Contents

1 Description of the system

1.1 Artificial chemistry

Basically, we want to use a system based on artificial chemistry to investigate the conditions promoting the formation of complex objects—where, by "complex objects", we refer to molecular complexity. Specifically, we will be focusing our analysis on investigating the spatial topology of the environment, to try to determine how it can program the emergence of complexity. We will utilize Assembly Theory to quantify the complexity formed therein, and in doing so this model will constitute a new test environment. Among other things, this could help refine our methods of calculation for the Assembly Index, and inform our knowledge on how the Assembly Index responds to various parameters and/or is correlated to life-like properties.

Artificial chemistry implementations can be categorized along two axis: how abstract (or realistic) they represent "real" chemistry, and whether they aim to model early or late evolution (see Fig. 1).

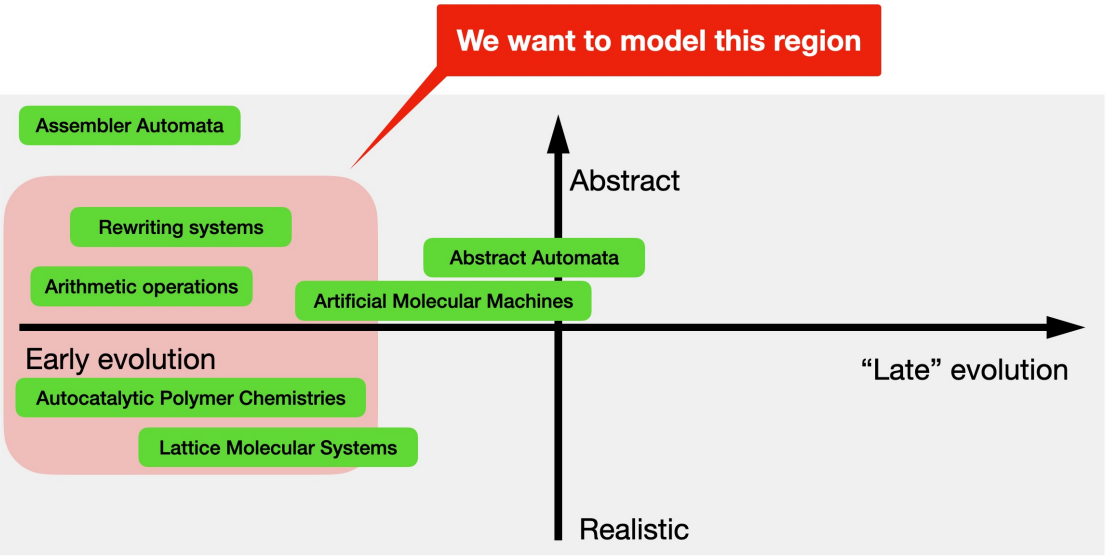

Figure 1: Classification of AC approaches along two axis: abstract vs realistic, and early vs late evolution. In the context of this project we will be focusing on the early evolution, using a balanced approach between more abstract and more realistic models.

The underlying (artificial) chemistry is based on several transformations: formard reactions, backward reactions, diffusion (Table 1). These reactions are parametrized by rate coefficients (k_f, k_b, k_d). Several other parameters can be defined, such as simulations (temporal) length τ_{max}, the number of reactors N, the number of inflows n, total mass M, etc.

1.2 System of reactors

These transformations are applied on a population of integers. These integers form a well-mixed system, inside a reactor/chemostat (Fig. 2a). The number of integers #1 is fixed, which translates into a fixed in-flow in the first reactor (as these integers are used to form other, more complex integers, or diffuse to the next chemostats). The system as a whole consists of several of these reactors, coupled together via in- and out-flows (defined by the corresponding rates), in a specific

Forward reaction	$A + B \xrightarrow{k_f} C$
Backward reaction	$C \xrightarrow{k_b} A' + B$
Diffusion	$C \xrightarrow{k_d} \varnothing$
	$C_i \xleftrightarrow{k_d} C_j$

Table 1: Reactions occurring in the artificial chemistry implemented in this system. Forward (constructive) reactions happens at rate k_f, backward (destructive) reactions happen at rate k_b and diffusion (where a molecule is shifted either to the next chemostat or out of the system) happens at rate k_d. Mass M is fixed, so the in-flow is coupled with the diffusion parameter.

topology. For example, one of the simplest topology is that of the regular lattice (Fig. 2b). When diffusion is very high, the whole system becomes well-mixed.

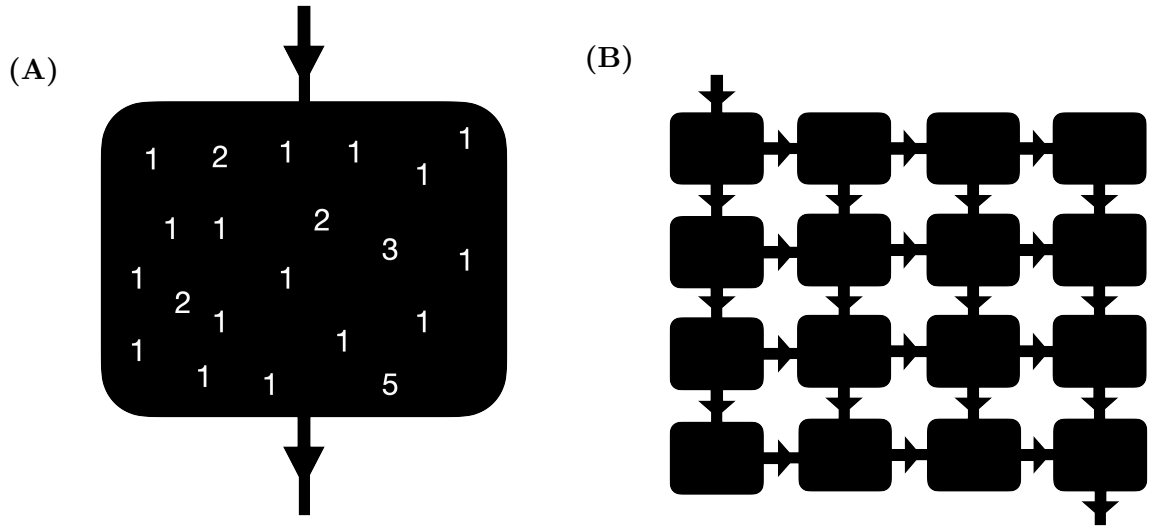

Figure 2: **(A)** Single reactor, inside of which a bunch of integers react. In-flow is fixed, i.e. the # of ones is fixed, and these react according to the reactions described in Table 1. An out-flow connects the reactor to subsequent chemostats. **(B)** Network of reactors/chemostats. Displayed is a regular lattice (note that diffusion happens both ways). There is a single in-flow, and a single out-flow. Other topologies are possible (see Fig. 3 below.)

1.3 Topologies

Several topologies can be used to connect the reactors together (Fig. 3). Examples include: path, lattice, Erdos-Renyi (random), regular (uniform node degree) or Barabasi-Albert (power-law degree distribution). This is especially relevant given that we're studying living systems, whose networks have been shown to possess certain specific properties deriving from their topology (e.g., resilience from BA networks, etc.) Other topologies (e.g. lattice, random) can be used as control/neutral.

1.4 Integer chemistry and Assembly Theory

We'll be using Assembly Theory to quantify the complexity emerging from the mixture, therefore we'll use the Assembly Index (A). Cole has calculated A for integers < 10000. The Assembly Index

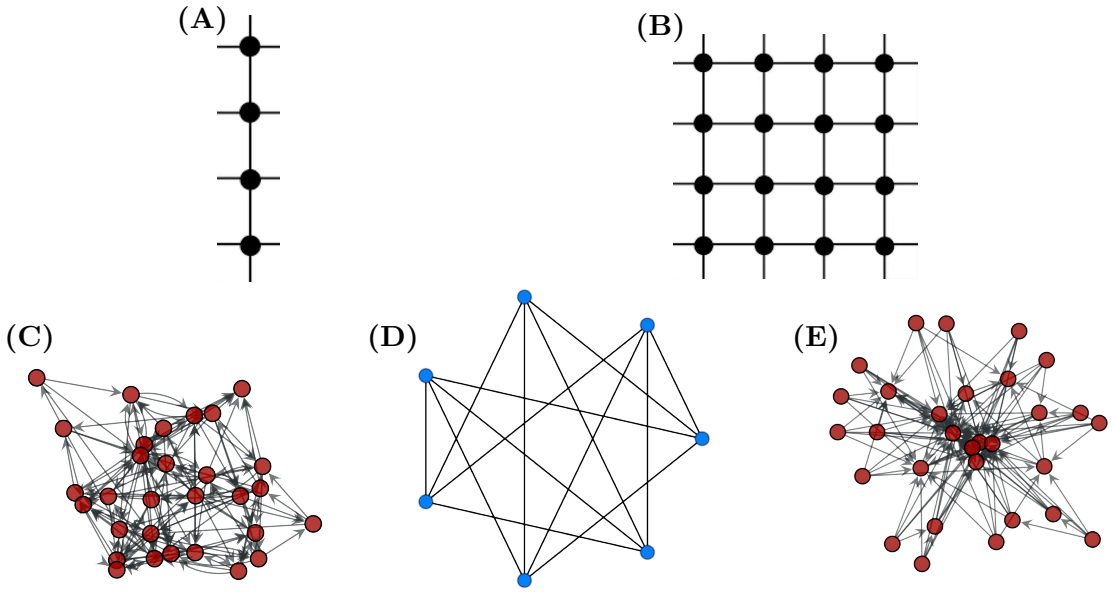

Figure 3: Illustrations of the different topologies we'll be exploring (at least, in the near future). The chemostats can be connected according to any of these topologies, which we presume will affect the construction of molecular complexity in the mixture. **(A)** Simple path lattice. **(B)** Lattice. **(C)** Random (Erdos-Rényi) graph. **(D)** Regular graph (all nodes have the same degree). **(E)** Scale-free (Barabasi-Albert) graph, with the distribution of nodes follow a power law.

increases as the logarithm of integer values (Fig. 4). We'll however probably be using both, as some relationships are better illustrated using y-axis as integers, and others with the y-axis as the Assembly Index.

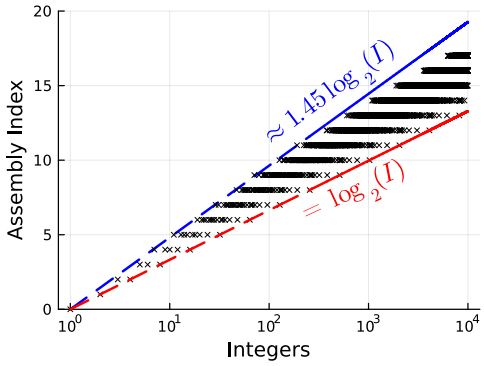

Figure 4: Assembly Index A for integers < 10000. A increases as the logarithm of integer values.

2 Experiments

2.1 Preliminary experiments (Sep-Oct 2024)

A first result we can plot using this system is the time series of the populations (shown on Fig. 5). The layout of this figure follows the same structure as the one on Fig. 2b: each panel shows the evolution of integers $I \in 1...10$ through time. The blue curve shows ones, we can see the number of ones is fixed on the first panel (top left). Diffusion pushes molecules to the next chemostats that are connected to the first one (where the in-flow is) so we're seeing initial transients/peaks at different times. The higher we increase diffusion, the faster these peaks happen, and the shorter is the initial transient.

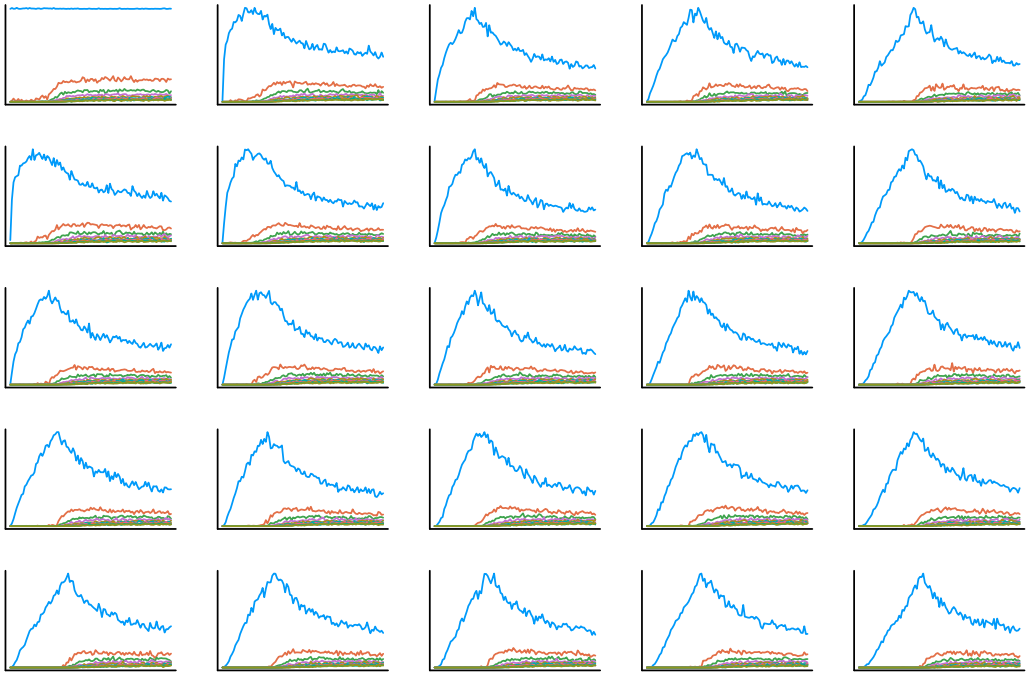

Figure 5: Time series for a simulation of 25 chemostats connected in a regular lattice. The first panel (top left) corresponds to the chemostat with an in-flow connected to it. The ten first species (integers $I \in 1...10$) are shown.

2.1.1 Varying the diffusion rate

A first parameter we're exploring is the diffusion rate. We've sampled simulations with diffusion rates $k_d \in [10^{-6}, 10^2$ and plot the first ten species (excluding ones) on Fig. 9. Panel A shows the average populations across reactors, i.e. taking all the chemostats together as one single population and calculating the average for species S at the last iteration of the simulation τ_{max}. We've done this calculation for 100 values of k_d (without using statistical ensembles for now). The shaded area on Panel A represents the standard deviation across chemostats, and is also illustrated on Panel B. We've also added a vertical (dashed) line to indicade the forward rate k_f.

One thing we can readily notice from Fig. 6 is the presence of transitions. On Panel A we have four regimes: low values in the populations of complex (i.e. $I > 1$) species for lower values of the diffusion coefficient ($k_d < k_f$), then increasing complexity ($k_f < k_d < 1$), then a dip in complexity

(k_d 10), and finally complexity increasing again for very high diffusion rates ($k_d \gtrsim 10^2$). Panel B shows an increase in σ until k_d 10^{-1} then a sharp decrease ($10^{-1} < k_d < 10$).

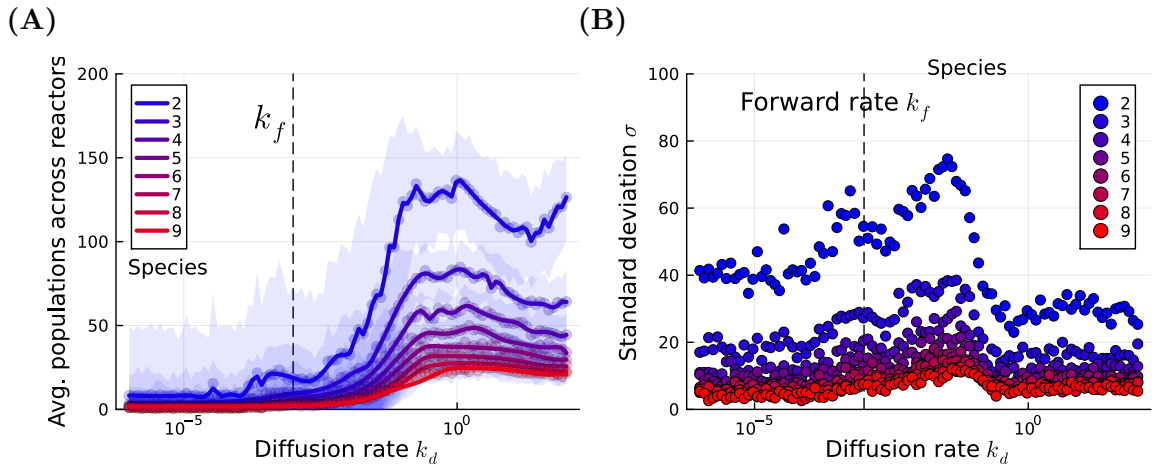

Figure 6: **(A)** Average populations, calculated across all chemostats, at τ_{max} for the first ten integers excluding ones. Shaded area represents the standard deviation (also pictured on Panel B). **(B)** Standard deviation for the same species shown on Panel A.

Another thing we can calculate from these systems is the total mass, i.e. $I_i \times f_i$ (where I_i is the integer of species i and f_i the number of identical copies or "copy number"). Shown on Fig. 6 is the total mass M against the same variation in coefficient discussed above. M stays constant for most values of the diffusion coefficient but increases exponentially for the highest values of k_d.

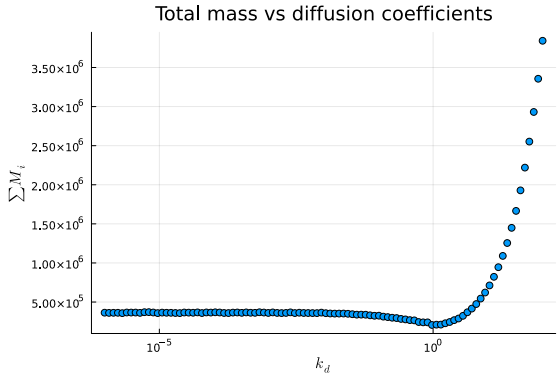

Figure 7: Increase in total mass as we increase diffusion. Total mass M is calculated by multiplying integers I_i with their copy number f_i.

Putting this all together: There seems to be three regimes based on what we're seeng in Figs 6-7: initially, we have a well-mixed system in the first chemostat. The complexity in chemostat #1 increases, but all other chemostats remain empty, which is why we're seeing low average values in Fig. 6 (they're calculated over all chemostats). Then, as we increase diffusion the other chemostats are progressively filled up: this is why the standard deviation increases. Finally, as all chemostats are fully filled and we continue increasing the diffusion, the system returns to a well-mixed state and the mass increases rapidly (because complexity now builds up in all the chemostats).

2.1.2 Measuring distance from source

One additional experiment we've done was to investigate how the mean assembly index (per reactor) varied with the distance from source D. For example, the distance D between the source (in green) and the target chemostat (in red) shown on Fig. 8a would be six. We've evaluated how the complexity varied with D for three diffusion regimes: low diffusion ($k_d = 10^{-4}$), intermediate diffusion ($k_d = 10^{-2}$) and high diffusion ($k_d = 10$). Complex molecules remain in the chemostats that are located close to the source $D = 1, 2$ when diffusion is low (Fig. 8, blue curve) or intermediary (red curve). When diffusion is increased (green curve), complexity spreads to the other reactors.

However, there remains a degeneracy when we examine the complexity contained in chemostats located at identical distances from the source (Fig. 8, dashed box surrounding data points at $D = 5$). Multiple chemostats at $D = 5$ have different (average values of the) assembly index. This is clearly evident when looking at the inset in Fig. 8 that displays the density plots (i.e., reconstructed PDF): the heavy tails extending to high values of A clearly differ from one chemostat to another. Consequently, D alone does not fully determine the quantitative complexity of the mixture.

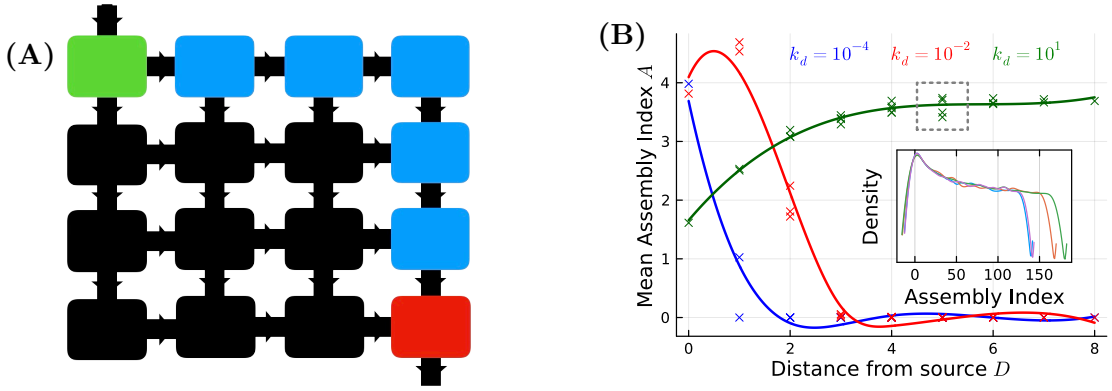

Figure 8: **(A)** Here we investigate how complexity varies as a function of the distance D from the source. For example, the distance between the in-flow (green) and the target chemostat (red) is six. **(B)** Complexity as a function of the distance from the source, for three diffusion regimes. Inset: density plots for the four chemostats having $D = 5$. Clearly, D alone does not fully determine the complexity of the mixture.

2.1.3 Measuring detection thresholds

Next, we want to relate these experiments to observational missions by evaluating thresholds in complexity that we could detect. We can therefore evaluate the abundance of the different species (i.e., populations as a fraction of total # of molecules or total mass—here we chose the former) and determine the most complex molecule we could detect using an instrument capable of measuring concentrations of 10^{-5} (i.e. 1 molecule in a system having a total of 10^5 molecules), then 10^{-4}, and so on. Fig. 9 how the complexity of detectable molecules varies as we change the diffusion coefficient.

In Fig. 9, on Panel A we plot the raw integer values against k_d whereas on Panel B we're showing the assembly index instead. We're using five different detection thresholds ($10^{-5}, \cdots, 10^{-1}$). Each data point indicates the integer value (or assembly index) of the most complex molecule that's detectable using this specific threshold. Going back to what we mentioned previously about transitions and regimes, we can now clearly see the three regimes we talked about in Fig. 9: for

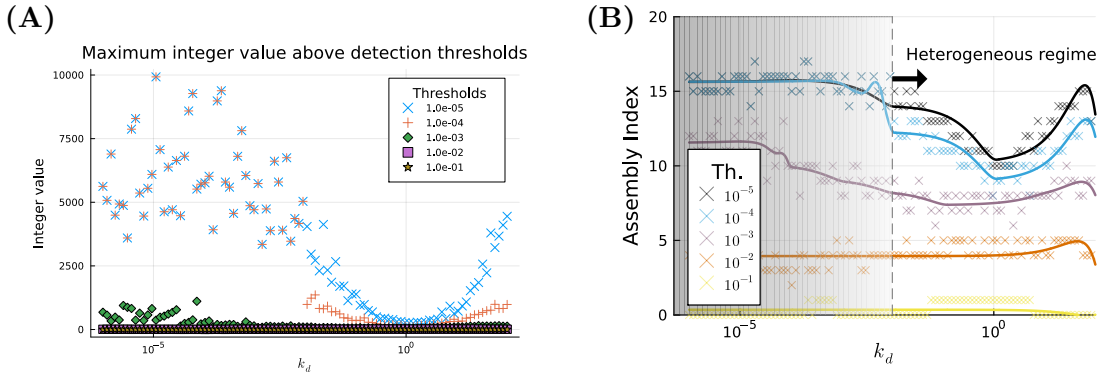

Figure 9: **(A)** Most complex detectable molecule for different thresholds (indicated in the legend) as we vary the diffusion rate. Raw integer values are shown. **(B)** Same as Panel A, but the Assembly Index is used instead to indicate the complexity of the molecules.

lower values of k_d, all the complexity is contained in the first chemostat and we're just sampling the heavy tails, as we increase diffusion the molecules spread in the other chemostats and complexity suddenly decreases, finally as we further increase k_d complexity rises again.

Basically, we have two competing phenomena: (1) increasing complexity as we increase k_d, but also (2) decreased ability to detect the molecules when they spread initially (which decreases their concentration)—until they have filled all the chemostats and their concentration increases again. This is why we're seeing a minima on Fig. 9: this point represents the area of the parameter space where the molecules are in the process of filling up the chemostats.

2.1.4 Upcoming experiments

Some things we are considering:

- making a figure of the "cones" which define assembly spaces for graphs/integers/molecules (similar to the one in Fig. 10 below)
- decoupling in-flow and diffusion: right now in-flow is implicit, i.e. we have a fixed number of ones in chemostat #1, and when we increase diffusion we push molecules to the next chemostats (which amounts to refilling chemostat #1, since the number of ones is fixed). This prevents us from increasing the in-flow independently from the diffusion coefficient, which could increase complexity without increasing diffusion so much (making the whole system a giant well-mixed system, and also making the simulations much longer to compute)

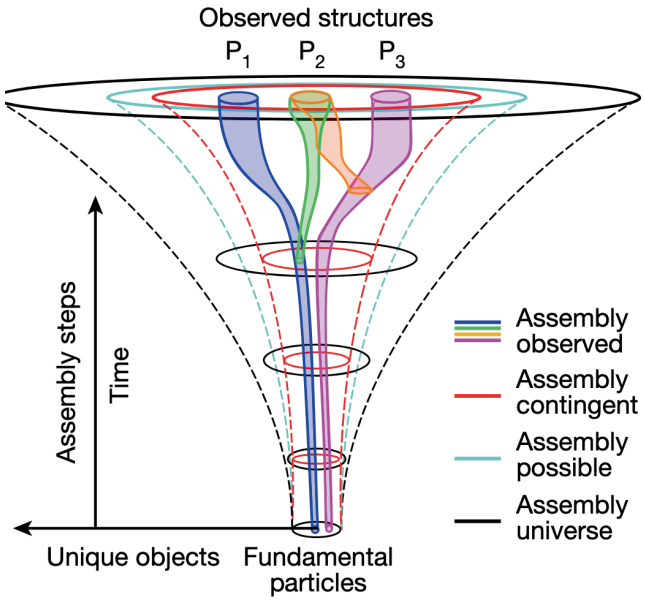

Figure 10: Cones defining the different assembly spaces. Taken from Sharma et al. (2023).

3 References

References

Sharma A., Czegel D., Lachmann M., Kempes C. P., Walker S. I., Cronin L., 2023, Nature, pp 1–8