



Quantum Generators: Integrating AI into Single Cell RNA – Sequencing for Improving on the Consistency of Cell Types

Poondru Prithvinath Reddy

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

April 5, 2024

Quantum Generators: Integrating AI into Single Cell RNA – Sequencing for Improving on the Consistency of Cell Types

Poondru Prithvinath Reddy

ABSTRACT

Quantum Generators is a means of achieving mass food production with short production cycles and when and where required by means of machines rather than land based farming which has serious limitations. The process for agricultural practices for plant growth in different stages is simulated in a machine with a capacity to produce multiple seeds from one seed input using computational models of multiplication (generating multiple copies of kernel in repetition). In this respect, we present profiling the gene expression program at the single cell transcription level by using key technology of single-cell RNA-sequencing (scRNA-seq). Therefore, a key task in single-cell RNA-seq (scRNA-seq) data analysis is to accurately detect the number of cell types in the sample, which can be critical for downstream analyses such as cell type identification, extracting underlying biological knowledge and for quantum generation from the large amount of data generated by scRNA-seq technology. For this scRNA-seq data clustering, we designed divisive clustering algorithm with k-means and has been recursively designed to automatically estimate the number of cell types through optimising the number of clusters in a dataset measured by robotic and on-line spectroscopy. The results of the study on estimating the number of cell types in variety of environment settings to create scRNA-seq data from a robotic system with varying number of cell types, with different input proportions enables us to assess the performance of the algorithm, from covering various aspects using a simulated dataset in absence of real data and we checked our approach by simulated data and found to have clustering concordance of cells to their predefined cell types. In this way, a group of computing units assisted synthesizer with reconfigurable system consisting of edge computing that is part of CellSynputer(a modular platform automating cell synthesis which embodies synthesis abstraction) is feasible for automated experimentation and in that respect an implementation of Clustering k-means algorithm for managing cell type groups as a part of robotic system based on small

model is presented. Although the platform model with cell type intelligence as modular unit based CellSynputer given us a method of automating and optimizing cellular assemblies however, this need to be tested using natural crop cells for quantum generation.

INTRODUCTION

A **Quantum** (plural quanta) is the minimum amount of any physical entity (physical property) involved in an interaction. On the other hand, **Generators** don't actually create anything instead, they generate quantity prescribed by physical property through multiplication to produce high quality products on a mass scale. The aim of Quantum Generators is to produce multiple seeds from one seed at high seed rate to produce a particular class of food grains from specific class of **seed** on mass scale by means of machine rather than land farming.

The process for agricultural practices include preparation of soil, seed sowing, watering, adding manure and fertilizers, irrigation and harvesting. However, if we create same conditions as soil germination, special watering, fertilizers addition and plant growth in different stages in a machine with a capacity to produce multiple seeds from one seed input using computational models of multiplication(generating multiple copies of kernel in repetition) then we will be closure to achieving mass food production by means of quantum generators(machine generated) rather than traditional land based farming which has very serious limitations such as large space requirements, uncontrolled contaminants, etc. The development of Quantum Generators requires specialized knowledge in many fields including Cell Biology, Nanotechnology, 3D Cellprinting, Computing, Soil germination and initially they may be big occupying significantly large space and subsequently small enough to be placed on roof-tops.

The Quantum Generators help world meet the food needs of a growing population while simultaneously providing opportunities and revenue streams for farmers. This is crucial in order to grow enough food for growing populations without needing to expand farmland into wetlands, forests, or other important natural ecosystems. The Quantum Generators use significantly less space compared to farmland and also results in increased yield per square foot with short production cycles, reduced cost of cultivation besides easing storage and transportation requirements.

In addition, Quantum Generators Could Eliminate Agricultural Losses arising out of Cyclones, Floods, Insects, Pests, Droughts, Poor Harvest, Soil Contamination, Land Degradation, Wild Animals, Hailstorms, etc.

Quantum generators could be used to produce most important *food* crop like rice, wheat and maize on a mass scale and on-demand when and where required. Computers and Smartphones have become part of our lives and Quantum Generators could also become very much part of our routine due to its potential benefits in enhancing food production and generating food on-demand wherever required.

METHODOLOGY

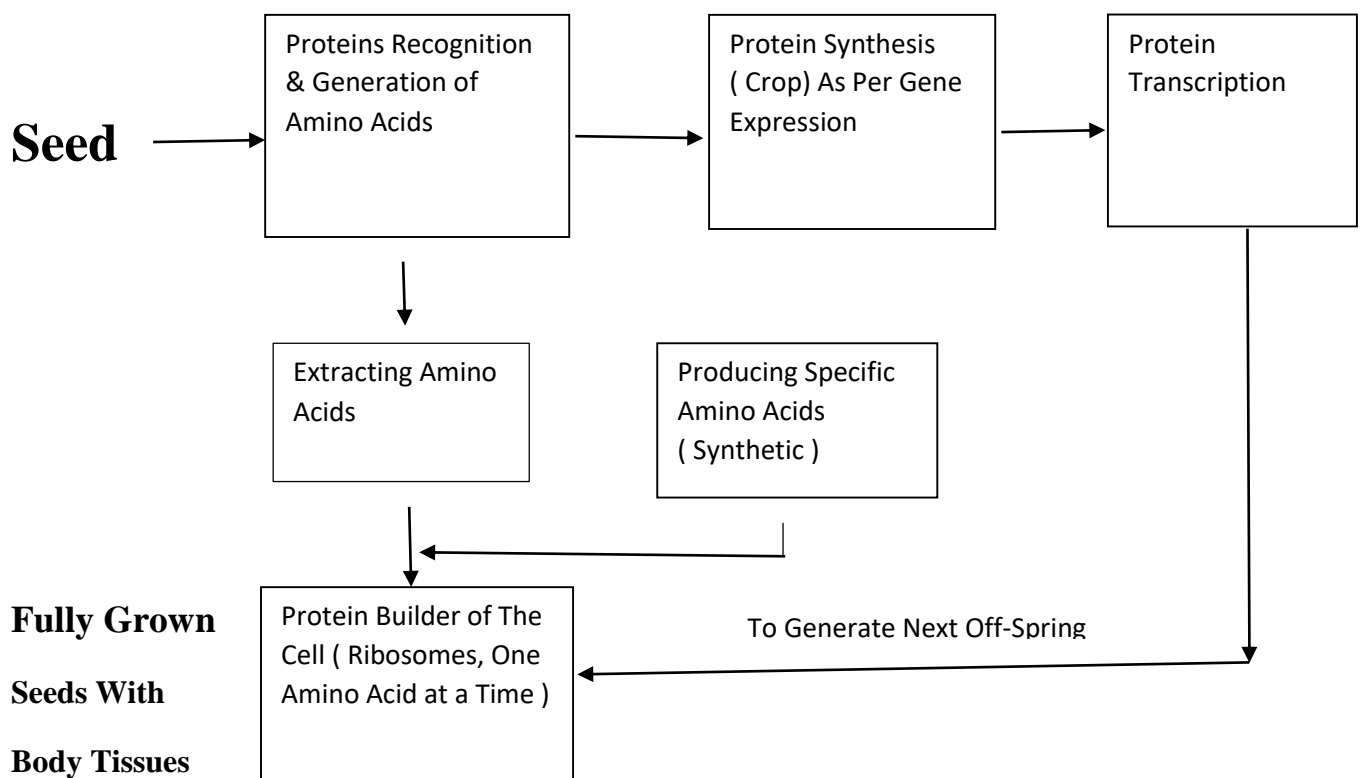


Fig 1. Process Flow Diagram of Seed Builder

Protein from input seeds is broken down into individual amino acids which are reassembled by Quantum Generating ribosomes into proteins that Crop cells need to be generated. The information to produce a protein is encoded in the **cell's** DNA. When a protein is produced, a copy of the DNA is made (called mRNA) and this copy is transported to a ribosome.

Protein **synthesis** is the process used by the QG (Quantum Generator) to make proteins. The first step of protein **synthesis** is called Transcription. It occurs in the nucleus. During transcription, mRNA transcribes (copies) DNA.

Body tissues **grow** by increasing the number of cells that make them up. Every **cell** in the crop body contains protein. The basic structure of protein is a chain of amino acids. We need protein in our diet to help human body repair cells and make new ones.

The major steps in protein synthesis are:

- DNA unzips in the nucleus.
- mRNA nucleotides transcribe the complementary DNA message.
- mRNA leaves nucleus and goes to ribosome.
- mRNA attaches to ribosome and first codon is read.
- tRNA brings in proper amino acid from cytoplasm.
- a second tRNA brings in new amino acid.

The journey from gene to **protein** is complex and tightly controlled within each cell. It consists of two major **steps**: transcription and translation. Together, transcription and translation are known as gene expression. Transcription is the transfer of genetic instructions in DNA to mRNA in the nucleus. Translation occurs at the ribosome, which consists of rRNA and proteins.

Ribosomes are the sites in a **cell** in which **protein** synthesis takes place. Cells have many ribosomes, and the exact number depends on how active a particular cell is in synthesizing proteins. **Ribosomes** are the protein builders or the protein synthesizers of the cell. They are like construction guys who connect one amino acid at a time and build long chains.

Ribosomes, large complexes of **protein** and ribonucleic acid (RNA), are the cellular organelles responsible for protein synthesis. They receive their “orders” for protein synthesis from the nucleus where the DNA is transcribed into messenger RNA (mRNA).

During the **process** of transcription, the information stored in a gene's DNA is passed to a similar molecule called RNA (ribonucleic acid) in the cell nucleus. A type of RNA called transfer RNA (tRNA) assembles the protein, one amino acid at a time.

Amino acids can be produced by breaking down proteins, known as the extraction method. However, the amount of amino acids in the source protein limits the amount of amino acids made. Extraction is not good for making mass quantities of specific amino acids. So Synthetic Methods of making amino acids is necessary in protein synthesis.

The Quantum Generator contains pre-programmed Protein Synthesizer relevant to specific Crop/Tissue which essentially reassembles ribosomes (Sites in a Cell) into proteins that your crop cells need. The sequence and information to produce a protein is encoded in the synthesizer of Quantum Generator.

Robotics for Automation and Optimization in Cell Synthesis

We believe that the potential of rapidly developing technologies (e.g., machine learning and robotics) are more fully realized by operating seamlessly with the way that synthetic biologists currently work. To reproduce this fundamental mode of operation, a new approach to the automated exploration of biological space is needed that combines an abstraction of biological synthesis with robotic hardware and closed-loop programming.

As there is a growing drive to exploit rapidly growing robotic technologies along with artificial intelligence-based approaches and applying this to cell synthesis requires a holistic approach. Here, we outline an approach to this problem beginning with an abstract representation of the practice of cell synthesis that then informs the programming and automation required for its practical realization. Using this foundation to construct closed-loop robotic synthesis engine, we can generate new syntheses that may be optimized, and repeated entirely automatically. These robots can perform synthesis reactions and analyses much faster than that can be done by other means. As such, this leads to a road map whereby molecules can be synthesized, optimized, and made on demand from a digital code.

The ability to make small molecules autonomously and automatically will be fundamental to many applications, including quantum generators. Additionally, automated synthesis requires (in many cases) optimization of reaction yields; following optimization, the best conditions can be fed to the synthesis robot to increase the overall yield. There are different approaches to automated yield optimization, and as optimization of reaction conditions requires live feedback from the robotic system, many different detectors are required to monitor progress of the reactions, including benchtop nuclear magnetic resonance spectroscopy, Raman

spectroscopy, UV-Vis spectroscopy, etc. Harvested data are then fed to optimization algorithms to explore often the multidimensional parameter space.

Robotics & Machine Learning towards Biological Space Exploration

Machine learning approaches are fundamental to scientific investigation in many disciplines. In biological studies, many of these methods are widely applicable and robotics/automation is helping to progress cell synthesis through biological space exploration. For our study, the yield of a synthetic reaction can be predicted using **machine learning** in the multidimensional space obtained from robotic automation to map the yield landscape of intricate synthesis following synthesis code. Meanwhile, our emphasis is on automation of synthesis, which is controlled by robots/computers rather than by humans. Synthesis through automation offers far better efficiency and accuracy. In addition, the machine learning algorithms explore a wider range of biological space that would need to be performed purely automated random search to fast-track synthesis. This brings the development of automation, optimization, and molecular synthesis very close.

Figure 2 shows a graphical representation of workflow for joining automated retrosynthesis with a synthesis robot and reaction optimization. The retrosynthetic module will generate a valid synthesis of the target that can then be transferred into synthesis code that can be executed in a robotic platform. The optimization module can optimize the whole sequence, getting the feedback from the robot.

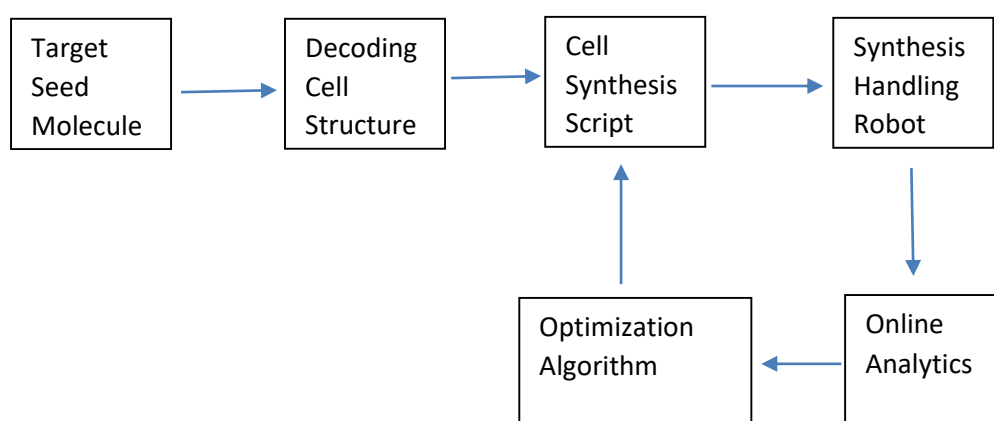


Fig. 2 Architecture of Robotic Synthesis of Crop Cells in a Quantum Generator

The methodology is essentially fundamental for getting the quantum generators as autonomous as possible and also as fast & optimized and the aim is to design processors both CPU and GPU to represent computations and their structural patterns and also controls required for the microcontroller in synthesizer unit from generator in realizing the desired quantity. Therefore, we use circuit extraction process from the CPU and desired IC's required in GPU and also final control generation required for microcontroller for the structural formation. The CPU and GPU are required to be trained separately and also microcontroller is to trained independently using reinforcement learning algorithm to arrive at the designs that can easily be adopted and customized from the environment in quantum generators.

Divisive K-Means Clustering

In machine learning, clustering is the unsupervised learning technique that groups the data based on similarity between the set of data. There are different types of clustering algorithms in machine learning.

Connectivity-based clustering: This type of clustering algorithm builds the cluster based on the connectivity between the data points.

Hierarchical clustering is a connectivity-based clustering model that groups the data points together that are close to each other based on the measure of similarity or distance. The assumption is that data points that are close to each other are more similar or related than data points that are farther apart.

Hierarchical Divisive clustering - It is also known as a top-down approach. This algorithm also does not require to pre-specify the number of clusters. Top-down clustering requires a method for splitting a cluster that contains the whole data and proceeds by splitting clusters recursively until individual data have been split into singleton clusters.

K-Means Clustering is a machine learning algorithm that is used to solve the clustering problems in a given data set. The k-means clustering tries to group similar kinds of items in form of clusters. It finds the similarity between the items and groups them into the clusters. It is an iterative algorithm that divides the unlabeled dataset into k different clusters in such a way that each dataset belongs only one group that has similar properties and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm. . The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

The k-means clustering algorithm mainly performs two tasks:

- Determines the best value for K center points or centroids by an iterative process.
- Assigns each data point to its closest k-center. Those data points which are near to the particular k-center, create a cluster.

Hierarchical Divisive Clustering Algorithm:

- given a dataset ($d_1, d_2, d_3, \dots, d_N$) of size N
- at the top we have all data in one cluster
- the cluster is split using a flat clustering method eg. K-Means, etc
- repeat
- choose the best cluster among all the clusters to split
- split that cluster by the flat clustering algorithm
- until each data is in its own singleton cluster

ARCHITECTURE

Platform Design in Cell Synthesis

Methodologies for the automation of cell synthesis, optimization, and crop yields have not generally been designed for the realities of crop-based yields, and that the potential of rapidly developing technologies (e.g., machine learning and robotics) are more fully realized by operating seamlessly with the way that synthetic biologists currently work. This is because the researchers often work by thinking backwards when planning a synthetic procedure. To reproduce this fundamental mode of operation, a new universal approach to the automated exploration of cell synthesis space is needed that combines an abstraction of cell synthesis with robotic hardware and closed-loop programming.

Automation Approach

There are different automation approaches for cell synthesis these include block based, iterative, multistep however, we considered CellSynputer which is integration of abstraction, programming and hardware interface, which is given below depicted as in Fig 3.

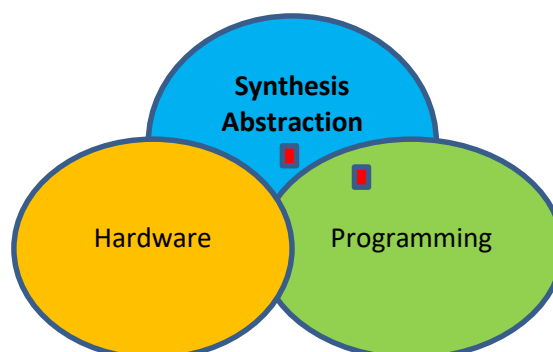


Fig. 3 Approach – Cell Synthesis Automation

Synthetic biologists already benefit from algorithms in the field of cell synthesis and, therefore, automation is one step forward that might help biologists and chemists to plan and develop biological space more quickly, efficiently, and importantly, CellSynputer is a platform that employs a broad range of algorithms interfacing hardware and abstraction to solve synthesis-related problems and surely can very well be established for quantum generation.

Synthesis via Programmable Modular System: ‘The CellSynputer’

We presented a modular platform for automating cell synthesis, which embodies our synthesis abstraction in ‘the CellSynputer’. Our abstraction of cell synthesis contains the key four stages of synthetic protocols: recognition, gene expression, transcription, and protein builder that can be linked to the physical operations of an automated robotic platform. Software control over hardware allowed combination of individual unit operations into multistep cell synthesis. A CellSynputer was created to program the platform; the system creates low-level instructions for the hardware taking graph representation of the platform and abstraction representing cell synthesis. In this way, it is possible to script and run published syntheses without reconfiguration of the platform, providing that necessary modules are present in the system.

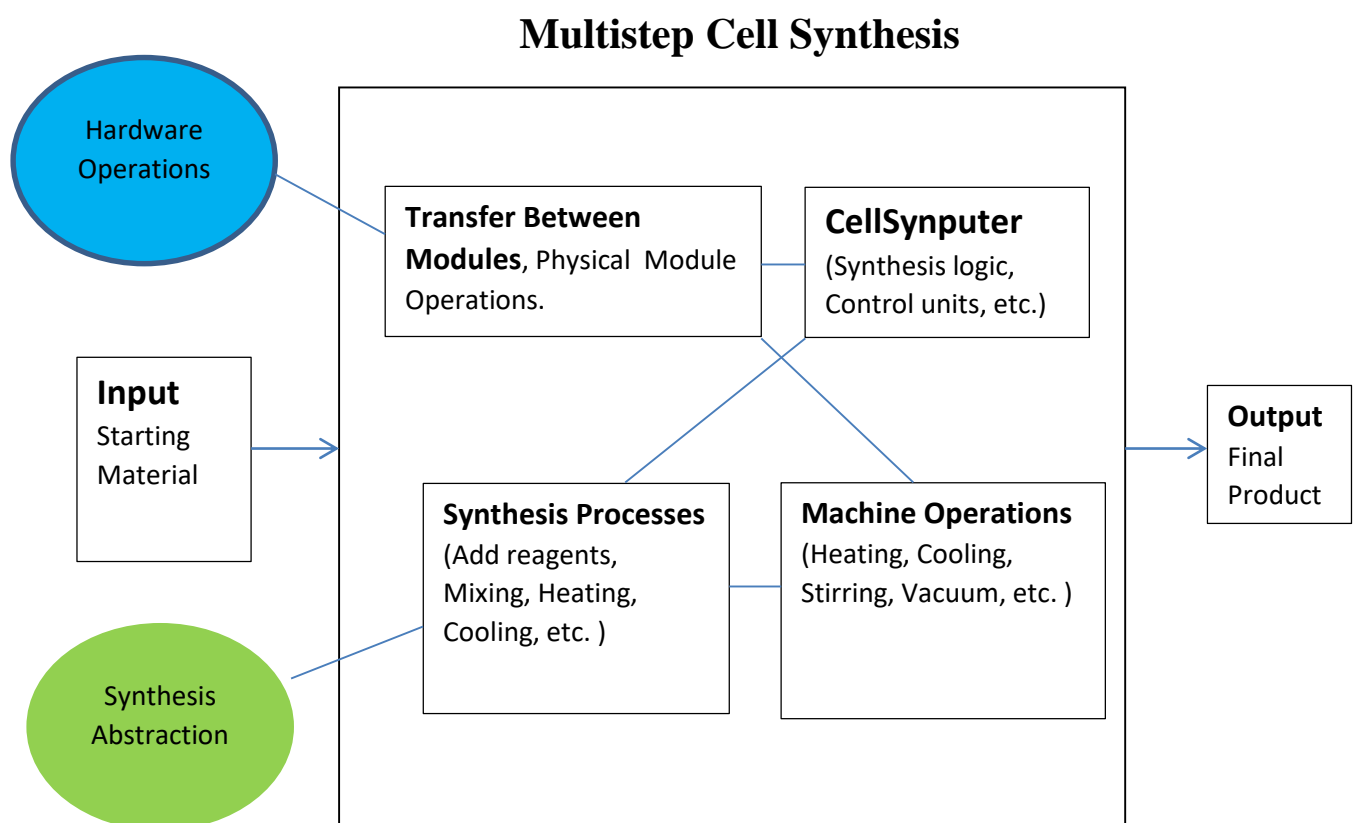


Figure 4. CellSynputer Operational Architecture

Finally, by combining CellSynputer platform and robotic systems with AI, it is possible to build autonomous systems working in closed loop, making decisions based on prior experiments. We already presented a flow system for navigating a network of synthesis reactions utilizing an infrared spectrometer for on-line analysis and as the sensor for data feedback. The system will be able to select the suitable starting materials autonomously on the basis of change in the infrared spectra.

Parallel Synthesizers

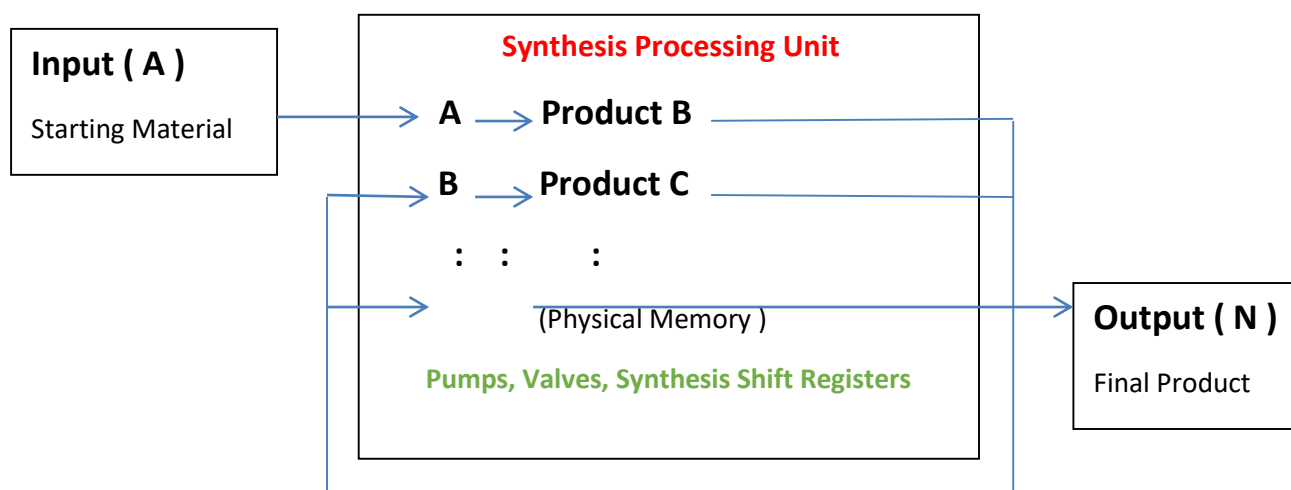
Parallel Synthesizer is a high yielding multiple synthesis systems consisting of parallel processing units & multiple synthesizers and these automated multistep units are used as parallel synthesizers for high yield applications. Parallel synthesis with cell synthesis processes is a way to use the advantages of combinatorial synthesis and this results in a smaller, more concentrated set of molecules, making the process of unit level synthesis easier.

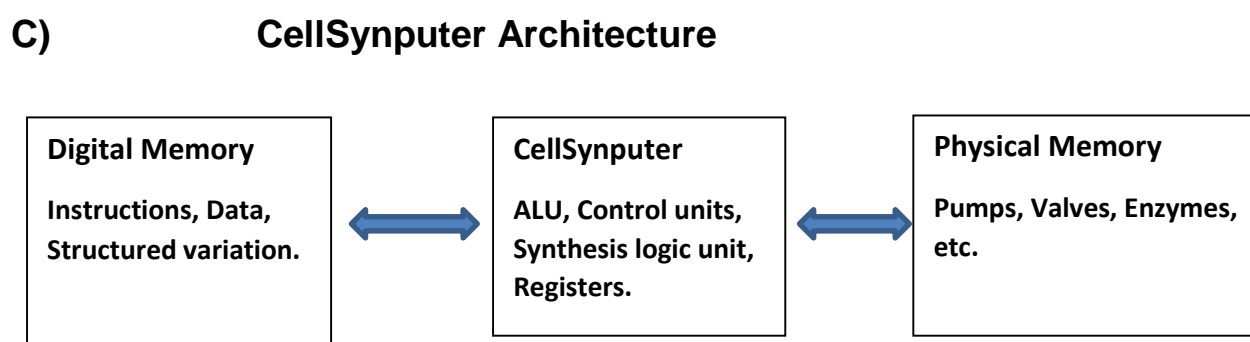
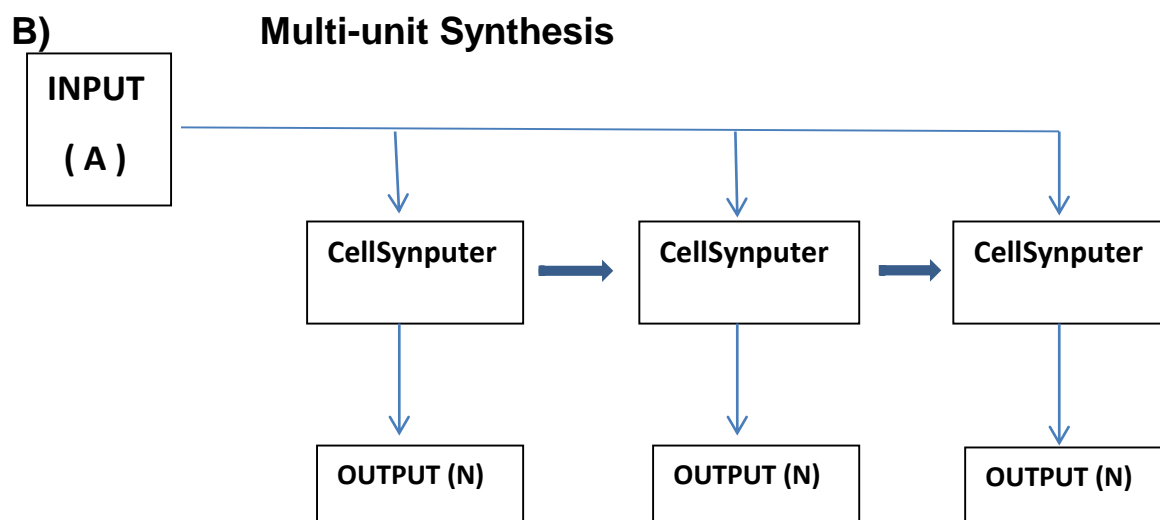
The following are the attributes of parallel synthesizer:

- Based on multi-unit concept
- Configurable at unit level
- High throughput
- Small scale at unit level
- Limited to individual synthesis scope
- Embodies multistep procedure

We give below automated cell synthesis using parallel synthesizer in pictorial format:

A) N-Step Cell Synthesis





Neural Networks in Exploring Synthesis Space

The automated synthesis could also make use of analysis and combination of starting materials for planning the synthesis routes to achieve the target molecules. There are many approaches to automated cell synthesis, and the one seems to be particularly promising as it employs neural networks and AI and it uses Monte Carlo tree search and symbolic AI to discover target molecule via different synthesis routes. The neural networks are required to be trained on all possible reactions in cell synthesis for a particular crop. The trained AI system allows cracking for many target molecules, faster than the traditional computer-aided search method, and this approach allows for faster and more efficient synthesis combination and analysis than any other well-known method. Figure 5 shows a workflow for joining automated synthesis of a target molecule of a desired crop with a synthesis robot and reaction optimization. The synthetic process module will generate a valid synthesis of the target that can then be transferred into synthesis code that can be executed in a CellSynputer/robotic platform. The optimization module can optimize the whole sequence, getting the feedback from the robot.

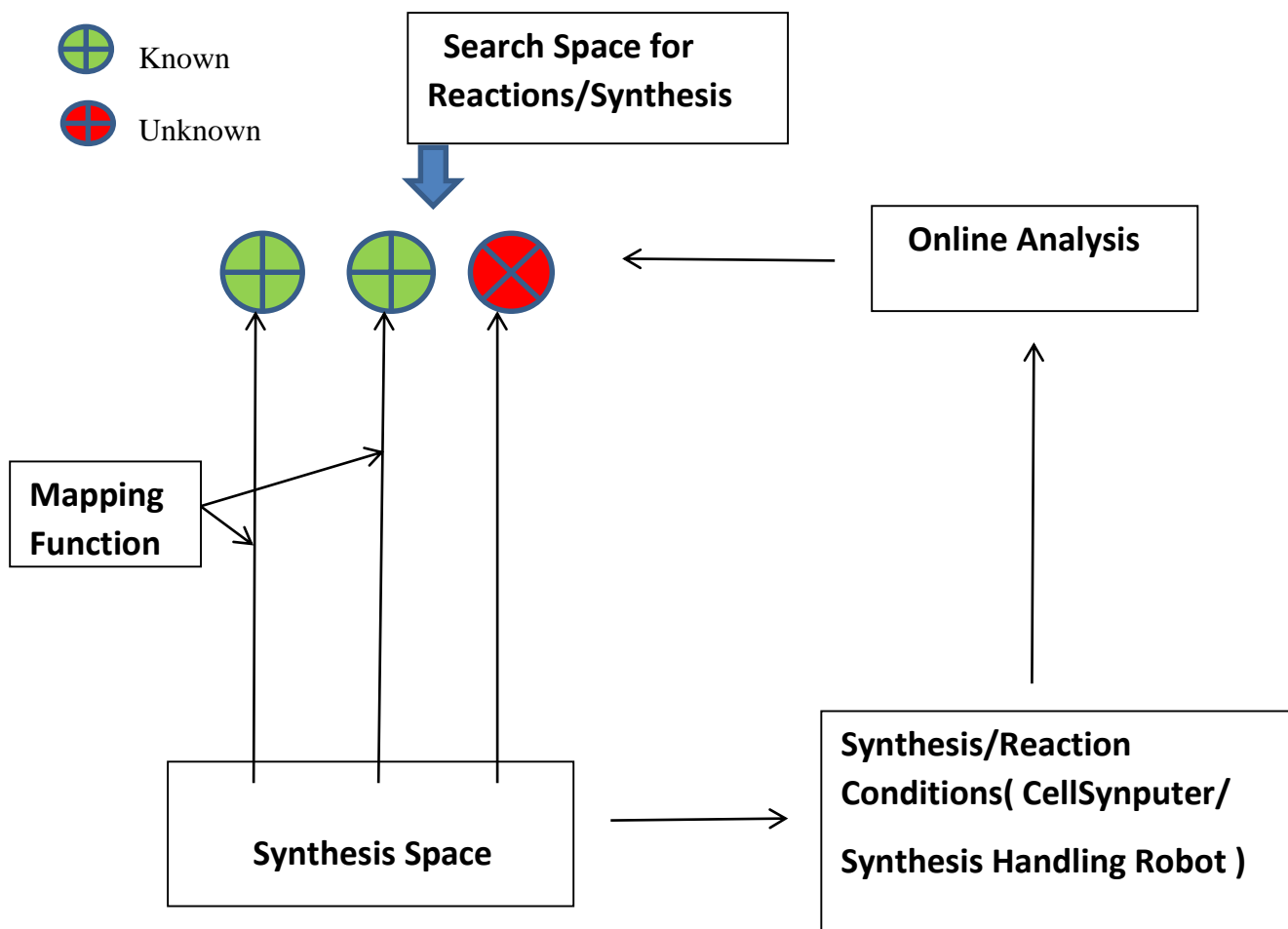


Figure 5. Exploring the Synthesis Space of Experiments with Neural Networks.

The platform operates in a closed loop with a machine learning algorithm; the machine learning algorithm suggest the most promising combinations and reactions that were then conducted and analysed automatically within the platform. The results of each experiment are automatically interpreted and the data are then used to update the machine learning model. The use of machine learning allows for autonomous exploration of synthesis space.

A standard crop grain composition parameter (like fibre, protein, carbohydrates, etc.) dataset is the first step and the data need to be collected from different subjects of variety. And also the dataset need to split into training (70%) and test (30%) sets based on data for subjects.

Robotic Microcontroller

A **microcontroller** is a compact integrated circuit designed to govern a specific operation in an embedded system. A typical microcontroller

includes a processor, memory and input/output (I/O) peripherals on a single chip.

A robot microcontroller is basically the brain of the robot. It is used to collect the information from various input devices such as sensors, switches and others. Then it executes a program and in accordance with it controls the output devices such as motors, lights and others

Microcontrollers are used in automatically controlled products and devices, such as automobile engine control systems, implantable medical devices, and other embedded systems and one of the main application of Microcontroller is sensing and controlling (process control) devices and this feature will be used in automatically controlled flow in CellSynputer.

Working of Divisive K-Means

The execution of the K-Means algorithm is explained in the below steps:

- 1) Select the number K to decide the number of clusters.
- 2) Select random K points or centroids.
- 3) Assign each data point to their closest centroid, which will form the predefined K clusters.
- 4) Calculate the variance and place a new centroid of each cluster.
- 5) Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.
- 6) If any reassignment occurs, then go to step-4 else go to FINISH.

Choosing the value of "K number of clusters"

One of the algorithms used to perform divisive clustering is recursive k-means. As the name suggests, you recursively perform the procedure of k-means on each intermediate cluster till you encounter all the data samples in the system or the minimum number of data samples you desire to have in a cluster

So, this recursive K-means is **a. based on the divide and conquer strategy**. The major difference is that, divide and conquer normally partitions the problem into predefined number sub-problems at each stage, whereas, the proposed recursive K-means decides the number of sub-problems.

The performance of the K-means clustering algorithm depends upon highly efficient clusters that it forms. But choosing the optimal number of

clusters is a big task. There are some different ways to find the optimal number of clusters, but here we are discussing the most appropriate method to find the number of clusters or value of K. The method followed is variations within a cluster by measuring **Cluster Sum of Squares**.

Implementation of K-means Clustering Algorithm

Before implementation, let's understand what type of problem we wish to solve here. So, in absence of real cell type data of desired crops, we have created a dataset of **University_Students**, which is the data of students who study in the university and get educated there.

In the simulated dataset, we have **Student_Id, Nationality, Gender, Age, Course, Subject and Subject Score** (which is the value of how much a student has scored in the subject, the more the value, the more the student has performed). From this dataset, we need to calculate some patterns, as it is an unsupervised method.

The following steps are followed for the implementation:

- **Data Pre-processing**
- **Finding the optimal number of clusters**
- **Training the K-means algorithm**
- **Visualizing the clusters**
- **Measuring the Output**

The above K-means based approach for estimating the number of cell type clusters is implemented using Python.

RESULTS

We systematically run k-means clustering algorithm on estimating the number of student types in a variety of settings by sampling from the university student data to create sequential datasets with a varying number of student types, varying number of students in each student type, and different student type proportions. After running the algorithm and examining the output, we could see the optimal number of clusters depicting the student types in each of the criterion. Therefore, the large number of datasets enables us to assess the performance of the algorithm, covering different broad categories of approaches, from various aspects using a panel of criteria for achieving cell sequencing

and this further enables cross-comparing the performance on datasets with high cell numbers.

CONCLUSION

Quantum Generators (QG) creates new seeds iteratively using the single input seed and the process leads to a phenomenon of generating multiple copies of kernels in repetition. A key task in single-cell RNA-seq (scRNA-seq) data analysis is to accurately detect the number of cell types in the sample, which can be critical for downstream analyses such as cell type identification and for quantum generation. For this scRNA-seq data clustering, we designed divisive clustering algorithm with k-means and has been recursively designed to automatically estimate the number of cell types through optimising the number of clusters in a dataset measured by robotic and on-line spectroscopy. The lack of studies with realistic data, however, complicates the choice of the relevant datasets and we checked our approach by developing simulated data and found to have clustering concordance of cells to their predefined cell types. Although the study given us a method of automating and optimizing single type cell assemblies however, this need to be tested using natural crop tissues and it could be promising in quantum generation.

REFERENCE

1. Poondru Prithvinath Reddy: **“Quantum Generators: A Platform for Automated Synthesis in a Modular Robotic System Driven by Cell Programming”**, Google Scholar.