Creating DNA from Scratch for DNA-based Data Storage

SDDEC23-05

Anna Hackbarth, Astha Upadhyay, Colin Frank, Caden Wortman, Evan Walters

Table of Contents

Introduction

1.1 PROBLEM STATEMENT

In our modern world, we have an ever-growing need for data storage, protection, and privacy. DNA is at the forefront of solving these problems as it has an expansive capacity for information storage. While there have been tremendous advances in DNA synthesis technologies, it is still challenging to create sequences with high accuracy. Our goal is to continue and modify a prior team's approach to print hundreds of thousands of short DNA oligomers, which are the building blocks for larger segments.

We hope to develop the current system further by improving the GUI and software to operate a Photon Mono 3D printer. When continuing another team's work in a field where our team has minimal experience, which in this case was biology, it is critical to know the ins and outs of the prior work completed. Most of our work this semester revolved around research and understanding the big picture of this project. Our goal is to integrate the various components of the project to run autonomously under this GUI. The desire is that our system design will be able to meet technical constraints, and user needs, and satisfy a new medium of digital storage to keep up with increasing demand at an affordable price.

1.2 REQUIREMENTS & CONSTRAINTS

While DNA is an impressive solution, why haven't we seen mainstream implementations? While DNA data storage is relatively sustainable and research shows incredible storage capacities within DNA. By creating DNA from scratch, ethical dilemmas of standard DNA manipulation are avoided.

These constraints and requirements become clear when discussing the technical aspects of the project. Liquid Crystal Display (LCD) is an important aspect of the project because it displays the image of DNA and what binds the nucleotides to the DNA within the microarray. The current problem with the LCD is light leakage. Potential solutions for the said issue will be discussed later in this document.

Another constraint was found as the previous group for the project had working code which we now have access to, however, they did not ever integrate their pieces together to make them easily usable. We are going to need to interconnect the microcontroller to the fluid system and create an interface that we can use with both systems, which will be done through the front end that we are going to create. When we finish the front end, we should be able to control the system without having to manually run specific code and increase the overall useability.

1.3 ENGINEERING STANDARDS

Systems and Software Engineering – Life Cycle Management Part: 1, Guidelines for Lifecycle Managment is one that is quite prevalent within our project. Since we are continuing an existing but incomplete project, Life Cycle Management is very important. .

Systems and Software Engineering – Software Testing Part: 2, Test Processes was chosen since we need to integrate the existing micro-fluidic programming with our own software, this Engineering Standard discusses using multiple testing processes to assist in the creation of complicated software.

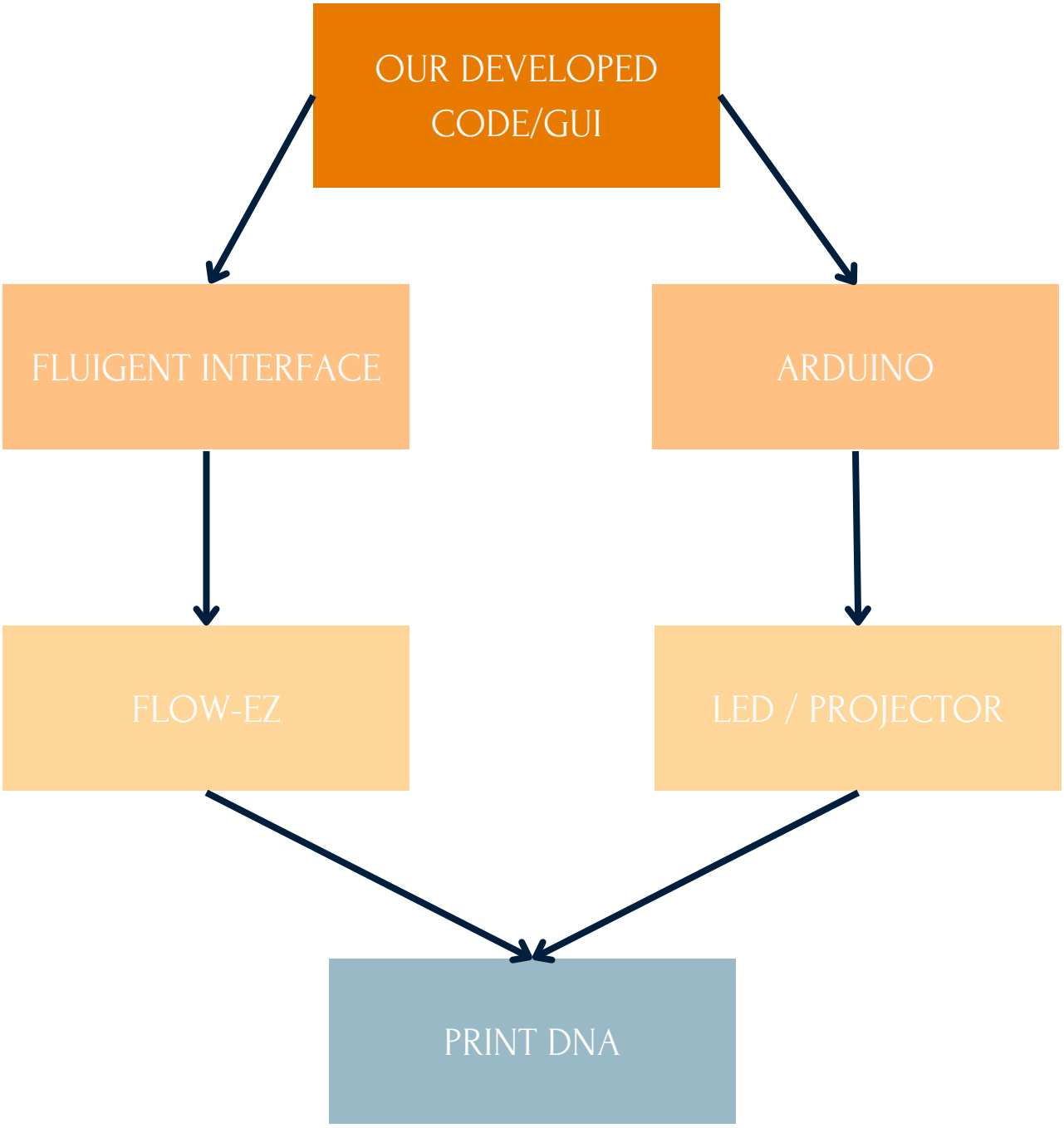
Systems and Software Engineering – Life Cycle Management Part: 6, Systems and Software Integration describes the integration between software and hardware. This applies to our project because we have separate teams working in both sections. This Engineering Standard provides us with guidelines on how to approach the cooperation and integration between the hardware and software sides of the problems.

1.4 USERS & INTENDED PURPOSES

Our team believes that there are stages to our project, so our users and clientele base will shift and grow. Starting out now, due to some constraints we listed earlier – our user base is limited to entities with a stake in DNA data storage such as research facilities or educational institutions such as Iowa State or MIT. They could also be corporations, one notable example is Microsoft, which has a tremendous amount of data that needs to be stored and shares concerns about the limited capacity we currently have for data storage.

Down the line, our team believes that as DNA data storage becomes a quicker, more efficient, and less costly practice – we will see this being implemented by a wider consumer base. We could see within five to ten years that DNA-based storage becomes available for general consumer purchase. Technology evolves so fast and we are hopeful that within our lifetime we can solve the storage problem.

1.5 WHOLE SYSTEM SYNOPSIS



2.1 PROJECT MANAGEMENT

We will use a more agile waterfall approach as our team's roles have us divide up several waterfall approach as our team's roles have us divide up several different projects to work on concurrently while still reaching milestones as a team. For example, Anna works on improving the flow control system while Colin and Evan work on improving software, and Astha and Caden improve the 3D printer and backlight screen. These are projects that occur simultaneously and will still allow us to achieve goals in a linear fashion.

Our group will use Slack, Discord, Figma, and our shared online notes document to track progress, ask questions, and plan out the next steps. We plan to maintain contact with the client and faculty advisor throughout the summer using our official Slack channel.

2.2 TASK DECOMPOSITION

Within this project, there are three divisions of labor as listed above: software, hardware, and flow control. The next page details each area of work and subtasks within that field.

2.2.1 Flow Control Tasks + Subtasks

- Familiarize self with flow control system
    - Gather dimensions
    - Compile parts and materials list
- Make more/better flow cells (single-use)
    - Learn about how the prior team built the flow cells
    - Practice using plans from last year to replicate flow cell
- Regression Testing
    - Think of ways to improve pressurized piping/identify the problem areas - test with ink
    - Construct pressurized piping using last year's designs
    - Test with ink under different conditions heat, wear, etc.
- Coding
    - Using Software Development Kit and OxyGEN to test the system
        - Write code for controlling Fluigent system
    - The flow control system comes with software, we want to combine commercial software provided with our developed code

2.2.2 Software Tasks + Subtasks

- Create a skeleton GUI
    - We plan to use the GUI to automate most of the progress, so this is the largest portion of the project.
- Work with the Flow Control team to integrate and test
- Create a user-friendly front end for consumer, researcher, and institution use

2.2.3 Hardware Tasks + Subtasks

- Research current systems and potential upgrades
- Implement new display system/projectors
- Perform unit and integration testing

2.3 MILESTONES AND EVALUATIONS

2.3.1 Software

Study the C# code from the S22 team to generate a pattern on a monitor screen (March).

- Understand the overall frame of the code
- Update the GUI user input and image generation function

Add the code to control the Fluigent pump and switch (April)

- Include the Fluigent SDK
- Write a loop to run 30-40 cycles of liquid injection, incubation, wash

Add the control of LED light source (May)

- Directly comminate via the USB port

2.3.2 Hardware

LCD panel update (March/April)

- Study the display interface
- MIPI/LVDS
    - (https://focuslcds.com/the-mipi-and-lvds-display-interfaces/)
- Optical characterization of the LCD panels
- Transmission measurement
- Black/white transmission ratio

LED on/off switch (May)

- Use the software to turn the LED on/off
- Add the function to adjust LED intensity

2.3.3 Flow Control

Fabricate flow cell (March/April)

- Use acrylic sheets to assemble 5-10 flow cells
- Add inlet and outlet connector

Test the flow cell stability (April)

- Run 30-50 cycles of the flow
- Observe and fix potential problems (clogging of the pipe, trapped air bubble, broken connectors)

Surface functionalization of the flow cell (May)

- This step is to immobilize the DNA molecules on the surface

Create GUI code for the flow control system (April/May)

2.4 RISK MANAGEMENT

Below are the details of risk factors that we anticipate facing in this project alongside the probability of occurrence, and our mitigation plans.

Risk Factors	Probability	Risk mitigation plan
Continued issues with dark spots in the new LCD screen	.5	We may need to add a more complex and efficient heat sink to the bottom of the LCD
The high-power LEDs not being responsive using the given API	.2	
Issues integrating the previous groups code together with itself/old bugs in the code	.7	We will keep our options open on needing to rewrite code if there are unforeseen issues with using the code of the previous group
Issues controlling the fluigent pump and switch automatically with code	.3	
Leakage in the connection points of the Flow Cell	.2	
Clogging of the pipe in the flow system	.4	

2.5 PERSONNEL EFFORT + TIMELINE

Task	Software	Hardware	Flow System
March	Look over and understand previous teams code, get better intuition on programing language, run it to make sure it works, then update the code GUI to make it work better.	Study the display interface, learn how it all works and what needs to be done to improve it	Use acrylic sheets to assemble 5-10 flow cell. Build better flow cell sheets to improve how the DNA gets dispersed from the system
April	Include the fluigent sdk and write the loop to help automate the system and bringing it all together	Optical characterization of the LCD panels (Transmission measurement and Black/white transmission ratio). Need to make the LCD better, reduce light leakage and flow heat away from system so it doesn't mess with final DNA image.	Run 30-50 cycles of the flow and Potential problem (clogging of the pipe, trapped air bubble, broken connectors) After making them, test them and make sure all fluid gets cleared out after A,T,G,C and that those get to microarray properly.
May	Set it all up to get the picture array more consistent with the code and connecting it to the USB port to connect to the computer to display the picture	Use the software to turn the LED on/off and add the function to adjust LED intensity. Connecting software and hardware together to help automate system	This step is to immobilize the DNA molecules on the surface. Combing it all together to get the DNA to flow properly to where its supposed to be and making sure that it gets cleared properly so there is no contamination.
Total Hours:	45 hours for software	45 hours for hardware	45 hours for flow system

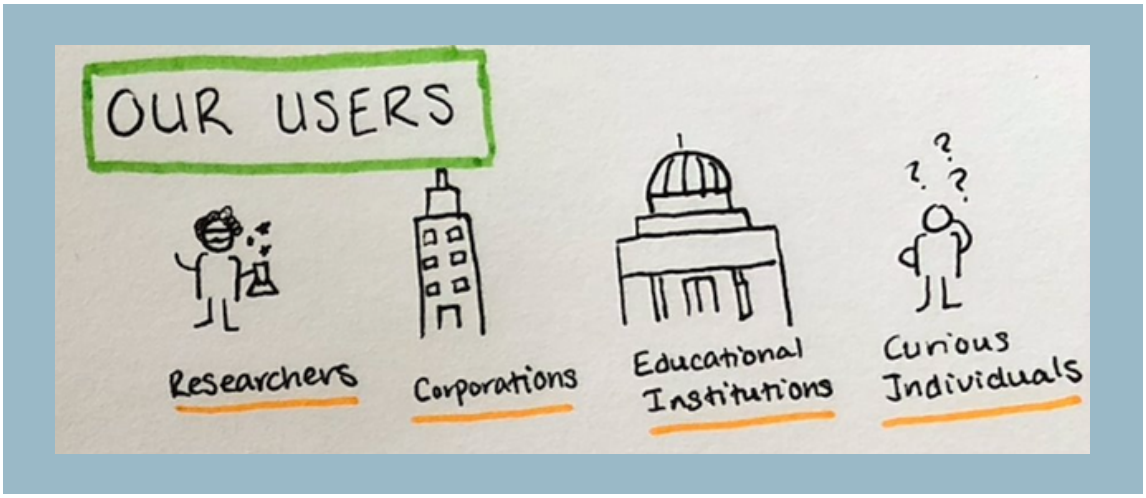
3.1 DESIGN CONTEXT

3.1.1 Broader Context

Area	Description	Examples
Public health, safety, and welfare	This project has a direct effect on the shareholder groups in this because everyone is going after the same goal of using DNA as storage and everyone is facing the same challenges.	Creating jobs / opportunities for research when money is dumped into it.
Global, cultural, and social	The project is culturally appropriate because we are trying to figure out a way to store data in DNA and create longer nucleotides of the DNA without any mutations. This would affect everyone in some way because of how big this research is and the possibilities of this.	Could potentially say once this gets figured out, I don't know how all cultures would feel about replacing someone's DNA strand (Take out bipolar gene), so genetics doesn't keep passing it down.
Environmental	There wouldn't be any environmental impacts because there are no harmful chemicals / unsustainable practices being used when doing research.	The amount of energy needed to create the long strands nucleotides / extract the DNA to get the data is unknown.
Economic	The economic impact of this is significant because this could change the way we store data in a more efficient way. Eventually data storage will get more and more expensive.	Every business would store data this way because it would be cheaper and more efficient.

3.1.2 User Needs

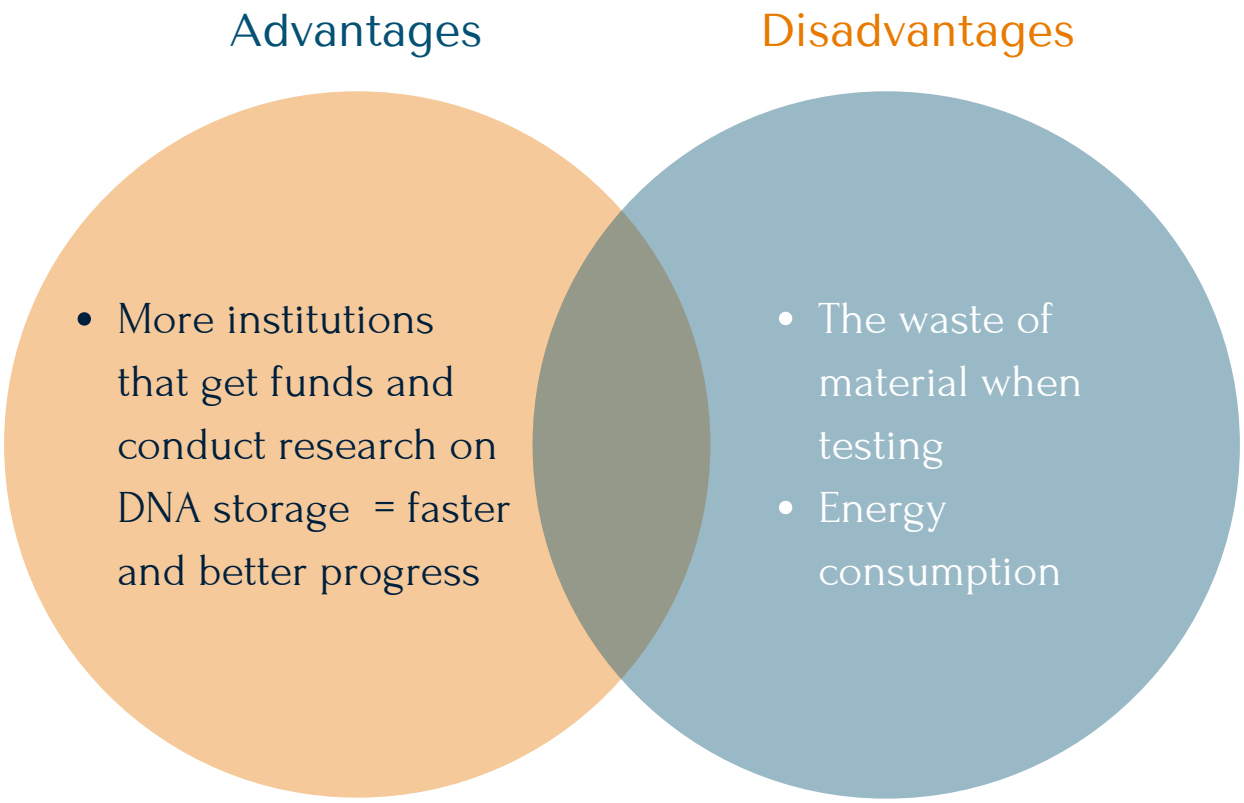
1. Researchers and STEM Institutions
    a. Researchers need a way to affordably experiment with DNA pairs because it will allow them to advance their research faster.
2. Educational Institutions
    a. Educational Institutions need a way to decrease the lead times of outsourcing DNA printing because it allows for faster access to the DNA and more flexibility.
3. Private Corporations
    a. Private Corporations need a way to increase efficiency in storing data because it will eventually become increasingly expensive to do so.
4. Curious Individuals
    a. Curious Individuals need a way to affordably print their own DNA pairs because it is currently far too expensive for a (normal) singular individual to fund.



3.1.3 Background Information

Relevant background/literature review provided by our advisor, Professor Meng Lu

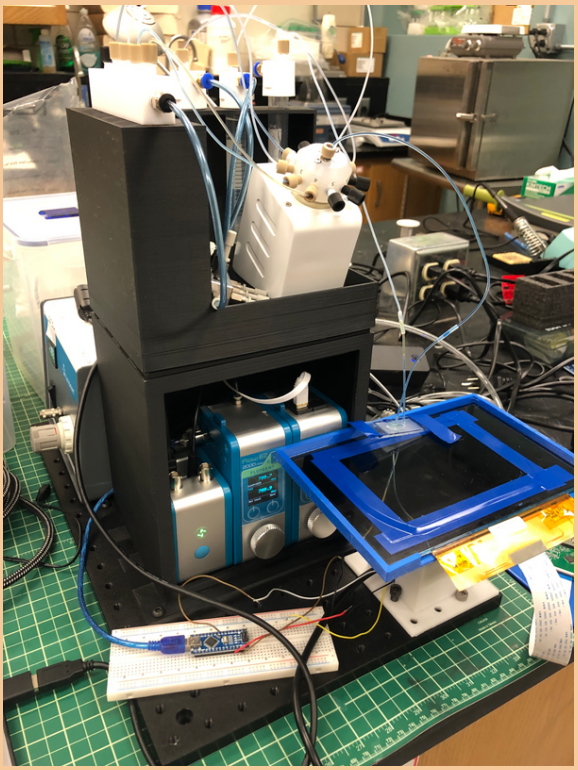
- Phosphoramidite Chemistry:
    - https://www.twistbioscience.com/blog/science/simple-guide-phosphoramidite-chemistry-and-how-it-fits-twist-biosciences-commercial#:~:text=Phosphoramidite%20chemistry%20is%20the%20gold,200%20base%20pairs%20in%20length
- Oligonucleotide Synthesis
    - https://en.wikipedia.org/wiki/Oligonucleotide_synthesis
- ADS Codex
    - The Los Alamos National Laboratory has developed technology that enables molecular storage. The Adaptive DNA Storage Codex (ADS Codex) translates data files from the binary language of zeroes and ones that computers understand into the four-letter code biology understands.
    - ADS Codex is a part of the Intelligence Advanced Research Projects Activity (IARPA) Molecular Information Storage (MIST) program
        - https://www.scientificamerican.com/article/dna-the-ultimate-data-storage-solution/



3.1.4 Technical Complexity

Our project contains three main branches which we have to work on: Software, Hardware, and the Flow System. Individually each branch is complex on its own. The Software Team needs to create a fully integrated program connecting with both the flow and hardware along with implementing it into an easily usable user GUI. The Hardware Team needs to work with LCDs, and projectors, and optimize them so it keeps up with the rate we need them to without the heat interacting with the DNA. The Flow System Team has to build a system that quickly, reliably, and repeatedly sends chemicals through a custom-made cell and countering any potential problems working with fluids that we have never been taught to work with.

This is all being built on top of the work of a previous group, therefore we are continuing a project from other people and prior teams, this reflects an important aspect of what career engineers need to do in the field.



Figures show the current printer, flow control system, and various other components

3.2 DESIGN EXPLORATION

3.2.1 Design Choices

1. Software development:
    a. GUI, hardware interface, DNA input/analysis, and pattern generation
2. Hardware
    a. LCD panel, UV/blue LED source
    b. Midsemester change to projection
        i. Prevents light leakage and other issues
3. Flow System
    a. Flow cell fabrication, Fluigent system

3.2.2 Ideation

We explored a few methods for identifying our potential options and we settled on the SCAMPER method. The SCAMPER method is a creative thinking technique that can be used to generate new ideas or solutions by exploring different aspects of an existing idea or product. The acronym SCAMPER stands for: Substitute, Combine, Adapt, Modify, put to other uses, Eliminate, Rearrange. By using the SCAMPER method, we can approach problems or ideas in a more structured and systematic way and generate a variety of creative solutions.

We also considered brainstorming, mind mapping, six thinking hats, SWOT analysis and reverse thinking. We were heavily inclined towards reverse thinking based on the nature of our project, but SCAMPER method aligned more with the design decision process.

3.2.3 Decision Making

The SCAMPER method and the other techniques we listed are all valuable tools for generating new ideas and solutions in a team setting. The acronym provides a clear framework for exploring different aspects of an idea or product, which can help to guide the team's thinking and ensure that all possible avenues are explored. On the other hand, some of the other techniques we listed, such as brainstorming and mind mapping, may be more effective for teams that prefer a more open-ended, free-flowing approach to idea generation.

3.3.1 Visuals and Description

Currently, our design is fully based on the last group's progress. It includes a high-powered LED, an LCD, a microarray, a micro-fluidic system, a microcontroller, and a laptop to use to run the code and work as the GUI. The LED and LCD are housed below the micro-array to be used to solidify the A, T, G, and C solutions in their proper locations. The microfluidic system will wash over the microarray with the desired solution and then when complete, use a cleaning fluid to dispose pf the unused solution. Everything is housed within a 3-D printed housing unit. There is also an air compressor that will be providing the microfluidic system with pressurized air to be used in moving the liquids around the system.

3.3.2 Functionality

The question we need to answer is: *How well does the current design satisfy functional and non-functional requirements?* The previous team has made enough progress in making the current design work, now we are trying to make it work more efficiently and get the bugs fixed in the code and upgrade the hardware to optimize the potential of the machine.

3.3.3 Areas of Concern

The difficulties and concerns come into play with the integration of software and hardware. There is separate software that manages the flow control system, separate software for the 3-D printer and LED board, and countless other hardware pieces that need to work together. Our largest priority and concern is creating a well-integrated machine out of the individual pieces and components that the last team created.

We got a later start than most groups with our project. The first six to eight weeks of the project was mostly review and reading since we are dealing with biological components that many other ECpE senior design students didn't need to consider. This along with late semester design changes have us concerned we won't meet all our goals. We plan to address this by working throughout the summer to further the project.

3.4 TECHNOLOGY CONSIDERATIONS

We are going to primarily focus on the technology in use for DNA synthesis. Our team is utilizing oligonucleotide synthesis which traditionally uses column-based or microarray synthesizers. We will be using a column-based synthesizer that follows a four-step progress:

1. Deprotection
2. Coupling
3. Capping
4. Oxidation

A brief description of each stage of single-strand synthesis is given below.

1. Deprotection
    a. Removal of DMT enables a strand to build together and form.
2. Coupling
    a. A phosphoramidite, or a bonding agent is added to the solution/liquids.
3. Capping
    a. Acetylation occurs, which prevents further uncontrolled chain extension.
4. Oxidation
    a. Exposure to oxygen and aqueous iodine causes a chemical reaction (capping)

PROS:

- WELL RESEARCHED
- UPDATES

CONS:

- OLDER AND SLOWER PROCESS

3.5 DESIGN ANALYSIS

One observation that we've mentioned previously is the LED board. While initially, we wanted to improve the LED with another similar model - it became clear that this was not the proper avenue. We have since switched to utilizing a projector which lessens heat and light leakage which damages the DNA.

Testing Plan

4.1 UNIT TESTING

- 3D Printer
  - Running test prints using software and code we created/improved
- Flow Control
  - Using Fluigent's software OxyGEN and ink to run a standard test cycle
- LED Board
  - The current LED board has been in talks to be upgraded; the board (current or new) will need to be tested with the current Arduino and hardware systems.
- Projection
  - Since the decision was made to invest in using a projector rather than a board, we would like to utilize the resources in the ECpE Grad Student Office to test how the lighting will affect the process and development of the DNA.

4.2 INTERFACES

- The 3D printer and the flow control system
- LED board / projector with the Arduino
- Our developed GUI with SDK of the flow control
- All code with the total hardware build

4.3 INTEGRATION TESTING

Integration of Units is critical to our project as one of our requirements was combining systems designed by the prior year's team. We will test this by utilizing the above listed interfaces such as the integration of the Arduino, software, LED board, and the flow control system.

4.4 SYSTEM TESTING

We will probably use performance/usability tests on each unit to make sure they perform well under stress and are functional. We will start using usability tests on the prior team's work to see if changes need to be made before we begin the whole system integration. This will give us a chance to improve individual units such as the code and perform regression testing to make sure our changes do their intended purpose. We can then start with smaller integration tests to confirm pieces work together with the code, for example, we would perform a test to run the flow control system using only code developed for the GUI.

4.5 REGRESSION TESTING

Regression testing evaluates the functionality of pre-existing systems, then again tests the functionality of those systems with modifications or changes. One idea we have to track changes is to begin leaving comments in the code that dictates changes, and the purpose of old sections, and maintain general notes and thought processes behind new additions. The documentation will allow us to follow the code and debug more efficiently should something break.

We've had a lot of regression testing and research to perform this semester which has made us feel behind in the project since we have had less time to implement and test our desired changes such as upgrading to a projector instead of an LED board.

4.6 ACCEPTANCE TESTING

Fortunately, our client is also our advisor for the project. Professor Lu will be with us for every testing stage prior to acceptance testing and therefore should be easy to involve. We don't believe that we will complete the project, we can simply improve what was given to us and pass it on to the next team as this is a multi-year investment. Based on this observation, acceptance testing for this project might take the form of beta testing or integration testing, since the prior team never got to integrate units.

We also plan to heavily involve our client and advisor Professor Lu in the acceptance testing stage with our developed GUI and projector implementation. The prior's years team did not make much progress on the GUI, so we plan to work closely alongside our faculty advisor to ensure that our code is doing precisely what it needs to.

4.7 RESULTS

Unit from Tests	Result	Solution (if needed)
Piping, Flow Control	Clogging, kinks, inconsistent flow	Replace piping, allow for enough slack to prevent kinks
LED Board	Light leakage causes DNA to burn	Switch to a projector unit, change design and testing as needed
Software	No integration or performance tests thus far	Use debugger, begin testing code in manageable sections

4.8 CONCLUSIONS

Accomplishments:

- Flow Control testing and design changes
- The foundation of the GUI was built
- Issues diagnosed with LCD, designs updated

Goals for Summer:

- Utilize the summer months to finish GUI
    - Finish flow control GUI component
- Complete necessary paperwork to access Grad Student Office for next stage projector testing
- Begin primarily integrations, Fluigent SDK with GUI

Goals for Fall 2023:

- Implement hardware Flow Control changes
- Begin integration testing with our software and the physical system.
- Switch LCD to the projector system
- Whole system testing

In conclusion, most of our work this semester involved laying the groundwork for a productive second semester. We needed to conduct a great deal of research into DNA synthesis, perform regression testing on the work done by the prior team, and modify the design.

Due to the amount of research necessary to set up a strong second semester, we got a later start than most other teams. For that reason, many of our initial aspirations were not met and adaptations to our scheduling, end goals, and processes had to be made. These adaptations include working over the summer, adjusting team roles, and assigning different priority to tasks.