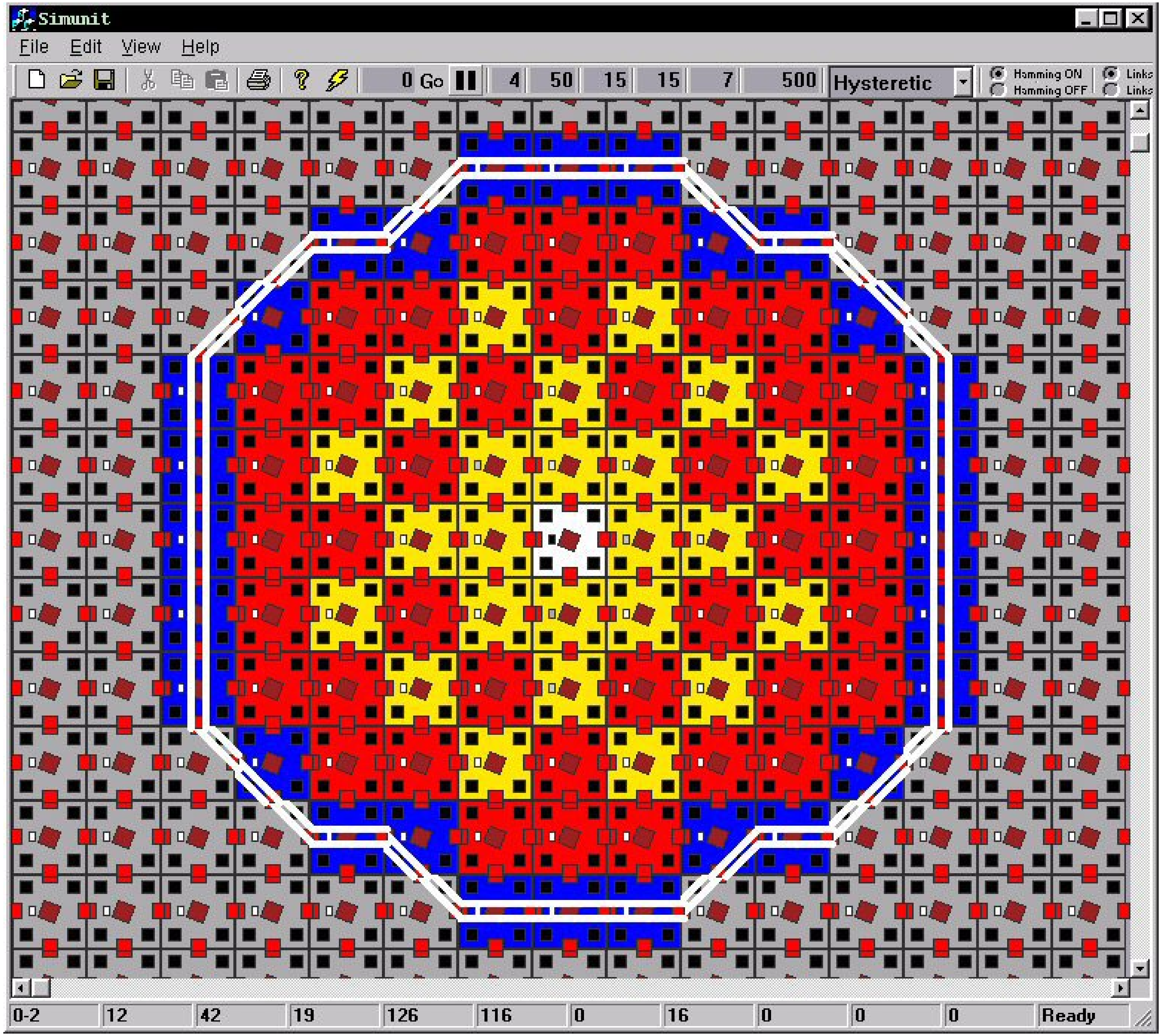
Distributed AI

Genetic algorithms



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

“Morphs”: large boundaries & short memory



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

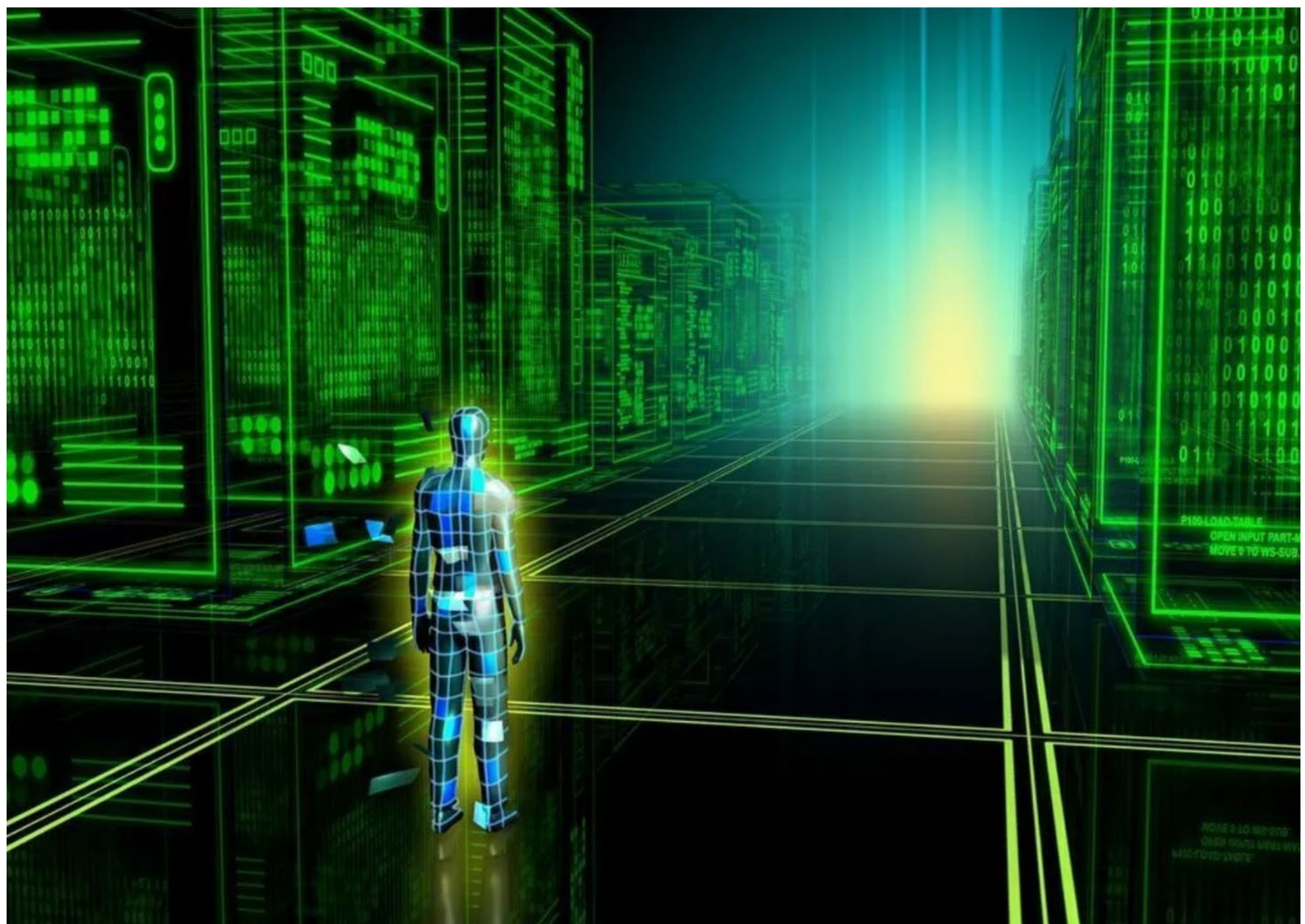
Questions

?

History of AI: main achievements and setbacks

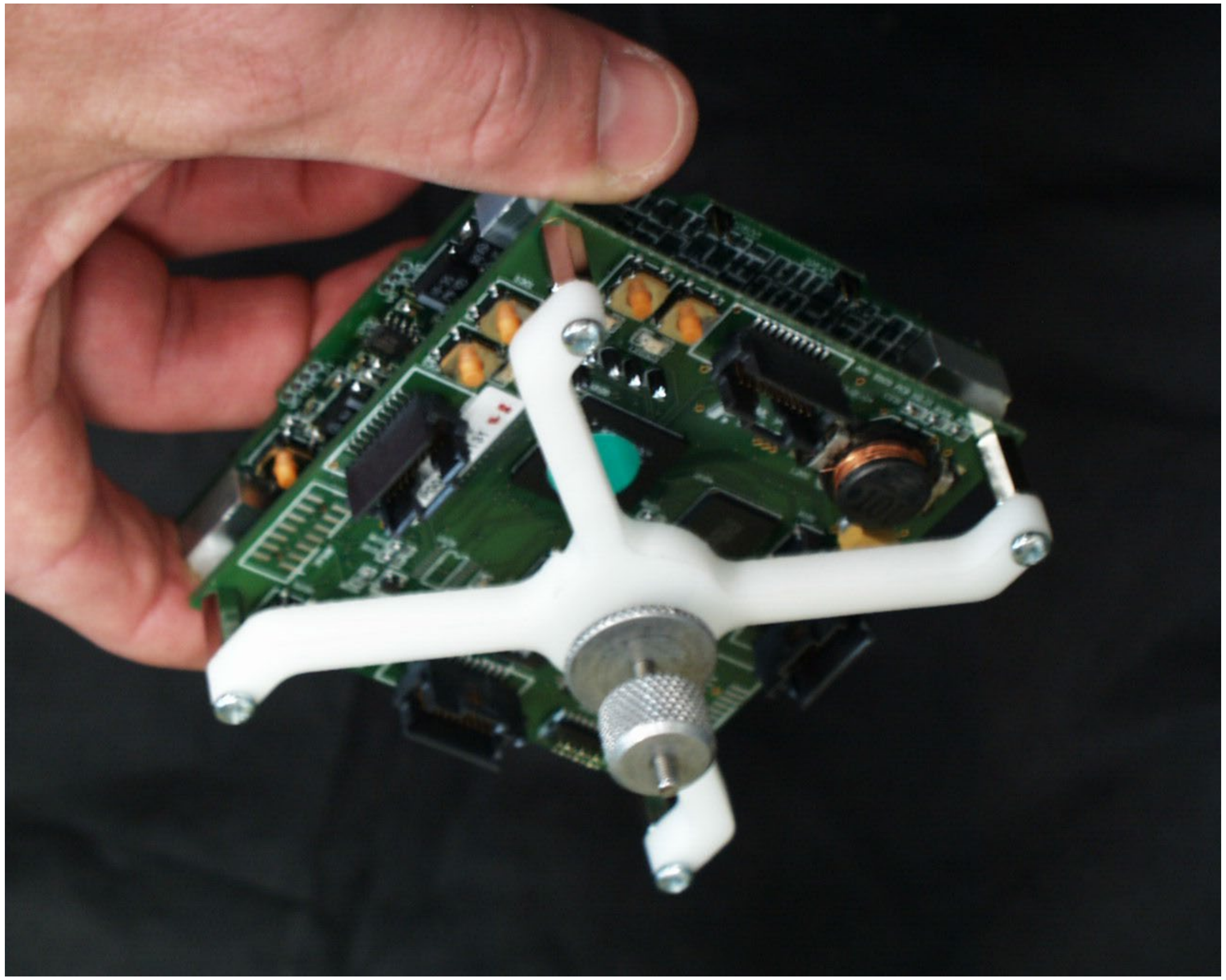
How to overcome Moravec's paradox (abstract reasoning needs very little computation, but perception and sensorimotor skills require large computational resources)

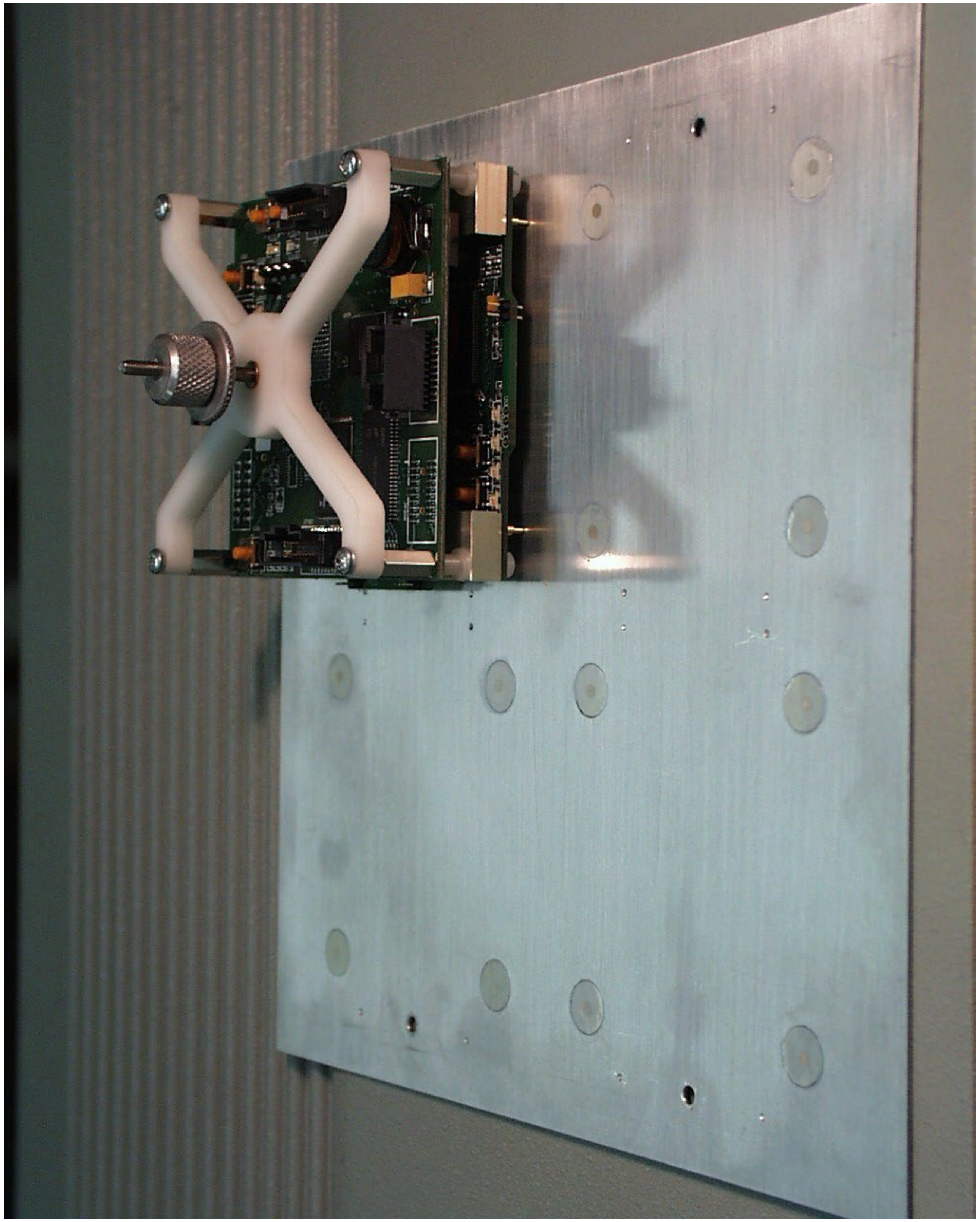
- ➢ Simulators can replicate real-world scenarios, helping AI to learn and adapt more efficiently



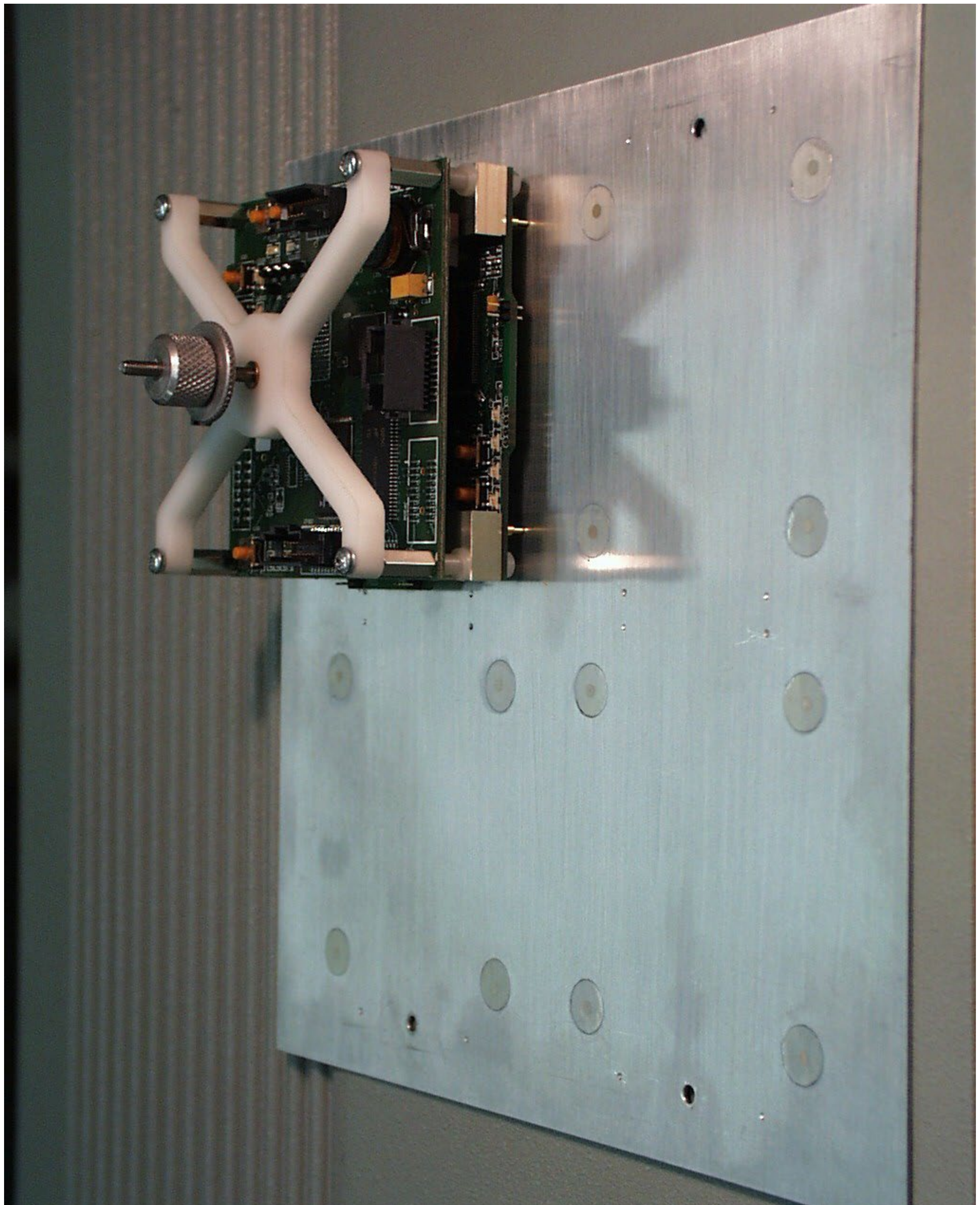
https://www.linkedin.com/pulse/unraveling-moravecs-paradox-enigma-robots-struggling-everyday-sharma
https://medium.com/simudyne/simulation-the-bedrock-of-ai-12153eaf7971

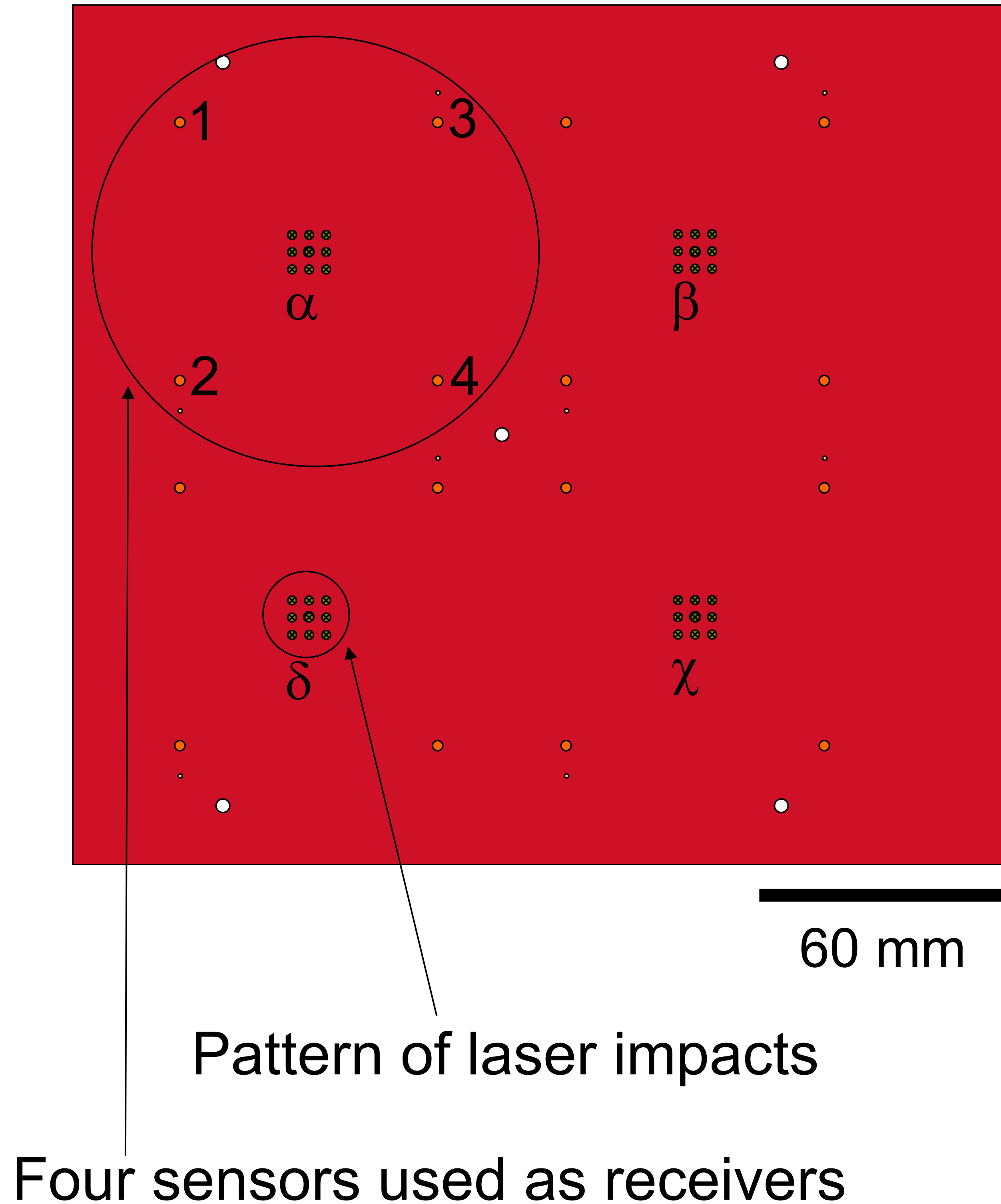
AAV cells and embedded sensors



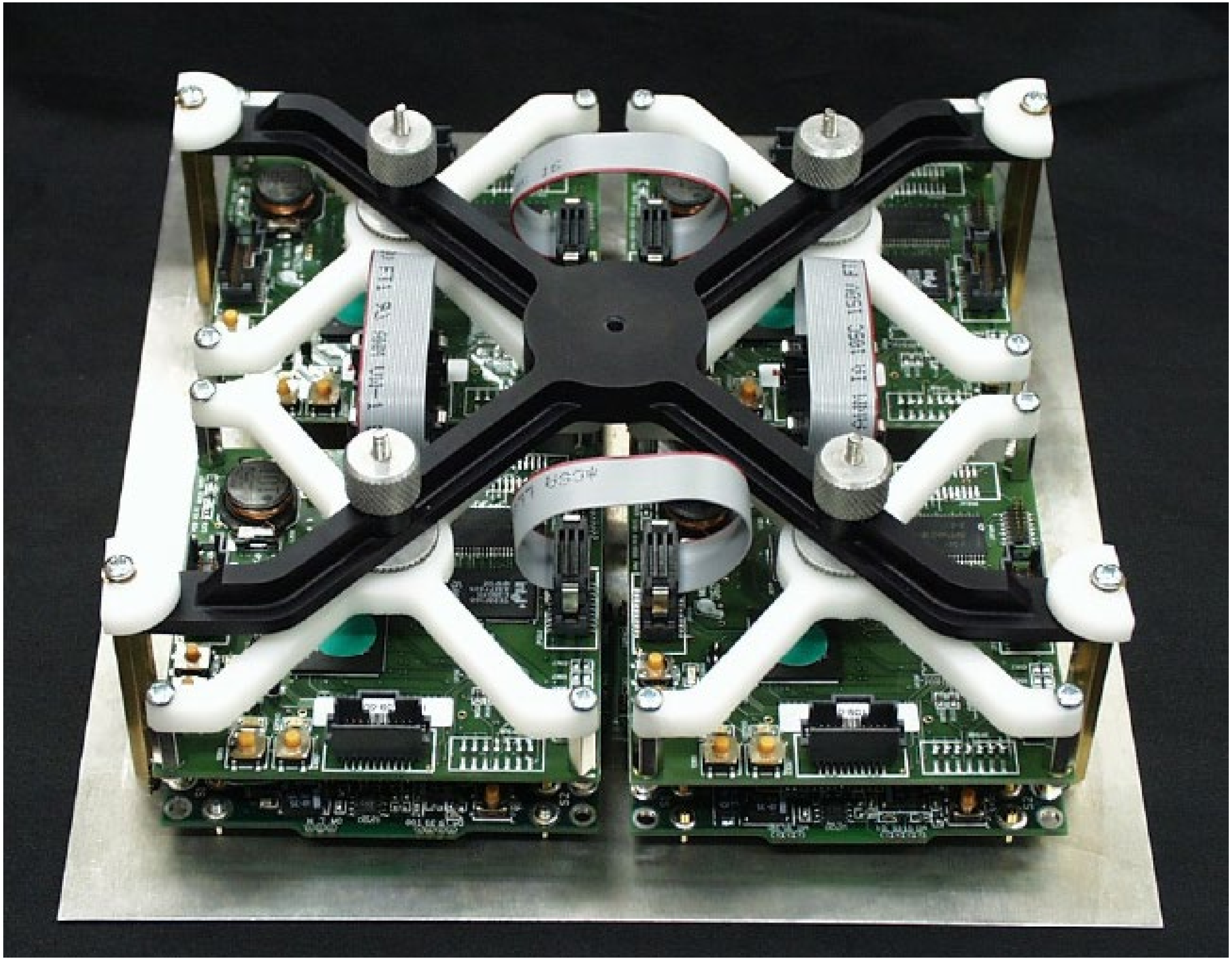


Piezo-electric polymer Acoustic Emission (AE) sensors



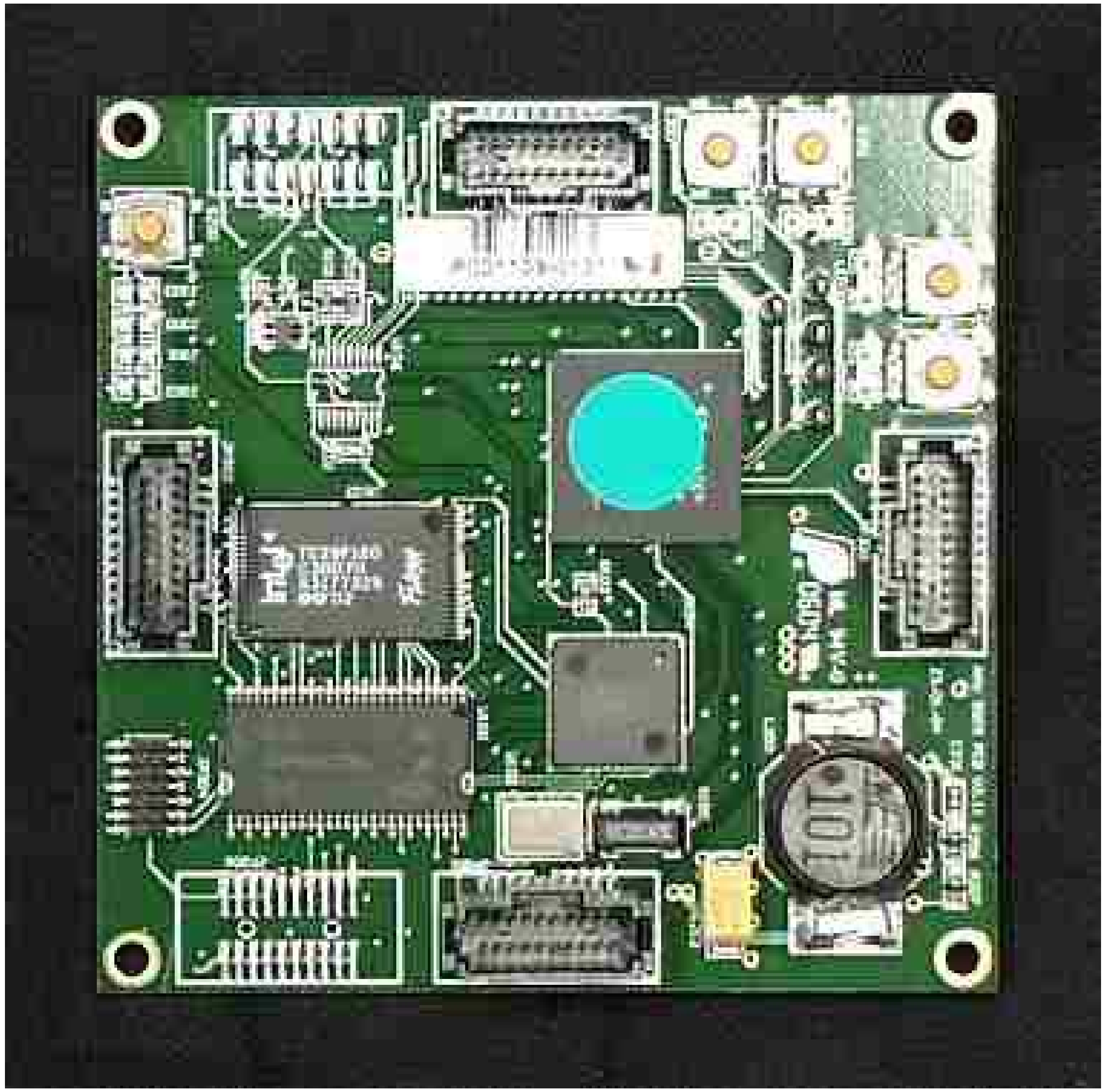


AAV panel



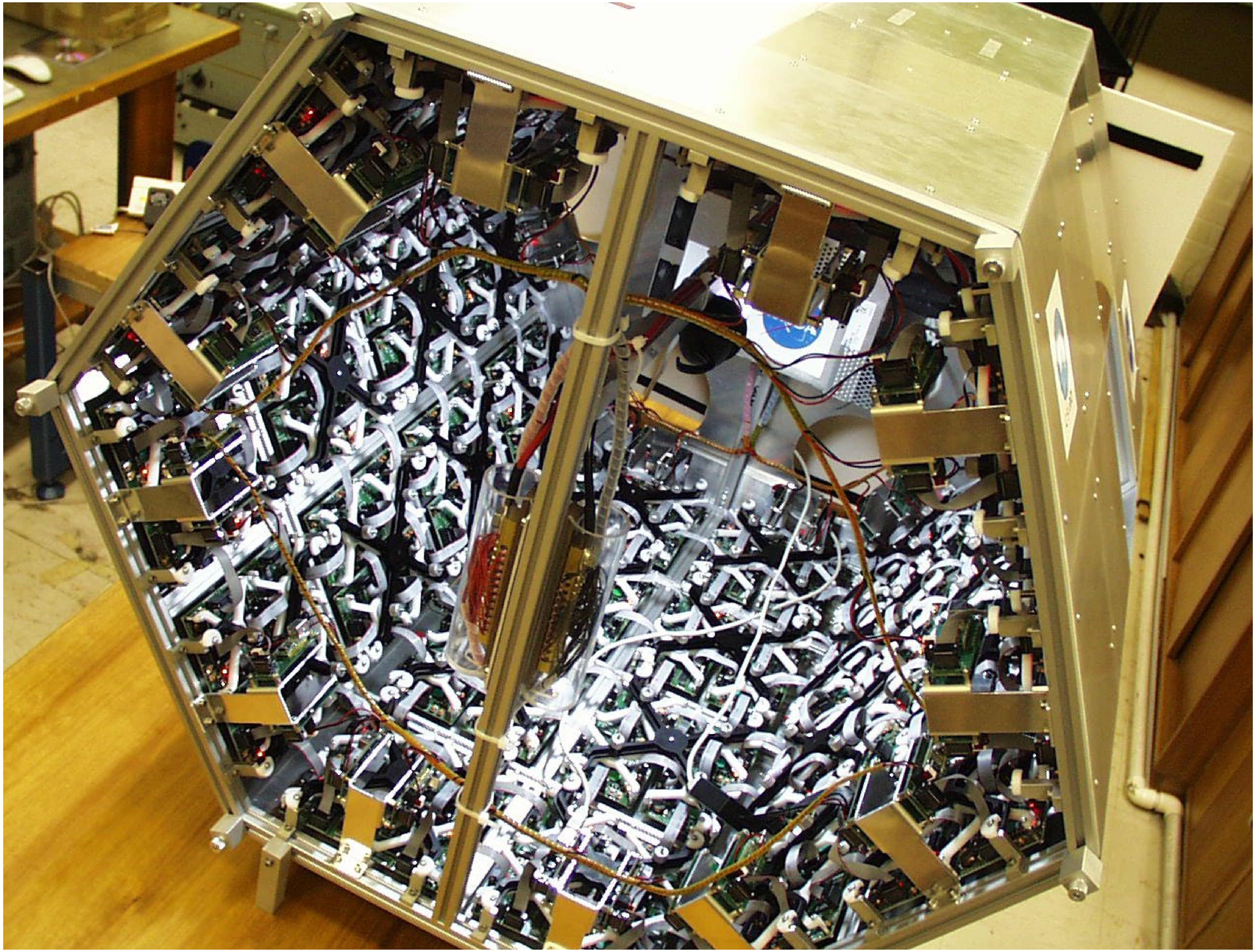
AAV boards

Data Acquisition board



Network Application board

AAV concept demonstrator

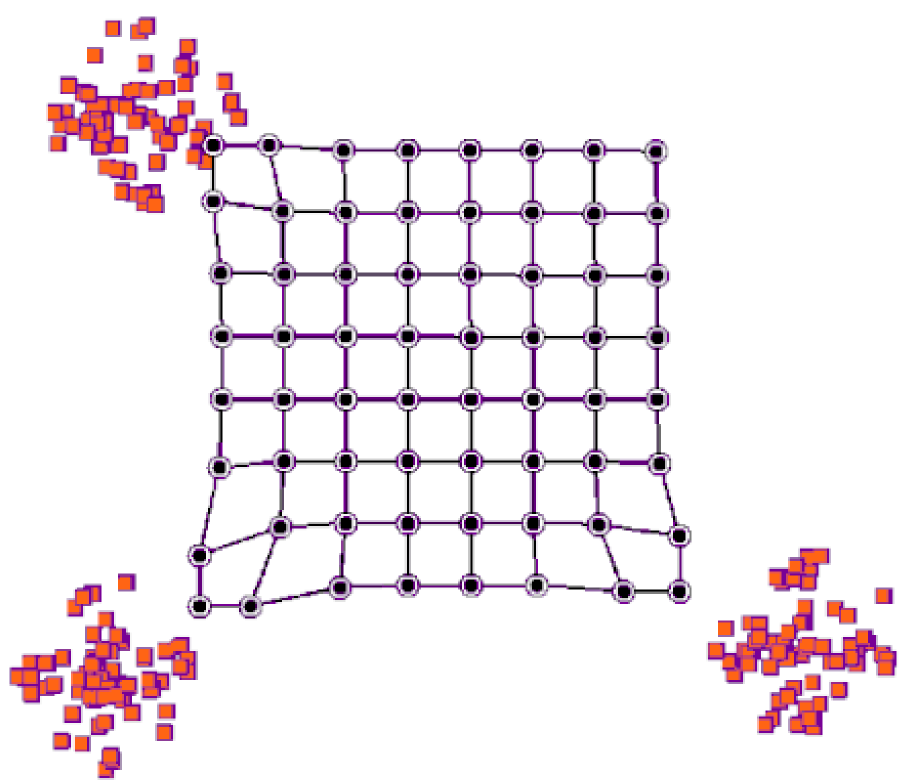


Questions

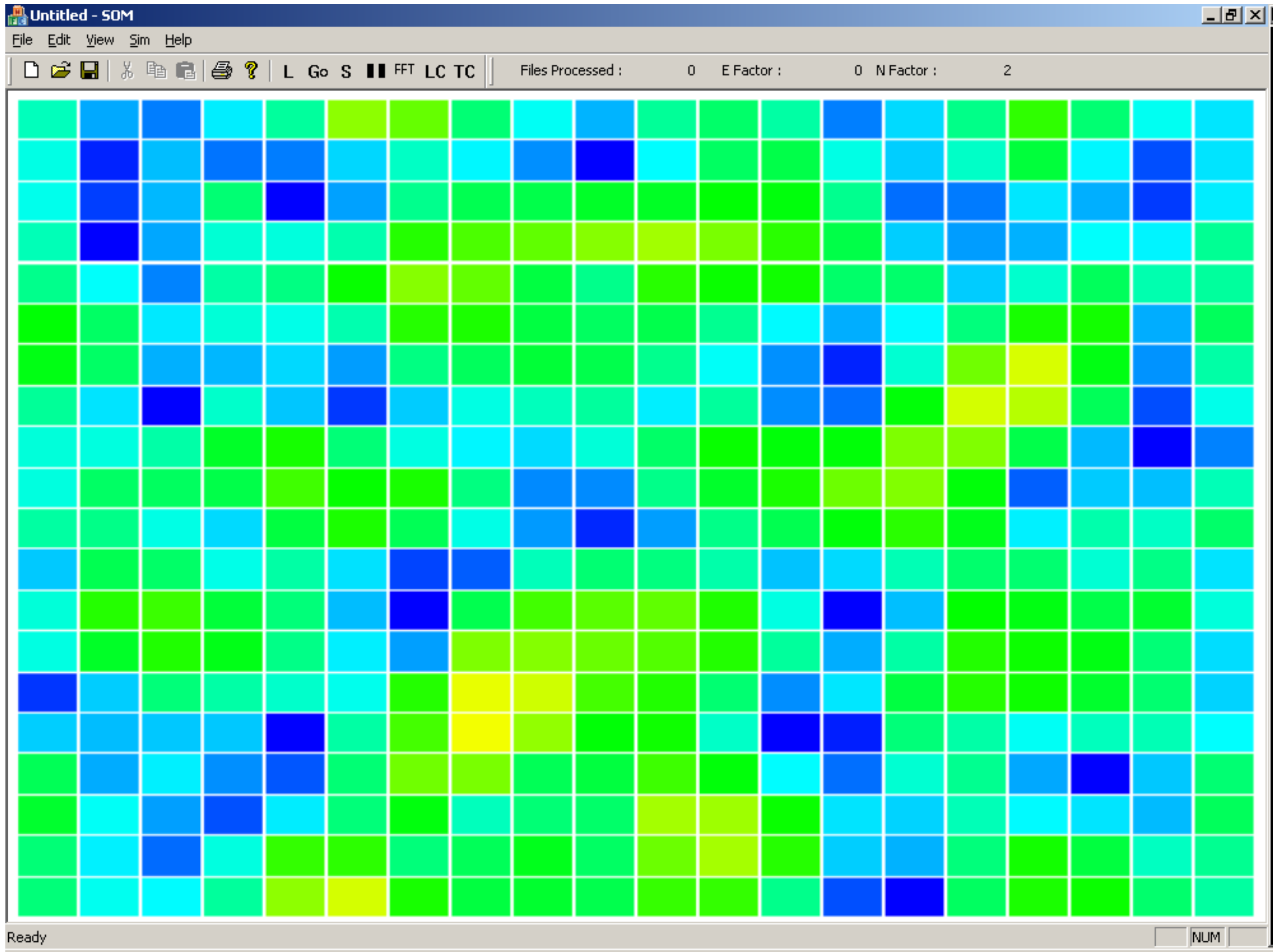
?

AAV diagnostics: self-organising (Kohonen) maps

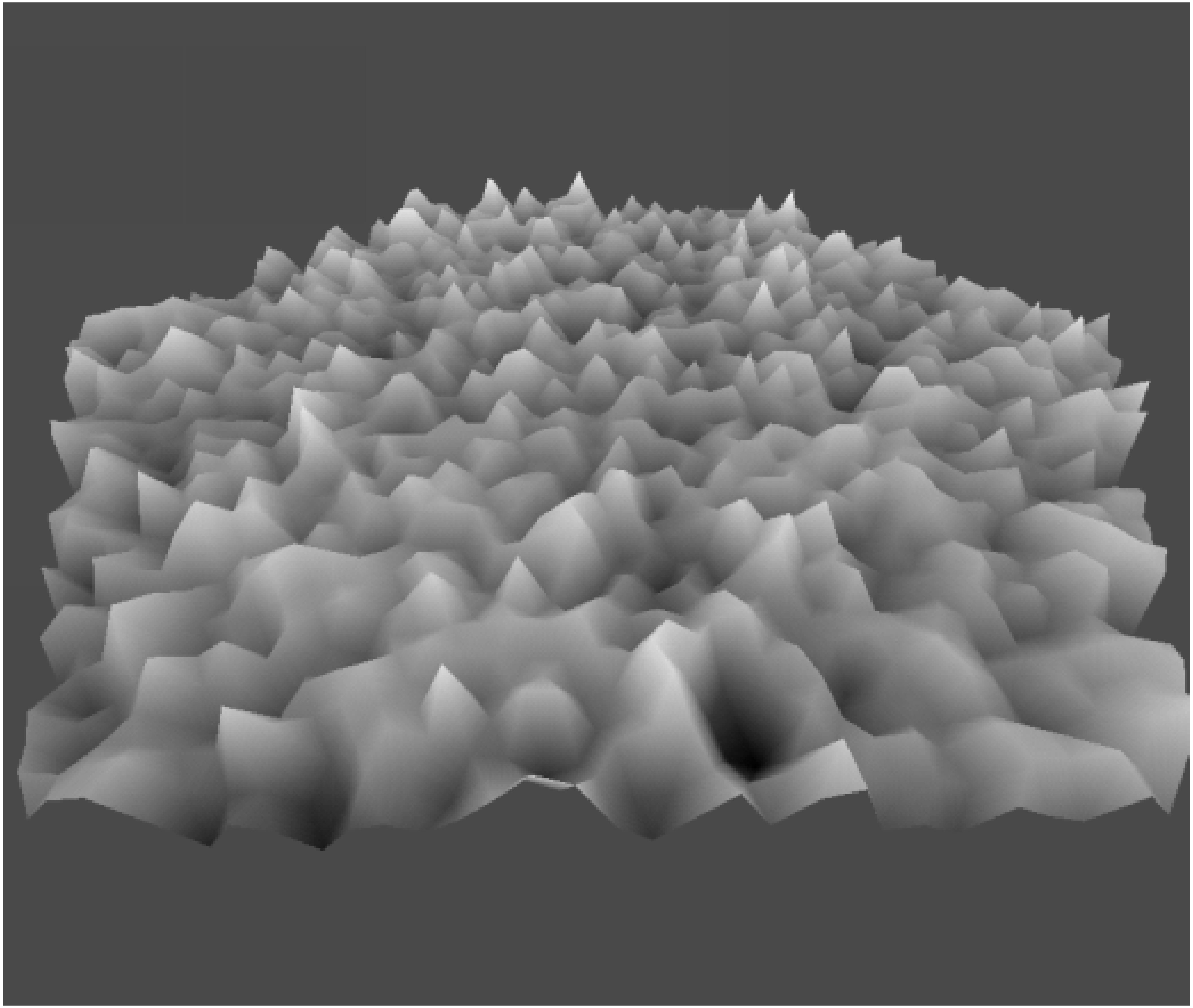
- **Input**
  - a number of *feature vectors*
- **Competitive learning**
- **Output**
  - neural network, *map* of *nodes positioned on a grid*
  - each node has a *codebook vector* of the same dimensionality as the input vectors



https://en.wikipedia.org/wiki/Self-organizing_map

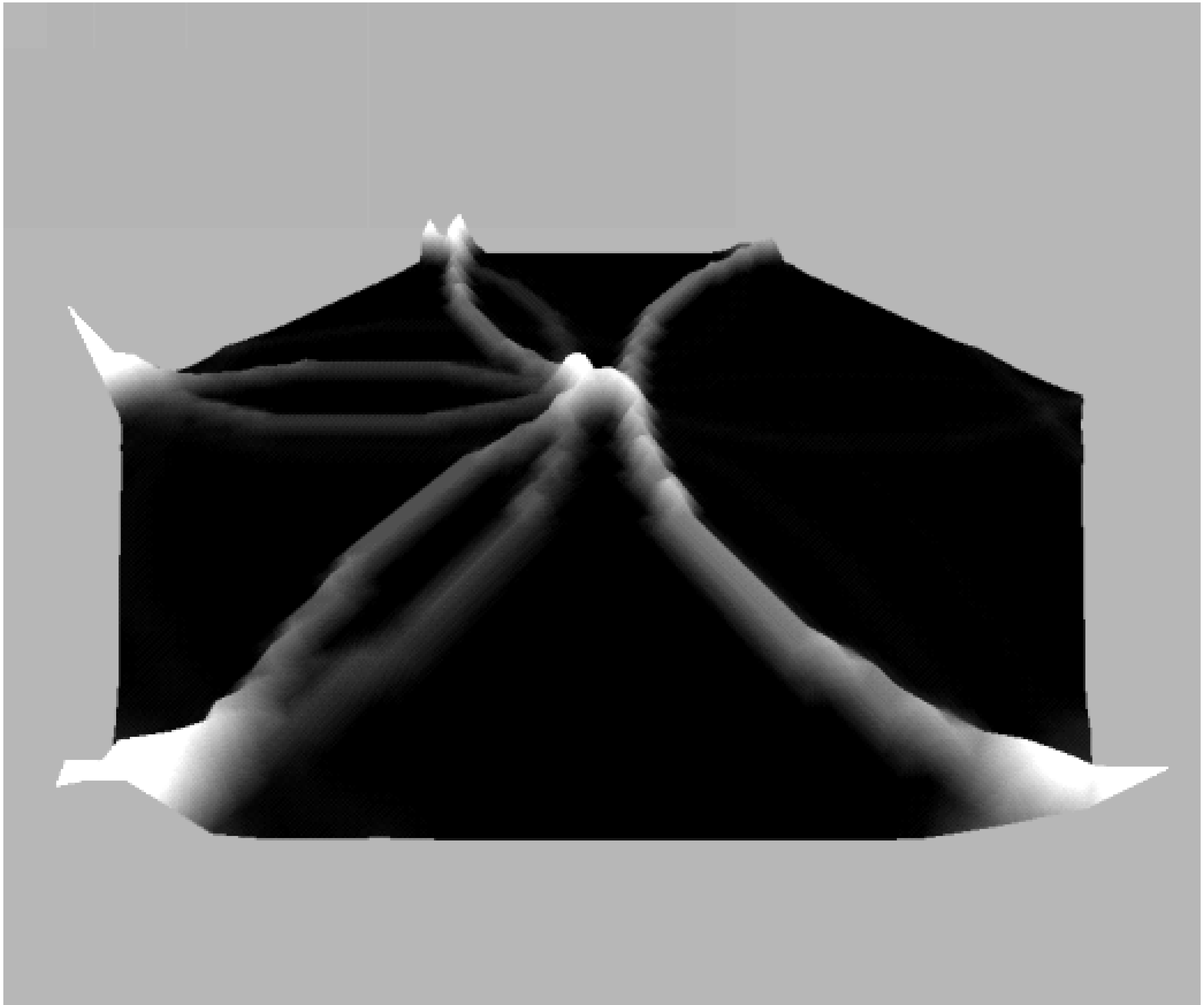


AAV diagnostics: self-organising (Kohonen) maps



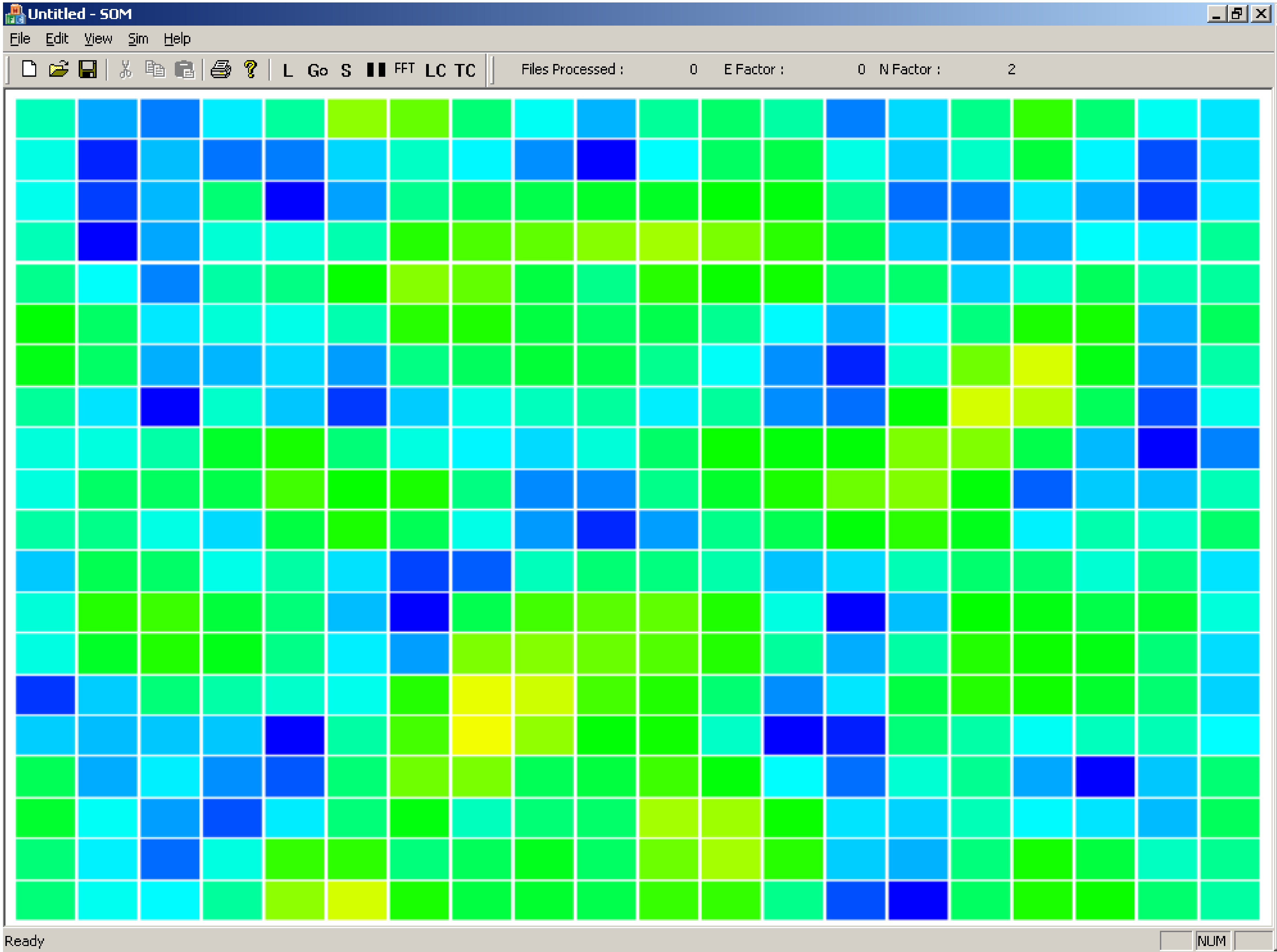
Visualisation of an initial unordered SOM

AAV diagnostics: self-organising (Kohonen) maps

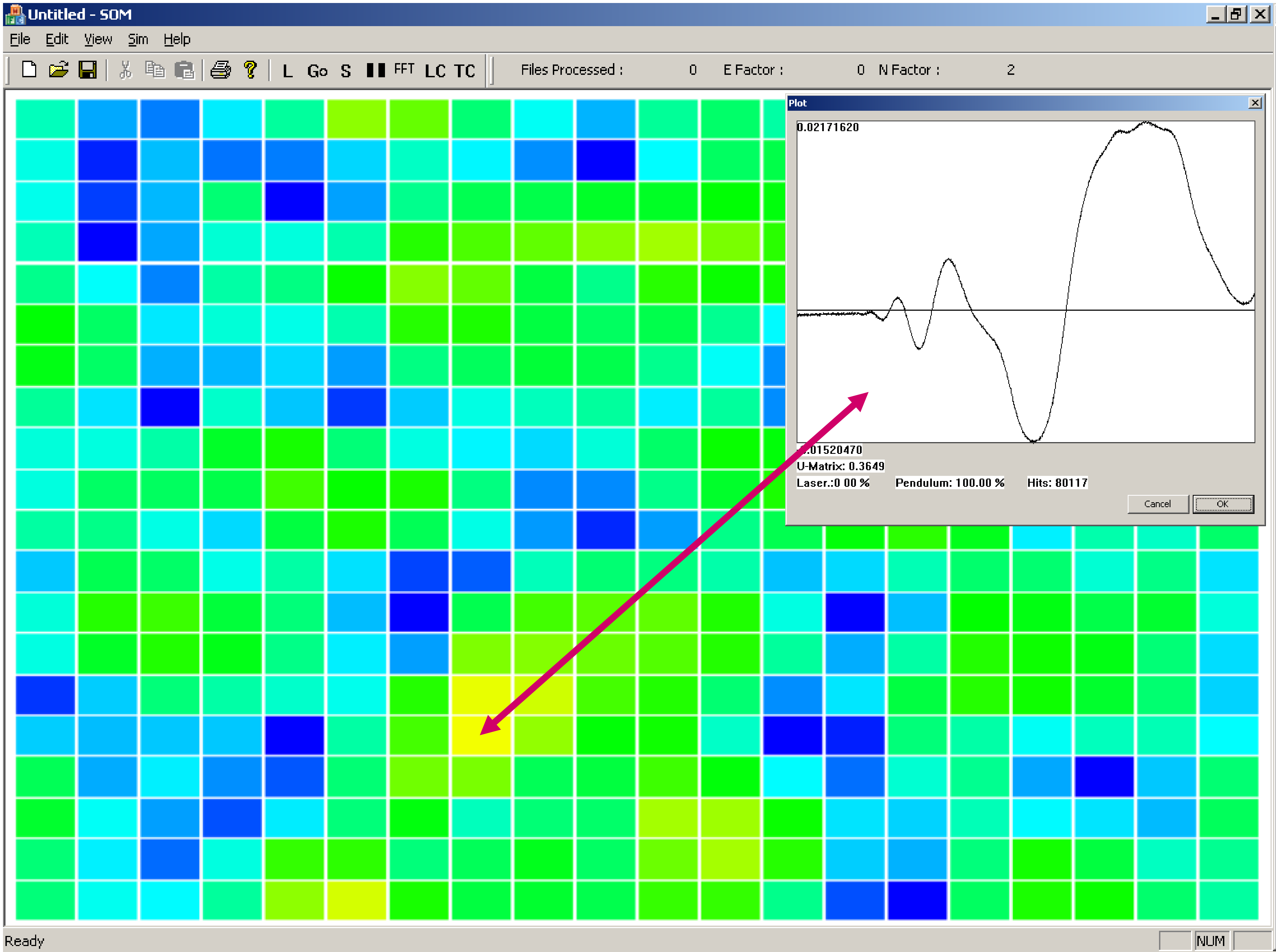


Partially ordered SOM after 5 time steps

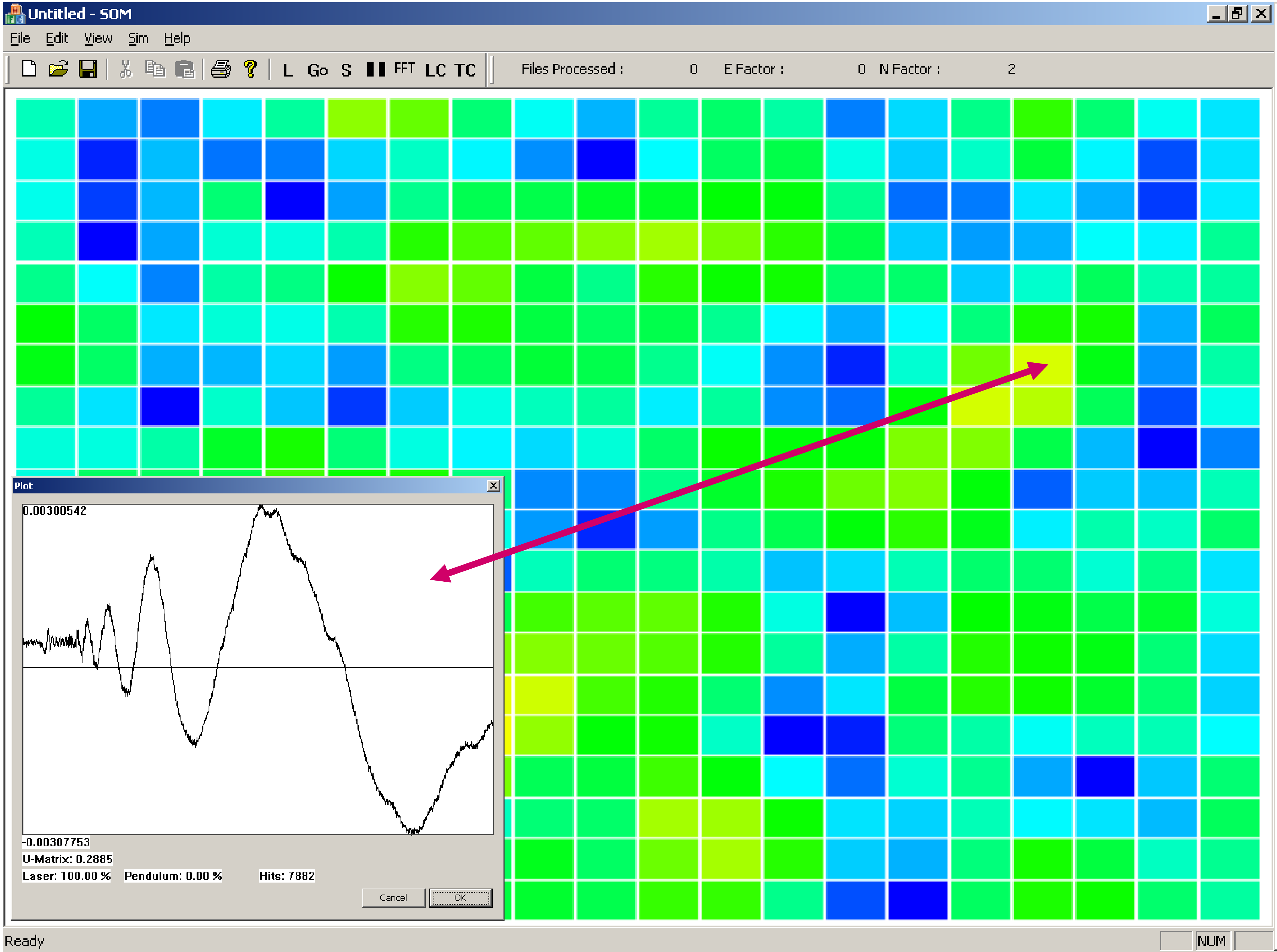
AAV diagnostics: self-organising (Kohonen) maps



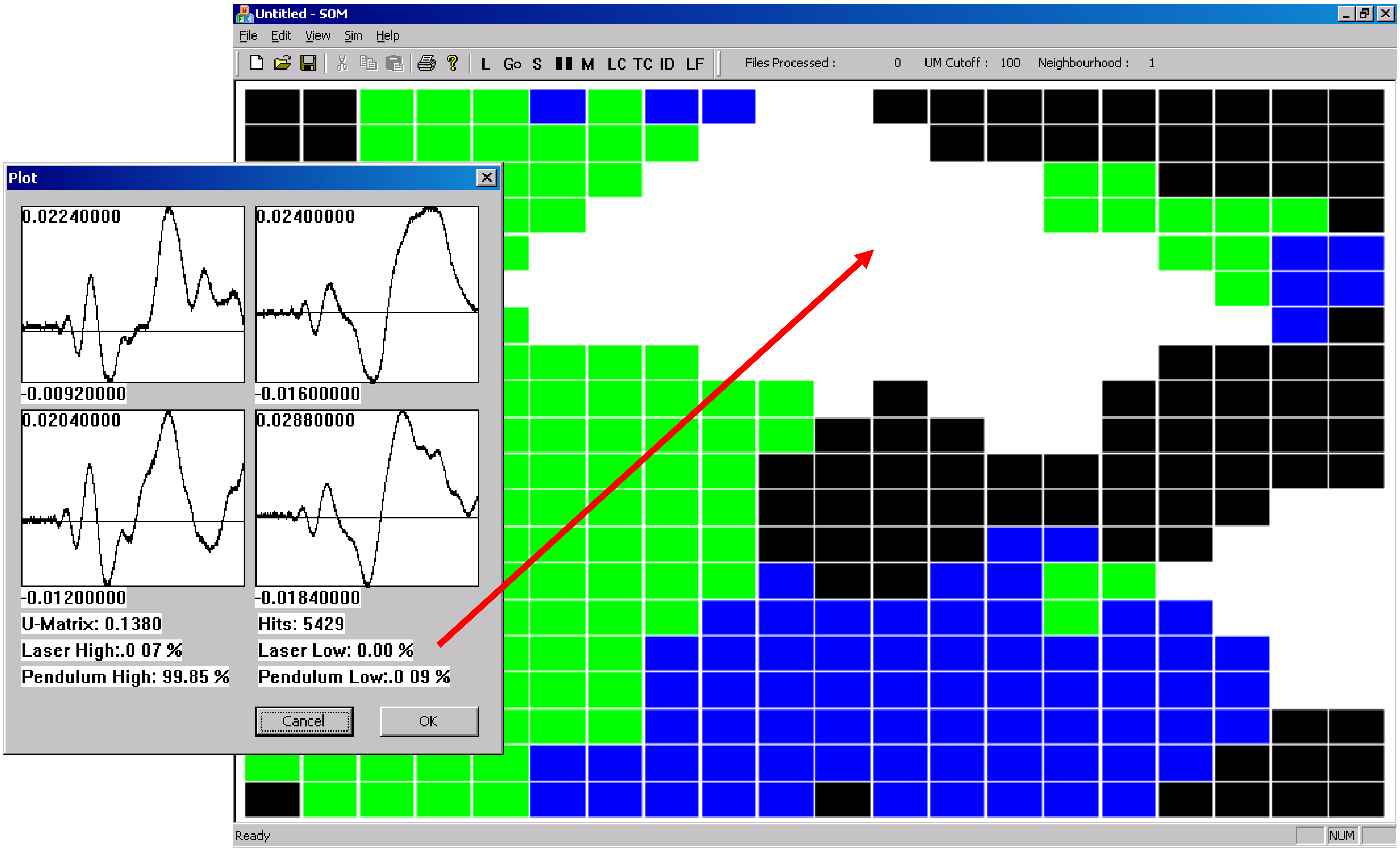
AAV diagnostics: self-organising (Kohonen) maps



AAV diagnostics: self-organising (Kohonen) maps



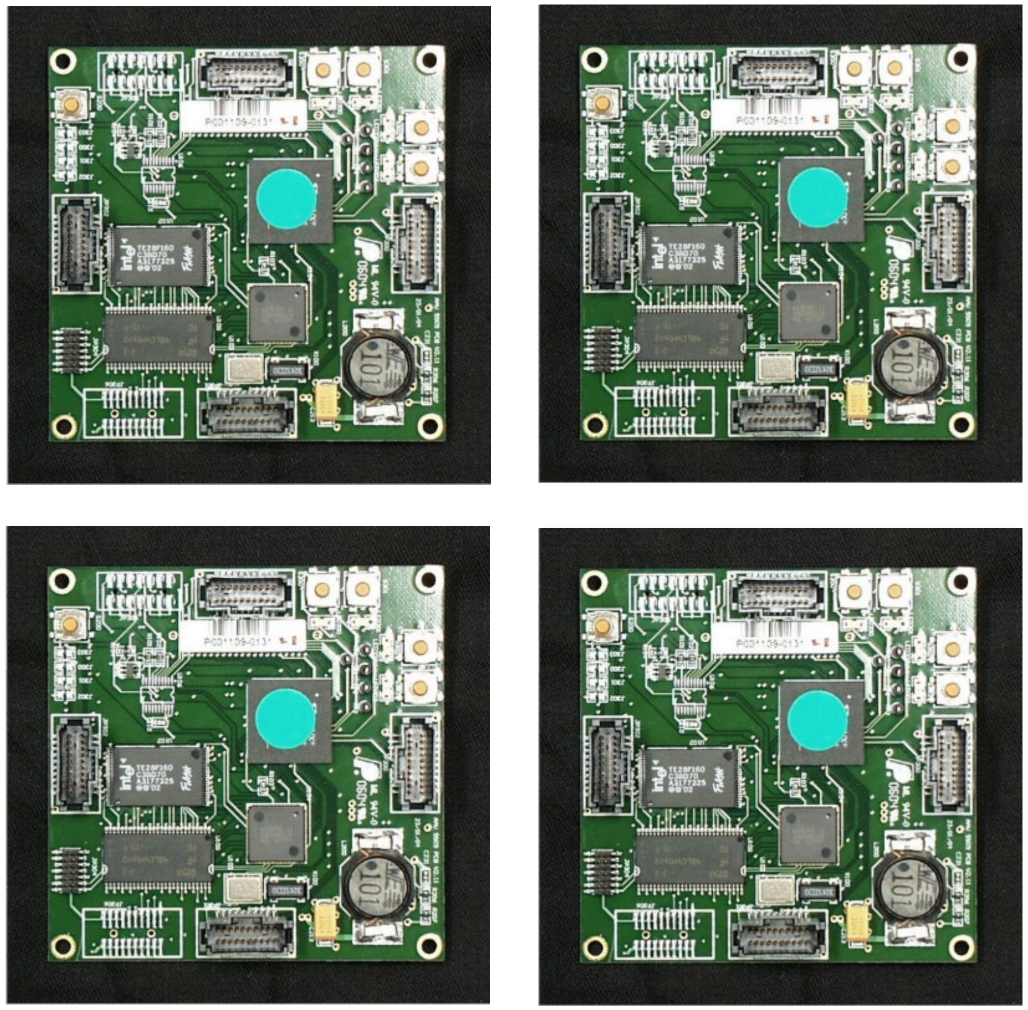
AAV diagnostics: self-organising (Kohonen) maps



Break

?

Design vs Self-organisation



design

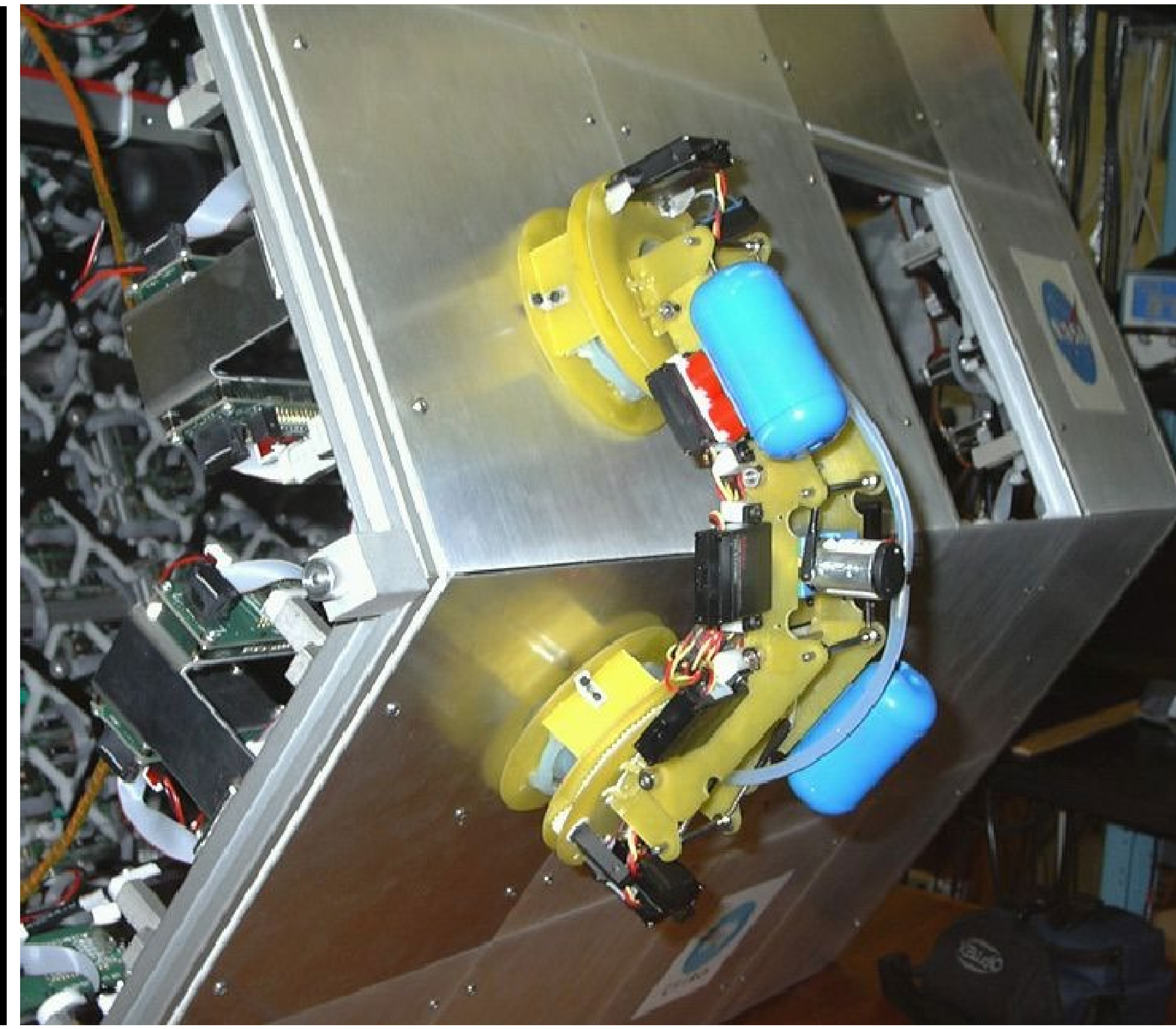
ideally: self-organisation

Autonomous agents
Local interactions

Multi-agent system
Self-organisation

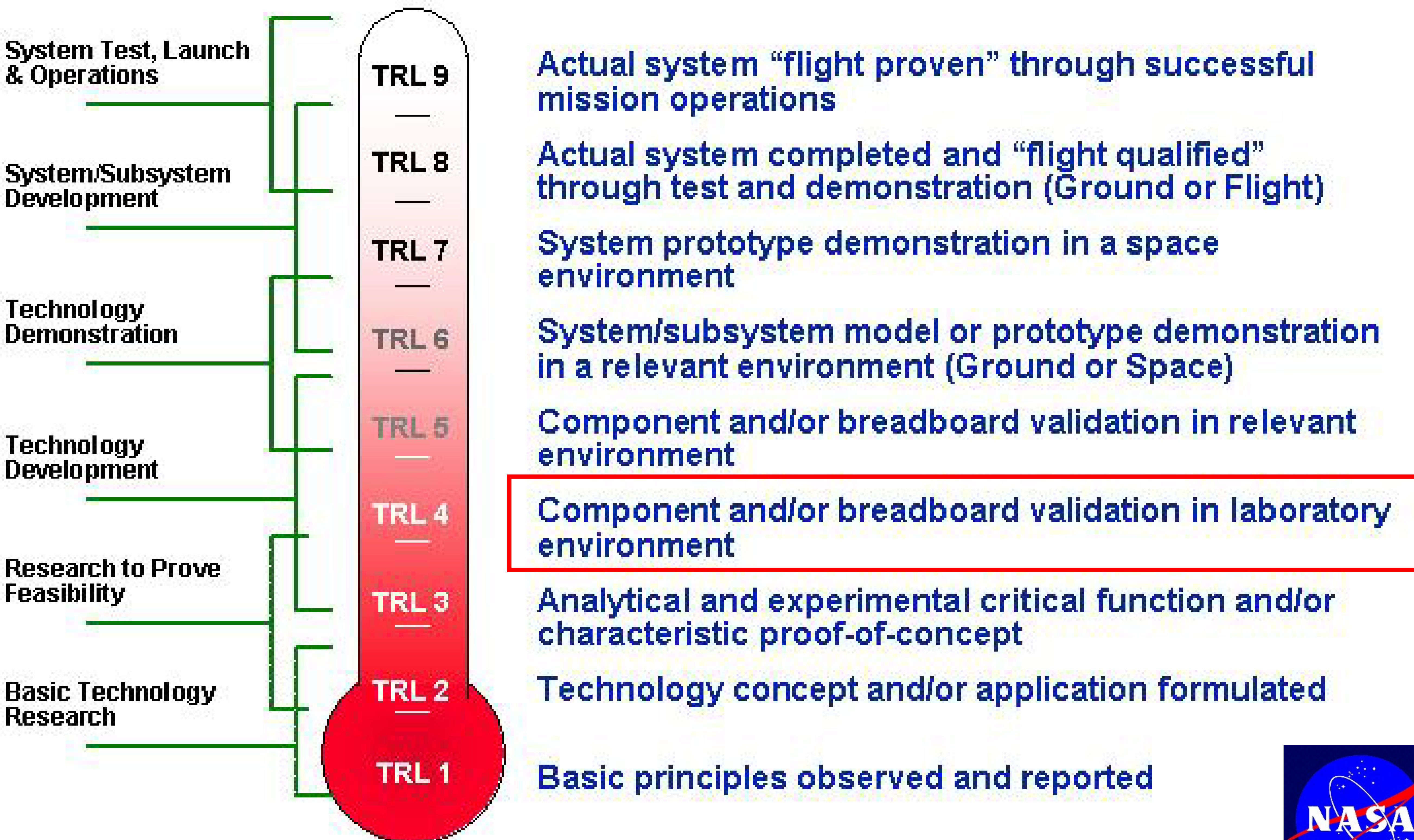
Design vs Self-organisation

Swarm robots in a loop with embedded SHMM networks



AAV project (2005)

Technology Readiness Levels (TRLs)



Benefits of bio-inspired architectures

- efficiency: more flexible structures (example: “smart skin”)
  - robustness to multiple concurrent component (“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

Questions

?

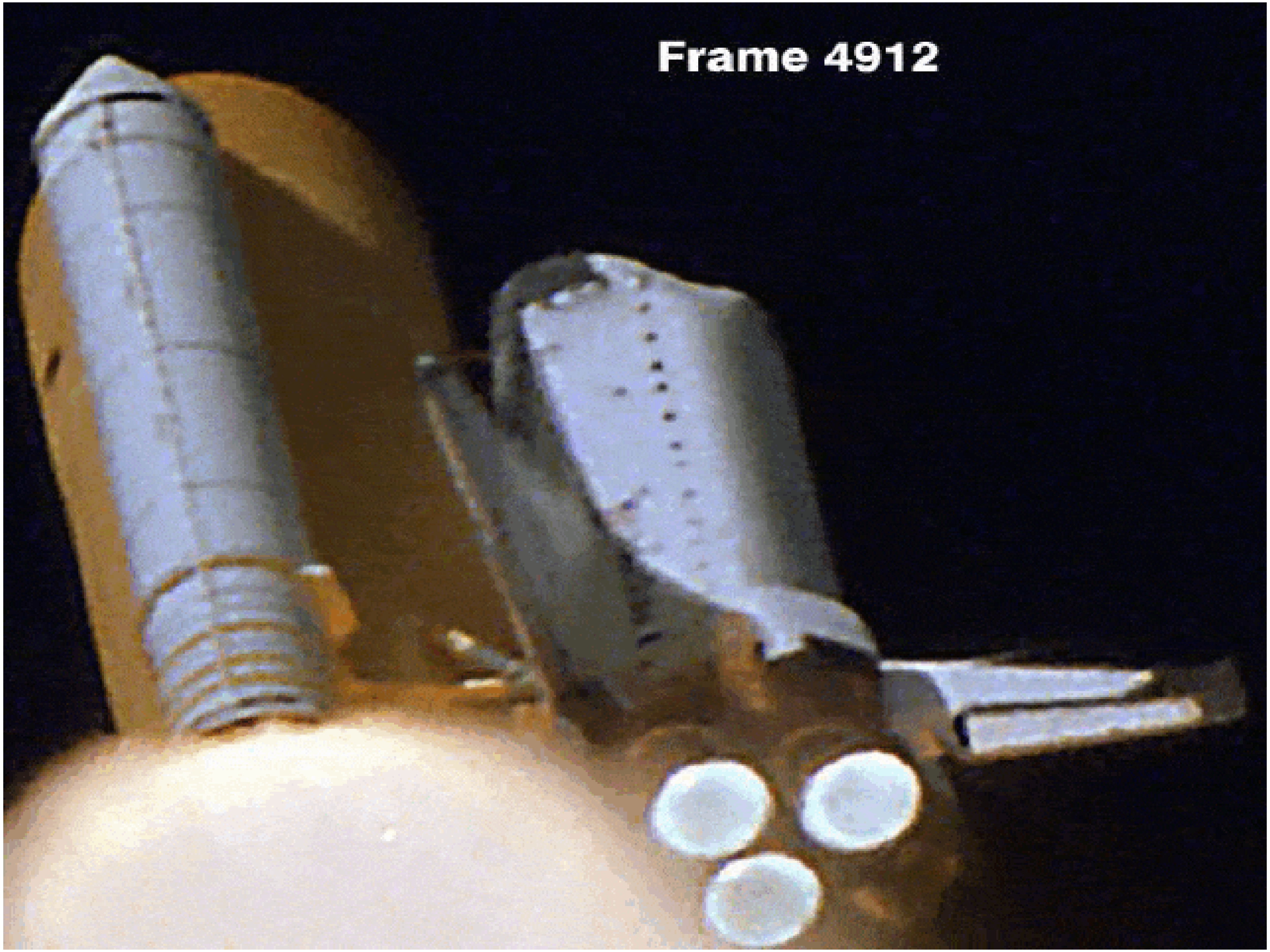
Columbia shuttle accident

https://www.dailymail.co.uk/sciencetech/article-13078549/space-shuttle-Columbia-Disaster-graphic.html

Columbia shuttle accident investigation

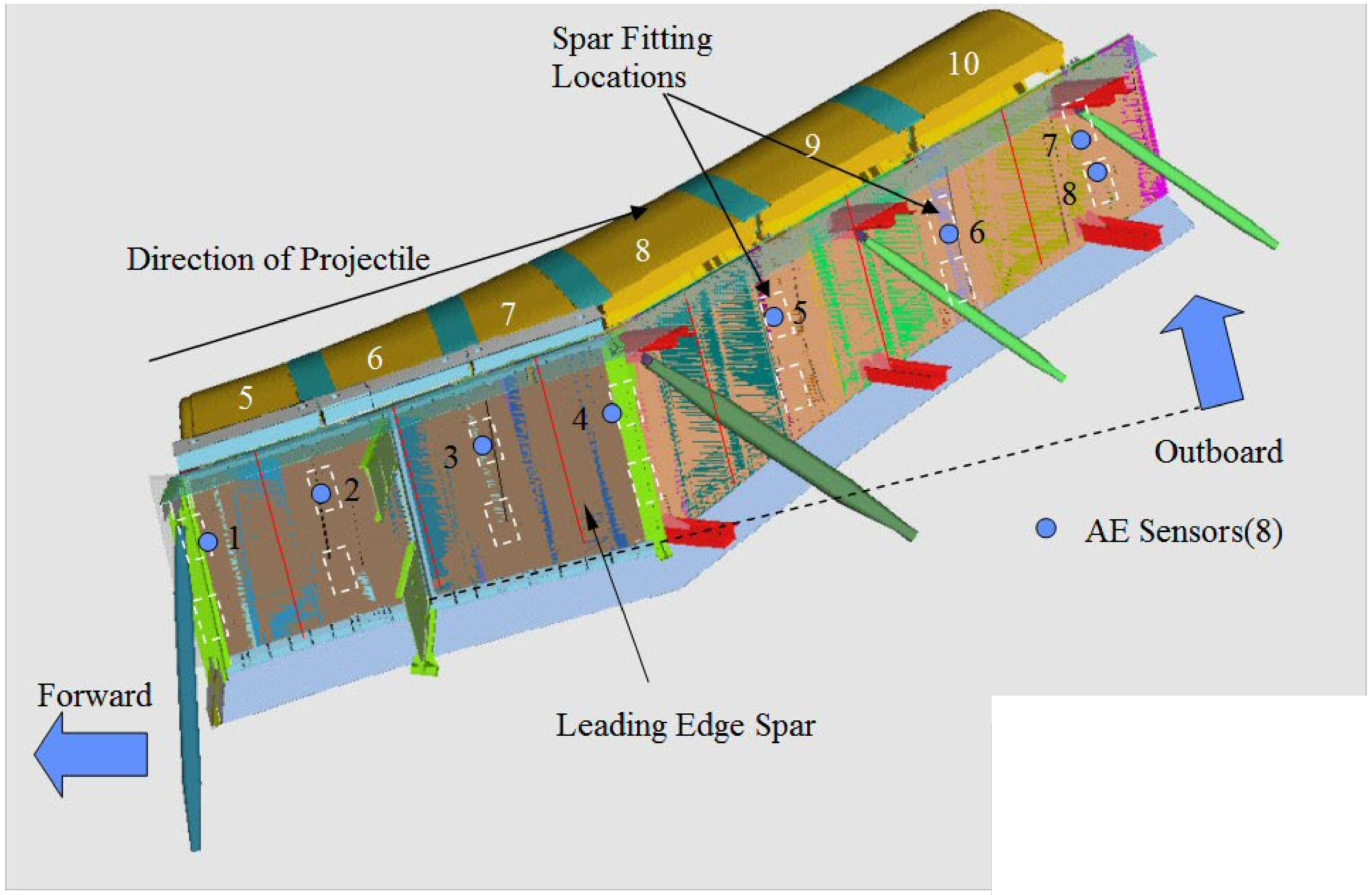
https://en.wikipedia.org/wiki/Space_Shuttle_Columbia_disaster

Columbia shuttle accident investigation



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

Columbia shuttle accident investigation



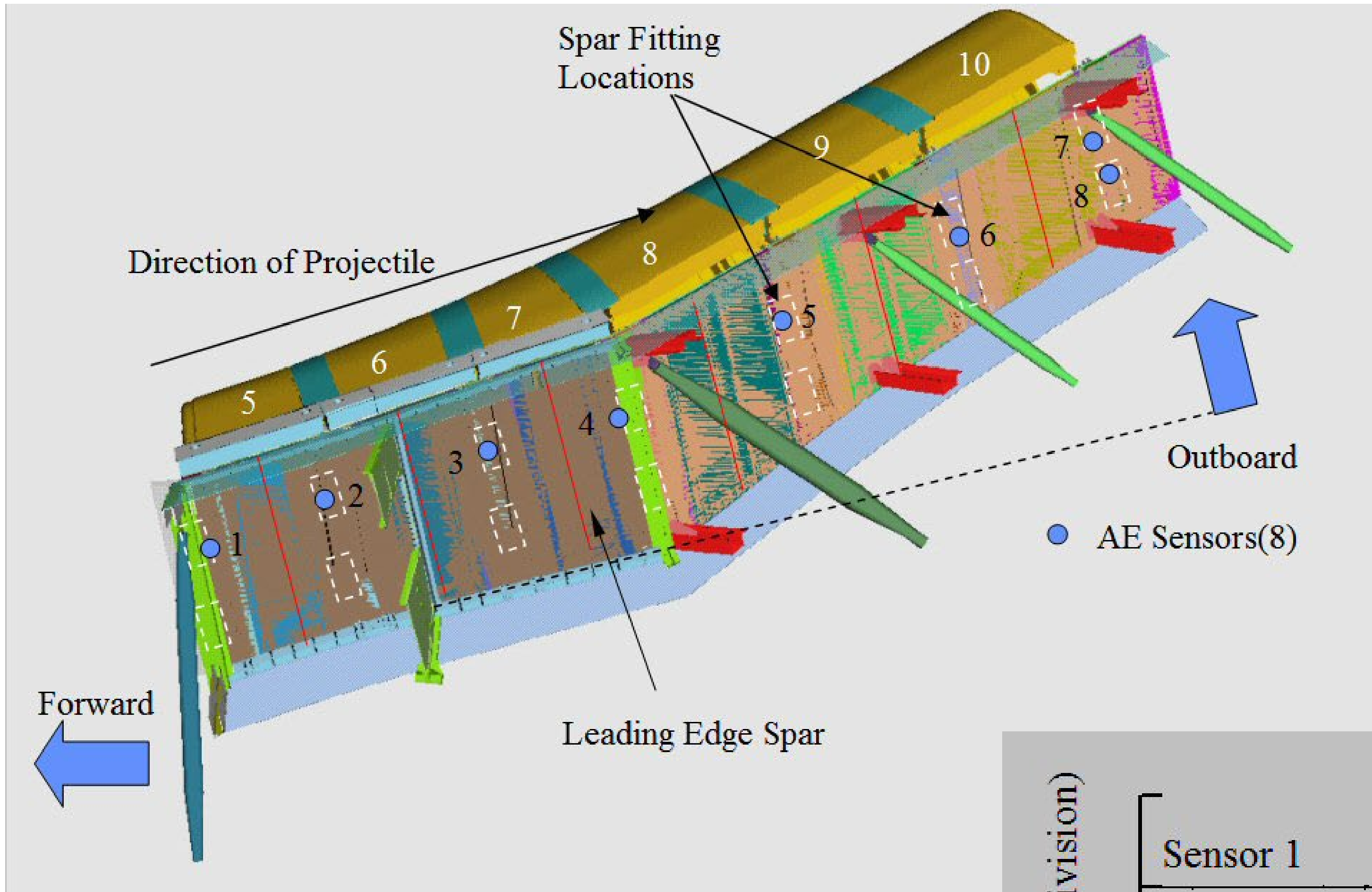
Schematic of Shuttle wing leading edge test article showing RCC panels, location of impact and position of AE sensors

AE signals from foam impact on Shuttle RCC wing leading edge

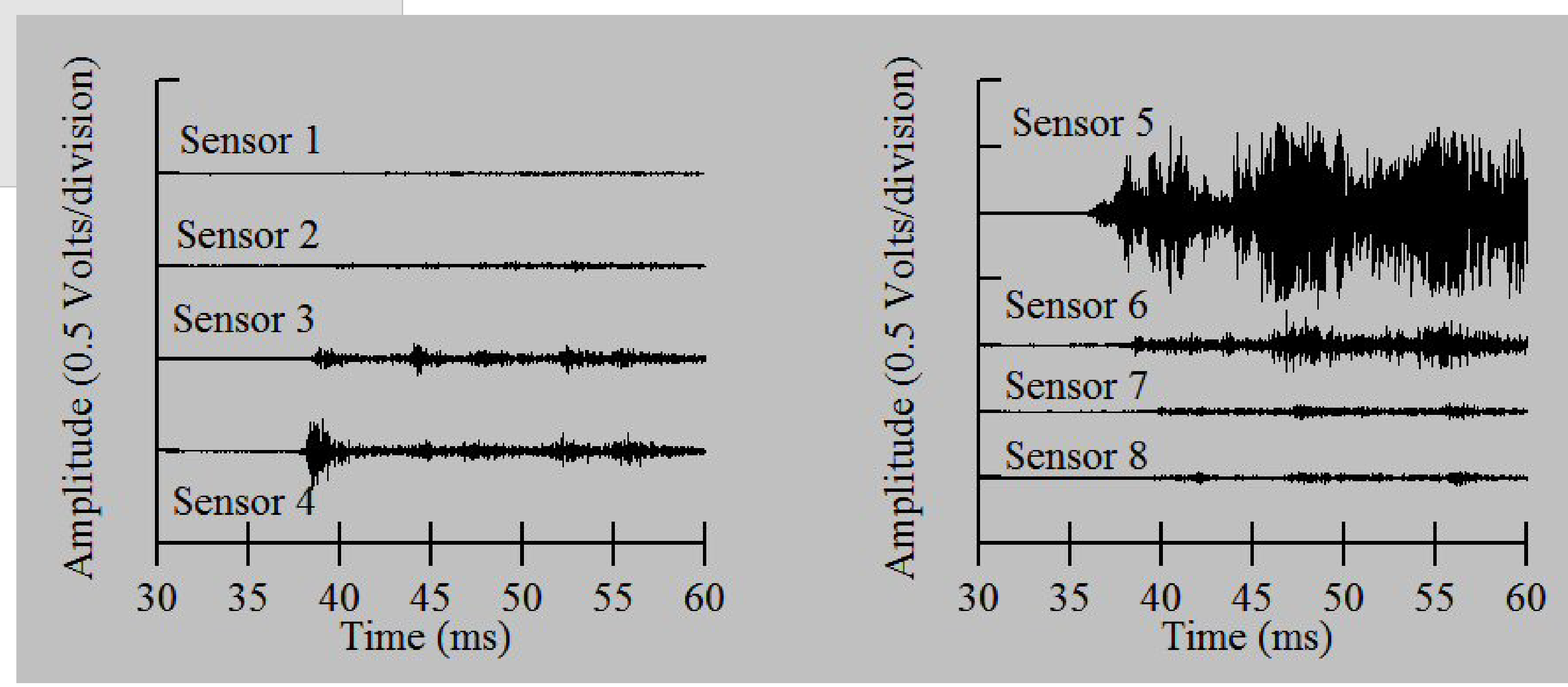
?

W. H. Prosser, S. G. Allison, S. E. Woodard, R. A. Wincheski, E. G. Cooper, D. C. Price, M. Hedley, M. Prokopenko, D. A. Scott, A. Tessler, J. L. Spangler (2004). Structural Health Management for Future Aerospace Vehicles, *The 2nd Australasian Workshop on Structural Health Monitoring (2AWSHM)*, Melbourne, Australia, December 2004.

Columbia shuttle accident investigation



Schematic of Shuttle wing leading edge test article showing RCC panels, location of impact and position of AE sensors



AE signals from foam impact on Shuttle RCC wing leading edge

W. H. Prosser, S. G. Allison, S. E. Woodard, R. A. Wincheski, E. G. Cooper, D. C. Price, M. Hedley, M. Prokopenko, D. A. Scott, A. Tessler, J. L. Spangler (2004). Structural Health Management for Future Aerospace Vehicles, *The 2nd Australasian Workshop on Structural Health Monitoring (2AWSHM)*, Melbourne, Australia, December 2004.

Questions

?

Weak (Narrow) AI

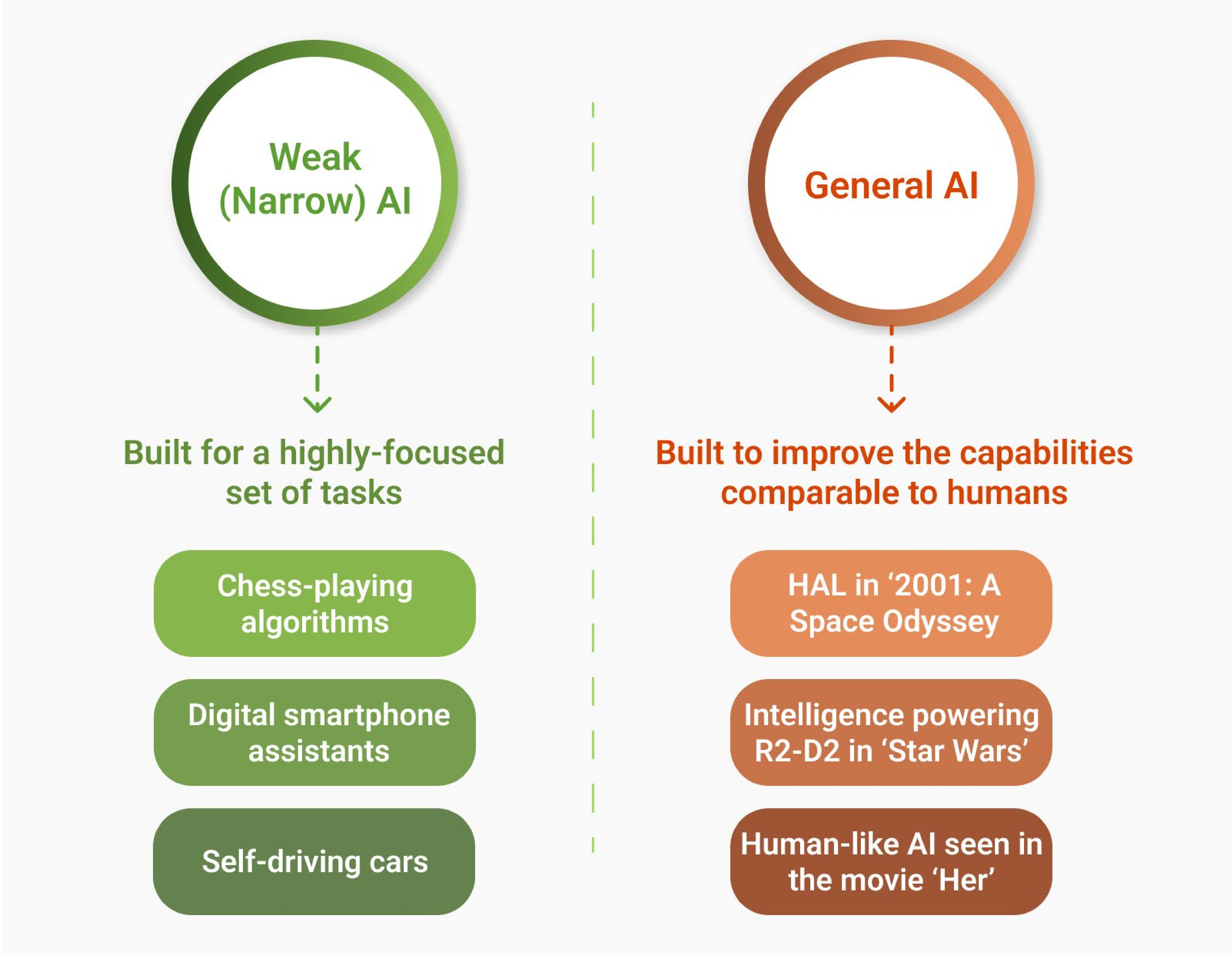
Built for a highly-focused set of tasks

- Chess-playing algorithms
- Digital smartphone assistants
- Self-driving cars

?

Built to improve the capabilities comparable to humans

- HAL in '2001: A Space Odyssey
- Intelligence powering R2-D2 in 'Star Wars'
- Human-like AI seen in the movie 'Her'



https://www.plugger.ai/blog/general-ai-vs-narrow-ai-2022-guide

Narrow AI	Artificial general intelligence (AGI)
AI focused on a specific, singular or limited task	Not fully realized, with some developers questioning if it will be possible
Examples include image recognition, hyperpersonalization, chatbots, predictive text	Seeks machines that can handle a range of cognitive tasks with little oversight
Trained on specific tasks by data scientists	The ability to learn, generalize, apply knowledge and plan for the future
Correlates questions or assignments to a specific data set to accomplish a task	Must consistently pass the Turing Test
No self-awareness, consciousness, ability to think	Single, general intelligence that possesses common sense and creativity and expresses emotions

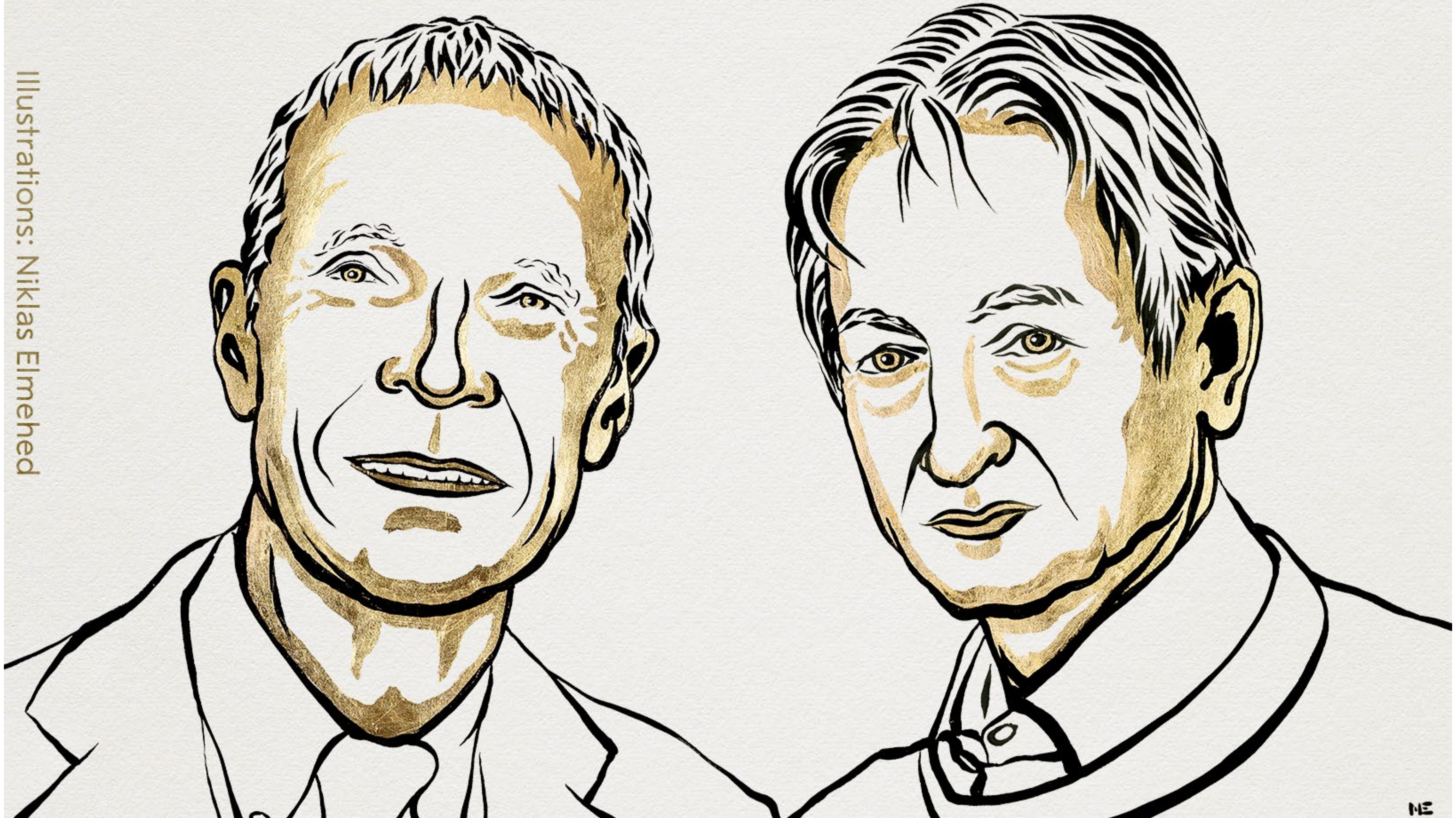
https://www.techtarget.com/searchenterpriseai/feature/General-AI-vs-narrow-AI-comes-down-to-adaptability

Questions

?

THE NOBEL PRIZE IN PHYSICS 2024

Illustrations: Niklas Elmehed



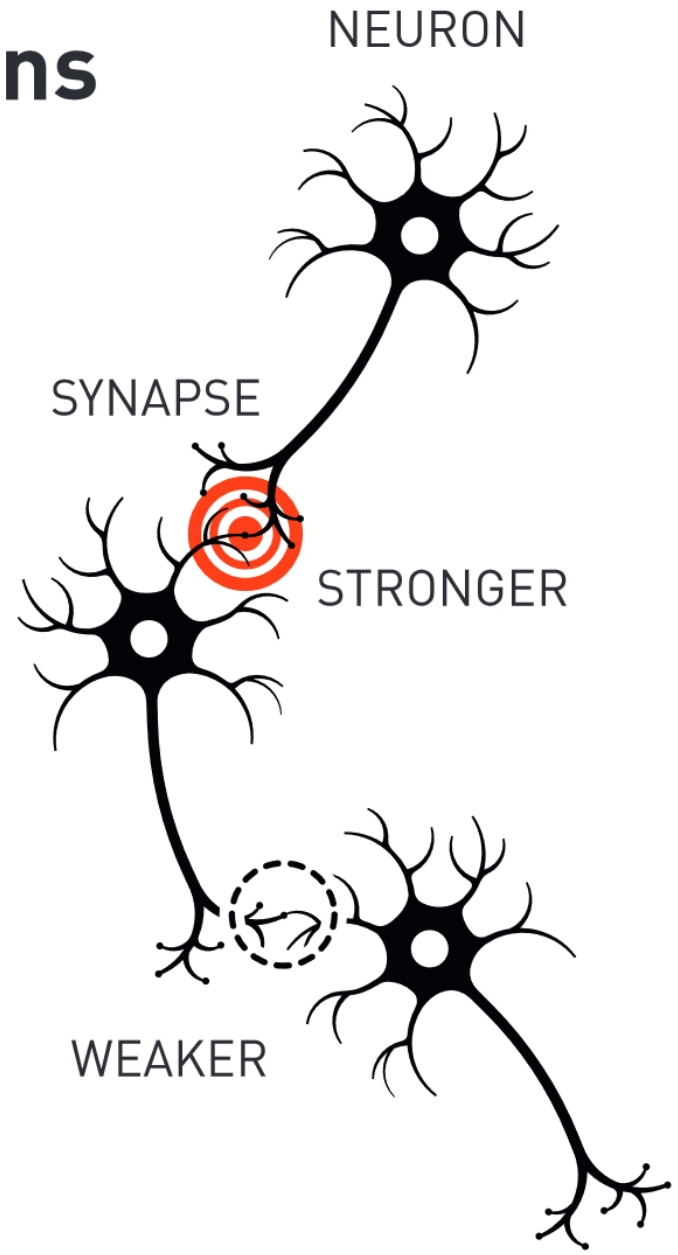
John J. Hopfield   Geoffrey E. Hinton

"for foundational discoveries and inventions
that enable machine learning
with artificial neural networks"

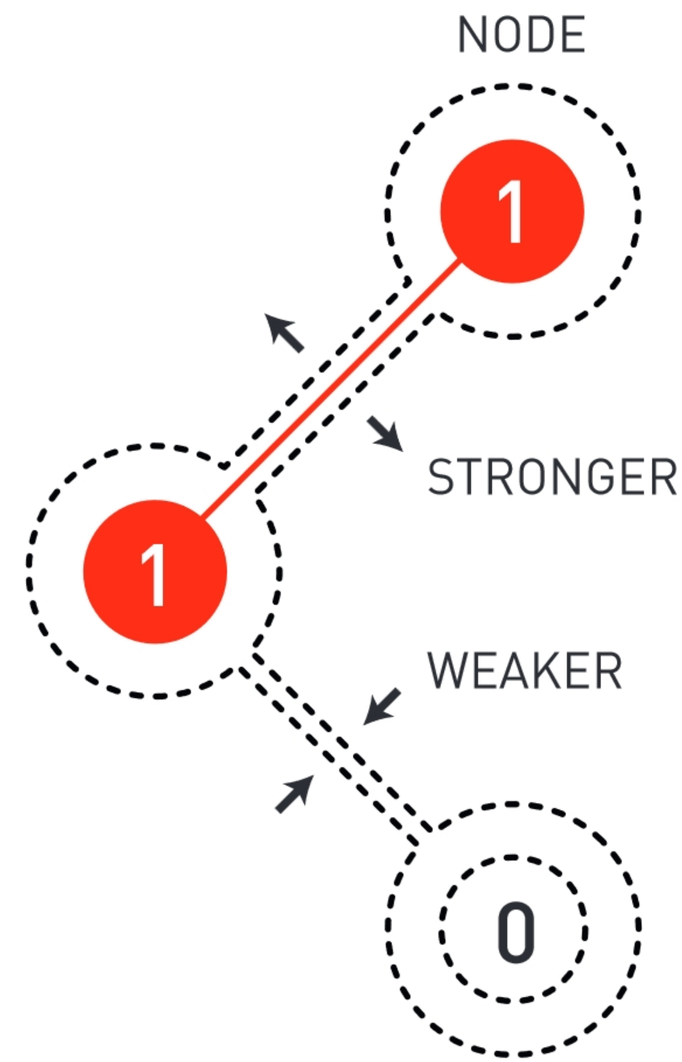
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Natural and artificial neurons

The brain's neural network is built from living cells, neurons, with advanced internal machinery. They can send signals to each other through the synapses. When we learn things, the connections between some neurons gets stronger, while others get weaker.

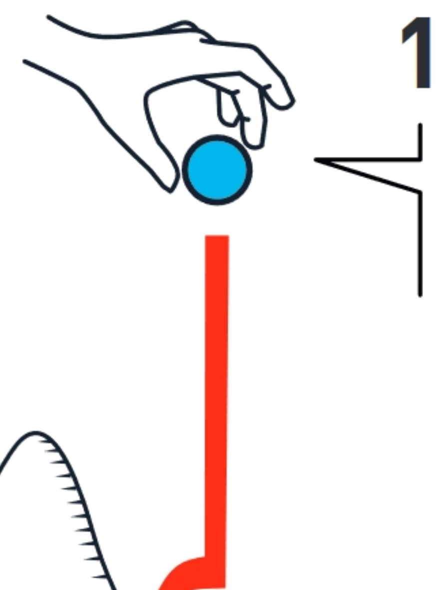


Artificial neural networks are built from nodes that are coded with a value. The nodes are connected to each other and, when the network is trained, the connections between nodes that are active at the same time get stronger, otherwi-se they get weaker.

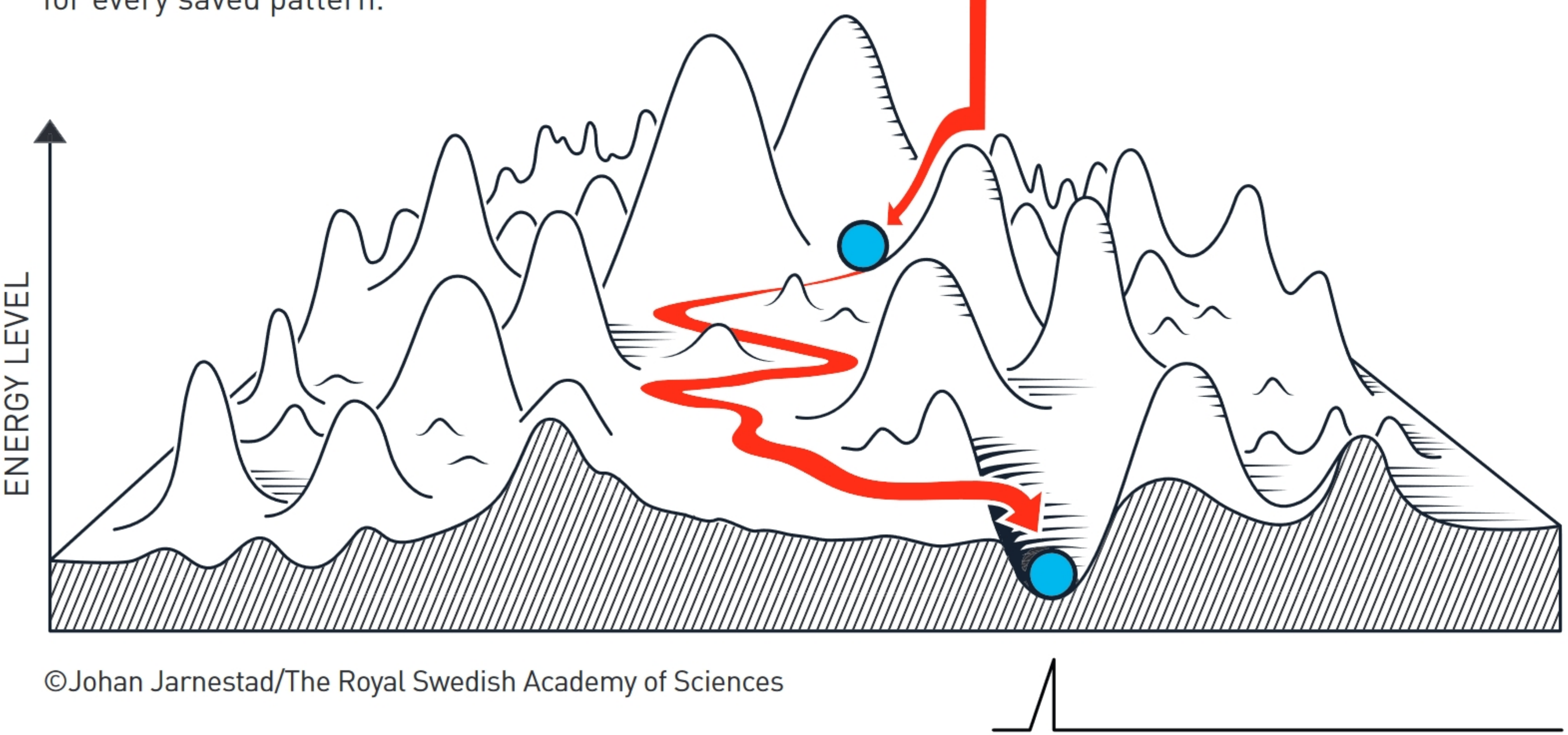


Memories are stored in a landscape

John Hopfield's associative memory stores information in a manner similar to shaping a landscape. When the network is trained, it creates a valley in a virtual energy landscape for every saved pattern.

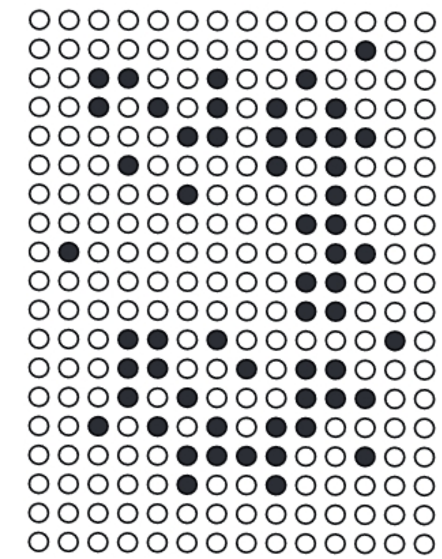


1 When the trained network is fed with a distorted or incomplete pattern, it can be likened to dropping a ball down a slope in this landscape.

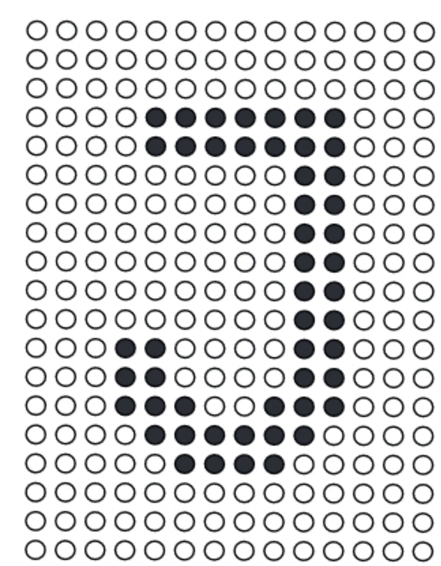


2 The ball rolls until it reaches a place where it is surrounded by uphills. In the same way, the network makes its way towards lower energy and finds the closest saved pattern.

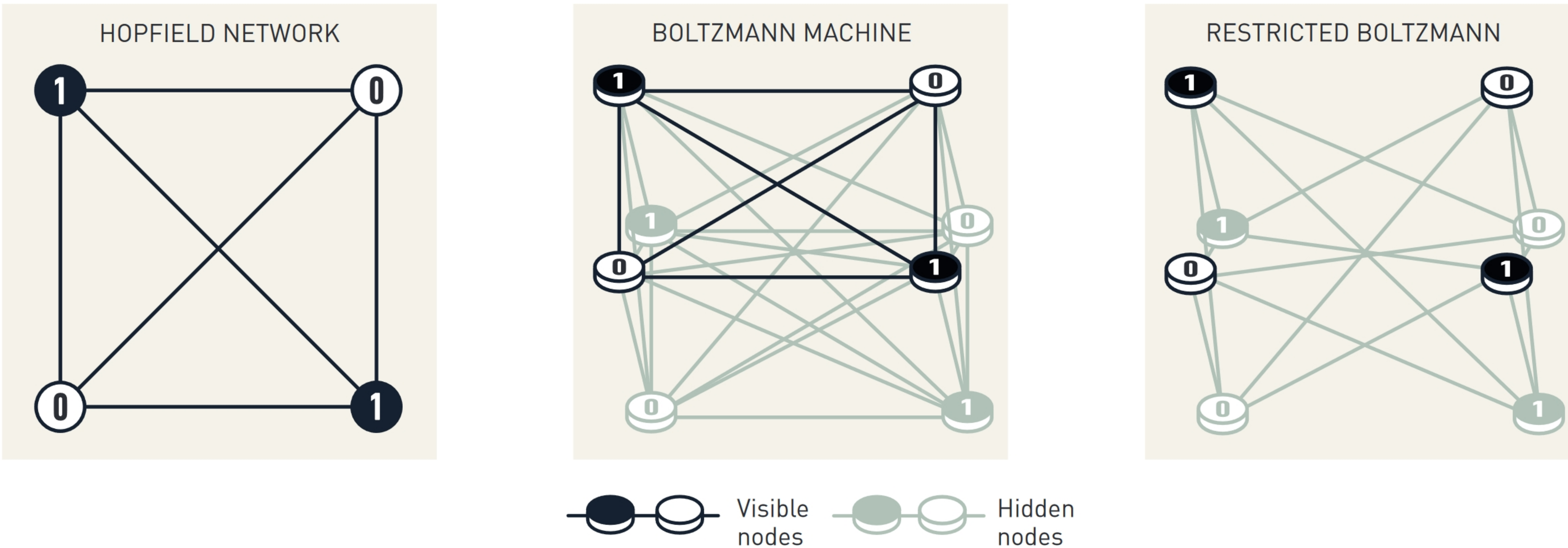
INPUT PATTERN



SAVED PATTERN



Different types of network



John Hopfield's associative memory is built so that all the nodes are connected to each other. Information is fed in and read out from all the nodes.

Geoffrey Hinton's Boltzmann machine is often constructed in two layers, where information is fed in and read out using a layer of *visible* nodes. They are connected to *hidden* nodes, which affect how the network functions in its entirety.

In a restricted Boltzmann machine, there are no connections between nodes in the same layer. The machines are frequently used in a chain, one after the other. After training the first restricted Boltzmann machine, the content of the hidden nodes is used to train the next machine, and so on.

THE NOBEL PRIZE IN CHEMISTRY 2024

Illustrations: Niklas Elmehed

David Baker

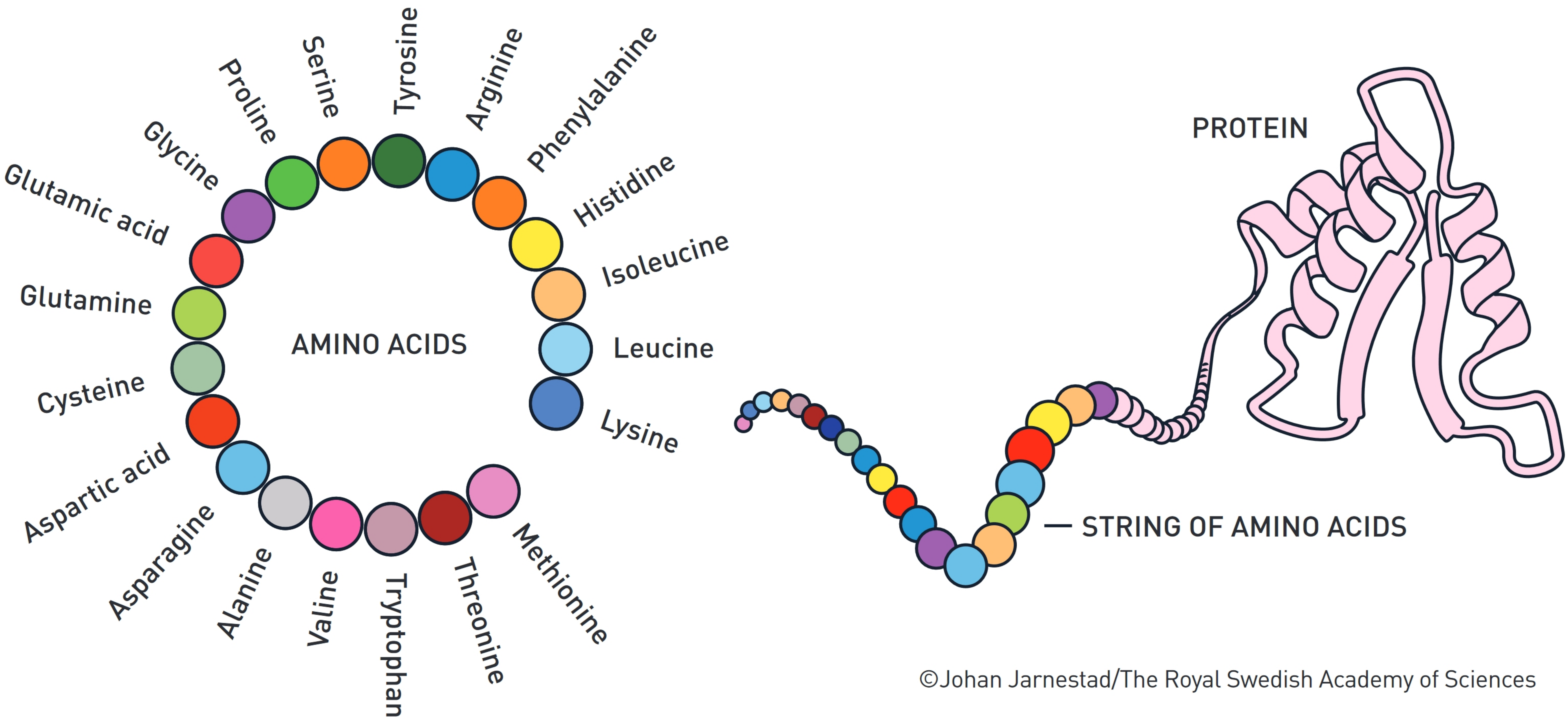
"for computational protein design"

Demis Hassabis

John M. Jumper

"for protein structure prediction"

THE ROYAL SWEDISH ACADEMY OF SCIENCES



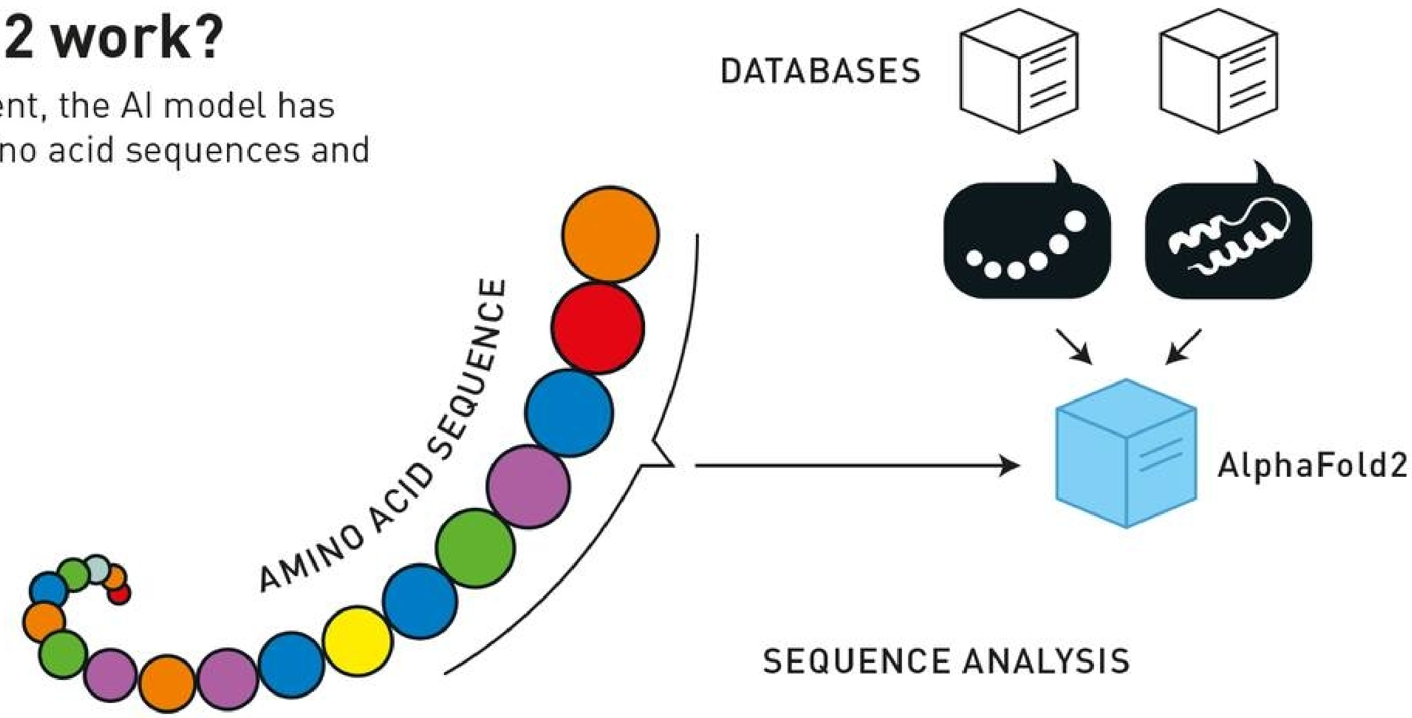
AMINO ACIDS
Tyrosine
Arginine
Phenylalanine
Histidine
Isoleucine
Leucine
Lysine
Methionine
Threonine
Tryptophan
Valine
Alanine
Asparagine
Aspartic acid
Cysteine
Glutamine
Glutamic acid
Glycine
Proline
Serine
PROTEIN
STRING OF AMINO ACIDS

How does AlphaFold2 work?

As part of AlphaFold2's development, the AI model has been trained on all the known amino acid sequences and determined protein structures.

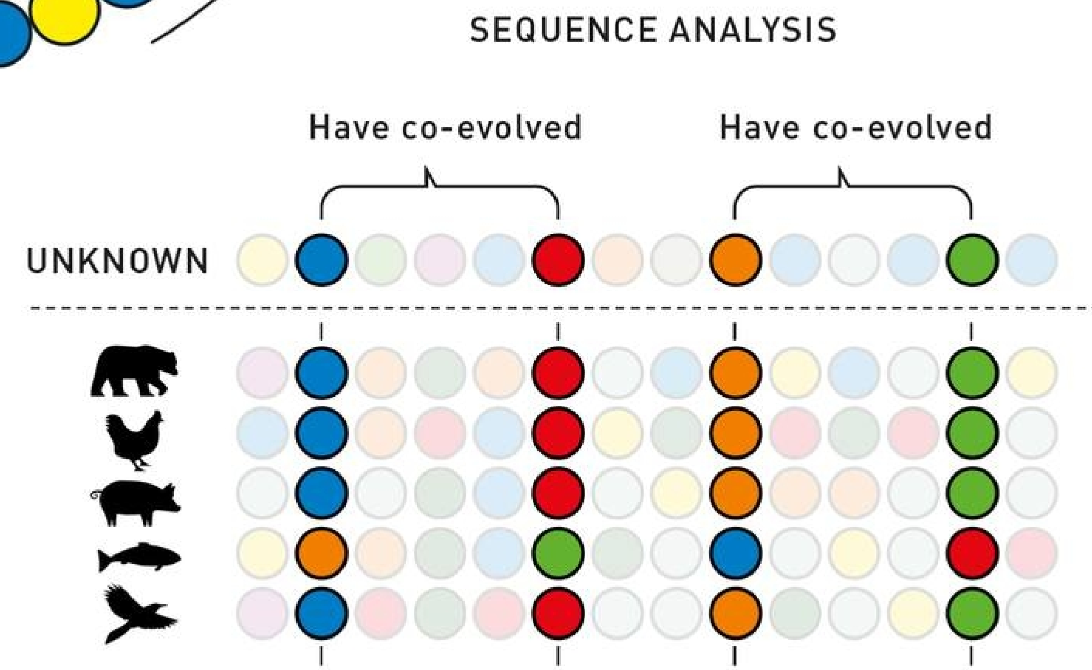
1. DATA ENTRY AND DATABASE SEARCHES

An amino acid sequence with unknown structure is fed into AlphaFold2, which searches databases for similar amino acid sequences and protein structures.

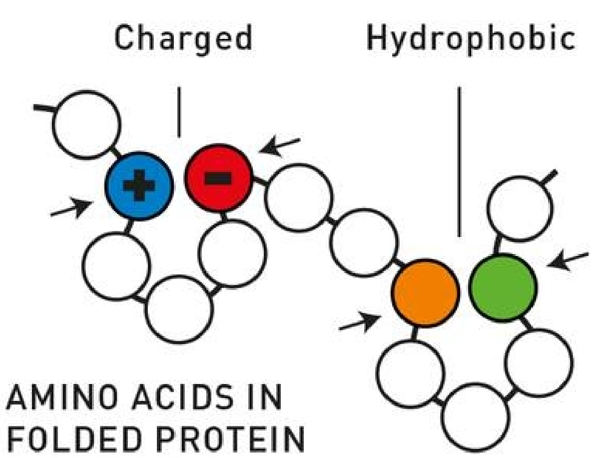


2. SEQUENCE ANALYSIS

The AI model aligns all the similar amino acid sequences – often from different species – and investigates which parts have been preserved during evolution.

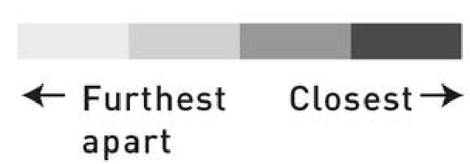


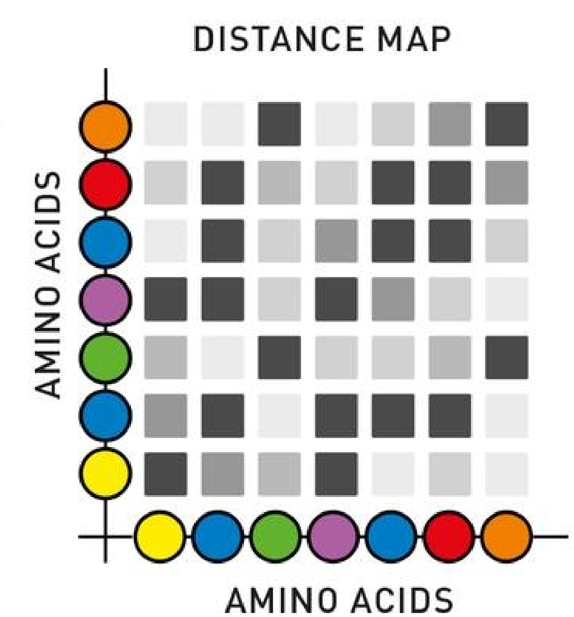
In the next step, AlphaFold2 explores which amino acids could interact with each other in the three-dimensional protein structure. Interacting amino acids co-evolve. If one is charged, the other has the opposite charge, so they are attracted to each other. If one is replaced by a water-repellent (hydrophobic) amino acid, the other also becomes hydrophobic.



AMINO ACIDS IN FOLDED PROTEIN STRUCTURE

Using this analysis, AlphaFold2 produces a distance map that estimates how close amino acids are to each other in the structure.



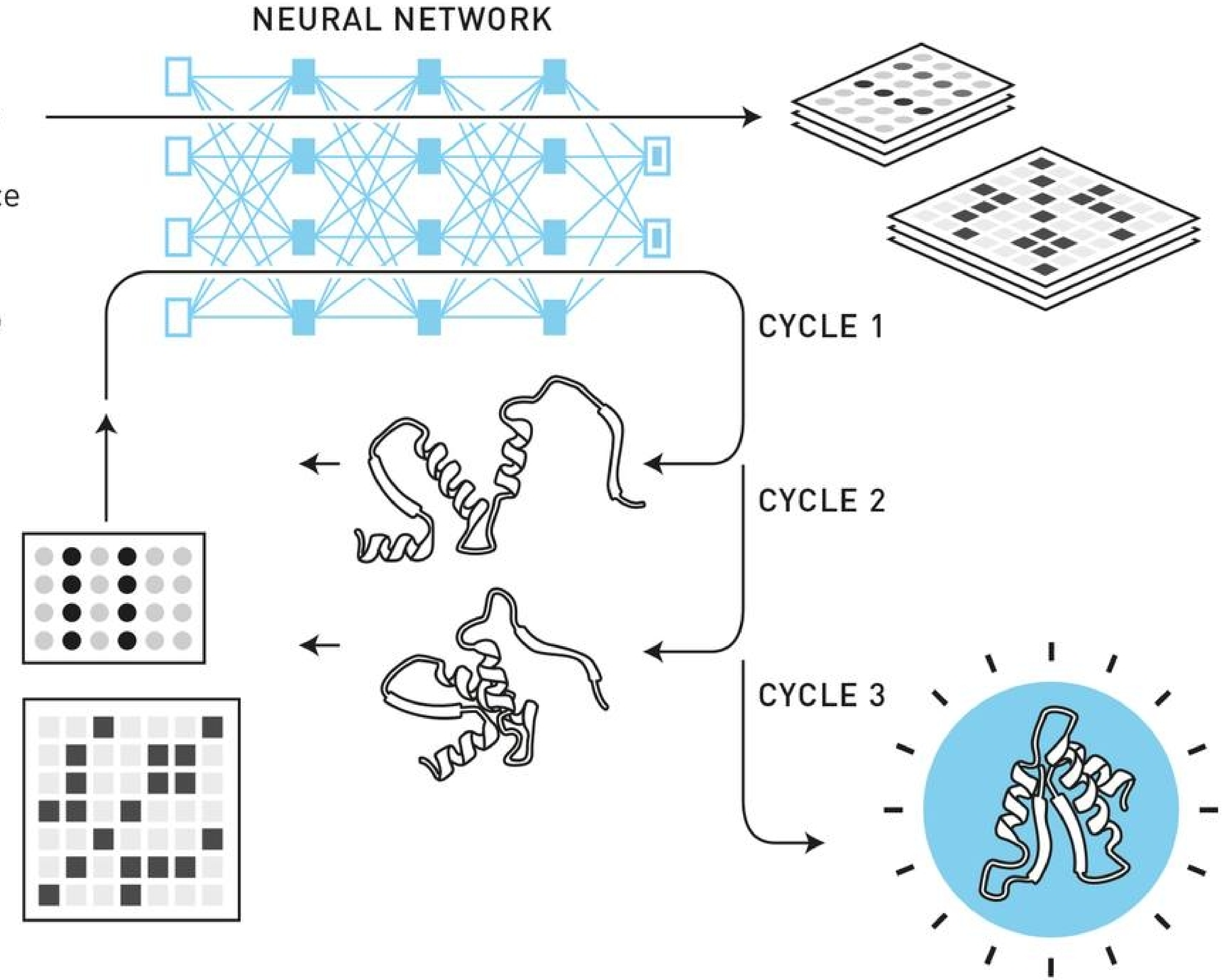


3. AI ANALYSIS

Using an iterative process, AlphaFold2 refines the sequence analysis and distance map. The AI model uses neural networks called transformers, which have a great capacity to identify important elements to focus on. Data about other protein structures – if they were found in step 1 – is also utilised.

4. HYPOTHETICAL STRUCTURE

AlphaFold2 puts together a puzzle of all the amino acids and tests pathways to produce a hypothetical protein structure. This is re-run through step 3. After three cycles, AlphaFold2 arrives at a particular structure. The AI model calculates the probability that different parts of this structure correspond to reality.



Questions

?

Assignments (example)

There are 3 assignment activities for this subject:

(1) Individual Report (30%)	– week 7	55/100 → 16.5 / 30
(2) Group Presentation (40%)	– week 11	64/100 → 25.6 / 40
(3) Individual Report (30%)	– week 14	78/100 → 23.4 / 30
Total (100%)		65.5 ≈ **66**

❖ Text matching software (Turnitin) will be used for plagiarism detection in (1) and (3)

Assignment #2

Four cases

1. IBM Watson

Watson famously won the quiz show Jeopardy! in 2011, showcasing its ability to understand natural language questions and provide accurate answers, even when the questions are ambiguous or pun-based.

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

2. iTutor Group's recruiting AI

In August 2023, tutoring company iTutor Group agreed to pay $365,000 to settle a suit brought by the US Equal Employment Opportunity Commission. The federal agency said the company, which provides remote tutoring services to students in China, used AI-powered recruiting software that automatically rejected female applicants ages 55 and older, and male applicants ages 60 and older.

https://www.gtlaw.com/en/insights/2023/8/eeoc-secures-first-workplace-artificial-intelligence-settlement

Four cases

3. Google's DeepMind partnership with Moorfields Eye Hospital

In 2018, an AI system developed by Google's DeepMind partnership with Moorfields Eye Hospital to assist in the diagnosis and treatment of eye diseases outperformed human experts in diagnosing eye diseases from retinal scans, showing promising potential for improving healthcare outcomes.

https://deepmind.google/discover/blog/a-major-milestone-for-the-treatment-of-eye-disease/
https://deepmind.google/discover/blog/using-ai-to-predict-retinal-disease-progression/

4. Tesla Autopilot Accidents

While Tesla's Autopilot feature is an advanced driver-assistance system that utilises AI, there have been instances where it failed to detect obstacles leading to accidents.

https://www.wired.com/story/tesla-autopilot-risky-deaths-crashes-nhtsa-investigation/

Assignment #2 (group presentation)

- introduction with an overview of key issues of all four studied cases
- comparative analysis of
  - reasons for their success and/or failures
  - their main novelties and gaps across 4 cases, highlighting common reasons for success or failure
  - their societal impacts
  - their limitations, biases and risks
- discussion and conclusion

Assignment #2 (group presentation)

Prepare a presentation video (10 minutes), upload it to a platform of your choice, and send an email with a video link to mikhail.prokopenko@sydney.edu.au

Upload presentation slides to Canvas (with your Group ID in the file name)

Deadline for slides upload: **23 October, Thursday, 11:59 pm (week 11)**

Q&A: 24 October (in class): 8 minutes Q&A for each group

There will be 20 groups in total: 16 groups with 5 students, and 4 groups with 4 students

Marking group presentation (case studies)

	Weight	0-49%	50-64%	65-74%	75-84%	85-100%	Mark
Structure and organisation	22.5%	No clear or coherent structure or organisation	Poor or incomplete structure and organisation	Structure and organisation are partially appropriate	Structure and organisation are appropriate	Structure and organisation are appropriate and effective	9
Topic development	37.5%	Development of content limited; may be incomplete or unclear. Inadequate supporting information, illustration or explanations. Societal impact and limitations described inadequately or mis-identified	Development of content adequate but lacks clearly stated propositions, argument or supporting information. Societal impact and limitations described incompletely. Some use of data, graphical or other visual material to support content but meaning and rationale for use may be questionable.	Clear and complete development of content and support for propositions and arguments. Societal impact and limitations described, but without sufficient detail, structure or argumentation. Content partially supported with relevant and well explained data, graphical or other visual material.	Full development of content and support for propositions and arguments, with purpose, relevance, focus, explanations. Societal impact and limitations described in sufficient detail, but without strong argumentation. Content supported with some data, graphical or other visual material.	Full and rich development of content and support for propositions and arguments, with purpose, relevance, focus, explanations. Societal impact and limitations described adequately, with sufficient detail, structure and argumentation. Content supported effectively with relevant and well explained data, graphical or other visual material.	15
Presentation (slides)	15%	Lack of fluency in expression makes it difficult to comprehend	Lack of fluency in expression but still relatively easy to follow	Good level of fluency and clarity in expression and evidence of logical progression of ideas	High level of fluency and clarity in expression and evidence of logical progression of ideas	Sophistication in fluency of expression	6
Q&A	25%	As above	As above	As above	As above	As above	10
Total	100%						40

In accordance with University policy, penalties apply when work is submitted after 11:59 pm on the due date (Sydney time):

- deduction of 5% of the maximum mark for each calendar day after the due date
- after ten calendar days late, a mark of zero will be awarded
- text matching software (Turnitin) will be used for plagiarism detection

Questions

?