

THE LONDON
INTERDISCIPLINARY
SCHOOL

Capstone Report

Multiscale simulation: a case study of plant morphogenesis and neural cellular automata towards the complex modelling challenges of collective intelligence

Draft

Alexander Clarke | MSc 2024
London Interdisciplinary School

Foreword

‘Collective intelligence’ in all its forms, has been a clearly identifiable part of the research zeitgeist in the last century. It links together many of the current fears and ambitions in technology, whilst remaining an elusive and ill-defined phenomenon in science. Bridging theory from biology and computer science, the two fields that are arguably best-equipped to tackle this investigation, this paper probes just one form of it, growing tissues, to learn how we might better understand it. When collections of cells proliferate in their billions, they solve a high-dimensional problem with striking efficiency, achieving global coherence with local communication. In this cacophony of interactions, they develop a robust anatomical and functional complexity we still struggle to fully explain. Morphogenesis is an exemplary study for the interdisciplinary study of complexity that invites contributions from biology, chemistry, mathematics, computer science, and physics.

The problem of scale and scalability in modelling is difficult to address. It is a given that interactions happen across all scales of space and time, but the way they are linked is not at all trivial. In physics, for example, we have created sophisticated models that range from the way electrons orbit an atomic nucleus, to the way stars cluster in galaxies, but trying to piece such disparate models together is simply unfathomable. Furthermore, the kinds of interactions at our ‘meso’ scale in the universe, between the micro and macro, are often particularly pertinent and perplexing, requiring us to look out for causation from above and below. ‘Multiscale coupling’ describes how disparate system components can interact or interfere with one another, which we know leads to significant feedback, non-linearity and often, chaos.

The ability to engineer systems that self-organise, self-assemble or self-repair will allow us to design entirely new forms of intelligence that grow to solve macro-scale problems with relative ease, as MIT engineer Eric Drexler imagines in his prescient 1986 book *Engines of Creation* melding science fiction, nanotechnology and cutting-edge molecular biology. I will discuss how we can take the first step towards this using the powerful paradigm of computational modelling, simulation. Through a better understanding of how to grow complex phenomena, we may design more adaptive architecture, more adept forms of artificial, collective intelligence, and more targeted medical interventions with nanotechnology. Interpretability through visualisation will also be paramount in the era of scientific assisted by machine learning. Santiago Ramon y Cajal was a pioneering neuroscientist who laboriously recreated beautiful early depictions of neurons and neural networks using primitive microscopy. I highlight Drexler and Ramon y Cajal as two inspirational figures who practise a highly creative form of science.

Contents

Foreword.....	2
Contents.....	3
1 Introduction.....	5
Towards Multilevel Models.....	5
A Qualitative, Pattern-oriented, Machine Learning Approach.....	6
Background: Reductionism is Rife.....	7
Why Simulate?.....	8
2 Complexity and Cellular Automata.....	10
Proteins as Sub-Units for Self-Assembly.....	10
The Cellular Automata Modelling Strategy.....	12
3 Leaf Morphogenesis.....	13
The Plant Cell.....	13
Leaf Formation.....	13
Division and Elongation Patterns.....	14
Conformal Growth.....	15
Polarity Patterns Provide Synchronisation Cues.....	15
A Complex Interplay of Biochemical Interactions.....	16
Phyllotaxis.....	17
Mechanical Factors Also Influence Growth.....	17
4 Simulation Design and Development.....	19
Design Principles.....	19
Related Works.....	21
Prototype Design.....	22
Neural Cellular Automata Implementation.....	23
Key Findings From Early Experiments.....	25
Utilising Biological Datasets.....	28
Prototype Development.....	29
Ongoing Limitations.....	31
5 Results.....	32
Pattern Growth and Stability.....	32
Growth Dynamics.....	33
Training Progression.....	34
Pattern Development During Training.....	35
6 Evaluation.....	36
Efficacy and Realism of the NCA Approach.....	36
Concluding Remarks.....	37
References.....	39
External Datasets.....	44

Code Availability.....	46
NCA Experiment.....	46
Key Definitions and Abbreviations.....	46

Draft

1 | Introduction

Towards Multilevel Models

Complex systems, from the tissues and organs in our bodies to the sprawl of cities, present a common obstacle to our understanding: they are composed of many diverse agents interacting over many scales. Their sub-components, whether that is minuscule proteins, or groups of people, have the ability to set in motion patterns many orders of magnitudes larger than themselves, more often than not, exhibiting forms of collective intelligence. Multiscale patterns in the structure and behaviour of systems represent an unprecedented modelling challenge, which is the subject of this report. Their inner workings remain highly unpredictable and perplexing to us, despite the fact that their overall appearance is familiar, like the way we quite easily classify clouds as ‘cirrus’ or ‘cumulus’ even though we fail to understand their inner workings. Specifically, our remaining obstacle “is to understand how the spatial organisation of the components of a system can influence overall temporal development”, and vice versa (Rascher et al., 2001, p.1).

I outline how current, siloed modelling efforts may be translated into a more integrative, modular simulation framework to overcome their current issues of generalisability, reusability and scalability, leveraging fundamental biological and software architecture principles; the discrete modelling technique of cellular automata (CA); and nascent research from machine learning (ML). By scaffolding a multilevel simulation architecture mimicking closely the adaptive organisation of the living systems we intend to model, we may engineer more complex forms of synthetic, collective intelligence with greater functional and structural realism. Enforcing deep ‘pattern’ realism like this is vital when using ML approaches, the common complaint being their lack of explainability, moving towards ‘grey box’ artificial intelligence. A thorough interdisciplinary review of developmental biology is melded with practical insights from computational modelling and machine learning, focusing on the key question in botanical science, “how do leaves get their shape?”. Drawing upon key concepts like *pattern-oriented modelling*, and *multiscale analysis* from complexity science, a proposal for a modular, discrete and image-based simulation framework consisting of flexible simulation sub-units is developed. A functional prototype of a ‘multiscale-ready’ simulation module is built with neural cellular automata (NCA) that learns to grow biological images with self-organising behaviour. In this experiment, the capabilities for cellular artificial intelligence to reproduce goal-based patterns from time lapses of real specimens (such as a simplified leaf vein pattern) are conceived using the accessible machine learning framework PyTorch, demonstrating clear potential for this multiscale modelling approach.

A Qualitative, Pattern-oriented, Machine Learning Approach

Models are a ubiquitous strategy across science and engineering for the qualitative and quantitative probing of real world phenomena, being loosely synonymous with ‘toys’, which are a vehicle for human understanding and knowledge exchange through play. We need simulations that we can interact with, ideally in real-time, since this permits equally-important qualitative “what-if” analysis (D’Angelo, Ferretti & Ghini, 2017), an intuitive basis on which more rigorous quantitative investigations can eventually follow. In recognition of the vast technical challenges associated with multiscale modelling, the development of a flexible qualitative simulation leveraging image-based machine learning is prioritised, since the premature addition of overly-mechanistic constraints can lead these models to become single-use, lacking generalisability and flexibility.

Given our limited observational and measurement abilities, we may use sparse and incomplete datasets to try and piece together a patchwork model. Some technologies like fluorescence microscopy, drone photography and satellite imagery are now allowing us to observe patterns beyond our usual limits, at increasingly high fidelity, providing data in the form of false-colour images and 3D scans, but it is also makes sense that we link theory and datasets across scales, in more comprehensive virtual environments, so we can better validate and augment model realism. Often scientific discoveries are found through resolution of competing ideas, and subjecting models to shared environments provides realistic or unseen conditions that lead to better detection of spurious results: researchers keen to prove a hypothesis may overfit a model to a limited experimental dataset by tuning model parameters and mechanisms as such. Equally simulations may “reproduce the right patterns for the wrong reasons” (Wang et al., 2018, p.942), insofar as minute interaction rules are oversimplified, insufficiently defined, or incorrectly combined, yet still lead to a convincing global pattern under some narrow test conditions. In line with pattern-oriented modelling (POM) principles, a “set of multiple, weak patterns, which are not necessarily matched precisely, can be a more powerful filter to identify the most appropriate parametrization and model structure than a single optimally fitted pattern” (*ibid.* p.957). This is especially important in biology, specifically genetics, where many genes are expressed simultaneously to provide regulation macro-scale phenotypic features like the shape of leaves with high error-tolerance (Terebieniec, 2019). The flawed (and convenient) assumption we have made in the historical, isolated study of these phenomena is that they embody one, singular, pointwise goal, rather than a competition of many factors that can fluctuate over time.

The use of extensive data to sieve underlying patterns is also in line with the latest paradigm in science, broadly that of machine learning, but specifically concerning the developments of deep-learning and attention; pre-training models; and most recently neural operators, involves using neural networks (NNs) to learn hidden, high-dimensional functions for novel scientific insights. To define and refine plausible models for real world phenomena, we would like to use sparse and seemingly disparate data to calibrate

interaction rules. Newer architectures have proven capable of implicitly learning fundamental laws like gravity or predict medium-range weather patterns, doing so many orders of magnitude faster than traditional, numerical methods (Azizzadenesheli et al., 2024; Cranmer, 2023). The important remaining difficulty is their explainability, which the image-based approach presented here aims to mitigate.

Background: Reductionism is Rife

For a long time in science, we have operated under the tenuous assumption that we can learn all there is to know about systems by performing sufficiently deep dissections; we break unknowns down and study their parts, freezing and isolating specimens on a petri dish. In plant science, for example, to cope with the complexity of genetics research, we have understandably opted to study ‘mutants’, that is, a variant of a species with a singular mutation corresponding to a gene we want to understand. However, as before, plants in particular have developed genetic code with high resilience against environmental stresses, meaning that many genes are co-expressed, and it is not always easy to identify a one-to-one mapping of genotype to phenotype. Reduction is an inevitable and convenient strategy, since being able to summarise what we observe in simplified form prevents an overwhelm of sensory input. It is only a problem when it narrows our objective lens so far that insurmountable disciplinary divides form. We find in scientific research, increasingly, that overly-reductive approaches limit our understanding of biological systems. For example, it is established that tumour growth and metastasis is greatly dependent on the ‘microenvironment’, meaning tumours grow differently on a petri dish, *in vitro*, than within bodily tissue, *in vivo*, since overcrowding leads cells on the border to divide faster than the ones in the centre of the tumour (Brú et al., 2008; Seyfried, & Huysentruyt, 2013). Naive analytical approaches have previously tried to fit simplistic population models like exponential growth, which initially pass but fail to capture the long term behaviour, then leading to gross under or overestimates in the efficacy of treatments. To give an example from systems biology, authors like Noble and Noble (2023) stress the importance of acknowledging edge cases to the well-accepted but streamlined view of genetics known as the Central Dogma, with that of epigenetics, which shows there are hereditary cell-level differences in gene expression - we cannot treat cells as identical. For complex systems therefore, reductive analytical methods can over-generalise and neglect diversity or the microenvironment, attributing linearity or acyclicity where there is conversely a strong interconnectedness and heterogeneity among agents.

Complex systems science seeks to overcome the limitations of linear, piecewise modelling strategies. In classical physics or engineering it might often be easier to approximate bodies with point masses, rigid bodies and perfectly smooth surfaces; biological cells cannot be aptly modelled this way. When biologists watch cells for example, they talk about cell lineages, which trace the many interactions that led to the cell in its current state. Fields as diverse as biology, psychology, neuroscience, urban planning and

economics are exemplary for the application of complexity because they all deal with the study of collectives and networks, requiring agent-based modelling strategies that allow for a degree of self-directedness, but also making them far more irreducible. The alternative strategy I explore here is the multiscale analysis of systems, where it might instead be possible to reduce the complexity of their patterns considering their relation to other systems across scales. Put simply, taking seemingly disparate systems, like fluid flow or people moving through a crowded corridor, we may spot structural and functional similarities. The apparent structural nesting, modularity and self-similarity of (biological) complex systems, where “nature duplicates and reuses existing parts and design principles again and again” (Agnati et al., 2009, p. 554) promises an orthogonal reductionist approach.

Why Simulate?

Simulation is the development and embedding of models in a *computational* setting, a powerful way to perform experiments in a virtual petri dish. In government, simulation may provide swift responses to disaster scenarios by enumerating all potential outcomes for a given parameter space. In engineering one may generate and test prototypes towards a set of real-world constraints and scenarios, recovering optimum solutions or parameters (inverse design). In bioengineering, computer and complexity science, one shared goal is to ‘engineer emergence’ (Bhalla, Bentley, Vize, & Jacob, 2012) or to design ‘synthetic life’ (Blackiston et al., 2021), by figuring out ways not only to mimic, but guide the processes of self-organisation, self-assembly, self-repair and self-replication in and beyond real, living examples that lead to multilevel patterns of coalescent, collective intelligence.

For simulation to have utility, it must avoid becoming totally abstracted, offering “a pipeline to move between the physical and the virtual world, where results in one environment can inform the other” (Blackiston et al., 2021, p.8). Biotechnology research without simulation, “would be impossible biologically because of the time and effort involved” (*ibid.*). This means effective simulators are steered towards the creation of ‘digital twins’, virtual models of real systems that are continually fed real-time data from their physical counterpart, in order to create a tighter digital-physical integration that can serve up accurate live forecasts, allowing for more agile, adaptive, autonomous interventions at scale.

It might seem pointless to recreate living systems in a virtual environment, when our problems are in the physical world. However, big questions almost always necessitate leaping ahead of the linear and logical scientific process before working our way back sequentially. The generative, evolutionary and creative process of speculation is what simulation is well-suited for. Due to sparsity and poor integrity of ancient specimens, fields like the origins of life naturally require a more ‘archaeological’, patchwork, or reconstructive approach, where simulation can validate plausible mechanisms for self-organising and self-repairing systems might have come to be.

More tangibly, simulation will (continue to) play a pivotal role in the design of future technologies, from smart healthcare, efficient food production and adaptive urban architecture, to disaster response and ecological conservation. Digital twins provide a platform to design targeted interventions, like nano-scale robotics inside the body, when paired with developments in genetic modification (Drexler, 1987). The Advanced Research and Invention Agency (ARIA), part of the UK government, has launched an initiative called ‘Programmable Plants’, aiming to rapidly “design, write and build fully synthetic crop genomes” inspired by the fast-tracking of the Covid-19 vaccine (Burnett, 2024).

Simulation, when accompanied with machine (deep) learning, has already proven its ability for applications like automated drug discovery or protein folding with DeepMind’s AlphaFold. In 2024 in the UK alone, pilot digital twinning projects have been launched for applications as disparate as environmental monitoring, responding to threats like coastal flooding (UKRI, 2024) and patient-tailored digital heart models for patients with critical heart conditions (NIHR Imperial BRC, 2024).

Digital twin technologies like these are made possible by our growing ability to sample the microenvironment, collecting data at scale and high resolution, using distributed sensor networks. A lot of progress has been made in scalable data collection with the advent of microcontrollers and mesh networks that underpin Internet of Things (IoT) technologies. Availability of lower cost, portable and increasingly automated sensor and imaging technologies enables challenges like surveying the health of an entire field or forest to be tackled in a (semi-)automated fashion with drones equipped with 3D scanning technologies like LiDAR and photogrammetry. In a study of modelling of urban forestry, Chen et al. (2024), however identify the scalability problems that remain for fine-grained data collection, proposing a novel approach balancing structural accuracy of scanned trees with efficient data storage and processing. The prevailing challenge is in the coordination of these devices and efficient processing and analysis of these data streams. D’Angelo, Ferretti and Ghini (2017) point out the growing difficulties in system design and management, as a result of the vast ecosystem of heterogeneous IoT devices in communication. In their example, performing proximity-sensing of users walking through a ‘smart market’ would require efficient filtering of dense device-device interactions. In the case of IoT modelling, there is an inherent need to model fine-grain device interactions in populated areas or but also over great distances, accommodating rural areas which are often neglected (D’Angelo, Ferretti & Ghini, 2017). Often we end up creating many different models at different levels of detail. An integrative approach would ideally operate with ‘adaptive spatiotemporal granularity’. This echoes the paradigm of Dynamic Data Driven Applications Systems (DDDAS) which is already used extensively to create large-scale models that can adapt to incoming data on-the-go, reducing their running time and improving their accuracy (Fujimoto et al., 2018).

2 | Complexity and Cellular Automata

Complexity in systems is associated with the capability for emergence, which means the combination of simple parts and rules interacting in a system may lead to structures and behaviours on the whole that cannot be reduced or recreated using simple, linear logic. The flexible, evolving hierarchies of these systems are what make them interesting, but also hard to study without considering dynamics. Their sub-components, or ‘agents’ tend to exhibit diversity and a degree of self-directedness, meaning their interactions can be non-trivial and preferential, their collections unique. They tend to exhibit *chaos* for this reason, in part because of these many competing goals and convoluted feedback loops. Dittrich (2014, p.20) aptly describes chaos as “an elevator permanently lifting information from small to large dimensions”, the opposite of which we might consider *dissipation*. This means we ought to consider how small effects or variations in the initial conditions can be propagated to far larger scales.

Proteins as Sub-Units for Self-Assembly

To give a concrete example relating to biology, I will briefly discuss proteins. Proteins are strings of amino acid ‘building blocks’ that bend, twist and fold on themselves over four different structural levels. As a result of molecular forces arising between their parts, they form a huge range of shapes with different functional properties. Proteins are coded by genes in our DNA, and assembled by machinery called ribosomes (that, in a recursive way, are themselves made out of proteins and genetic material). Proteins often function as enzymes, which catalyse other reactions and fundamentally allow systems of life to be (self-)sustaining. Herein lies the remarkable feature of proteins, the tight intrinsic mapping between gene and geometry, software and hardware. To give one example, whilst visualising protein sequences for this project (*Figures 1-3*), I spotted a family of plant proteins with a characteristic repeating pattern in their sequence (Cheng et al., 2017). Incidentally, it is this repeated string of ‘proline’ units that gives the EXT-11 protein a characteristic spring-like shape. In turn, this structural property informs its function, where it provides selective elasticity in the plant cell wall. Proteins can then form complexes of millions of sub-units, such as those found in cell membranes, which further demonstrates this ‘upwards cascading’ ability for self-assembly, when short genetic codes are combined with biophysical interaction dynamics. As detailed before, chaotic, non-linear behaviour can be induced by feedback loops and chain reactions, which is the case for gene regulatory networks, where a two-step control process (transcription and translation) means a single mRNA template can be used to create many protein copies.

Figure 1. *Arabidopsis Thaliana* protein sequence with proline-rich regions
Araport11 Dataset (Cheng et al., 2017)

```
MSLVPPPLILSPPSSNSSTAPPPLQTQPTTPSAPPVTPPPSPPSPPVSSSPPPPVSPPSSSSPPPPVITSP
PPTVASSPPPPVVIASPPPSTPATTPPAPPQTVSPPPPDASPSPAPTTNPPPKPSPSPPGETPSPPGETPSPPKPS
STPTPTTTSPPPPATSASPPSNPTDPSLAPPPTPLPVVPREKPIAKPTGPASNNGNNTLPSSSPGKSEVTGGIVA
IGVIVGLVFLSLFVMGVWFTRKRKRKDPGTFGYTMPPSAYSSPQGSDDVLFNSRSSAPPKMRSHSGSDYMYASSDSGMV
SNQRSWFSYDELSQVTSGFSEKNLLGEGGFGCVYKGVLSDGREVAVKQLKIGGSQGEREFKAEVIEIISRVHHRLVTLVG
YCISEQHRLLVYDVVPNNTLHYHLHAPGRPVMTWETrVRVAAGAARGIAYLHEDCHPRIIHRDIKSSNILLDNSFEALVA
DFGLAKIAQELDLNTHVSTRMGTFGYMAPEYATSGKLSEKADVSYGVILLELTGRKPVDTSQPLGDES LVEWARPLL
GQAIENEEFDELVDPRLGKNTIPGEMFRMVEAAAACVRHSAAKRPKMSQVVRALDTLEEATDITNGMRPGQSQVFDSRQQ
SAQIRMFQRMAFGSQDYSSDFDRSQSHSSWGSRDQSRFVP*
```

Figure 2. Visualised protein sequences with proline-rich regions

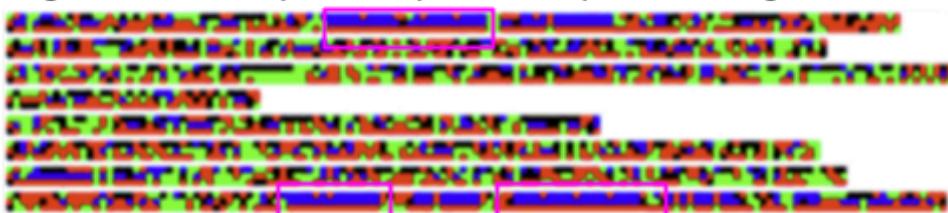


Figure 3. Approximation of EXT-11 protein in 3D, with a spring-like section



The Cellular Automata Modelling Strategy

Clearly biological systems have deeply nested, complex structure and behaviour. To adequately model them, we need a simulation strategy or ‘grammar’ that is equipped for local interactions that lead to much larger global patterns. Many different abstract modelling grammars have been developed that can be used, ranging from L-systems, which aptly capture (approximately) fractal growth rules like tree branches, or finite-element models (FEM) that use polygons to approximate expansion in tissues in a deformable mesh. Cellular automata (CAs) are a unique simulation grammar explored here, somewhat poised between particle-like agent-based models in complex systems science and lattice-like finite-element models in engineering. It is vital to study the spatiotemporal patterns, building models that capture both spatial organisation and structural nesting alongside behavioural dynamics, in line with Functional-Structural plant (FSP) strategies used in the modelling of plants (Wang et al., 2018). Mitsanis, Hurst and Tekinerdogan (2024) argue against solely data-driven models in the development of digital twins, because they lack generalisability, and “preclude biological explanation of the prediction”. CA provides a flexible and unbounded grammar on which to scaffold realistic biological rules.

The CA grid stores the state of individual cells, with the state at each step computed by transforming the last with a local update rule applied over a neighbourhood. Over many steps, global patterns may emerge, allowing the manageable study of complex or chaotic dynamical systems. This simulation strategy was developed to mimic the mechanism of cellular communication, rooted in the ‘artificial life’ intrigue of mathematicians, with early computer scientists like Alan Turing working on morphogenesis until his death in 1954, and John Von-Neumann and Stanislaw Ulam working on rules for machine self-replication at Los Alamos at the same time. The Santa Fe Institute (SFI), founded in the 1980s, in part by ex-Los Alamos scientists, continued to explore big questions surrounding the origins of life and artificial forms. Notably, the work of Chris Langton of SFI (*Langton’s Ant* and *Langton’s Loop*); John Conway, (*Conway’s Game of Life*); and Stephen Wolfram, (*Wolfram’s Elementary Cellular Automaton*) have seeded important developments in the models discussed here when coupled with cutting-edge machine learning research. It is worth mentioning that there is an inherent architectural similarity between cellular automata and convolutional neural networks, due to the way these NNs perceive by computing local filters on each input pixels and their neighbours, and this means they can be used to recreate CAs with rules that are learned from video inputs (Gilpin, 2019).

3 | Leaf Morphogenesis

In this section a summary of the important features of the case study, leaf morphogenesis, will be provided. Plants, being unable to move around, must use sophisticated regulatory mechanisms to alter their growth to highly-changeable environmental conditions, showing a high degree of plasticity, different to animals, even at the sub-individual level: leaves near the ground level that are frequently browsed by herbivores regrow to become more prickly in shape than the rest of the tree (Herrera & Bazaga, 2013). Growth stability is therefore balanced by highly-modular flexibility and diversity of sub-components, using mechanisms like epigenetic switching, or DNA methylation which serve as a kind of ‘memory’ of local conditions. Although appearing sessile, at the subcellular level there are in fact surprisingly non-trivial dynamics, such as the complete remodelling of the intracellular scaffolding in time periods as short as 30 minutes. Stable growth of their tissues at the cellular level involves carefully controlled patterns of cell division, specialisation and elongation, mediated by complex biochemical signalling pathways, as well as the lesser-understood mechanical factors.

Leaf Formation

Leaves emerge from the tip of the growing tip of the plant (*Figure 4*), a distinct region called the shoot apical meristem (SAM). Here cells proliferate in an undifferentiated state, as stem cells. This dome-shaped region has three distinct zones of growth (Armenta-Medina & Gillmor, 2019, p.519):

- Organising centre (source of slow-dividing stem cells)
- Central zone (rapidly-dividing stem cells from organising centre)
- Peripheral zones (where new leaves form)

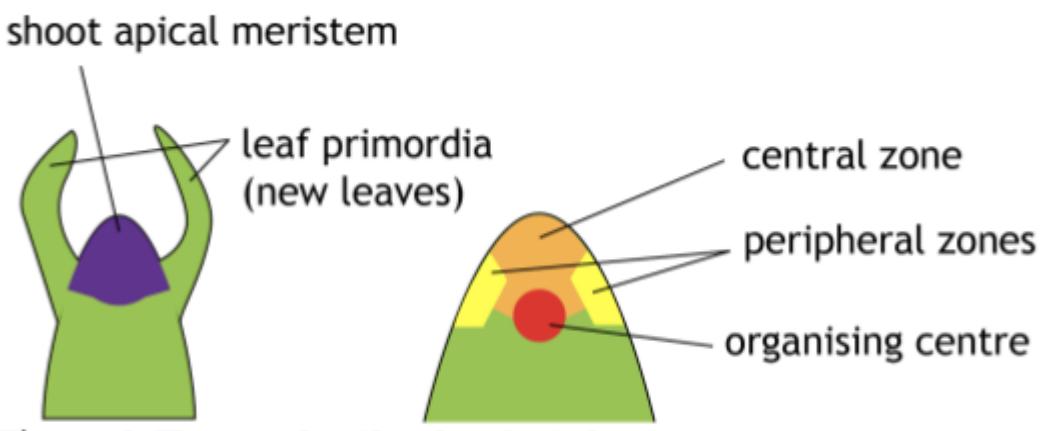


Figure 4. The growing tip of a plant shoot.

Division and Elongation Patterns

Plant cells, unlike animal cells, tend to expand in one direction, meaning complex geometries are constructed by the precise coordination of their orientation as they divide and elongate. When plant cells divide, they typically follow the ‘shortest wall rule’ (*Figure 5*), dividing along their shortest axis (von Wangenheim et al., 2016). In general the division plane orientation is more tightly controlled in early plant development, occurring at alternating perpendicular and parallel patterns to the tip axis in lateral roots. In leaves, the precise timing of cell division and expansion phases, which occur somewhat sequentially around the leaf tissue, influence not only the size of individual cells, but also the overall leaf size and shape (Fox et al., 2018). Specific proteins like LIM1 act as a growth inhibitor, modulating this transition from cell division to cell expansion (Vuolo et al., 2018), such that mutants without it show decreased leaf complexity.



The ‘shortest wall rule’ for dividing cells

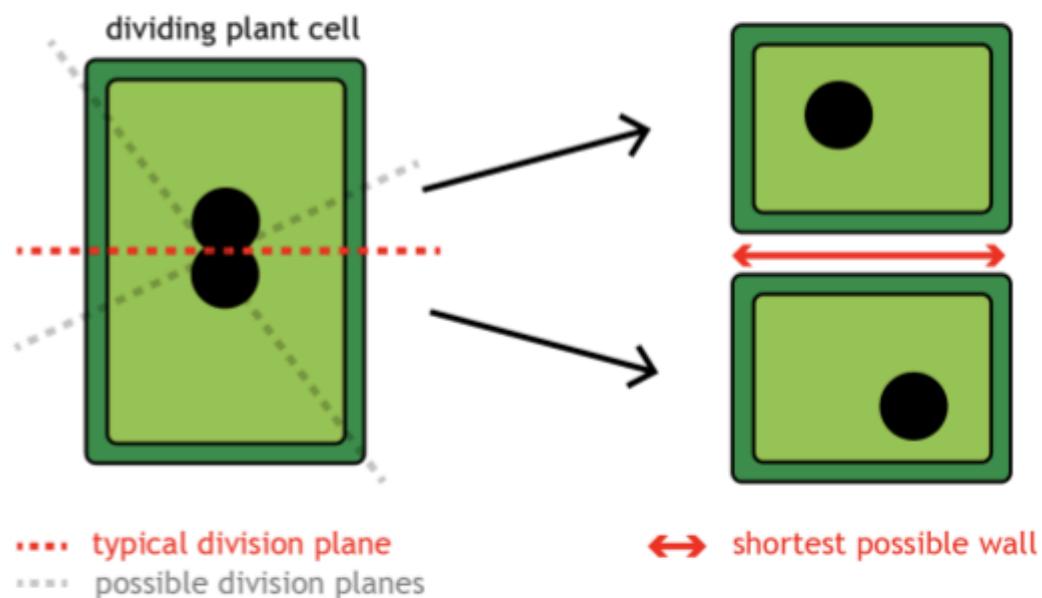


Figure 5. Dividing plant cell.

Conformal Growth

First hinted by physicist and computer scientist Stephen Wolfram in 2002, Mitchison (2016) shows that *Arabidopsis* leaf growth can be explained with an accuracy of 89% in real leaves by a ‘conformal map’, where expansion is “locally isotropic (the same in all directions) but the amount of expansion can vary with position”. To satisfy the conformal growth model, local angles, but not distances have to be preserved in a particular kind of ‘stretching’ operation without shearing, meaning cells cannot shift around or reorder. By applying this geometric transformation to the leaf, macro-scale rotation is induced as they move apart, caused by slightly differing rates of expansion at the micro-scale. Although this 2D mathematical model presently neglects well-known underlying biochemical explanations like polarity patterns and auxin gradients, and finer details of real leaves like serrations, it is interesting to consider because it supports the fact that modulating local expansion through biochemical means, can propagate a smaller, tightly-organised arrangement like a leaf primordium to much larger scales simply by modulating expansion. It seems it would be more efficient for a leaf to ‘stretch’ a locally-established cell division pattern through subsequent cell expansion, rather than extending it in a piecewise fashion as we would presently construct a building.

Polarity Patterns Provide Synchronisation Cues

The development of intricate yet predictable leaf shapes through the division, elongation and specialisation of a great number of cells is only possible if they do so in a highly-coordinated manner. Cells therefore require local position and orientation ‘cues’ to globally synchronise their behaviour. In animals and plants, the concentration of various growth chemicals (or ‘morphogens’, as coined by Alan Turing) provide such instruction by influencing individual cells’ gene expression. The gradients of diffusing chemicals has the effect of establishing cell and tissue-wide ‘polarity’, a kind of cellular compass that ensures tissues consistently develop a coherent pattern.

The most important growth hormone across plants is a class called auxins, which are found in the shoots and roots of plants, stimulating their growth. The local concentration of auxin influences cell-level gene expression, in complex feedback loops that further direct its flow (Bhatia et al., 2016). Within cells, proteins like PIN1 and BASL are co-expressed, ‘localising’ or accumulating in particular regions of the cell, at various stages of growth, in response to the auxin gradient, as well as high mechanical stresses in the cell wall (Mansfield et al., 2018). These mechanisms together establish a very stable kind of directional cue for growth at a cellular level, enforcing a tissue-wide polarisation of all three leaf axes that enables them to correctly differentiate (Bhatia et al., 2016; Kierzkowski, et al. 2019). Interestingly, in the case of stomata, the pores on the bottom of leaves for gas exchange, polarisation is highly-responsible for their complex geometry, since the accumulation of BASL proteins to one side of the cell signals an unusually

asymmetric division, making the daughter cells specialise differently (Cammarata & Roeder, 2018).

Another kind of synchronisation is seen in the periodicity of metabolic cellular activities. We know that plants, and not just animals, have a kind of biological clock, or circadian rhythm, as a “spatiotemporal product of many weakly-coupled oscillators” (Rascher, et al., 2001, p.1). We know from organisms like fireflies, that collectives of individual oscillators can quickly achieve global synchronisation by adjusting their timing following only their closest neighbours’ cues. In leaf tissue, cells appear to continually transition between tissue-wide synchronisation and desynchronisation in their metabolic rate depending on light conditions, as a result of their interactions. These transitions between rhythmic, synchronised states and arrhythmic states with “strongly heterogeneous isolated patches” are also seen in cardiovascular patterns or slime mould signalling (*ibid.*, p.5). This can be seen as a collective intelligence mechanism for modulating the overall photosynthetic efficiency of the leaf and perhaps (more speculatively) an endogenous timing signal for other mechanisms in the tissue, considering mechanistically the importance of a global ‘clock’ for engineered systems like computers. Developmental plant models should therefore include external cues from other tissues or the environment for accuracy.

A Complex Interplay of Biochemical Interactions

Well beyond the scope of this report, a vast network of “transcription factors, microRNAs and hormones” influence cell division, elongation and differentiation alongside auxin (Terebieniec, 2019, p.5). As we see across biology, there tends to be multiple ways to get to the same result - the genetic pathways have high redundancy, otherwise individual mutations would be especially damaging. Studying mutants, or ‘edge cases’, plants with altered expression of particular gene which may exhibit (fatal) growth deformities, is a central part of genetics research, since it allows us to determine correspondences between micro-level genotypic and macro-level phenotypic variation, at least in theory. Isolating genetic effects remains challenging in reality, since these factors are so often co-expressed: as Terebieniec (2019, p.42) concludes “leaf shape variation is a highly complex, polygenic trait under the control of a very large number of small effect size loci”, meaning there is rarely singular causes for different leaf phenotypes. We should therefore be cautious seeking parsimonious or univariate explanations in genetics. Moreover, a significant portion of the research data available is towards the mapping of the model species, *Arababidopsis Thaliana* (a mustard-family weed), which is used for its rapid growth and fertility, and small genome for easy genetic modification. In the future, with improving imaging and genetic sequencing technology, it will be easier to collect and analyse morphological variation across a wide range of species, not least because a lot of arable crops we wish to understand better are monocots (plants with one initial leaf), not dicots (two initial leaves) like *Arabidopsis*. The current big data and deep learning trend

could provide powerful explanations of complex genetic mechanisms than mutants alone if they are interpretable.

Phyllotaxis

Leaves and flowers tend to form in predictable angular arrangements, spaced-out around the stem, a patterning mechanism called phyllotaxis, which has excited mathematicians for centuries with its mysterious regularity, and the occurrence of fibonacci numbers. In 1952 pioneering computer scientist Alan Turing's 1952 paper, introduced the 'reaction-diffusion' model dynamical models of chemicals diffusing to describe how developmentally how zebras get their stripes, or cheetahs their spots. This model of chemical signalling is still making waves in biology today. Fujita et al. (2011) now use it to recreate phyllotactic patterns in the SAM. At a cellular level, we know it is the polar distribution and transport of auxin (with positive feedback from the expression of PIN proteins in the cell membranes), that establishes periodic auxin 'hotspots' in the SAM, and consequently leaf outgrowths (Bhatia et al., 2016; Bhatia & Heisler, 2018; Yonekura et al., 2019). Equally, genes are expressed to mark the boundary between the growing leaf surface and the stem, providing a signal for the leaf to begin differentiating into a flat outgrowth, without which growth defects are known to occur (Terebieniec, 2019). As vascular tissue (veins) form within the leaf, the auxin gradient is strengthened, feedback that prevents the growth of other new leaves nearby. Further growth cues are known to involve the cellular sensing of mechanical stress in their walls, which leads cells to reinforce their boundaries with tensile fibres to counteract these forces. The orientation of the strengthening is consistent with the biochemical polarity pattern, suggesting these mechanisms are interlinked (Bhatia & Heisler, 2018).

Mechanical Factors Also Influence Growth

Alan Turing, in his early morphogenesis model, acknowledged that the overall growth process is "of formidable mathematical complexity", where the "interdependence of the chemical and mechanical data adds enormously to the difficulty", justifying his choice to focus to initially neglect cell mechanics (Turing, 1952, p.3). Unlike animal cells, plant cells have substantial internal turgor pressure, requiring a rigid cell wall to prevent bursting. For a plant cell to expand, its cell wall must be selectively weakened in spots. The inner volume of a cell, the cytoplasm is not just vacant fluid (cytosol), but also occupied by suspended organelles and a dense, supportive polymer meshwork called the cytoskeleton. This inner scaffolding is not at all fixed, but instead made up of highly-motile, dynamical elements, called microtubules, which perform essential, directed functions like pulling apart chromosomes in cell division; structural support against compression; and effective transport of large cellular materials (like wall-reinforcing substances) to the right locations faster than diffusion (Howard, Grill & Bois, 2011). They are common structures to animal, plant and fungal cells, and although their general dynamics are understood to be "multi-tiered process that involves coordination between the levels of the various

severing proteins”, they are involved in more specialised mechanisms too (Baas, Karabay, & Qiang, 2005, p.522). In plants they arrange in carefully-oriented cortical bands, giving the cells anisotropic mechanical properties and hence determining the direction of elongation; they also provide a fibrous substrate to grow new walls during division (Kost & Chua, 2002).

Mechanical sensing and stress distribution within and between cells, leading to their rapid remodelling, provides a means for directed tissue-wide signals to propagate and equilibrate. The tessellation of highly-lobed, interlocking puzzle cells that gives leaves their familiar surface pattern, is in fact a means to resist significant tensions in the tightly-packed epidermis (Sapala et al., 2018) (*Figure 6 & 7*). Recent research highlights the importance of low-level cytoskeletal reorganisation in coordinated growth, alongside biochemical factors. Some even suggest the overall tissue shape may show a kind of self-similarity to the organisation of cortical microtubules inside individual cells (Boudon et al., 2015), though models are rarely developed with sufficient granularity to adequately capture microtubule dynamics. The dynamic instability of microtubules means they are constantly built and broken down, which points to a fundamental low-level mechanism for robust self-assembly and optimisation. Goodson and Jonasson (2018, p.16) identifies that “the combination of random probing and selective stabilisation, as seen in the microtubule cytoskeleton, is a theme that occurs throughout biology”. This makes them an adaptive network that can “respond quickly to changes in the environment and the influence of regulatory proteins” (*ibid.*, p.12).

Figure 6. Vertical cross section of leaf.

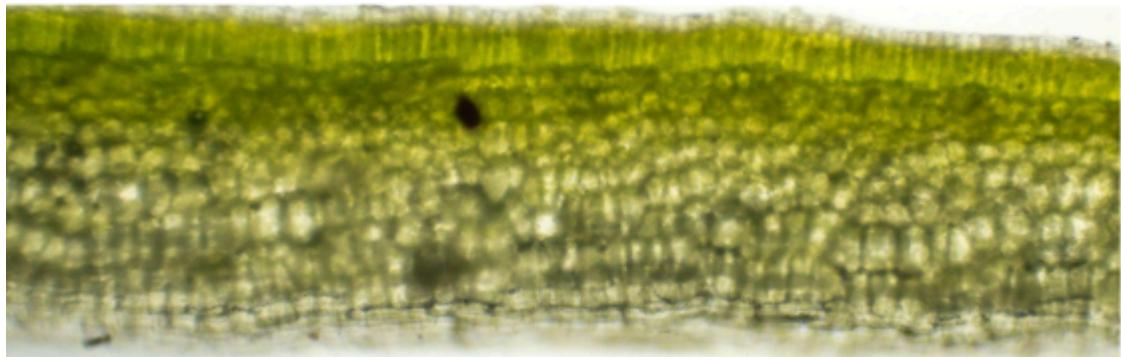
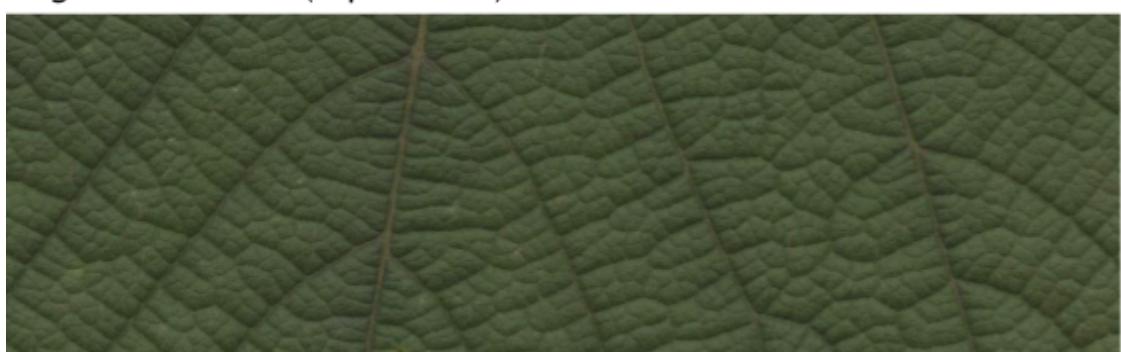


Figure 7. Adaxial (top surface) of a leaf.



4 | Simulation Design and Development

As discussed above, the interplay of micro-level biochemical and mechanical as well as macro-level environmental conditions have been shown to play an important role in plant and animal morphogenesis, whether that is tumour growth or the phenotypic variation in leaf shape due to climate. It makes sense to consider all parts of the organism, too, because in the case of plants, especially when situated in competitive environments, it is apparent that root conditions greatly affect the shoot growth (Evers et al., 2018). An efficient and extensible multiscale (and multi-modal) simulation is needed in tissue modelling to accommodate such overwhelming complexity and sparsity in available granular data. This could mean integrating a reaction-diffusion style biochemical simulation with a physics-based cell mechanical model (Runser, Vetter, & Iber, 2024). Uhrmacher, Degenring & Zeigler (2005, p.66) helpfully lay out four ‘axes’ of modelling considerations when approaching biological systems: “continuous and discrete, quantitative and qualitative, stochastic and deterministic”; and of course, micro and macro, that I will address in this section about design principles.

Design Principles

From the evidence outlined above, a discrete, pattern-oriented and qualitatively interpretable design is chosen. To enable model construction by a wide range of researchers qualitative modelling is preferable over quantitative, since it can be directly trained on biological samples, negating the development of sophisticated, verbose physics solvers and application-specific validation tools. A discrete approach (with a short-ranged continuous approximation) is utilised, since the level of quantisation can be set at various intervals (e.g. protein, cell, tissue, plant), but also simulations easily nested within one another in terms of the software architecture. In developing simulations that span a wide range of scales, continuous approaches would allow trivial, fluid interpolation between levels, where increasingly fine-grained discrete levels can certainly become computationally intractable. This is one clear benefit of the continuous approach, especially with the potential of novel machine learning techniques like neural operators, which are ideal for multiscale applications, since they can generalise and work with flexibly-sized inputs unlike traditional neural networks (Azizzadenesheli et al., 2024). To model many real-world systems, with a high degree of structural nesting, where there may be recurrent but not truly self-similar or fractal features, it makes more sense to use discrete methods. With adequate *adaptive optimisation*, assembling a multiscale discrete model with cellular automata allows a wide range of models to be integrated by different researchers in a piecewise manner. Such a flexible model would allow multiscale calibration, using the limited data we have on systems to constrain each level, in line with POM principles about the strength of many weak patterns. The wiring between levels enables short-range interpolation between simulation scales for continuity. Much like the attention mechanism of generative ML models, which mimics how we complete a sketch

by refining faint lines bit-by-bit, rather than a whole image at once (Gregor et al., 2015), this approach will be made computational performant with an *adaptive granularity* mechanism, allowing the modeller and solver to move between scales based on the mechanism in question. We may dynamically decide how much detail (at which spatial and temporal steps) we should calculate solutions, rather than assuming accuracy spatiotemporally consistent.

In this preliminary design, a modular neural cellular automata prototype is used to evaluate the viability of such an approach at a single scale, so that further studies may include diversity in agent-types, since each cell is self-contained, and therefore crucially extend to further scales. Current NCA models tend to apply a neural network to the whole grid, which decreases portability. More broadly, by taking a machine learning approach which is by nature more data-driven and pattern-oriented than rule-based, capturing only very simple and abstract morphogenetic rules, this might initially do away with structural realism for biochemical and biophysical models (e.g. quantifying contact forces) but its strong advantage is scalability towards the goal of engineered emergence and the opportunity to better assess which rules are really needed. Using software engineering principles like encapsulation and inheritance to design the NCA framework allows robust self-similarity and avoids code boilerplate in stacking models depthwise and breadthwise. The ability to compose or reuse individual features on the same interoperable base class is favourable, since some layers may be naive or rule-based CAs and others can be driven by neural networks, and the substrate channels can then be trivially extended to represent new and more realistic interaction pathways, quantities and states, like auxin concentration or turgor pressure, each with their own properties. Specialists may more easily tackle different facets of the modelling problem this way, too.

On the point of stochasticity, deterministic models are preferable for their traceability and therefore explainability. Randomness is often a premature assumption or placeholder in complex models, but would ideally be omitted, seeking explanation instead in the propagation of chaotic effects up and down model levels instead. Even suspiciously parsimonious abstract mathematical descriptions like fractals with chaotic features can end up falling back on naive randomness. In a well-known example, mathematician Michael Barnsley approximates the frond of ferns using fractal growth, which is contended by biologists for the fact these macro-scale features only generalise over a few scales unlike true fractals (Lev-Yadun, 2012). In their paper, which lacks biological reference, Barnsley, Hutchinson and Stenflo (2003) consequently overcome the rigidity and missing realism of their simulated fern leaf by injecting randomness into the fractal generation process, but then their fractal approach is somewhat capped in further insight. Nevertheless randomness can first reveal where more explanation or attention is needed. On a practical point, if randomness must be used, it can be made more reproducible by specifying fixed random seeds in Python libraries like Numpy and PyTorch, sticking to (chaotic) pseudo-random generators used sparingly.

Related Works

Cutting-edge simulation frameworks, such as Goo (Ruzette & Antoine 2024) and SimuCell3D (Runser, Vetter, & Iber, 2024) allow realistic modelling of animal cells in 3D, taking into account mechanical factors like cell-cell adhesion, elasticity, surface tension but also contact forces and tissue constraints that influence the overall geometry to form realistic sheets or tubes as well as balls of cells. Interestingly, SimuCell3D also provides the capability to model the effects of internal cell volumes like the nucleus: this is even more important in plant cells, which have large vacuoles under substantial turgor pressure. In conversation with Steve Runser, one of its creators, he described the computational demands of modelling the contact between cells as being the remaining bottleneck. This requires the use of computationally-intensive explicit numerical solvers where less accurate models neglecting these factors can afford to use implicit time integration with larger time steps, which would make an adaptive time-stepping mechanism relating to the amount of cell activity very beneficial (Fletcher, & Osborne, 2022). Uhrmacher, Degenring & Zeigler (2005, p.71) describe how a discrete model using “predicted threshold crossing rather than [fixed] time steps” can not only approximate the continuous behaviour of differential equations sufficiently, but achieve this in a much shorter time. Equally, continuous approximations have already been demonstrated in CAs with highly-realistic CA simulations like Lenia, which make use of continuous transition rules and smooth time-stepping (Chan, 2020). Runser had also mentioned the importance of collecting real datasets at cellular resolution, specifically granular force data for modelling mechanically stresses, the measurement of which admittedly remains a practical hurdle. More generally, when parameters are difficult to measure, we may also be able infer them from experimental data, calibrating models using techniques like Abstract Bayesian Computing (ABC) (Fletcher, & Osborne, 2022).

Considering specifically the use of CAs in tissue modelling, they have proved effective in simulating different types of tumour growth (Valentim, Rabi & David, 2023) or liver damage, where Adhyapok et al. (2020) notably use a hexagonal grid and transition rules with varying time scales and stochasticity, to enforce realistic, competing dynamics in damaged tissue. This model elegantly captures how cell stresses spread to healthy cells in places, whilst pockets of healthy cells may also proliferate to fill the spaces left by dying stressed cells, a kind of mechanical sensing. Multilevel CAs however remain relatively unexplored, particularly in biological modelling for the challenge they present. Authors such as Diaconescu, Tomforde & Müller-Schloer (2018, p.2) demonstrate an abstract ‘holonic’ CA architecture, explaining how multilevel CAs need top-down orchestration alongside bottom-up self-organisation to create true “macro-micro couplings”. Richardson et al. (2024, p.24) highlight multilevel NCAs as a worthwhile direction for exploration in their experiments, taking note of their current limitations, such as fixed time-stepping and coarse grids, but also the importance of facilitating top-down orchestration for systems “with non-local (or multiscale) interactions that cannot be elegantly explained with purely local rules”. In a preliminary study they demonstrate the ability of CAs to first learn

idealised systems of PDEs and Turing patterns, towards (biological) data-driven applications in the future.

Prototype Design

The final NCA prototype consists of a grid of hexagonal pixels approximating a collection of proliferating cells, each of which may only communicate with their nearest neighbours. At each time step of the simulation, an update is applied incrementally to each cell, a kind of non-trivial weighted sum, aggregating their own and their neighbouring cells' current states with some function that mimics biochemical intracellular activities, and more importantly, their intercellular signalling. Observing the simulation over many time steps, these numerous cell-to-cell interactions may lead to pattern explosion, collapse, or the evolution of stability, but only when the update function is calibrated correctly. An artificial neural network within each cell is used to learn the non-trivial update function which enables their successful coordination for pattern growth - much like a real cell's gene regulatory network.

It may seem trivial that each cell could 'recite' a stored pattern from memory, however, as with real genetic code, this would not only be very inefficient to store in a genome, but cells would need *absolute* location and orientation cues to know which part of the pattern to recreate. Instead, we can apply a significant compression, by defining growth rules *relative* to neighbours in feedback loops, in an iterative or *recursive* fashion, so that genetic code in theory consists more parsimoniously of algorithmic rules for their collaboration, correctly filtering and respond to various cues from neighbours, so that the desired pattern emerges as a by-product of their many interactions with feedback. Hence an overall biochemical pattern is unpacked and unfolded across a multicellular substrate, working in external constraints, so that it is only by evolving hundreds of interactions with thousands, millions or billions of other cells that the original pattern is decoded, even though the target pattern is unknown to the individual cell.

Neural Cellular Automata Implementation

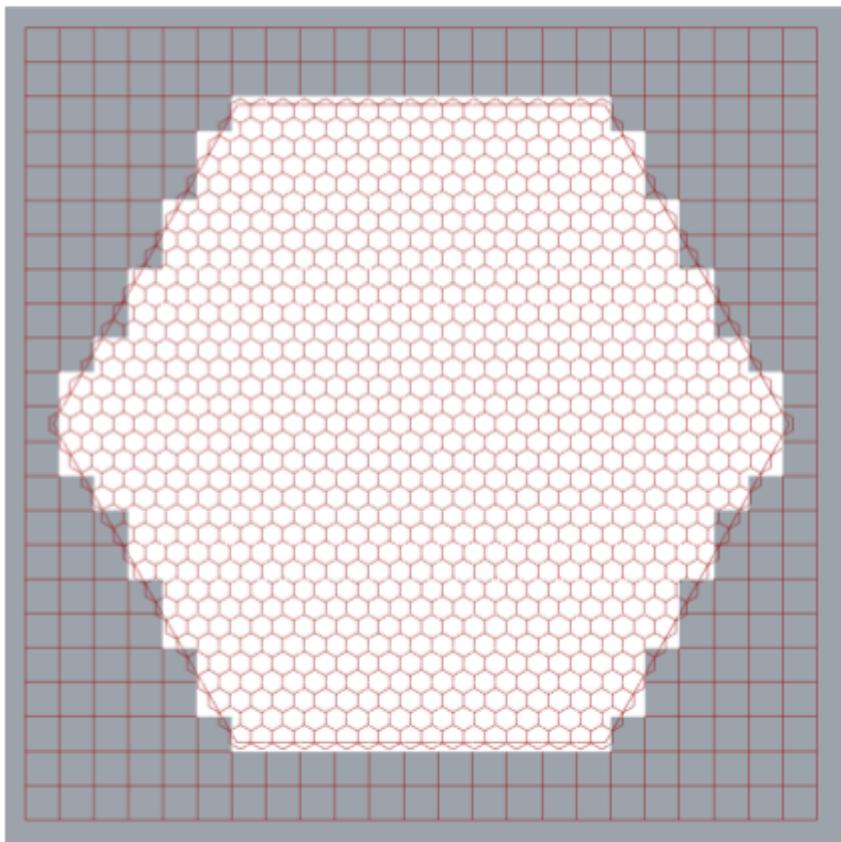
Neural cellular automata were actually proposed (in early form) as far back as 1988, with the context of implementing analog-like circuits in a digital setting for their unparalleled "real-time signal processing capabilities" (Chua and Yang, 1988, p.1), although their practical realisation is far more recent. By coupling a neural network with a CA grid, a (residual) recurrent neural network is essentially constructed (Mordvintsev et al., 2020). More broadly, this is an increasingly used artificial neural architecture that most accurately mimics our own, and surpasses convolutional neural networks in Turing completeness, since they allow for non-trivial feedback, cyclicity. This intuitively entails a kind of internal 'memory' or 'thought process', more easily visible to us on the CA grid rather than inside a 'black box' neural network, allowing us to yield functional insights directly from experiments, like the polarity establishment in the hidden channels, shown

later in this section. Recurrent NNs are perfect for time series data, but also image generation tasks where images are generated sequentially, by focusing on particular regions at a time with a kind of selective attention mechanism (Gregor et al, 2015).

The Python-based NCA architecture developed by Mordvintsev et al. (2020) provides the basis for the NCA prototype presented here. Their guidelines and architecture will be detailed briefly in this section. The simulation begins with a single living pixel, where a NN with a convolutional layer is trained to learn an incremental update rule applied to a random subset of cells at each CA step and eventually grow the pattern. The NN is initialised to do nothing at first. By applying an ‘update mask’, only some cells are actively communicating at a given time, so the neural network cannot rely on implicit global synchronisation between cells, having to learn with redundancy. Importantly, the image alpha channel is similarly used to create a ‘living mask’, such that only cells with living neighbours may communicate. The values of all other cells are zeroed during the update. The neural network uses sobel operators (an edge-enhancing image filter) to aid the cells’ *perception* of its neighbours’ signals across the 16 image channels in use, visible and hidden. Catrina et al. (2024) identify that successful NCA training is very dependent on the contrast provided by outlines in training images, since the ‘stop’ conditions to maintain stability and prevent explosion after the growth phase is a significant obstacle. Bold edge colours and edge-finding filters improve the NCA’s ability to reproduce patterns, without which (in my early experiments) only a solid colour target could be learned. A high-contrast, low resolution image of emojis provides the target for the RGB channels, and the neural network must learn how to use the other ‘hidden’ channels for cell-cell signalling in order to achieve stable growth. The goal of stability is reached by the subsequent use of an evolutionary pooling strategy (simplified genetic algorithm) alongside gradient descent (Catrina et al., 2024; Mordvintsev et al., 2020).

To improve robustness to damage, as real tissues would be capable of self-repair, regions are erased during training. The model is also made to be resilient or invariant to rotation by presenting many different orientations, as this the topology of the (square) grid can influence growth. It is found that patterns seem to grow fastest and preferentially diagonally, which is an interesting limitation seen in Game of Life too (Catrina et al., 2024). To model plant tissue, I experiment instead with hexagonal grid cells, thereby reducing the neighbourhood to 6 neighbours per cell (reduces the number of computations and memory usage slightly), halfway between a von Neumann and Moore neighbourhood, but have the advantage of “higher degree of symmetry and equality of distances to neighbouring pixels” (He et al. 2010, p.243), which limits growth reliance on the underlying grid topology. I therefore had to implement a hexagonal variant of the sobel operator for perception, which He et al. detail. Hexagons are a compactly-nesting shape found across nature, making it a good choice for problems where space is limited (*Figure 8*).

Figure 8. Hexagonal NCA grid



Key Findings From Early Experiments

Before working towards the hexagonal, modular NCA prototype, a more basic version was first implemented in PyTorch to learn more about the capabilities and limitations of NCAs, developing an intuition for suitable ‘hyperparameters’ that govern the NN training regime. Very simple time-series images were initially used as synthetic training datasets (*Figure 9*), in this case an animation of an expanding cell, generated using 3D modelling software *Grasshopper* in *Rhino*, and then downsampled to the CA grid dimensions.

Figure 9.

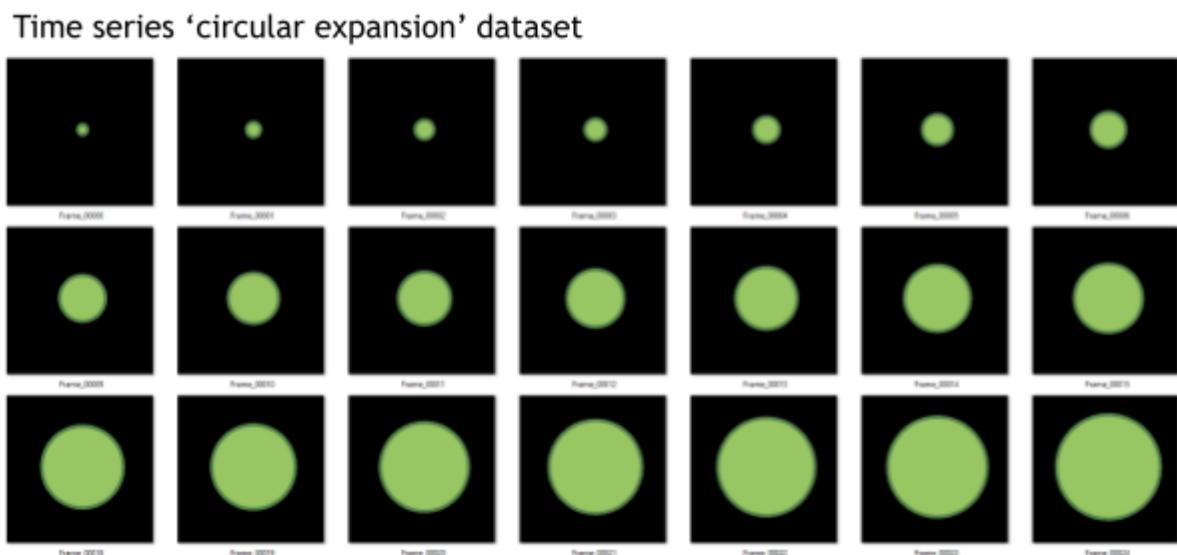
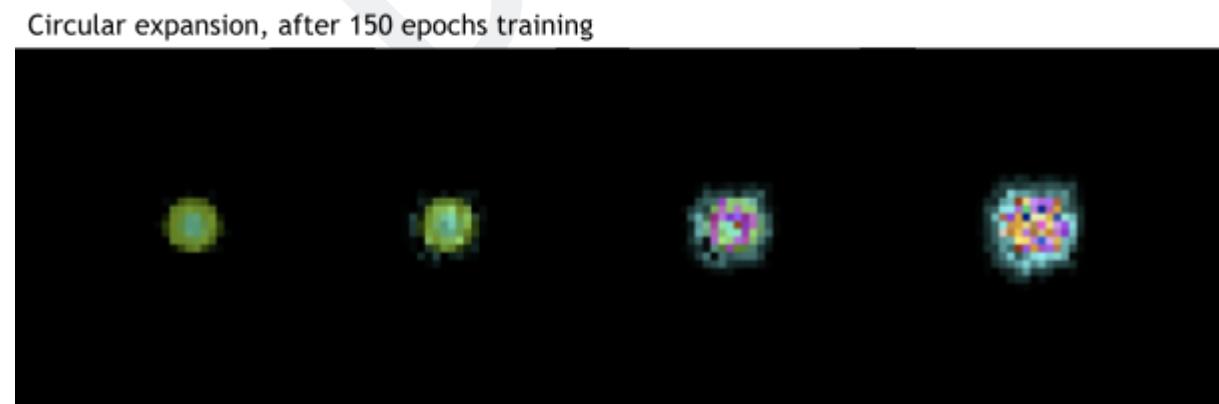


Figure 10.



Training quality (*Figure 10*) was found to be unexpectedly poor on smaller grids, perhaps because there are very sharp changes in colour values due to downsampling. The activations for machine learning models can have a big impact on training quality - here it leads to more directed, or diffuse growth. *ReLU* activation was found to be fast, but more unstable, in comparison to *tanh* or *GeLU*, the latter providing the most consistently high-quality results.

Each training run involves evaluating the simulation for a number of steps (within a random, predefined range), in forward passes of the neural network which decides how best to update each cell in the grid. After assessing the current loss, and performing backpropagation, this is repeated a few more times, before starting an entirely new simulation run. Frequent backpropagation, using small step sizes, leads to more stable training but was found to bottleneck overall growth accuracy and stability. Applying gradient normalisation after each step mitigated the exploding of gradients (and hence the loss), which allowed larger step intervals to be used (Mordvintsev & Niklasson, 2021). When the number of overall training time steps is sufficiently large (300-500) it was found that patterns had reasonable stability and damage tolerance (*Figure 11*), without the need for additional training protocols as other authors mention. In terms of the hidden channels, more were useful, but this entails significantly longer training times, since the model is bigger and has increased degrees of (combinatorial) freedom. It must eventually learn how to use them all in a targeted and coordinated manner. Visually, the hidden channels develop sharper and more stable local features as training progresses, becoming differentiated in functionality, whilst initially showing significant redundancy.

Figure 11. Simplified Leaf, after 3820 training epochs

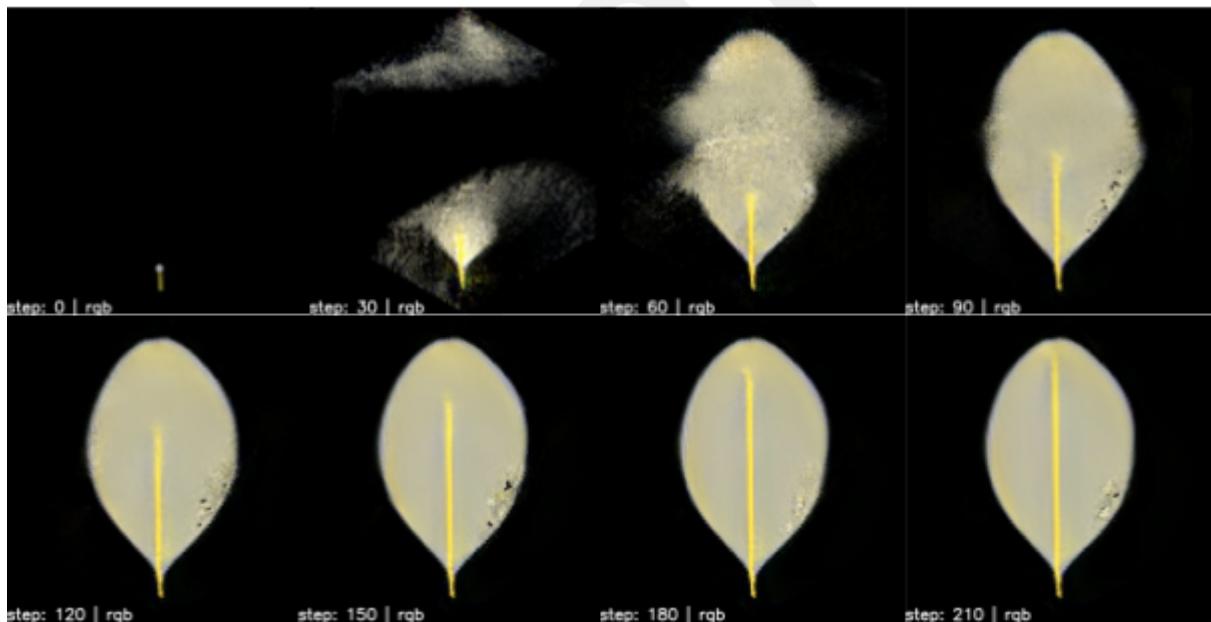
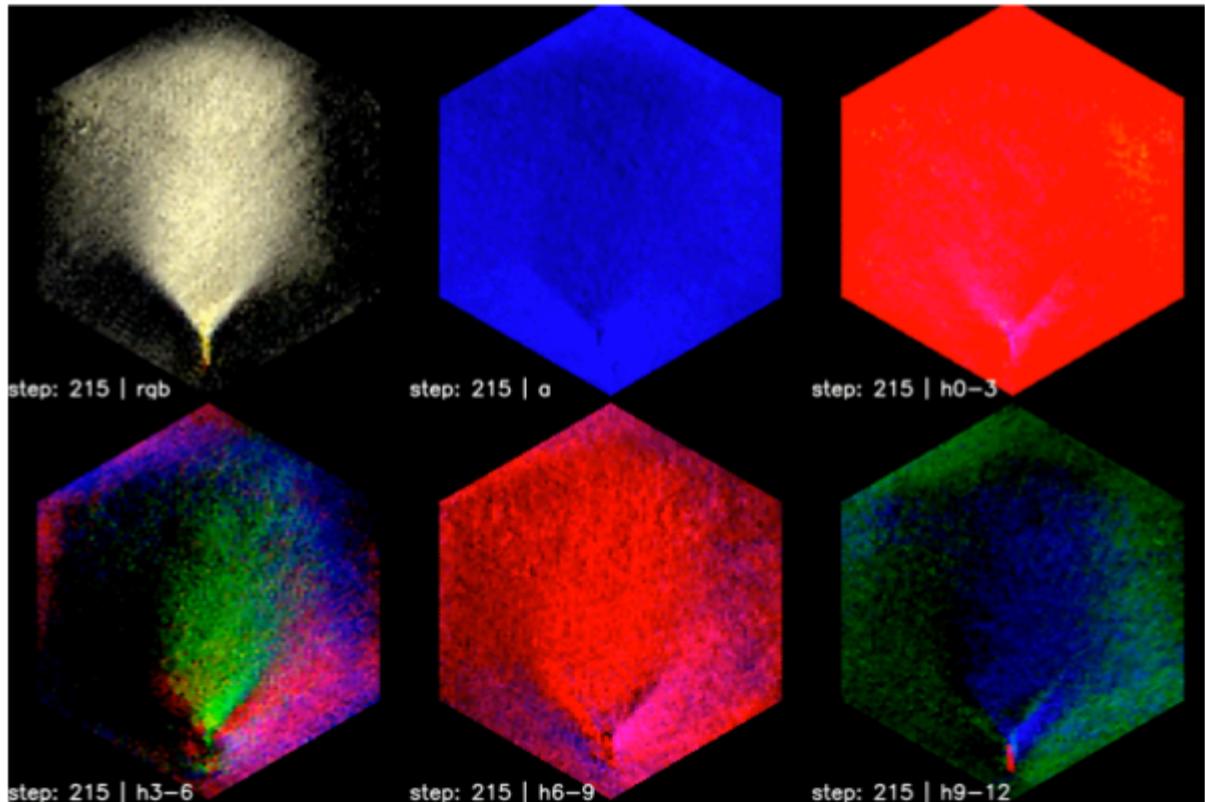


Figure 12. Image channels during simulation evaluation (including hidden)



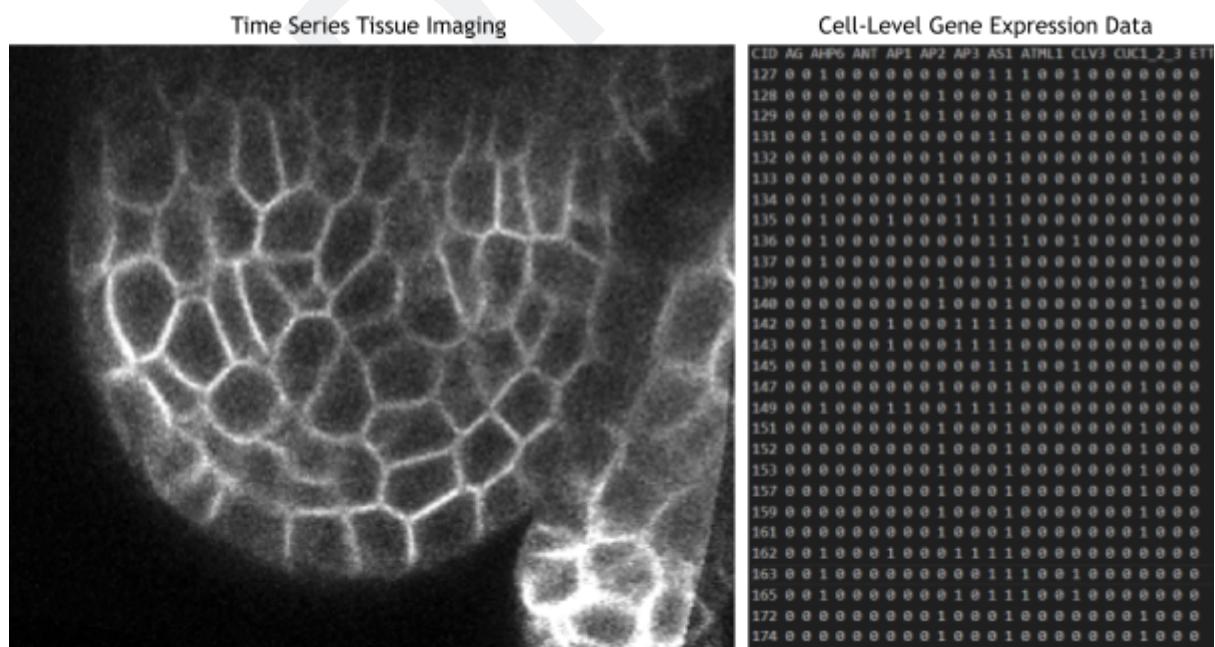
The discarded experiment above, reconstructing a simplified leaf, highlighted an issue in the specifics of the boundary conditions used by the model (as is often the case in finite element simulations), but nonetheless resulted in an interesting conclusion. As you can see from step 30 in *Figure 11*, and the alpha and hidden channels in *Figure 12*, the neural network learnt to rapidly proliferate cell material, ‘bouncing’ signalling cues off the container walls, rather than expanding slowly, as a real leaf made out of membrane-bounded cells would. This consistently provided the simulated tissue with a steep and unrealistic shortcut for learning a successful update rule. The reason for this is the ‘wrapping’ boundary condition, used by many CAs to provide continuity to the otherwise vacant neighbourhood of marginal cells (wrapping around diametrically opposite sides of the hexagonal container), and hence behaving more like an open system, rather than attenuating or reflecting border values, which can lead to pattern collapse or explosion. It is interesting to speculate the biological importance of membrane elasticity and pressure to constraint rapid outward growth. Indeed there might be neglected, transient ‘scaffolding’ components in the microenvironment of our tissue models that support early growth, like embryos developing in a seed, egg or womb. Many leaves begin in a folded, stacked or coiled state, with greater ability to sense its limits, and I recalled

the thin, dried casing of young fern fronds which fall away, along with their dense covering of hairs (trichomes) that might be further enhancing early coordination.

Utilising Biological Datasets

Considering the eventual goal of training NCAs on real images, I investigated potential datasets for future use, to steer early implementation decisions towards something future-proof. Improving biological imaging technologies, such as fluorescence microscopy, allow increasingly granular spatial but also temporal resolution (false-colour) images of tissues, even highlighting particular cellular processes of interest, like their individual gene expression. The ability to record 3D timelapses of tissue development means newer simulations can be driven far more directly by real growth data, making digital twins of tissues more plausible. In *Figure 13*, tissue from a growing flower in *Arabidopsis Thaliana* is captured in stacks of 2D slices (TIFF stacks) that comprise 3D timelapse frames (Refahi et al., 2021). Deep-learning techniques then allow for cell segmentation (identification and tracking of cells between frames (Vijayan et al., 2021), which can then be used to extract time series gene expression data for entire cell lineages. The continued development of openly available software tools like MorphoGraphX, and online dataset repositories, like The *Arabidopsis* Information Resource (TAIR) will foster far wider, crowd-sourced simulation efforts. Eventually it may be possible to integrate real gene expression data into a CA growth model, providing realistic constraints and equally allowing more direct genetic insights to be drawn from their evolving structural and behaviour.

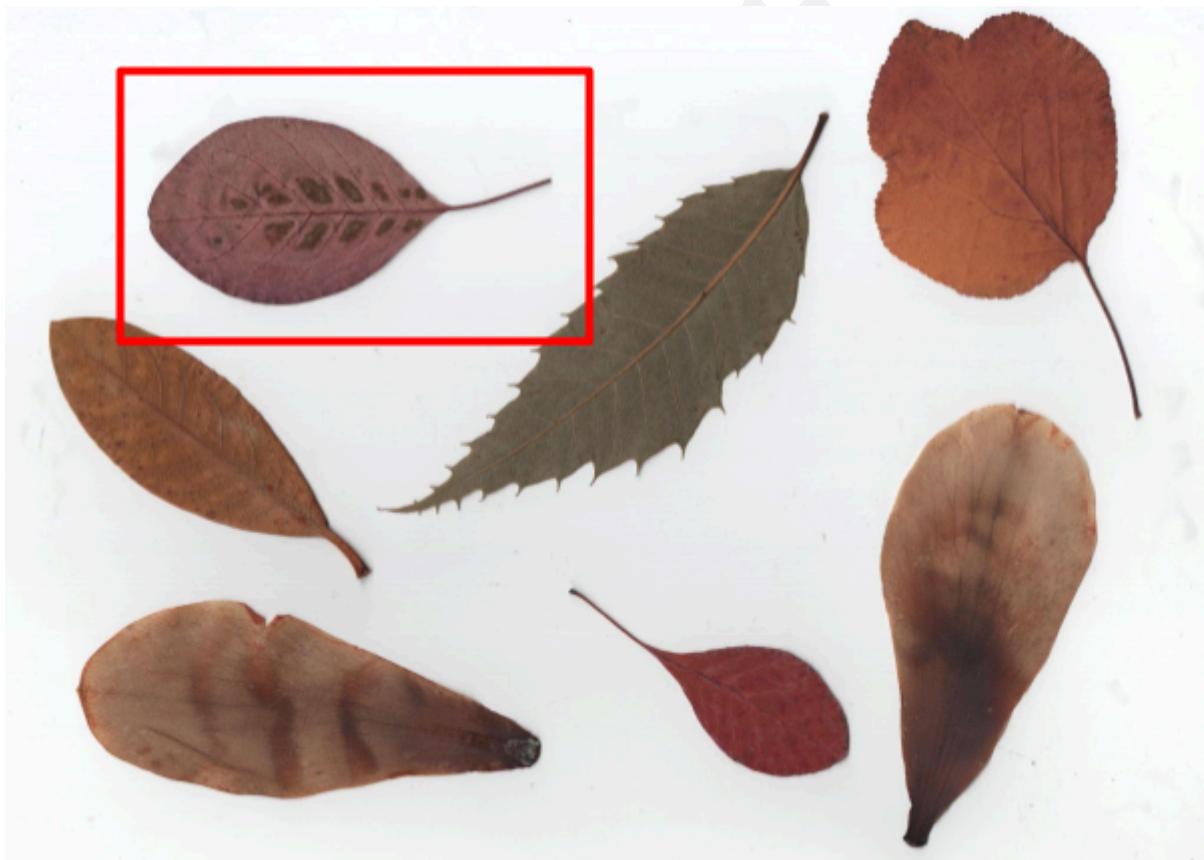
Figure 13. Tissue image and gene expression data from time lapse (Refahi et al., 2021)



Prototype Development

As expected, real images of leaves (*Figure 14*) were found to have an exceedingly high informational content, in part due to image artefacts, growth irregularities and their highly unique venation patterns. This presents a well-known challenge for under-parameterised, generative machine learning models, which require an extensive dataset to generalise (produce results with variation) and prevent overfitting, that is, the exact reproduction of singular training images. Given that existing NCAs have only been trained on high contrast, low resolution and relatively low entropy¹ images like emojis, already embodying significant computational demands, this basic NCA architecture cannot be trained on biological leaf samples, but it may still be possible to derive simplified datasets from them.

Figure 14. Scans of dried leaves



Using a consumer flatbed scanner, dried leaf samples were scanned and isolated from the background in Adobe Photoshop. The broad, pinnate leaf was selected for its simple margin and vein pattern. The solid leaf area was masked and the venation pattern

¹ Basic image entropy analysis was conducted, to estimate and compare image complexity of various training samples, though it was determined that more rigorous metrics need to be developed to conclude what kinds of images are more easily reproduced with NCA growth.

reconstructed approximately with red polylines. By creating a ‘seed image’ with a bare leaf stem (red vertical line with a small grey bud), realistic starting conditions, and a kind of global constraint from the rest of the plant are provided. The accompanying ‘mask’ image (*Figure 15*) is used to prevent the alteration of these pixels, so it is enforced that the model will use this marker as the origin to grow a much larger form.

It should be noted that the colour mapping used in simulation is arbitrary, as long as it is consistent and interpretable. The leaf tissue has been coloured with a grey fill to make equal use of all three visible channels (red, green and blue), under the intuition that real leaves certainly consist of the deposition of more than one material, and more than one regulatory pathway. The leaf veins in the target image are superimposed in ‘red’ with orange borders to clearly denote a different tissue type, using a single colour to demarcating vascular tissue, but also to provide a very crude analogue for auxin (or nutrient) availability, since grey material will only grow where red material is found in proximity. After prior training experiments started to show successful growth capabilities, an NCA model was trained more extensively (2500 epochs) to reproduce the “target” image (*Figure 16*). The simulation after 250 steps shows a clearly defined reddish-grey infill, with a defined midline vein that extends from the red leaf stem in the starting image.

Figure 15. ‘Leaf mono’ training set.

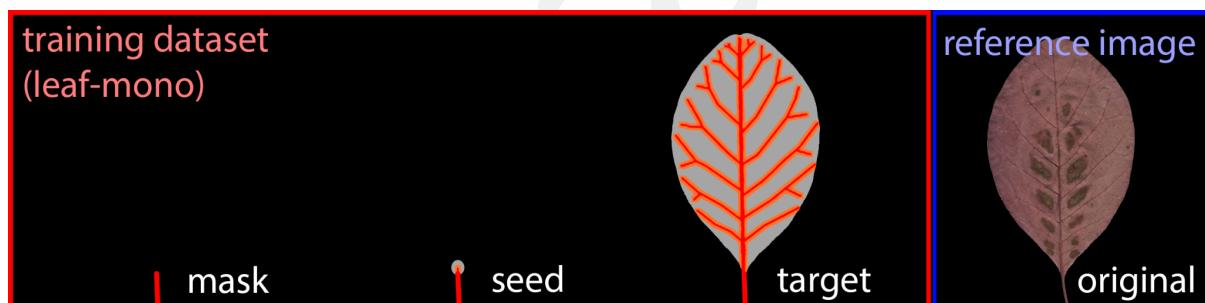
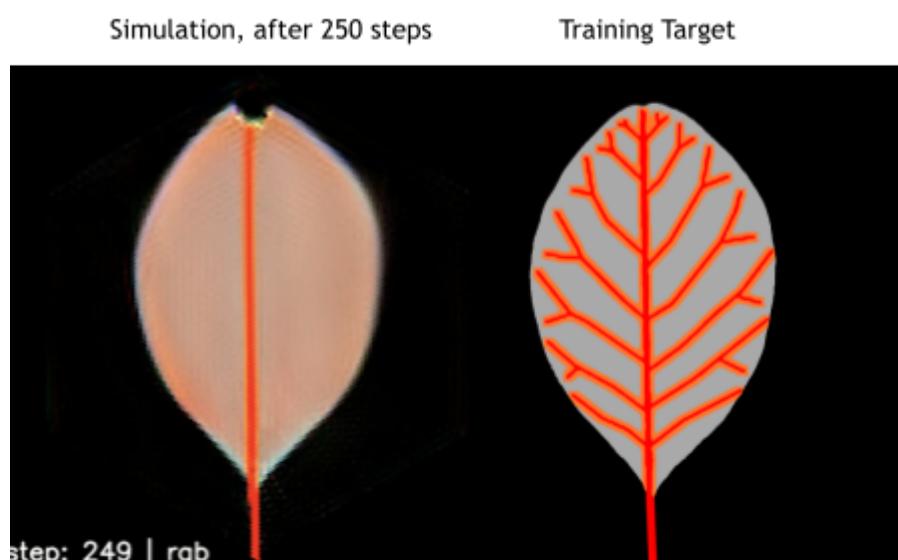


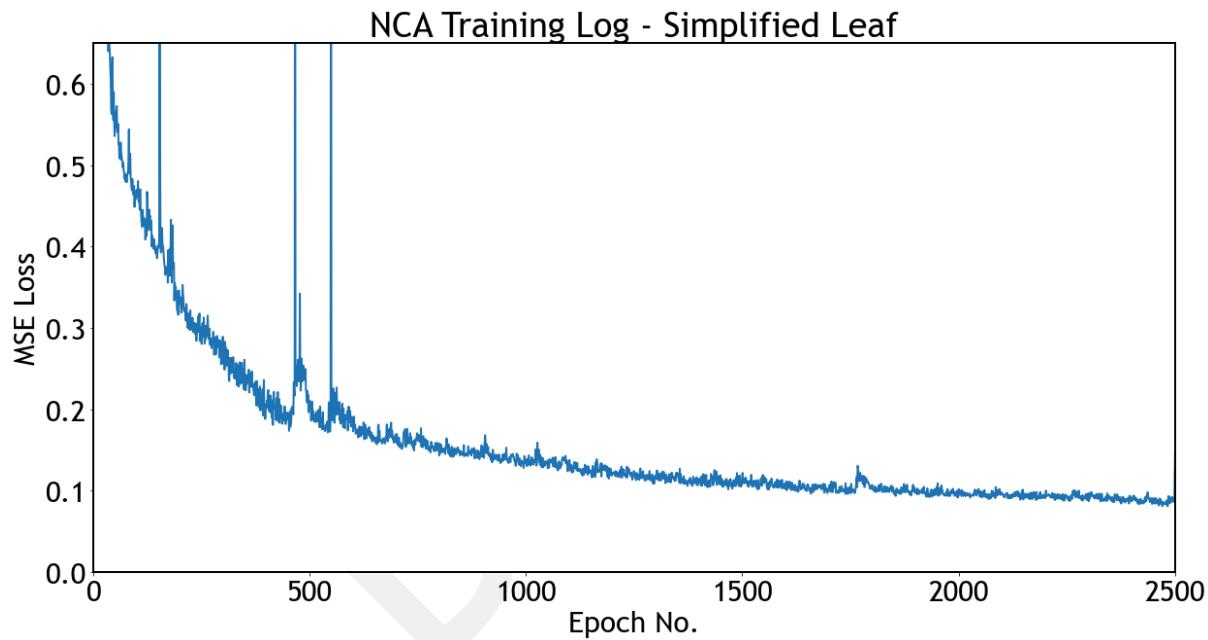
Figure 16. ‘Leaf mono’ simulation preview.



With a grid consisting of 9241 cells total, 16 image channels and 96 neurons in the neural network hidden layer; backpropagation performed 5 times per simulation run, every 80-90 steps, the simulation was successfully trained to reproduce the leaf pattern in 400-450 steps total.

Using a workstation with a 16-core 3500 MHz processor, 64Gb RAM, NVIDIA GeForce RTX 2080 Ti, training time was substantial, approximately 39 hours, or ~50 seconds per epoch, each epoch consisting unusually of only one run for brevity. The loss graph, estimating the growth accuracy, is shown in *Figure 17*. Training was stopped due to a sudden gradient explosion after ~2500 epochs.

Figure 17. ‘Leaf mono’ training log.



Ongoing Limitations

By training on a single image, the additional generalisation challenge of training is mitigated, though it is a known limitation that the model then learns to use the global orientation cue provided by the regular neighbourhood layout of cells to naively direct growth upwards, no matter the orientation of the leaf stem. This can be mitigated by using the ‘neighbour scrambling’ function each simulation run, to shuffle the effective orientations or internal polarity of individual cells, but this significantly increased training instability as expected, inline with the observation that real systems may use more auxiliary mechanisms (alongside diffusing biochemical gradients) to induce stable growth.

Training quality is much lower when neighbours are scrambled. Without underlying polarity, the overall MSE loss (in relative units) after 500 epochs remained at 25 vs 0.6 with

this additional cue. Nevertheless it still must learn self-organising behaviours, since the only positional cues are provided by the coloured pixels in the starting image.

Draft

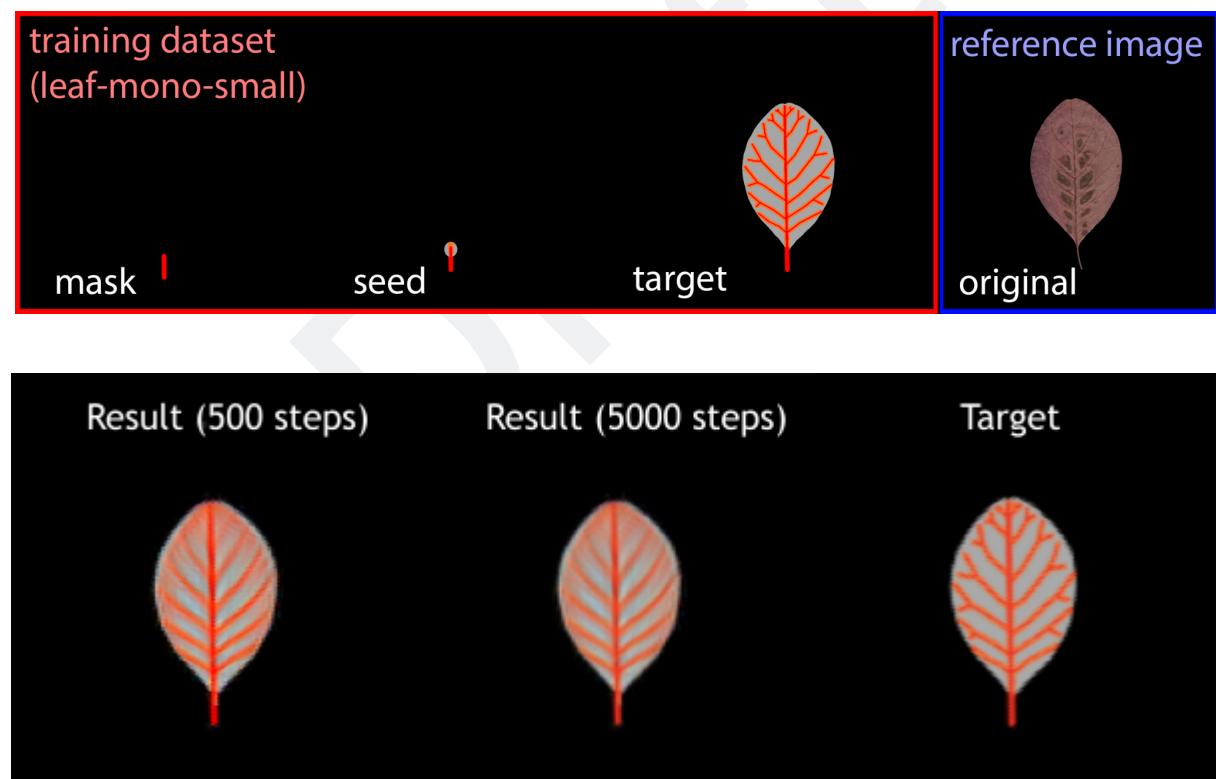
5 | Results

Further resources provided in ‘Code Availability’.

Pattern Growth and Stability

With the latest implementation, using a small, simplified leaf image, semi-realistic and stable growth behaviour was obtained. The target image, and the simulation result after 500 and 5000 steps are shown in *Figure 18*. Appearance of the leaf pattern is coordinated in a very reasonable 150-200 steps, after which it refines and stabilises to an acceptable tolerance of the target image, with the exception of the forked tertiary veins near the borders of the leaf. The pattern also remains stable and coherent well beyond the trained duration (500 steps) until at least 5000 steps.

Figure 18. ‘Leaf mono small’ training set and results.



Growth Dynamics

Figure 19 shows step-by-step growth of the simulated leaf using the best obtained parameters during training (epoch 4464 with minimum loss). There are no discernible (morphological) updates applied by the model after 240 steps.

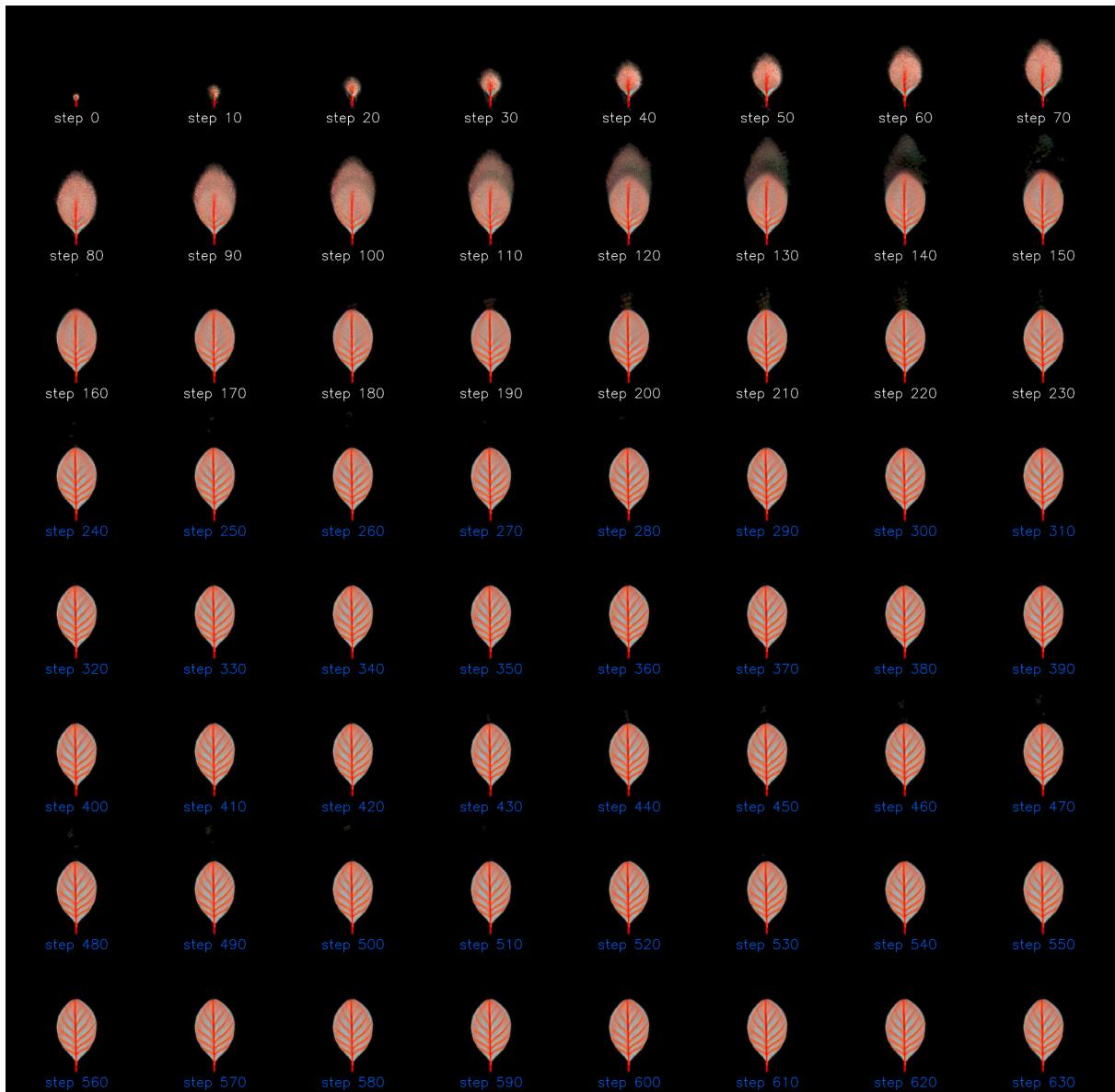


Figure 19. ‘Leaf mono small’ step-by-step analysis.

Growth occurs rapidly in the first 150 steps, with the enlarging of the red/grey region with a clear midvein and oval expansion. A faint plume unrealistically grows beyond the tip of the target leaf at step 100, but vanishes quickly by step 140. The secondary vein pattern develops more slowly from bottom to top, becoming much sharper and stable by step 240.

Training Progression

Training time was 32 hours total for 4500 epochs, ~26 seconds per epoch, a reduction from the earlier experiment. This suggests that the improved loss function incentivising just the growth of coloured cells in the target image, allowed for a significant improvement in overall evaluation and training efficiency. The loss graph (*Figure 20*) shows a shallow descent with occasional spikes, and the growth quality plateaued at 2500 epochs, when training was halted. For completeness, the learning rate (typically 0.001) was then manually decreased by a factor of 10 to encourage further progress (*Figure 21*), training for a further 2000 epochs. This resulted in (a diminishing) qualitatively significant improvement in reproductional accuracy. The minimum loss was located at epoch 4464.

Figure 20. ‘Leaf mono small’ training log 0-2500.

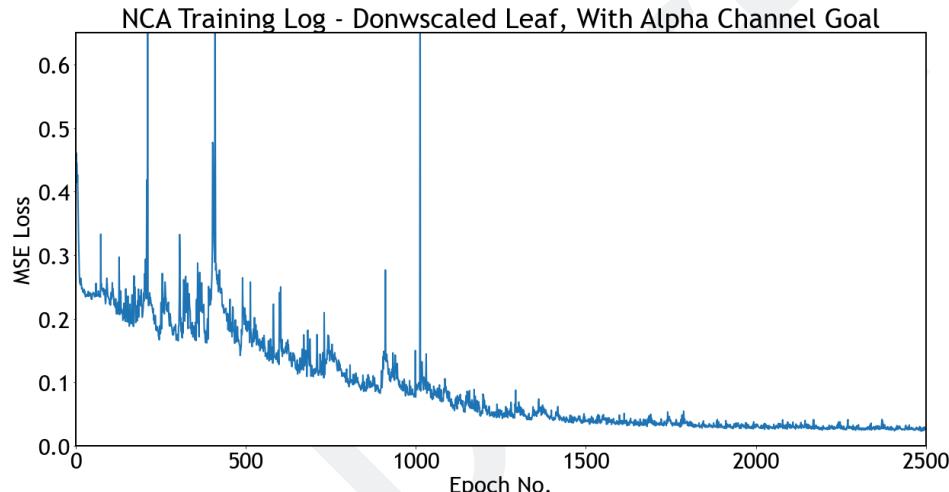
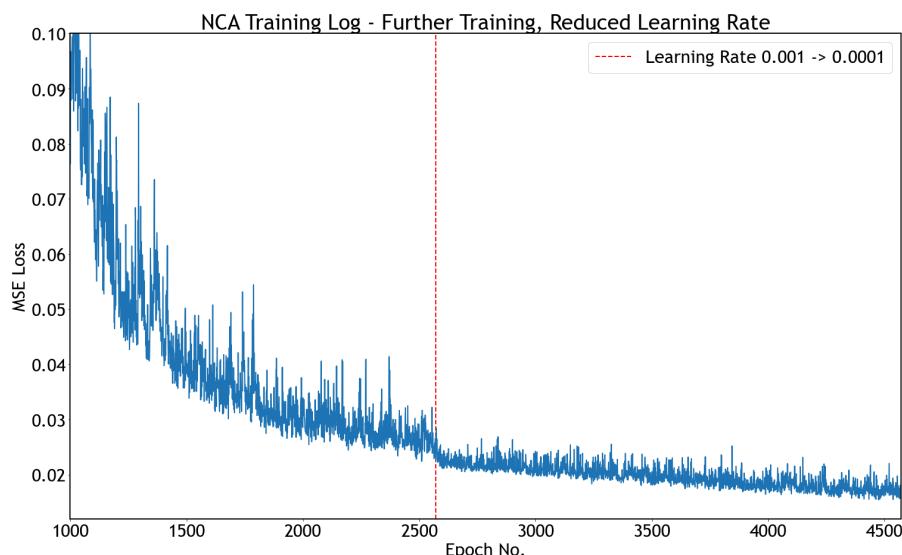


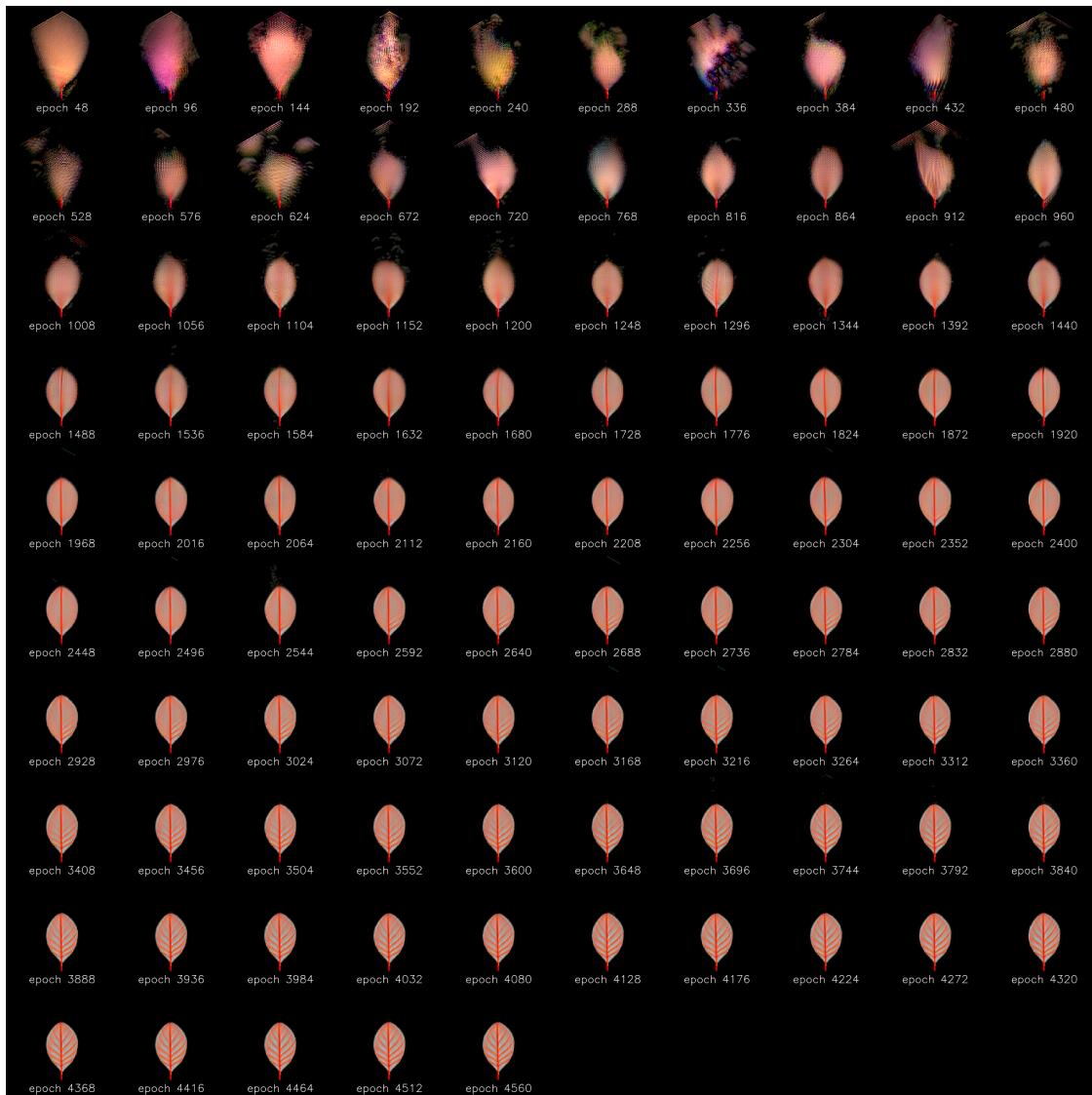
Figure 21. ‘Leaf mono small’ training log epochs 0-4500.



Pattern Development During Training

Finally, the epoch-by-epoch training progression is visualised in *Figure 22*, showing snapshots of the simulation result which steadily improves. As expected, the pattern quality develops from bottom of the grid to the top, due to the location of the fixed growth cue, the leaf stem. In the first 1000 epochs, the solid leaf area is reproduced, followed by a clear leaf midvein visible after 1500 epochs training. In the remaining epochs, the more intricate venation pattern develops from tip of the stem upwards, with far slower progress until the 4500 epoch, when training was halted.

Figure 22. ‘Leaf mono small’ epoch-by-epoch training analysis.



6 | Evaluation

Efficacy and Realism of the NCA Approach

The objective of this preliminary experiment was to show that an NCA simulation can indeed serve as a highly malleable simulation base for biological tissues, an analogue for the many complex local interactions that underlie collective intelligence and growth. With very crude biological growth constraints, the simulation was able to robustly grow a stable 2D leaf shape in a grid of nearly 10,000 cells total, provided only with a ‘seed’ and ‘target’ colour image. The model was given only a small positional cue of a leaf stem, proving how easily these simulations can adapt local behaviour to global cues, successfully learning to refine an intricate vein pattern that propagates from the stem. As pointed out before, the major shortcoming of this current model is overfitting: the fact it learns to grow in one particular orientation provides additional (false) constraints, and a less unstable pathway to self-organisation. The intelligence of real collectives is the ability to find flexible and adaptive ways to limit degrees of freedom. It should be ascertained next as to whether implementing plant-specific rule-based logic, such as: directional cell elongation and division rules; auxin-driven polarity and growth; mechanical stress remodelling and cell specialisation, outlined here for leaf growth, can adequately mitigate these training challenges by sufficiently narrowing the parameter space. The last of which, that of mimicking genetic transcription factors, can lead to impressive NCA models with far more capability for variation, where this model fell short (Hernandez, Vilalta, and Moreno-Noguer, 2021).

Another limitation of the current NCA approach, which can easily recreate the reaction and diffusion of chemicals, may be its missuitedness for modelling accurate mechanical interactions important for morphogenesis; these involve near-continuous physical constraints and energy conservation, a key driver for real living systems. In early experiments the neural network circumvented realistic growth by reflecting signals, using interference ‘across’ the container wall to limit outward growth. Even in the latest experiment, the biggest growth inaccuracy is centred in the fluid-like behaviour of simulation cells, lacking membrane material and internal pressure, despite showing a crude form of surface tension. NCAs are intuitive for discrete, microscale interactions, and we know fine-grained NCAs can approximate the behaviour of idealised systems like PDEs remarkably well (Richardson, Antal, Blythe & Schumacher, 2024), but I would nevertheless speculate that using neural operators or other continuous methods as a companion to discrete multiscale simulations could be more effective and interpretable for mechanical constraints. In failed experiments not shown here, succinctly expressing local logic for physical laws like pressure and elasticity proved challenging and inefficient.

Further limitation is the use of a 2D grid; more rigorous models would consider the extension of NCAs to 3D, explored in a game environment by Sudhakaran et al., (2021). In

a similar vein, plant cells are more amenable to a static, regular grid arrangement as they migrate less, are larger in size, and more regularly arranged. Animal cells are actually quite motile, meaning they can move around, adding additional complexity to their growth and regulatory algorithms. When these controlled replication and growth mechanisms malfunction, tumours form and spread to other parts of the body (Friedl & Gilmour, 2009), making this additional ability of vital importance to understand. This is a concern for the wider applicability of such a framework, though concepts like movable or graph cellular automata may be useful here, which allow flexible or extensible connections between cells (Zhikharevich, Hazdiuk, & Ostapov, 2019). Furthermore, much like the question of mechanical factors in plant cells, which may provide more immediate, and long-range cues for whole-tissue coordination, animals (with the exception of sponges) have nervous systems that allow for similarly high-speed, long-range signalling. This wider dynamic range in the interaction types means smaller time scales still need to be considered.

Concluding Remarks

It was not expected in the scope of this short, individual project that a working multiscale simulation prototype be built, but instead for a balanced appraisal of the problem space to be delivered that lays the all-important groundwork, grounded in deep theoretical insights and a functional, but rudimentary experimental prototype. Nevertheless, it would be interesting, in future work, to network many of these NCA modules to evaluate their scaling behaviour across multiple levels, and experiment with diverse cell types to mimic biological cell specialisation. There is clearly identifiable practical and theoretical evidence for data-driven simulations trained on multi-modal biological specimen images, in line with POM principles that underlie feasible multiscale models, and leveraging newer imaging technologies like 3D tissue timelapses and machine learning to process this high-dimensional data from the interactions of collectives, speculating plausible interaction rules in a highly-automated manner. Presently, machine learning workflows like the ones presented still require a lot of manual intervention, given their instability. With better underlying NN architectures, distant simulations that themselves may dynamically reorganise, under principles of DDDAS and attention for adaptive granularity, thus filtering against complication and ‘engineering emergence’. In biological modelling, it is clear the unaddressed challenge is the joining of biochemical and mechanical constraints for biological realism, particularly those microtubule dynamics with CAs (Alfonseca et al. 2015), will steer the development of future iterations. By placing isolated yet interoperable simulation entities together in a high-fidelity scenario, a simulation ecosystem populated with diverse models, we may provide external constraints and competing evolutionary goals that improve their accuracy and our understanding in a similar way to how generative adversarial networks operate.

Hence there are many exciting avenues for improvement. Performance optimisations will decrease their training time and foster more real time interactivity with models and their

ability to extend to digital twins. Before this, the more extensive and organised collection and distribution of multiscale datasets, on which future simulations may be trained, especially those that are open and crowd-sourced, will allow the derivation of clearer application-specific interaction rules, inform software design decisions and enhance the physical realism of these models. (Pre-)training and cross-validating models on such expansive datasets, beyond the narrow parameter space of single experiments, could drastically improve the trainability on difficult generative tasks (Cranmer, 2023; Wang et al., 2018). Lastly, in consideration of the high-level user interface design, a web-based modelling interface, making use of the recent and popular node-based paradigm in place of scripting, will allow for the continual development of more succinct and flexible rule-based grammars, by a wider range of researchers, which will enforce realistic and efficient constraints that reduces instability in model calibration. Web-based approaches, as Fletcher and Osborne (2022) importantly reflect, ensure an actively crowd-developed, documented, and maintained simulation implementation that will avoid the current habit of rebuilding the same simulation packages across science, in favour of modularity.

References

- Adhyapok, P., Fu, X., Sluka, J. P., Clendenon, S. G., Sluka, V. D., Wang, Z., ... & Glazier, J. A. (2020). A computational model of liver tissue damage and repair. *PLoS one*, 15(12), e0243451.
- Agnati, L. F., Baluška, F., Barlow, P. W., & Guidolin, D. (2009). Mosaic, self-similarity logic and biological attraction principles: Three explanatory instruments in biology. *Communicative & integrative biology*, 2(6), 552-563.
- Alfonseca, M., Ortega, A., De La Cruz, M., Hameroff, S. R., & Lahoz-Beltra, R. (2015). A model of quantum-von Neumann hybrid cellular automata: principles and simulation of quantum coherent superposition and decoherence in cytoskeletal microtubules. *Quantum Inf. Comput.*, 15(1-2), 22-36.
- Armenta-Medina, A., & Gillmor, C. S. (2019). Genetic, molecular and parent-of-origin regulation of early embryogenesis in flowering plants. *Current topics in developmental biology*, 131, 497-543.
- Azizzadenesheli, K., Kovachki, N., Li, Z., Liu-Schiaffini, M., Kossaifi, J., & Anandkumar, A. (2024). Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 1-9.
- Baas, P. W., Karabay, A., & Qiang, L. (2005). Microtubules cut and run. *Trends in cell biology*, 15(10), 518-524.
- Barnsley, M., Hutchinson, J. E., & Stenflo, Ö. (2003). V-variable fractals and superfractals. *arXiv preprint math/0312314*.
- Bhalla, N., Bentley, P. J., Vize, P. D., & Jacob, C. (2012). Programming and evolving physical self-assembling systems in three dimensions. *Natural Computing*, 11, 475-498.
- Bhatia, N., Bozorg, B., Larsson, A., Ohno, C., Jönsson, H., & Heisler, M. G. (2016). Auxin acts through MONOPTEROS to regulate plant cell polarity and pattern phyllotaxis. *Current Biology*, 26(23), 3202-3208.
- Bhatia, N., & Heisler, M. G. (2018). Self-organizing periodicity in development: organ positioning in plants. *Development*, 145(3), dev149336.

Blackiston, D., Lederer, E., Kriegman, S., Garnier, S., Bongard, J., & Levin, M. (2021). A cellular platform for the development of synthetic living machines. *Science Robotics*, 6(52), eabf1571.

Boudon, F., Chopard, J., Ali, O., Gilles, B., Hamant, O., Bouadaoud, A., ... & Godin, C. (2015). A computational framework for 3D mechanical modeling of plant morphogenesis with cellular resolution. *PLoS computational biology*, 11(1), e1003950.

Brú, A., Casero, D., de Franciscis, S., & Herrero, M. A. (2008). Fractal analysis and tumour growth. *Mathematical and computer modelling*, 47(5-6), 546-559.

Burnett, A (2024). Synthetic plants for a sustainable future. *ARIA: What we're working on*. <https://www.aria.org.uk/wp-content/uploads/2024/06/ARIA-Synthetic-plants-for-a-sustainable-future.pdf>

Cammarata, J., & Roeder, A. H. K. (2018). Development: Cell Polarity Is Coordinated over an Entire Plant Leaf. *Current biology : CB*, 28(16), R884-R887.

Catrina, S., Catrina, M., Băicoianu, A., & Plajer, I. C. (2024). Learning about Growing Neural Cellular Automata. *IEEE Access*.

Chan, B. W. C. (2020, July). Lenia and expanded universe. In *Artificial Life Conference Proceedings* 32 (pp. 221-229). One Rogers Street, Cambridge, MA 02142-1209, USA journals-info@ mit. edu: MIT Press.

Chen, C., Wang, H., Wang, D., & Wang, D. (2024). Towards the digital twin of urban forest: 3D modeling and parameterization of large-scale urban trees from close-range laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, 127, 103695.

Chua, L. O., & Yang, L. (1988). Cellular neural networks: Theory. *IEEE Transactions on circuits and systems*, 35(10), 1257-1272.

Cranmer, M. D. (2023). Interpretable Machine Learning for the Physical Sciences (Doctoral dissertation, Princeton University).

D'Angelo, G., Ferretti, S., & Ghini, V. (2017). Multi-level simulation of internet of things on smart territories. *Simulation Modelling Practice and Theory*, 73, 3-21.

Dittrich, T. (2014). ‘The concept of information in physics’: an interdisciplinary topical lecture. *European Journal of Physics*, 36(1), 015010.

Drexler, E. (1987). *Engines of creation: The coming era of nanotechnology*. Anchor.

Evers, J. B., Letort, V., Renton, M., & Kang, M. (2018). Computational botany: advancing plant science through functional-structural plant modelling. *Annals of Botany*, 121(5), 767-772.

Fletcher, A. G., & Osborne, J. M. (2022). Seven challenges in the multiscale modeling of multicellular tissues. *WIREs mechanisms of disease*, 14(1), e1527.

Fox, S., Southam, P., Pantin, F., Kennaway, R., Robinson, S., Castorina, G., ... & Coen, E. (2018). Spatiotemporal coordination of cell division and growth during organ morphogenesis. *PLoS Biology*, 16(11), e2005952.

Friedl, P., & Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. *Nature reviews Molecular cell biology*, 10(7), 445-457.

Fujimoto, R., Barjis, J., Blasch, E., Cai, W., Jin, D., Lee, S., & Son, Y. J. (2018, December). Dynamic data driven application systems: research challenges and opportunities. In *2018 Winter Simulation Conference (WSC)* (pp. 664-678). IEEE.

Fujita, H., Toyokura, K., Okada, K., & Kawaguchi, M. (2011). Reaction-diffusion pattern in shoot apical meristem of plants. *PloS one*, 6(3), e18243.

Gilpin, W. (2019). Cellular automata as convolutional neural networks. *Physical Review E*, 100(3), 032402.

Goodson, H. V., & Jonasson, E. M. (2018). Microtubules and microtubule-associated proteins. *Cold Spring Harbor perspectives in biology*, 10(6), a022608.

Gregor, K., Danihelka, I., Graves, A., Rezende, D., & Wierstra, D. (2015). Draw: A recurrent neural network for image generation. In *International conference on machine learning* (pp. 1462-1471). PMLR.

He, X., Wei, D., Lam, K. M., Li, J., Wang, L., Jia, W., & Wu, Q. (2010). Canny edge detection using bilateral filter on real hexagonal structure. In *Advanced Concepts for Intelligent Vision Systems: 12th International Conference, ACIVS 2010, Sydney, Australia, December 13-16, 2010, Proceedings, Part I* 12 (pp. 233-244). Springer Berlin Heidelberg.

Hernandez, A., Vilalta, A., & Moreno-Noguer, F. (2021). Neural cellular automata manifold. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10020-10028).

Herrera, C. M., & Bazaga, P. (2013). Epigenetic correlates of plant phenotypic plasticity: DNA methylation differs between prickly and nonprickly leaves in heterophylous *Ilex*

aquifolium (Aquifoliaceae) trees. *Botanical Journal of the Linnean Society*, 171(3), 441-452.

Howard, J., Grill, S. W., & Bois, J. S. (2011). Turing's next steps: the mechanochemical basis of morphogenesis. *Nature Reviews Molecular Cell Biology*, 12(6), 392-398.

Kierzkowski, D., Runions, A., Vuolo, F., Strauss, S., Lymbouridou, R., Routier-Kierzkowska, A. L., ... & Tsiantis, M. (2019). A growth-based framework for leaf shape development and diversity. *Cell*, 177(6), 1405-1418.

Kost, B., & Chua, N. H. (2002). The plant cytoskeleton: vacuoles and cell walls make the difference. *Cell*, 108(1), 9-12.

Lev-Yadun, S. (2012). Fern leaves and cauliflower curds are not fractals. *Plant signaling & behavior*, 7(5), 533-534.

Mansfield, C., Newman, J. L., Olsson, T. S., Hartley, M., Chan, J., & Coen, E. (2018). Ectopic BASL reveals tissue cell polarity throughout leaf development in *Arabidopsis thaliana*. *Current Biology*, 28(16), 2638-2646.

Mitchison, G. (2016). Conformal growth of *Arabidopsis* leaves. *Journal of theoretical biology*, 408, 155-166.

Mitsanis, C., Hurst, W., & Tekinerdogan, B. (2024). A 3D functional plant modelling framework for agricultural digital twins. *Computers and Electronics in Agriculture*, 218, 108733.

Mordvintsev, A., & Niklasson, E. (2021). μ NCA: Texture Generation with Ultra-Compact Neural Cellular Automata. *arXiv preprint arXiv:2111.13545*.

Mordvintsev, A., Randazzo, E., Niklasson, E., & Levin, M. (2020). Growing neural cellular automata. *Distill*, 5(2), e23.

NIHR Imperial BRC (2024). Digital heart modelling project could open up the path of virtual patient monitoring. *NIHR Imperial Biomedical Research Centre: Cardiovascular, Digital Health*.

<https://imperialbrc.nihr.ac.uk/2024/05/08/digital-heart-modelling-project-could-open-up-the-path-of-virtual-patient-monitoring/>

Noble, R., & Noble, D. (2023). *Understanding living systems*. Cambridge University Press.

Rascher, U., Hütt, M. T., Siebke, K., Osmond, B., Beck, F., & Lüttge, U. (2001). Spatiotemporal variation of metabolism in a plant circadian rhythm: the biological clock as

an assembly of coupled individual oscillators. *Proceedings of the National Academy of Sciences*, 98(20), 11801-11805.

Refahi, Y., Zardilis, A., Michelin, G., Wightman, R., Leggio, B., Legrand, J., ... & Traas, J. (2021). A multiscale analysis of early flower development in *Arabidopsis* provides an integrated view of molecular regulation and growth control. *Developmental Cell*, 56(4), 540-556.

Richardson, A. D., Antal, T., Blythe, R. A., & Schumacher, L. J. (2024). Learning spatio-temporal patterns with Neural Cellular Automata. *PLOS Computational Biology*, 20(4), e1011589.

Runser, S., Vetter, R., & Iber, D. (2024). SimuCell3D: three-dimensional simulation of tissue mechanics with cell polarization. *Nature Computational Science*, 1-11.

Ruzette, S. M., Antoine A; Megason, Sean M. (2024). *Goo: an open-source modular Python-based software to simulate cells, tissues and early embryos*. doi:10.5281/zenodo.10296203

Sapala, A., Runions, A., Routier-Kierzkowska, A. L., Das Gupta, M., Hong, L., Hofhuis, H., ... & Smith, R. S. (2018). Why plants make puzzle cells, and how their shape emerges. *Elife*, 7, e32794.

Seyfried, T. N., & Huysentruyt, L. C. (2013). On the origin of cancer metastasis. *Critical Reviews™ in Oncogenesis*, 18(1-2).

Sudhakaran, S., Grbic, D., Li, S., Katona, A., Najarro, E., Ganois, C., & Risi, S. (2021). Growing 3d artefacts and functional machines with neural cellular automata. In *Artificial Life Conference Proceedings* 33 (Vol. 2021, No. 1, p. 108).

Terebieniec, B. (2019). *Using systems genetics to explore the complexity of leaf shape variation in Populus tremula* (Doctoral dissertation, Umeå University).

Turing, A. M. (1952). The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London Series B*, 237(641), 37-72.

Uhrmacher, A. M., Degenring, D., & Zeigler, B. (2005). Discrete event multi-level models for systems biology. *Transactions on computational systems biology I*, 66-89.

UKRI (2024). *Digital twin projects to transform environmental science*. UKRI. <https://www.ukri.org/news/digital-twin-projects-to-transform-environmental-science/#:-:text=The%20digital%20twin%20pilot%20projects, and%20ecosystems%2C%20and%20natural%20hazards>.

Valentim, C. A., Rabi, J. A., & David, S. A. (2023). Cellular-automaton model for tumor growth dynamics: Virtualization of different scenarios. *Computers in biology and medicine*, 153, 106481.

Vijayan, A., Tofanelli, R., Strauss, S., Cerrone, L., Wolny, A., Strohmeier, J., ... & Schneitz, K. (2021). A digital 3D reference atlas reveals cellular growth patterns shaping the *Arabidopsis* ovule. *Elife*, 10, e63262.

von Wangenheim, D., Fangerau, J., Schmitz, A., Smith, R. S., Leitte, H., Stelzer, E. H., & Maizel, A. (2016). Rules and self-organizing properties of post-embryonic plant organ cell division patterns. *Current Biology*, 26(4), 439-449.

Vuolo, F., Kierzkowski, D., Runions, A., Hajheidari, M., Mentink, R. A., Gupta, M. D., ... & Tsiantis, M. (2018). LMI1 homeodomain protein regulates organ proportions by spatial modulation of endoreduplication. *Genes & development*, 32(21-22), 1361-1366.

Wang, M., White, N., Grimm, V., Hofman, H., Doley, D., Thorp, G., ... & Hanan, J. (2018). Pattern-oriented modelling as a novel way to verify and validate functional-structural plant models: a demonstration with the annual growth module of avocado. *Annals of botany*, 121(5), 941-959.

Yonekura, T., Iwamoto, A., Fujita, H., & Sugiyama, M. (2019). Mathematical model studies of the comprehensive generation of major and minor phyllotactic patterns in plants with a predominant focus on orixate phyllotaxis. *PLoS computational biology*, 15(6), e1007044.

Zhikharevich, V., Hazdiuk, K., & Ostapov, S. (2019). Simulation of Bio-Like Systems and Processes Using Movable Cellular Automata. In *CMIS* (pp. 796-809).

External Datasets

Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*, 89(4), 789-804

Available at:

https://www.arabidopsis.org/download/list?dir=Proteins%2FAraport11_protein_lists

Refahi, Y., Zardilis, A., Michelin, G., Wightman, R., Leggio, B., Legrand, J., ... & Traas, J. (2021). A multiscale analysis of early flower development in *Arabidopsis* provides an integrated view of molecular regulation and growth control. *Developmental Cell*, 56(4), 540-556.

Available at:

<https://www.repository.cam.ac.uk/items/d0aa97a3-04a8-4aef-8542-b4a0877010d7>

Code Availability

NCA Experiment

- ‘Modular NCA’ framework GitHub repository:
<https://github.com/paveworkshop/modular-nca>
- Pre-trained NCA for evaluation in a Jupyter Notebook:
https://github.com/paveworkshop/modular-nca/blob/main/eval_nca.ipynb
- Animated previews of ‘leaf-mono-small’ experiment:
<https://github.com/paveworkshop/modular-nca/tree/main/results/leaf-mono-small/animations>

Key Definitions and Abbreviations

CA - Cellular Automata. A ‘grid of pixels’ that can be used to model local interactions. Each pixel’s state or colour at each time step are derived from their neighbours’ states and a set of transition rules.

FSP - Functional-Structural Plant Modelling. A modelling technique in biology that captures both the spatial organisation of tissue and organs in an organism as well as the underlying interaction dynamics between these components.

ML - Machine Learning. Using computational methods to ‘learn’ or determine an unknown, underlying relationship in a dataset.

NCA - Neural Cellular Automata. A recent variant of CAs that use machine learning to determine desirable transition rules.

NN - (Artificial) Neural Network. Here referring to artificial neural networks used in machine learning. A model of highly interconnected neurons which, due to their wiring pattern and individual behaviours, can be used to store and process information through their firing patterns, after the sensitivity of each connection has been tuned carefully in training.

PDE - Partial Differential Equation. A mathematical formulation relating change between a set of dependent and independent variables in some unknown function, using their partial

derivatives. They are notoriously hard to solve, but often very useful to model real-world phenomena like the way heat conducts through materials, or how fluids flow.

POM - Pattern-oriented modelling. A modelling strategy that uses wide-ranging, multiscale experimental data as ‘filters’ to validate expected model behaviour.

RGB(A) - A common colour format used by devices like computers to store or reproduce the visual information about pixels in an image, where each pixel has varying intensities of three visible channels (red, green and blue) as well as optionally an alpha channel which allows for transparency.

SAM - Shoot Apical Meristem. Tip of the plant shoot where new leaves form.

SFI - Sante Fe Institute. Organisation established in the 1980s, dedicated to study of chaos and complex systems.