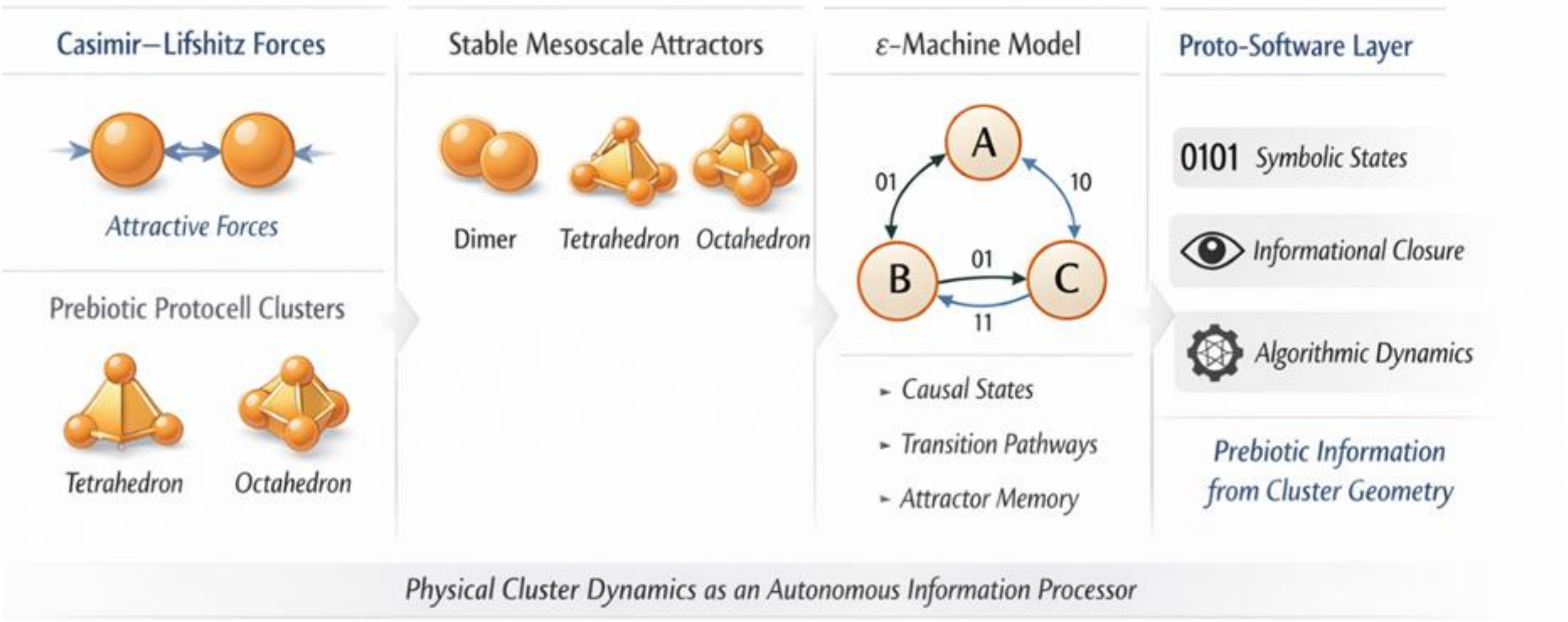
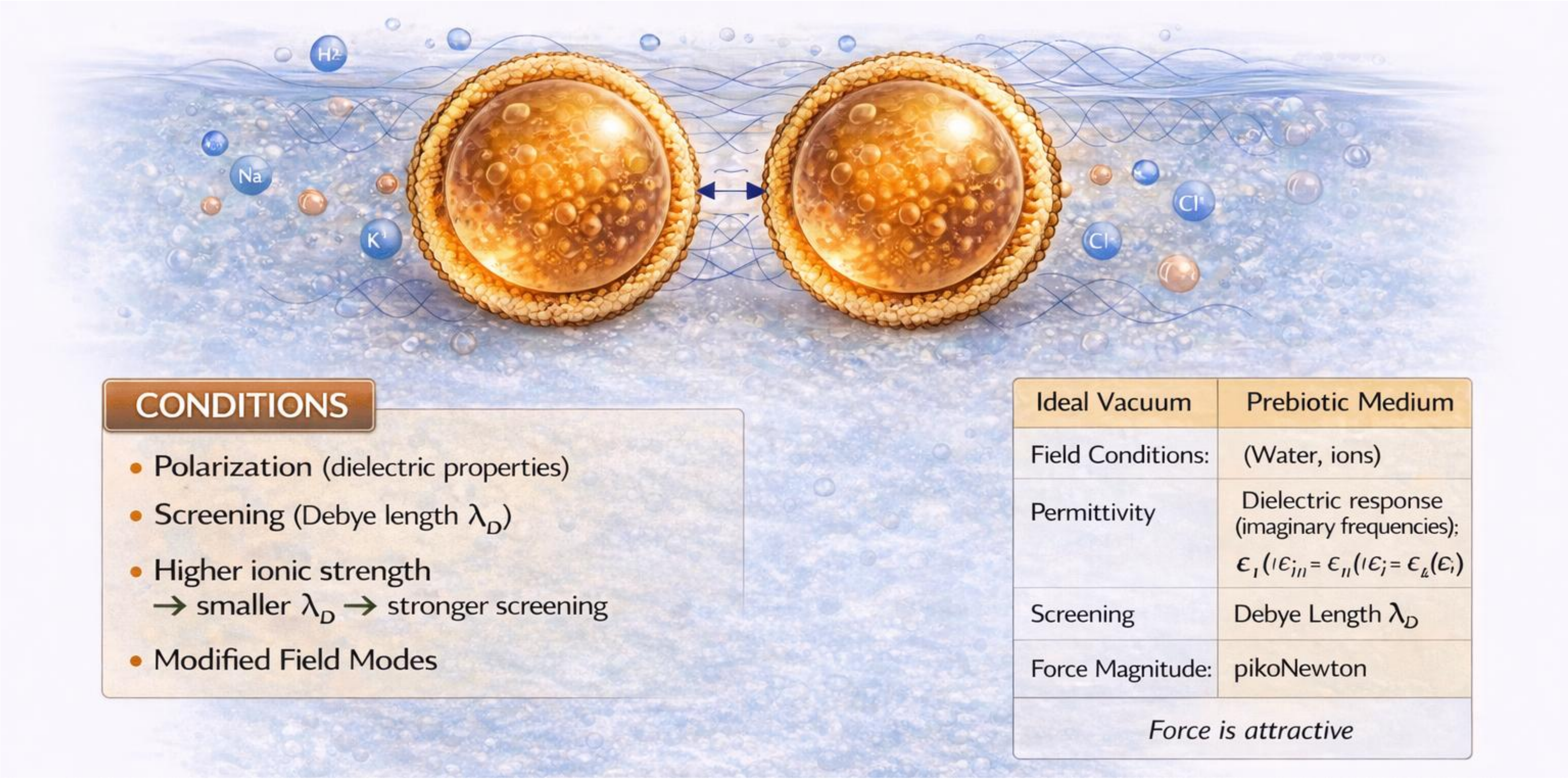
Emergent Information Formation in Prebiotic Protocell Clusters: A Computational Mechanics Framework of ε-Machines and Attractor Memory



Michael Massoth
Hochschule Darmstadt (h_da)
BIOTECHNO-2026 | Valencia

Casimir-Lifshitz Attraction in primordial soup



Ideal Vacuum	Prebiotic Medium
Field Conditions:	(Water, ions)
Permittivity	Dielectric response (imaginary frequencies): $\epsilon_I(i\epsilon_{in} = \epsilon_{II}(i\epsilon_i = \epsilon_L(\epsilon_i)$
Screening	Debye Length λ_D
Force Magnitude:	pikoNewton
Force is attractive	

Department of Computer Science

Why this matters: From Physics to Proto-Software

Question: How did **information show up** in a world without genes?

The standard narrative starts with polymers—RNA-like molecules and coding.

My proposal is simpler at first:
information can arise from stable, distinguishable macrostates and their reliable transition structure—
purely physical, before sequence-based encoding.

Research Question and Scientific Motivation

One of the central questions in origin-of-life research is:

How could information arise before genes, enzymes, or replication systems existed?

Classical models focus on:

- RNA-world scenarios
- polymer-based information storage
- chemical replication mechanisms

This work explores an alternative hypothesis:

Information may emerge from physical self-organization at the mesoscale.

Department of Computer Science

Computational Mechanics Framework by Rosas et al.

Computational mechanics models dynamical systems using **ε-machines**.

An **ε-machine** is a **minimal predictive state machine** that:

- identifies relevant system states
- represents transitions between them
- captures the predictive structure of the dynamics

The **internal states** are called **causal states**.

Rosas et al.: Software in the natural world (1)

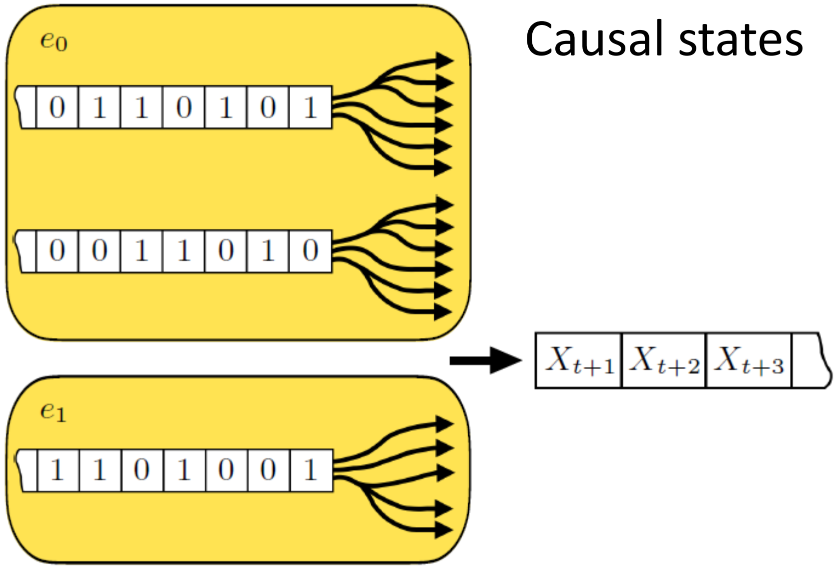
Software in the natural world: A computational approach to hierarchical emergence

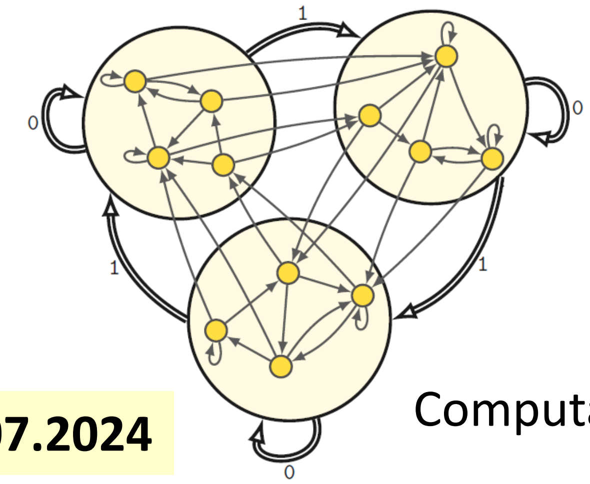
Fernando E. Rosas,[1,2,3,4,*] Bernhard C. Geiger,[5,6] Andrea I Luppi,[7]
Anil K. Seth,[1,2] Daniel Polani,[8] Michael Gastpar,[9] and Pedro A.M. Mediano[10,11]

[1] *Department of Informatics, University of Sussex*
[2] *Sussex Centre for Consciousness Science and Sussex AI, University of Sussex*
[3] *Center for Psychedelic Research and Centre for Complexity Science, Department of Brain Science, Imperial College London*
[4] *Center for Eudaimonia and Human Flourishing, University of Oxford*
[5] *Know-Center GmbH, Graz, Austria*
[6] *Signal Processing and Speech Communication Laboratory, Graz University of Technology, Graz, Austria*
[7] *Montreal Neurological Institute, McGill University*
[8] *Department of Computer Science, University of Hertfordshire, Hatfield, UK*
[9] *School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland*
[10] *Department of Computing, Imperial College London*
[11] *Division of Psychology and Language Sciences, University College London*

Understanding the functional architecture of complex systems is crucial to illuminate their inner workings and enable effective methods for their prediction and control. Recent advances have introduced tools to characterise emergent macroscopic levels; however, while these approaches are successful in identifying *when* emergence takes place, they are limited in the extent they can determine *how* it does. Here we address this important limitation by developing a computational approach to emergence, which characterises macroscopic processes in terms of their computational capabilities. Concretely, we articulate a view on emergence based on how software works, which is rooted on a mathematical formalisation of how macroscopic processes can express self-contained informational, interventional, and computational properties. This framework reveals a hierarchy of nested self-contained processes that determines what computations take place at what level, which in turn delineates the functional architecture of a complex system. This approach is illustrated on paradigmatic models from the statistical physics and computational neuroscience literature, which are shown to exhibit macroscopic processes that are akin to software in human-engineered systems. Overall, this framework enables a deeper understanding of the multi-level structure of complex systems, revealing specific ways in which they can be efficiently simulated, predicted, and controlled.

Causal states





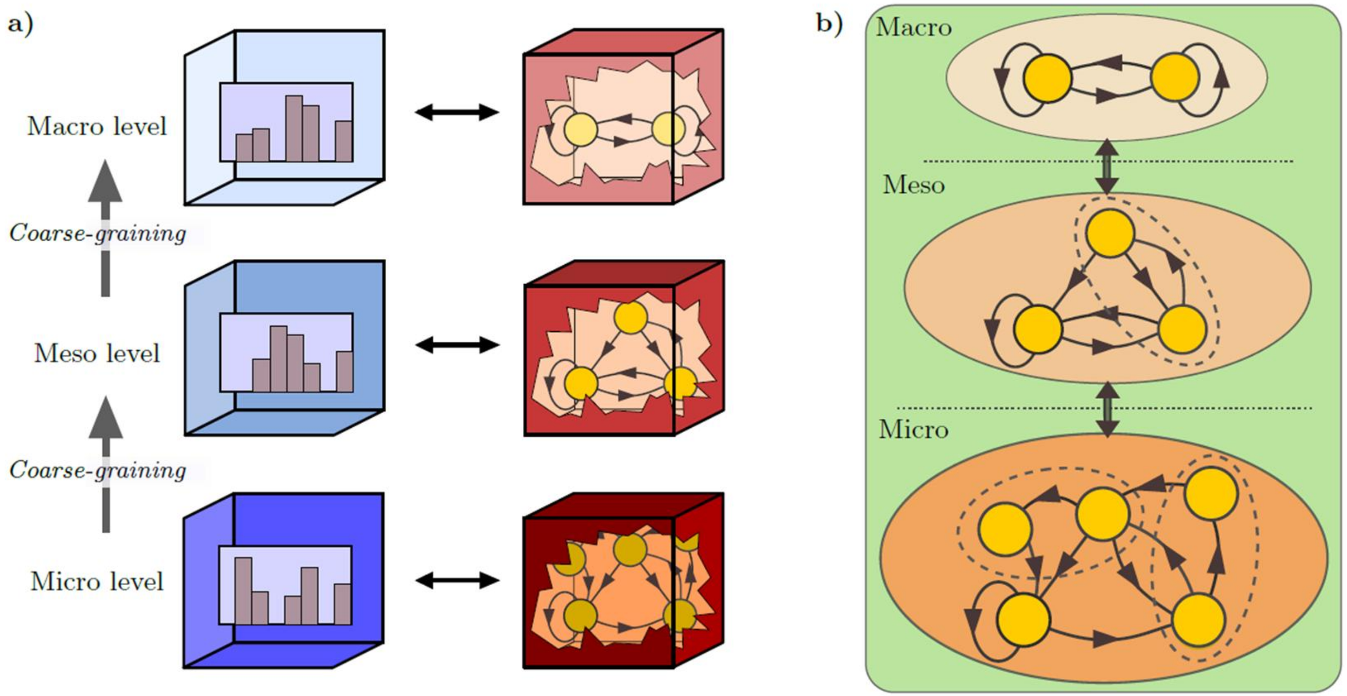
05.07.2024

Computational closure

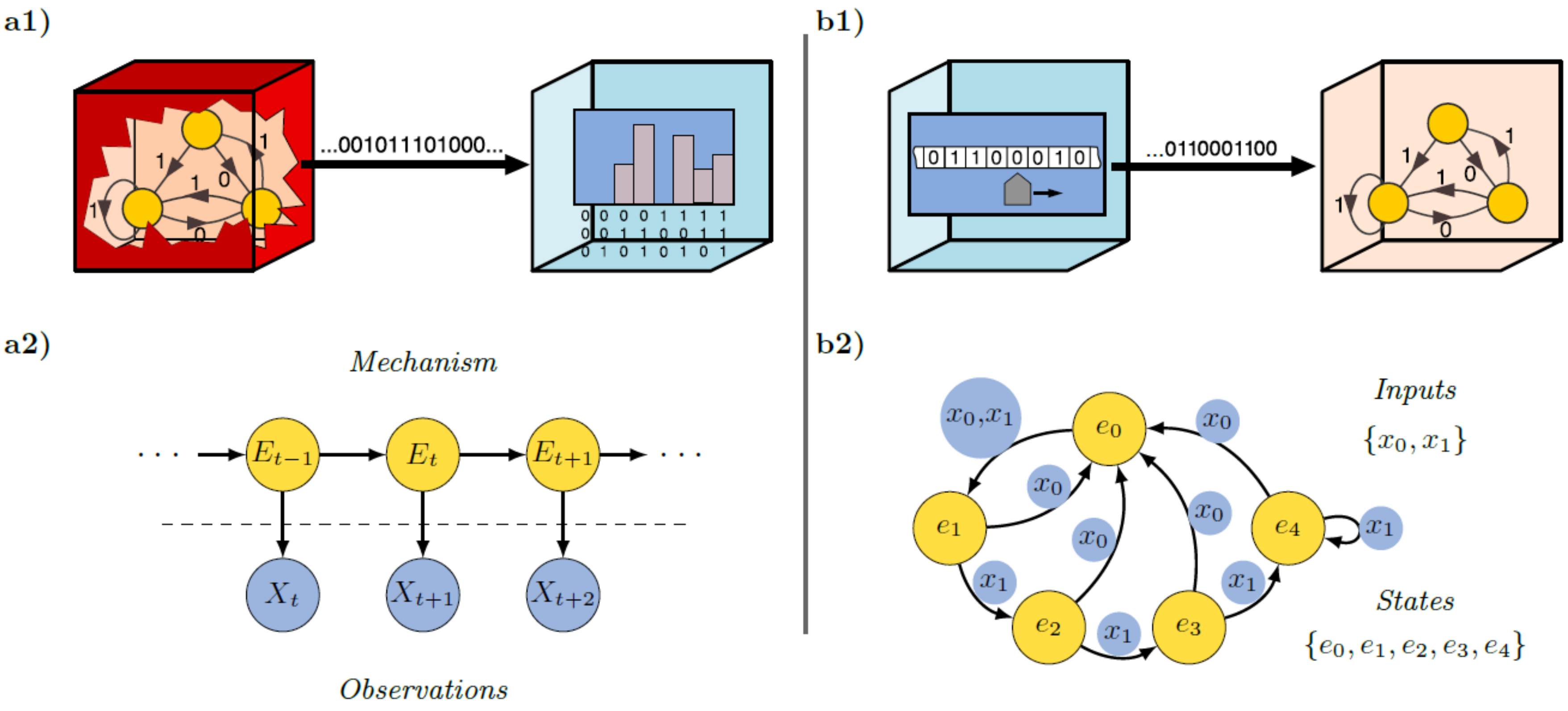
Rosas et al.: Software in the natural world (2)

TABLE I. Summary of main concepts

Term	Explanation
Micro level	Basic time-series
Macro level	A coarse-graining of a micro process
ϵ-machine	Model of a process built on the history at the same level
υ-machine	Model of a process built on the history at a level below
Causal states	Hidden states of ϵ- or υ-machines
Information closure	When best predictions of macro can be done from the same macro
Causal closure	When ϵ- and υ-machines coincide
Computational closure	When ϵ-machine of macro is a coarse-graining of ϵ-machine of micro
Lumpability	The ability of a Markov process to remain Markovian after being coarse-grained



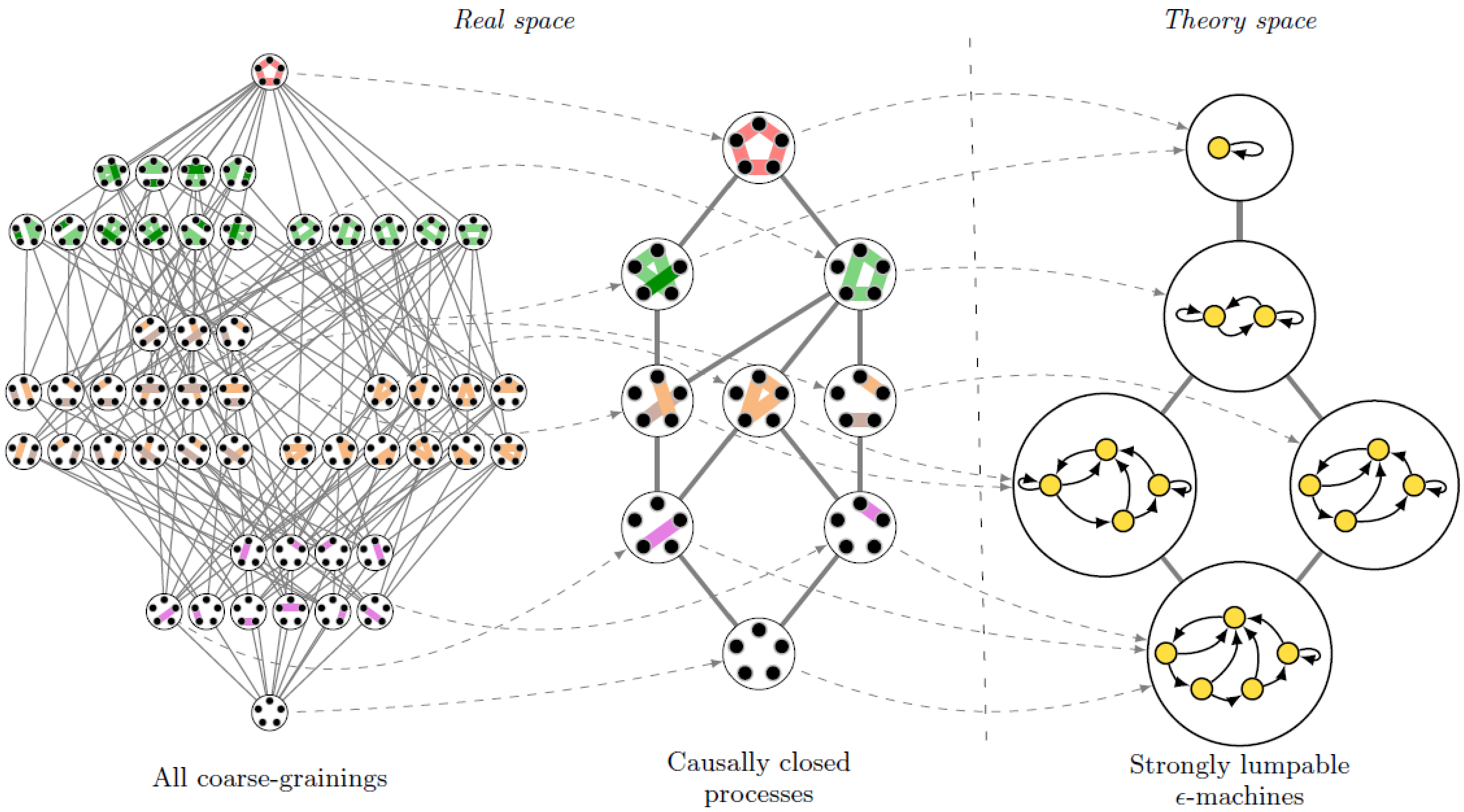
Rosas et al.: Software in the natural world (3)



(**a1**). Technically a hidden Markov process; dynamics that take place on variables Et on a latent state-space,
(**a2**) while generating the observable data Xt.
(**b1**) **ϵ-machines** as discrete automata, where the data corresponds to inputs given by a user driving the system between different states .
(**b2**) Technically a discrete automata with states ek, whose deterministic transitions are governed by the input data xi
Note that (**a1**) focuses on variables (e.g. Xt,Et), while (**b2**) portraits the states that those variables can take (e.g. x0, e0).

The two faces of ϵ-machines: Illustration of the dual interpretation of ϵ-machines that establish a bridge between causality and computation. a) **Causal face** and b) **Computational face**.

Rosas et al.: Software in the natural world (4)



Left : Lattice of all possible coarse-grainings, here illustrated for the case of a process that can take five possible values.
Center: Sub-lattice of only those coarse-grainings that are causally/informationally closed.
Right: Lattice of strongly-lumpable coarse-grainings of the ϵ-machine of the microscopic level.
Only the last lattice provides a minimal blueprint that highlights the distinct computational processes, and distinguishes which computations take place at what level.

The multiple hiearchies describing multi-level computations in a complex system.

05.07.2024

Physical Interaction Mechanism

Casimir–Lifshitz forces arise from **quantum** and **thermal fluctuations** of the electromagnetic field (QED) between material surfaces.

Key properties:

- operate between dielectric materials in aqueous environments
- effective at nanometer to sub-micrometer distances
- independent of molecular recognition

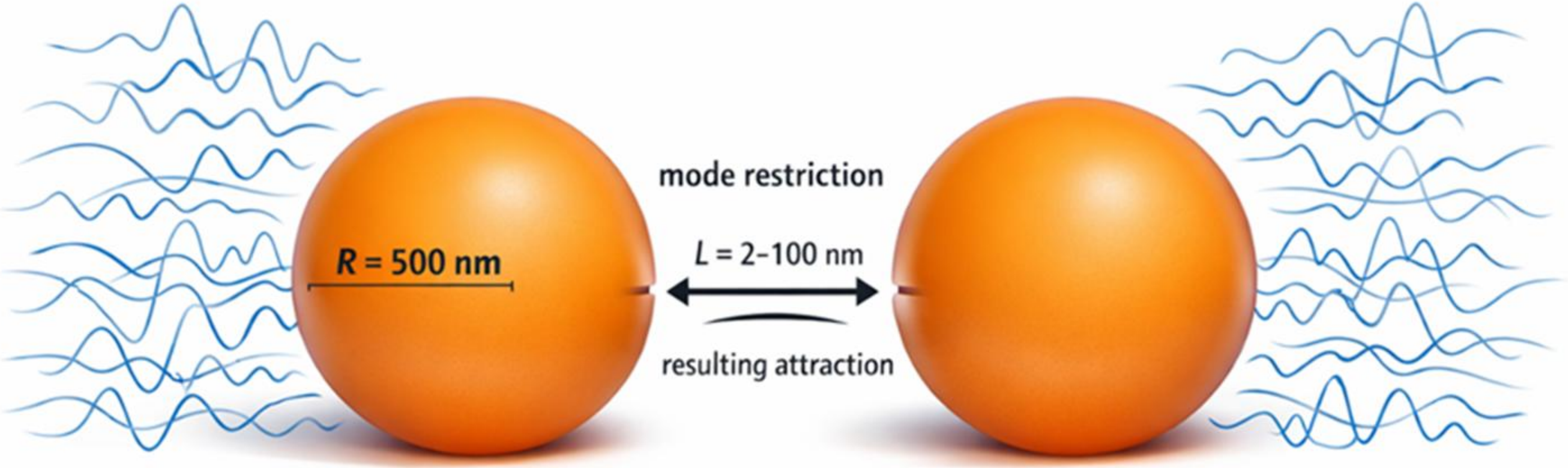
Under plausible prebiotic conditions these forces can stabilize interactions between protocell membranes.

The core physical claim: Interaction landscape

- In saline water (primordial soup), **DLVO repulsion collapses** (short Debye length)
- **Van der Waals** is too short-range for mesoscale stability
- Casimir–Lifshitz: **unavoidable**, long-range, algebraic decay

Interaction	Range	Salt sensitivity	Chemistry required
Hydrophobic	< 1 nm	–	yes
DLVO	< 2 nm	high	yes
Casimir–Lifshitz	2–100 nm	low	no

Minimal geometry: Two protocells (dimer)



$$F_{CL}(L) \approx -\frac{A_{eff}}{6} \cdot \frac{R_{eff}}{L^2}$$

with $R_{eff} = \dfrac{R_1 + R_2}{R_1 + R_2}$

A_{eff} = *effective Hamaker constant* $\approx 5 \times 10^{-21}$ Joule.

Mesoscale Protocell Clusters

Prebiotic Protocells with radii of roughly **R=200–1000 nm** can form **stable aggregates/clusters** driven by fluctuation-induced forces.

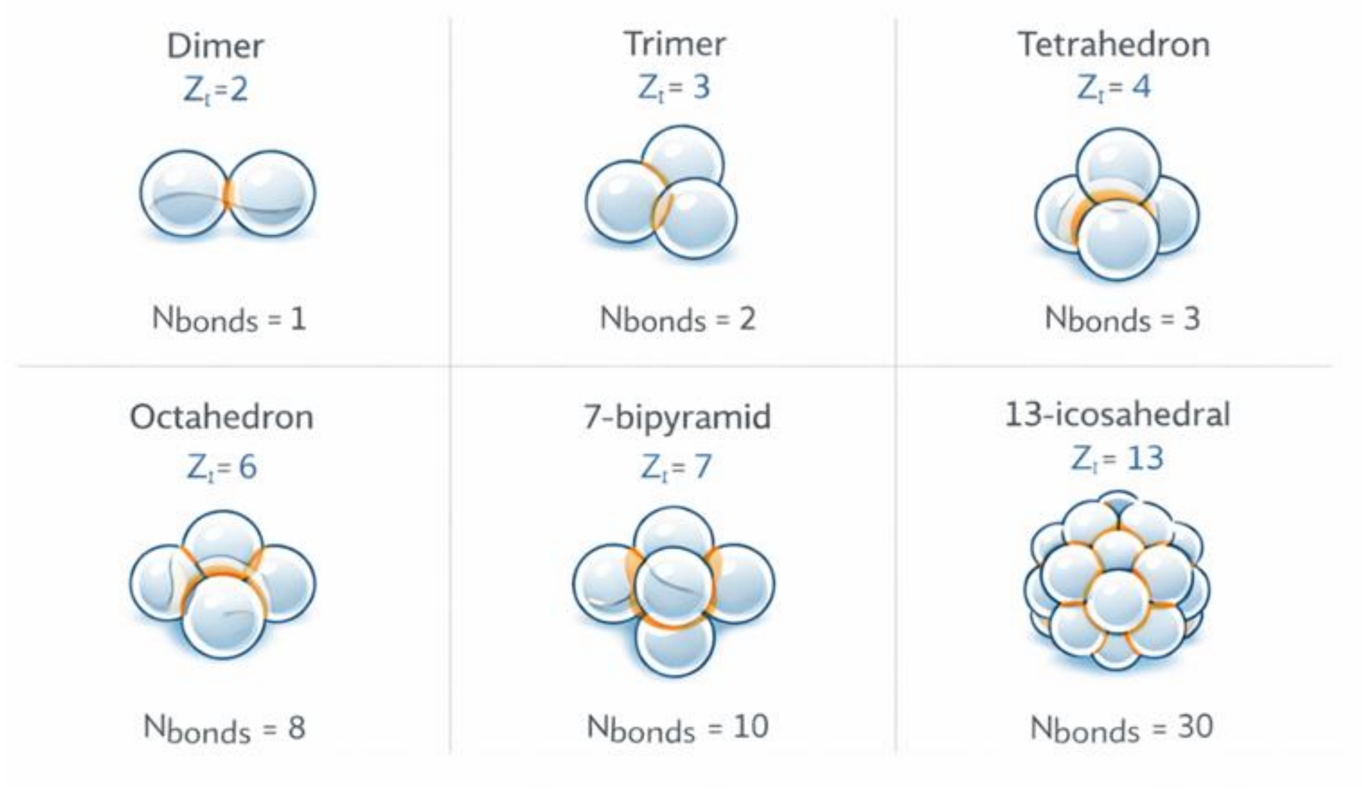
Typical cluster geometries include:

- dimers
- trimers
- tetrahedra
- octahedra

These configurations **maximize physical contacts** and **minimize energy**.

Department of Computer Science

Magic-Number Cluster Gallery: Discrete macro-states



N	Protocell Structure	N_{bonds}	$E_N/k_BT \approx$
2	Dimer	1	−5.1
3	Triangularer Trimer	3	−15.2
4	Tetrahedron	6	−30.5
6	Oktahedron	12	−61.0
7	Pentagonal Bipyramid	15	−76.2
13	Icosahedral 13-Cluster	42	−213.4

Energetic Stability

For realistic parameters the effective interaction energy **must exceed thermal fluctuations** (kBT).

Consequences:

- Brownian motion is partially suppressed
- contact times can reach minutes
- highly symmetric clusters become energetically favored

These conditions support robust mesoscale organization.

From Microstate to Macrostates

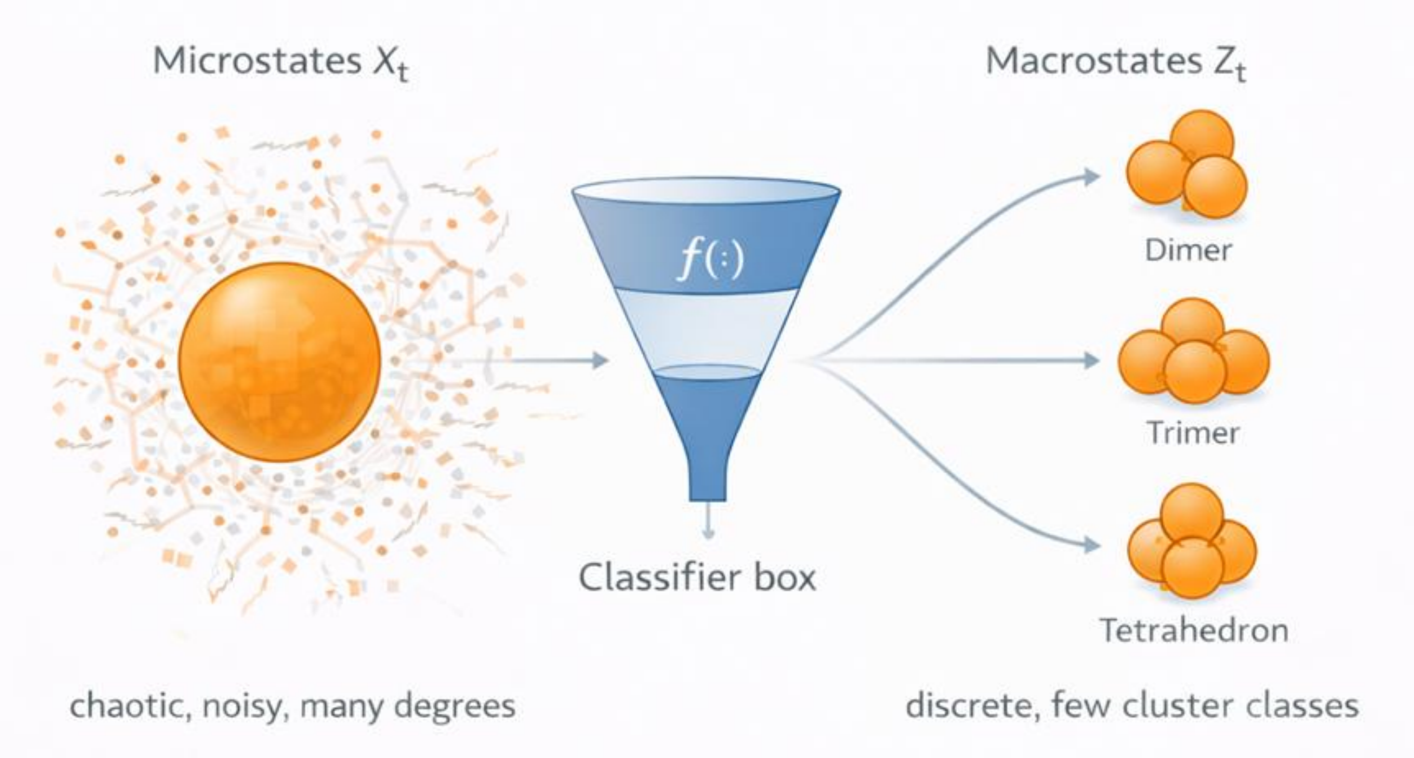
Microscopic protocell dynamics involve many variables:
positions, membrane fluctuations, forces, and hydrodynamic effects.

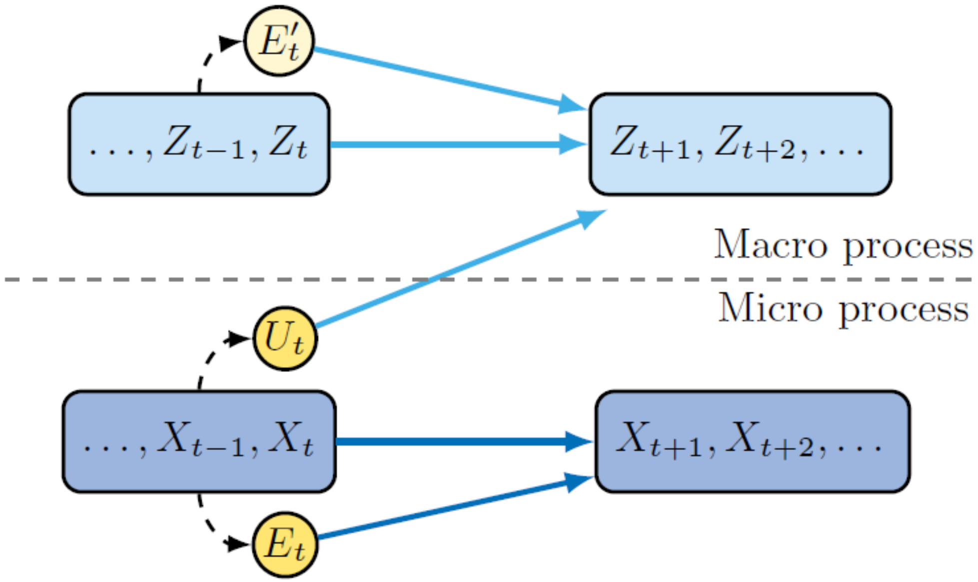
A coarse-graining approach maps these microstates to a smaller set of macroscopic system states.

Examples of macrostates:

- cluster geometries
- stable contact patterns
- characteristic neighborhood topologies.

Microstate → Macrostate Coarse-Graining: X_t to Z_t





Attractor Structures

Cluster configurations behave as attractors in configuration space.

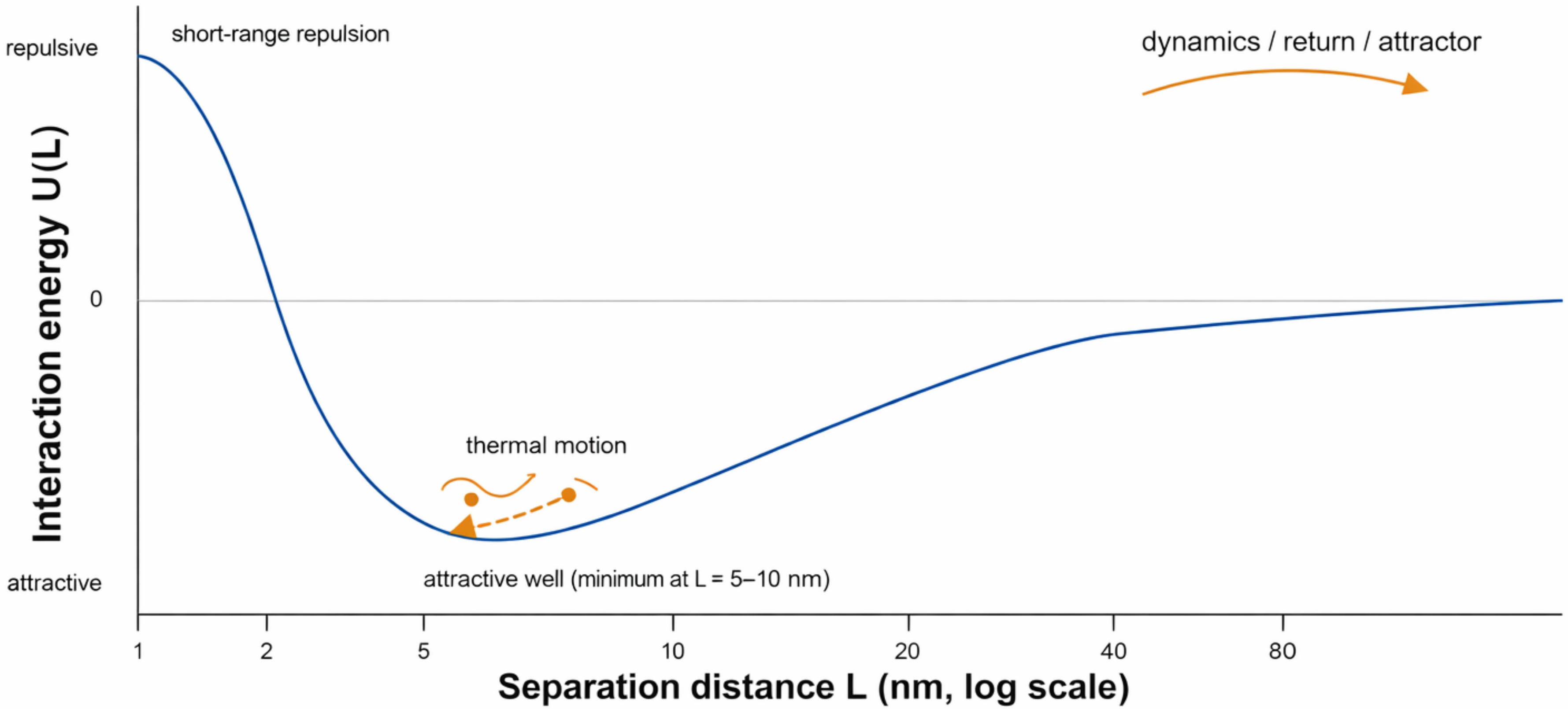
Key characteristics:

- stability against perturbations
- reproducible return trajectories
- a limited set of dominant system configurations

These attractors generate structured mesoscale dynamics.

Department of Computer Science

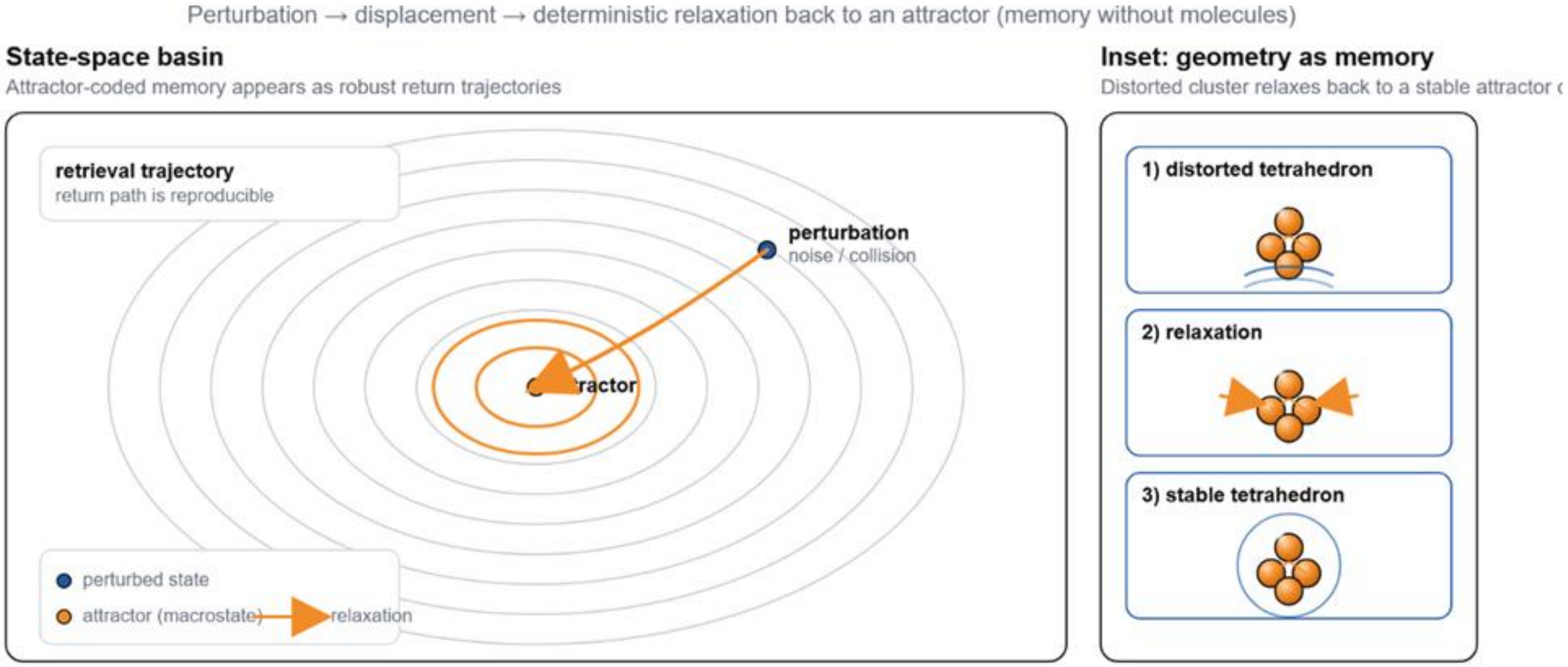
Casimir-Lifshitz interaction landscape



The potential well constrains motion while enabling local exploration

Department of Computer Science

Attractor Memory & Retrieval Trajectories:



Attractor-coded memory: perturbations decay along reproducible trajectories back to stable macrostates (retrieval without genes or polymers).

ε-Machine Representation of Cluster Dynamics

In this framework causal states correspond to stable cluster geometries.

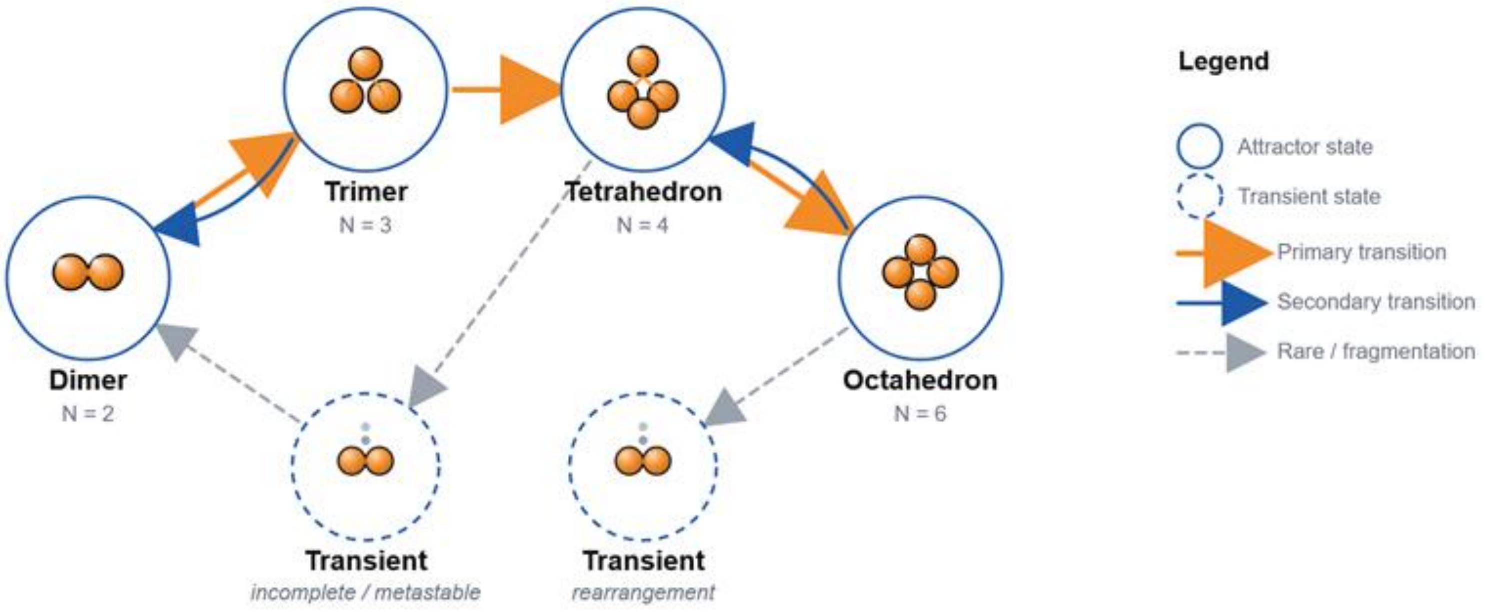
Examples:

- dimer
- tetrahedral clusters
- octahedral clusters

Transitions between these geometries define a finite-state dynamic system describing mesoscale evolution.

ε-Machine Graph of Cluster Dynamics

Tangential edges: arrowheads terminate at node boundaries and point to cluster icons



Primary edges follow energy descent (more bonds); secondary edges reflect reversible reconfiguration; dashed edges indicate rare fragmentation pathways.

Informational closure

A system is **informationally closed** when its **future macro-dynamics** can be **predicted from macro-states alone**.

Implication:
Microscopic details do not significantly improve prediction.

For protocell clusters this indicates the emergence of an autonomous mesoscale level of organization.

Department of Computer Science

Software-Like Dynamics in Natural Systems

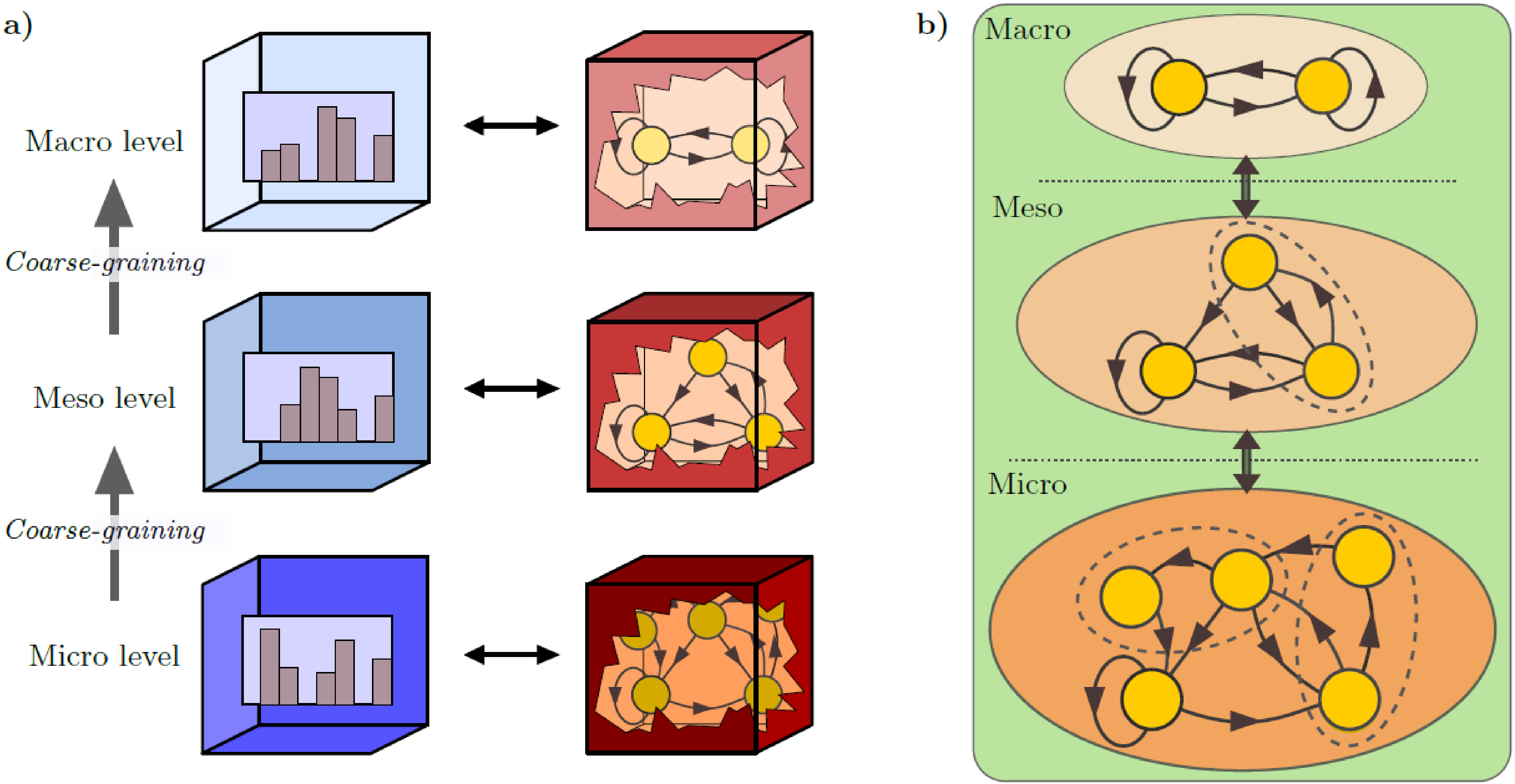
Recent theoretical work suggests that **causally closed macroscopic dynamics** resemble software running on hardware.

Analogy:
Hardware → physical protocells
Software → cluster state dynamics

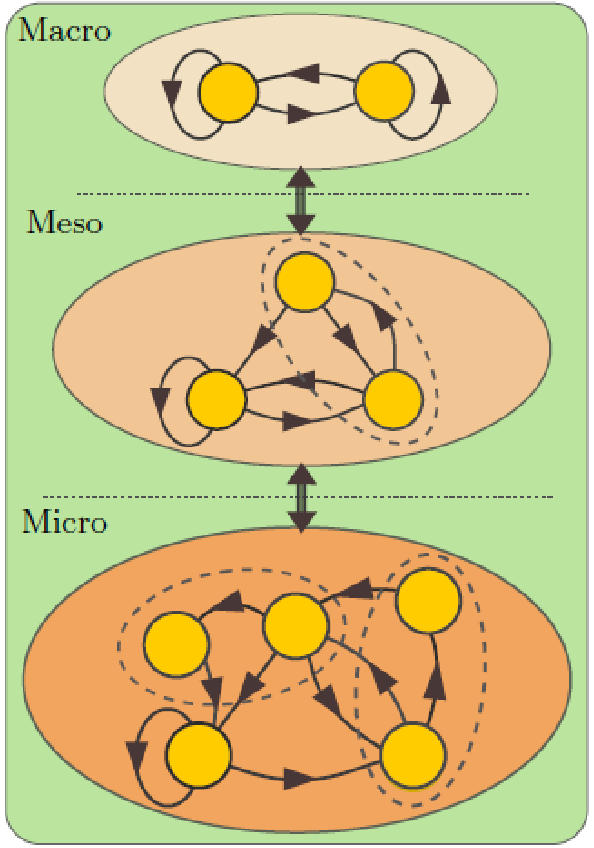
Cluster geometries therefore function as physical carriers of early information.

Department of Computer Science

Multilevel analysis via ϵ-machines:



Multi-level Hierarchy of Protocell Dynamics



Level	Description	Formalization
Micro	Forces, positions, thermal noise	Stochastic dynamics X_t
Meso	Cluster attractors	Macroprocess: $Z_t = f(X_t)$
Macro	Information dynamics	ε-machine E_t, transitions T_{ij}
Software layer	Attractor-coded memory, pattern retrieval	Causal / computational closure

Unified hierarchy: micro → meso → macro → proto-software

Micro: X_t: forces, positions, thermal noise
High-dimensional stochastic dynamics

Meso: $Z_t = f(X_t)$: cluster attractors
Low-dimensional macroprocess

Macro: ε-machine E_i and transitions T_{ij}
Minimal causal architecture

Layer: Attractor-coded memory & retrieval
Autonomous proto-software

This is the formal bridge from **physical clustering** to **prebiotic information dynamics**

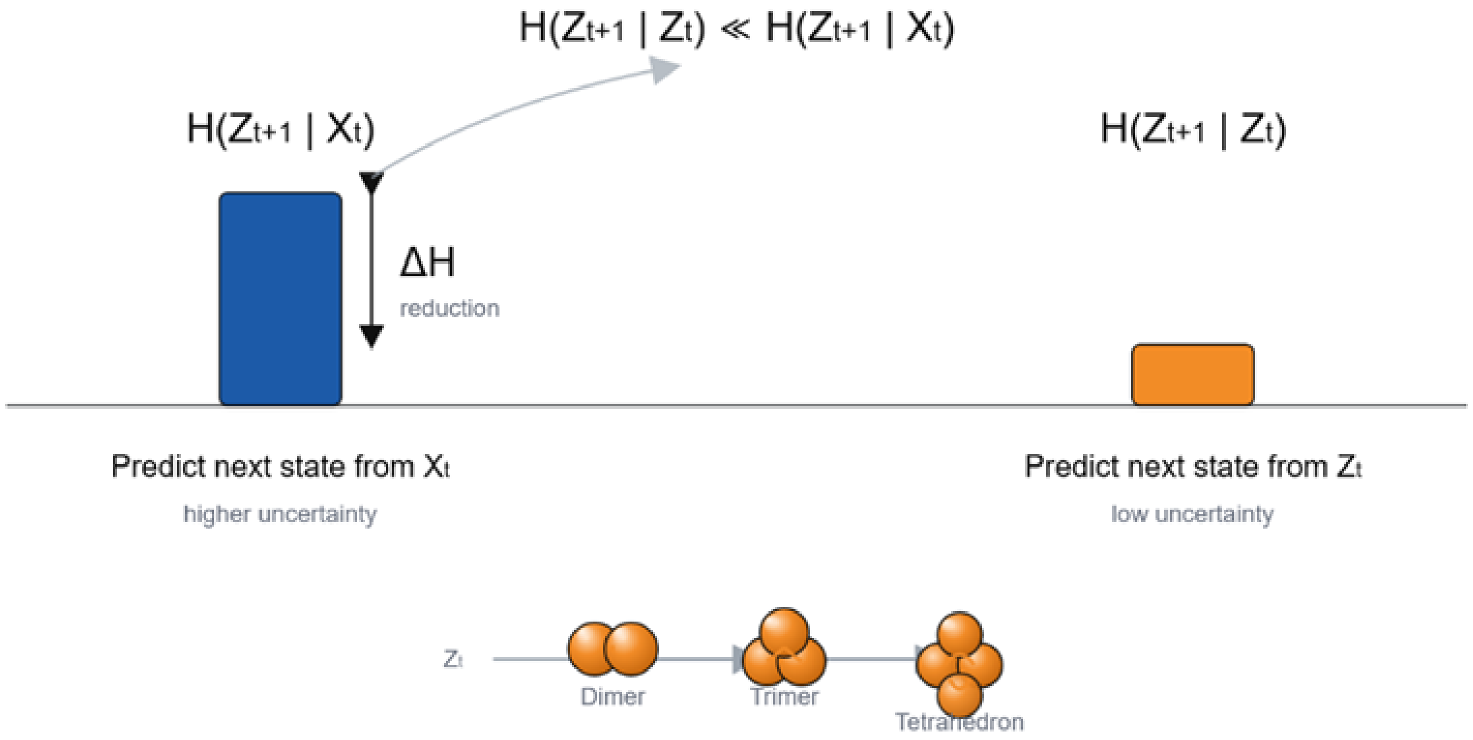
The Key Information Inequality:

Predictability increases after **coarse-graining** (dt. Vergröberung, es. granulación gruesa):

$$H(Z_{t+1} \mid Z_t) \ll H(Z_{t+1} \mid X_t)$$

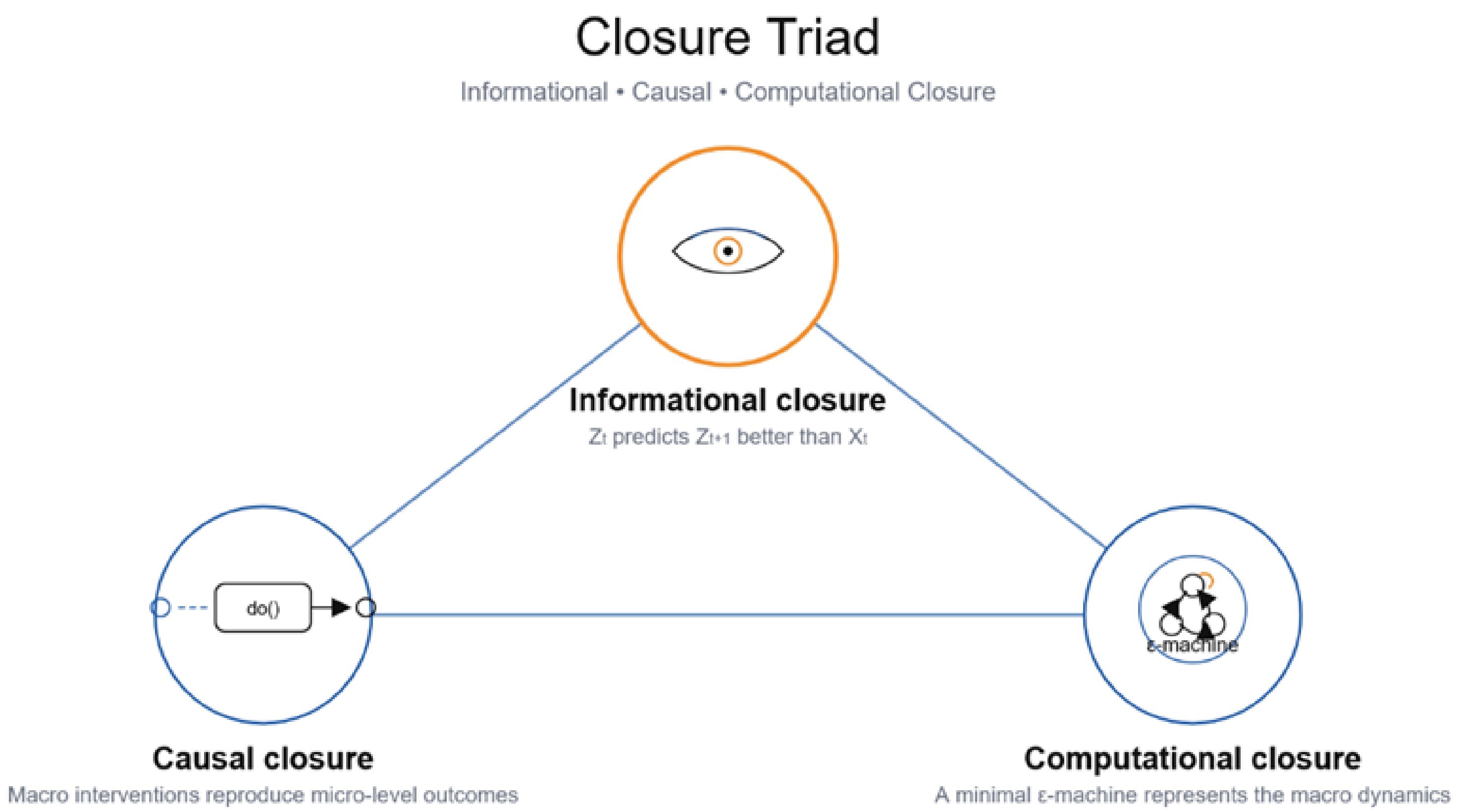
Interpretation: **macro-dynamics** become **law-like**
This is where “**information**” **starts to exist**.

Predictability Gain: Entropy / Information Inequality



Coarse-graining yields discrete macrostates with substantially lower conditional entropy.

Triad: Informational, Causal and Computational Closure



Informational closure: $H(Z_{t+1} \mid Z_t) \ll H(Z_{t+1} \mid X_t)$, showing that the macrostate predicts the future better than the microstate.

Causal closure: $P(Z_{t+1} \mid do(Z_t)) = P(Z_{t+1} \mid do(X_t \in pre(Z_t)))$, demonstrating that macro-interventions and equivalent micro-interventions yield the same transitions.

Computational closure: $\varepsilon(f(X_t)) = \pi(\varepsilon(X_t))$, where f is the macro-mapping and π the projection onto causal classes. This shows that coarse-graining produces the same causal architecture as full microanalysis.

Implications for the Origin of Biological Information

These results suggest that information may arise before genetic polymers emerge.

Key implications:

- information need not be polymer-encoded
- physical structures can function as memory states
- mesoscale organization may precede molecular evolution

This perspective provides a new physical pathway toward
the emergence of biological information.

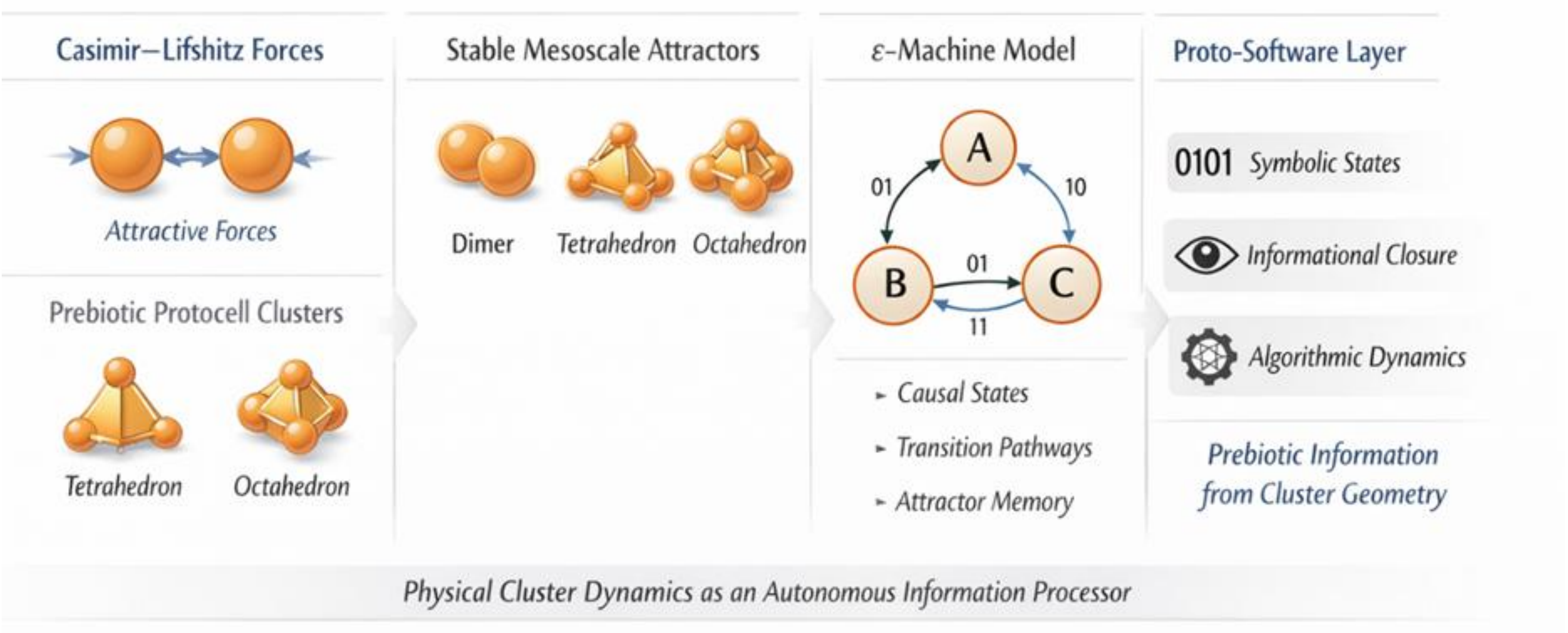
Department of Computer Science

Take-Home Message:

- CL forces stabilize robust mesoscale clusters → discrete macrostates.
- Coarse-graining yields **informational + causal + computational closure**.
- An **ε-machine** captures rule-level dynamics: a proto-software layer.
- **Bridge:** Paper 3 — From physical difference to meaning (constructor theory).

Department of Computer Science

Thank you very much for your attention.
Are there any questions?



Michael Massoth
Contact:
michael.massoth@h-da.de