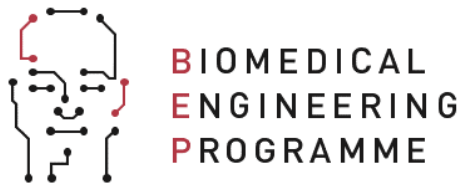


University of Pécs
Faculty of Engineering and Information Technology
Biomedical Engineering Master program

**The Design, Modelling, and Simulation of a Polymerase Chain Reaction
(PCR) Thermocycler Based on the Thermal properties of DNA.**



Author: Mhd Mouayad Ahmad Abdalla

Supervisor: Dr. Sári Zoltán

Pécs

2024

Abstract:

Polymerase chain reaction (PCR) is a revolutionary technique for DNA amplification, today it is considered the most important technique in biotechnology as its uses range from DNA and RNA cloning, DNA sequencing, medical forensics, etc., a PCR Thermocycler is a device that cycles and holds the temperature of samples on predefined conditions, it consists of a heating and cooling mechanism, a heat block that holds the sample which have DNA/RNA in them. Commercial thermocyclers use temperature sensors to measure the temperature of the heating block rather than measuring the temperature of the samples, and by using a controller the device can control the temperature of the system. This paper introduces a new possible technique for designing a more accurate thermal controller for PCR that can allow very specific primer binding temperature control, rather than a traditional block temperature feedback control. This can be achieved by controlling the temperature of the samples based on the thermal properties of the PCR test mixture. This can be achieved by using a modelling environment and analysis and control software such as COMSOL Multiphysics and Simulink.

Introduction:

From the beginning of the twentieth century and with the emerging studies about genetics, scientists began showing interests in genes and how information is copied and transferred between cell generations, and how a multicellular organism can theoretically replicate its cells for an infinite number of times, and how is this enormous amount of information is preserved in something as tiny as a singular cell [1].

In the 1940s after thorough research about genetics, scientists discovered that genetic information is largely instructions to make different proteins, proteins play many roles for living organisms as they play a role in gene expression, serve as building blocks for cell structures and enzymes, as well as cellular communication. Later it was discovered that hereditary information is carried on chromosomes, these chromosomes consisted of a series of Deoxyribonucleic acid (DNA) and protein, DNA is composed of a double stranded helix of deoxyribonucleic acids, the length of the sequence and the type of the pair identifies all living forms on earth, as each organism has its signature sequence [1].

The invention of new technologies has always transformed the way humanity thinks about a certain problem, In the case of microbiology having the ability to replicate, clone, and sequence various DNA and RNA sequences was a major challenge, the PCR method was considered as a breakthrough in the field of biology and biotechnology and according to Dr. Kary Mullis 'PCR changed the rules of molecular biology' [2-3].

Over the years the PCR technology dramatically changed, new more advanced PCR types were introduced such as the real-time PCR and RT-PCR, also, a huge advancement was made in PCR thermocycler, starting from water-bath based heating and cooling mechanisms, Peltier and Ohmic based heating systems, or even ultrafast microfluid mechanisms. Although PCR is an incredible technology, still it is prone to errors, one problem that has been present is the non-specific amplification of the DNA/RNA sample, to resolve this issue a new type of PCR thermocycle must be developed, using temperature micro-sensors implemented in the samples is an option, theoretically it is a valid, but because of the complexity of the design and the fact that the samples might evaporate if the sample tube is unsealed, hence, another type of thermocycler can be developed, by using a simulation that can predict the temperature profile of the samples and the thermal power input using the thermal properties of the sample which is what is being developed in this project [4].

Polymerase Chain Reaction (PCR):

Polymerase chain reaction was considered a radical leap in science, it is probably the most important discovery made in the field of genetics and biotechnology in recent time, it allows the amplification and replication of large amounts of DNA, RNA, and proteins, today every biotechnology research centre, laboratory, or medical diagnostic institutes must be equipped with at least one functional PCR thermocycler, PCR can be defined in simple terms as a basic -yet very powerful- method for selectively amplifying DNA and RNA using some a mixture of several components including [5-6]:

- Template DNA, representing the original DNA to be replicated.
- Oligonucleotide PCR specific primers, they come in the form of a pair, and they are directional and region specific, providing an initiation point to the amplification process.
- DNA polymerase is Very important enzyme in the reaction after the primary latches they start by building the new strands using the dNTPs they catalyse the reaction one of the most famous among them is Taq polymerase, it is widely used because of its thermal robustness and low replication error.
- Buffers, including $MgCl_2$ it is important to increase specificity, and control the chemical set of the medium they also include some additives such as (DMSO. Betaine solution, formamide solution, etc...).
- dNTPs (free nucleotides), they are the building blocks for the extended strands, for DNA there are 4 types of oligonucleotides adenine (A), thymine (T), cytosine (C), and guanine (G), in RNA thymine is replaced with uracil (U), via two hydrogen bonds adenine pairs with thymine (or uracil in case of RNA), and cytosine pairs with thymine.

The thermal cycler has a heating block usually made of aluminium or silver (where the samples are held in), and the block's temperature is monitored by digital temperature sensors and the temperature is controlled using a microcontroller, the temperature of the block is changed using a combination of heating and cooling mechanisms, and it is also changed according to predefined steps.

The first step is known as denaturation where the mixture is heated to a temperature above the melting point of the complementary DNA strands for the target DNA, the second step is called annealing the aim is to lower the temperature to allow the specific primer to bind to the target DNA to be replicated. The last step is known as extension or synthesis, at this stage the temperature is raised again so that the DNA polymerase can extend the primers by adding the free nucleotides to the developing DNA.

The previous three steps combined are known as a thermal cycle, in each cycle every DNA double helix is amplified, then new amplified strands act as new templates for even new amplifications, after cycling for 30-40 times the total number of DNA is exponentially increased. And it is approximated that a single DNA strand after 30 cycles can generate 130 million short products from the original molecule, it should be noted that the reaction ends if any of its components is depleted [5-7].

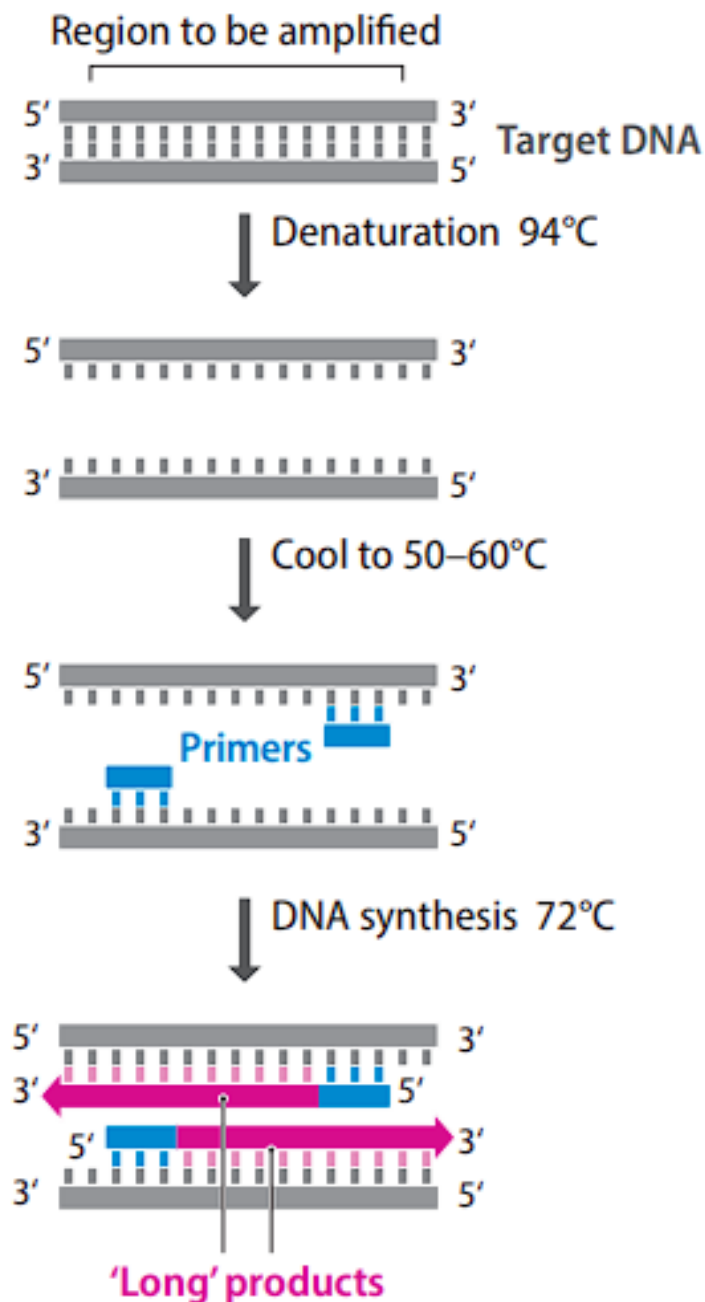


Figure 1 DNA amplification using PCR.

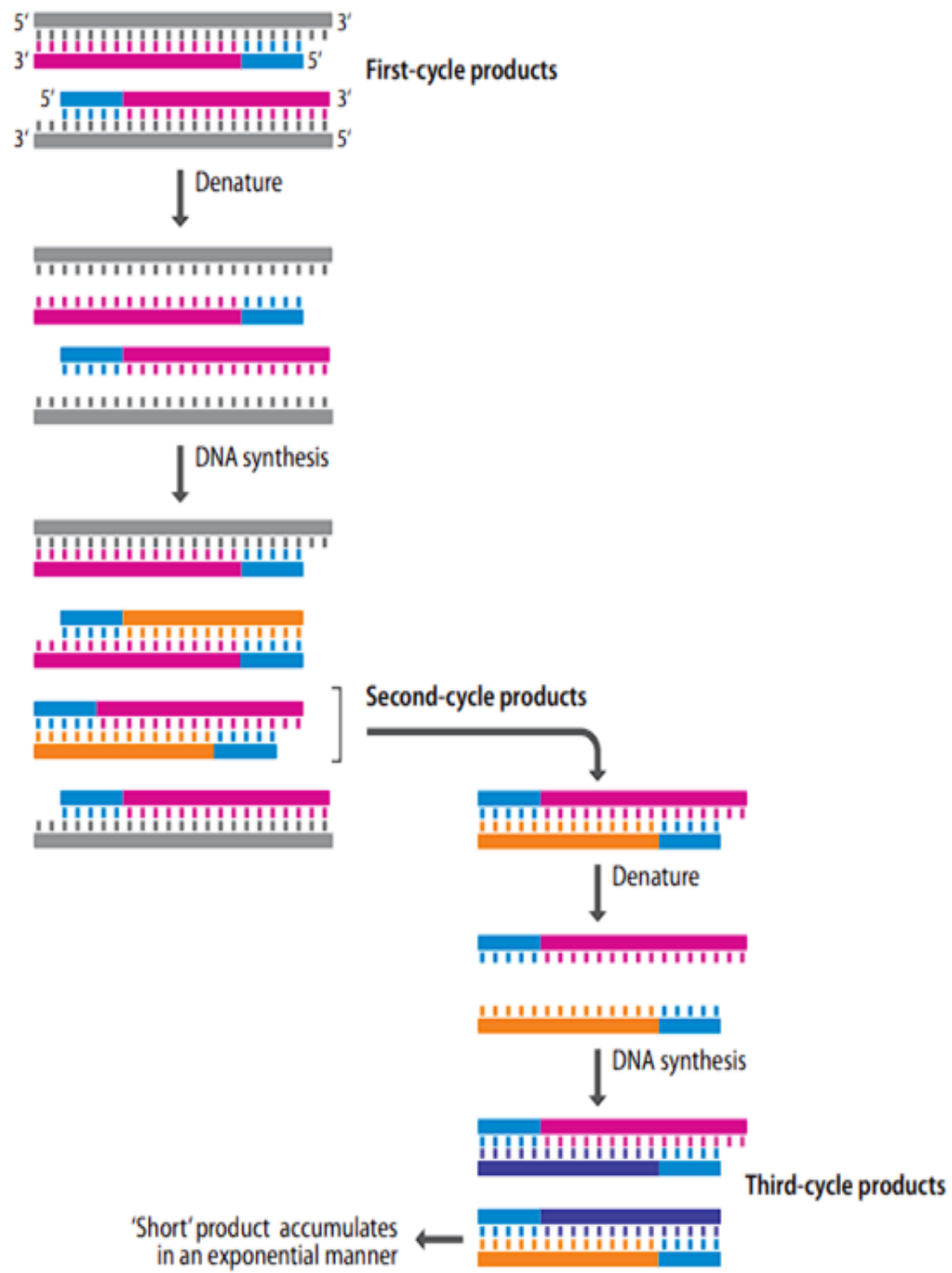


Figure 2 PCR amplification process.

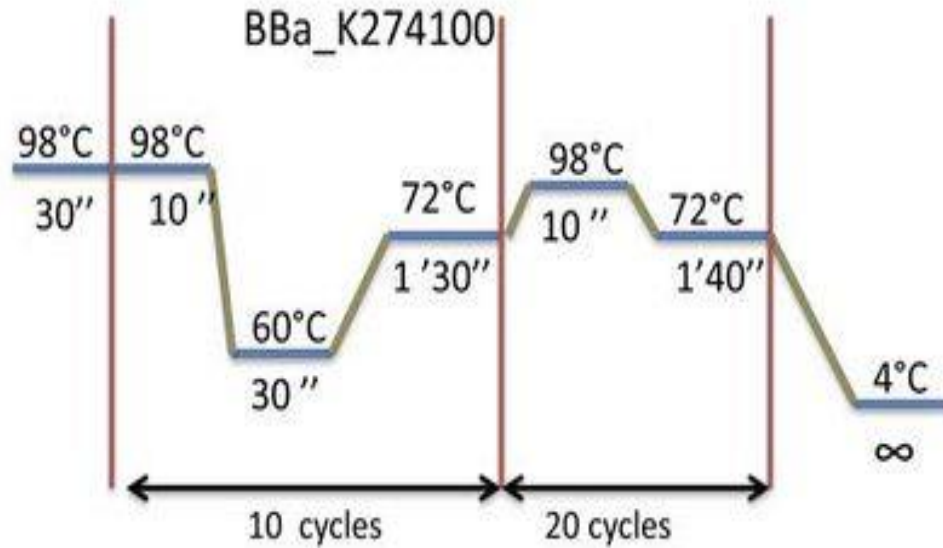


Figure 3 an example of a thermal cycle.

This graph represents a thermal cycle, a thermal cycle starts with the first temperature called the denaturation temperature, this temperature usually ranges between 94 to 98 degrees Celsius, it usually lasts between 30 seconds to two minutes, the second temperature is between 50 to 65 degrees Celsius it is called the annealing temperature this is where the PCR primer latches to the targets sequence, temperature selection is of extreme importance in this phase, it is regulated by the manufacturer primer design, and the duration for this phase is between 30 seconds to 1 minute. The last temperature is known as the extension temperature, and it is usually set to 72 degrees Celsius, in this phase the duration is set between 1-2 minutes. All three temperatures are set according to a protocol, or the required sample conditions by the operator clinician. The combination of the three temperatures and their respective times are known as a thermal cycle and the device repeats cycles between 25 to 40 times, heating upgrades and cooling rates are known as the ramp rates, and they usually range between 1°C for old generation PCR to 5°C which is the average for today's technology, but some newly developed types of microfluid PCR's can reach extremely high temperatures of 50 or even 100 °C as the volume of the sample is extremely small and the sample numbers are lowered and plate size decreased [7].

Thermal properties of DNA:

The science of studying the film of properties of molecules and Materials is a very vast knowledge, especially when we're talking about DNA as there are a lot of factors at play one of the most important factors when considering the thermal changes in the properties of the DNA is actually the length of the DNA sequence, One method of determining the thermal properties for materials is known as differential calorimetry scanning (DCS), understanding the thermal properties of DNA and RNA sequences are of extreme importance, weather for any thermal analysis of the sequence such as PCR or to fully understand The human genome especially after the results of the years-lasting research on sequencing the human genome [8-9].

'The heat capacity changes depend on the base composition and sequences and are generally limited in the range of -160 to $\cong -40 \text{ cal. mol}^{-1} \cdot \text{K}^{-1}$ per base pair' [8].

Differential Scanning Calorimetry:

Differential Scanning Calorimetry is widely used method for determining thermal characteristics for a given sample, it can be used to measure information such as heat capacity, heat flux, Enthalpy, peak temperature, ... etc. Its uses range from biological, physical, chemical, to engineering fields. One of the most important uses in biology is the study of protein thermodynamics specifically the change of temperature and heat capacity during the folding and unfolding o different proteins and biomolecules.

When it comes to DNA unfolding, unlike normal proteins where a two-state process between folding and unfolding happens, rather various DNA segments can start to melt at the same time, thus the Calorimetry Machine must have high sensitivity to small energy changes.

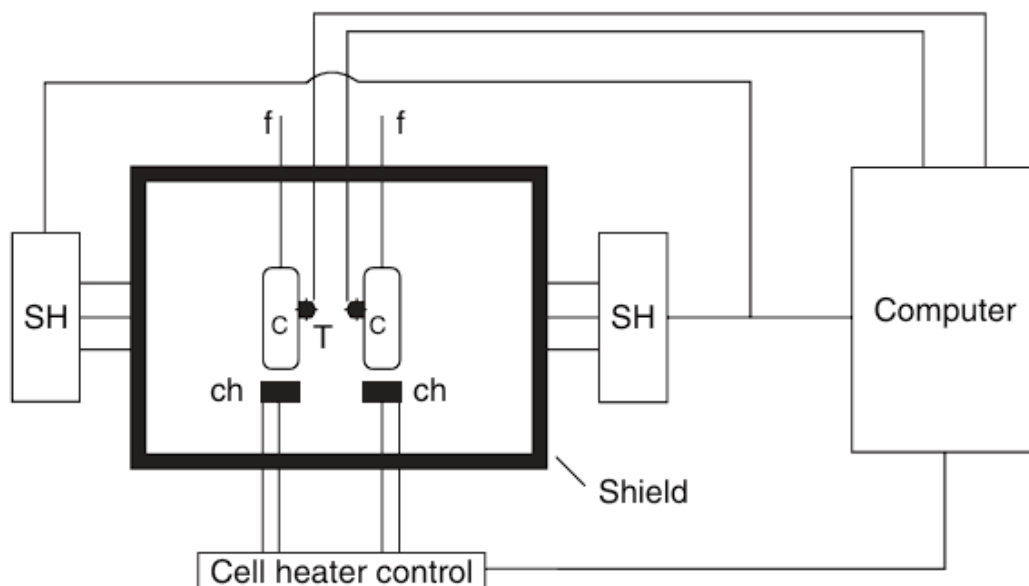


Figure 4 Diagram of differential scanning calorimetry (DSC).

The device has two chambers one for the sample and the other for the reference, the device is connected to a computer for data acquisition and analysis, the sample and reference are put in chambers or cell structures (C) and separate cell heaters heat both samples, a temperature sensor reads the temperatures and the difference is set to heat up the sample using the heaters, and it is controlled by the computer, the compensation energy per unit time is measured and recorded [10-11].

Heat capacity change of DNA:

It is found that heat capacity changes in relevance with the existing polar and nonpolar groups in the sample when this double strand is melted hydrogen bonds break down, and the heat capacity changes to become higher for the single strand [8].

an equation to Estimate the change of heat capacity according to the effect of hydrating the polar and apolar groups of proteins can be expressed in the equation:

$$\Delta C_p(25\text{ }^{\circ}\text{C}) = 2.14 * \Delta ASA_{\text{apolar}} - 1.27 * \Delta ASA_{\text{polar}} \quad (1)$$

A closer look at the melting of all-CG duplexes of different lengths, and thus melting temperatures, indicates that the assumption of a zero ΔC_p is indeed incorrect” and in fact, the heat capacity of the DNA is a function of base pair (bp) sequence, length, and temperature as well [9].

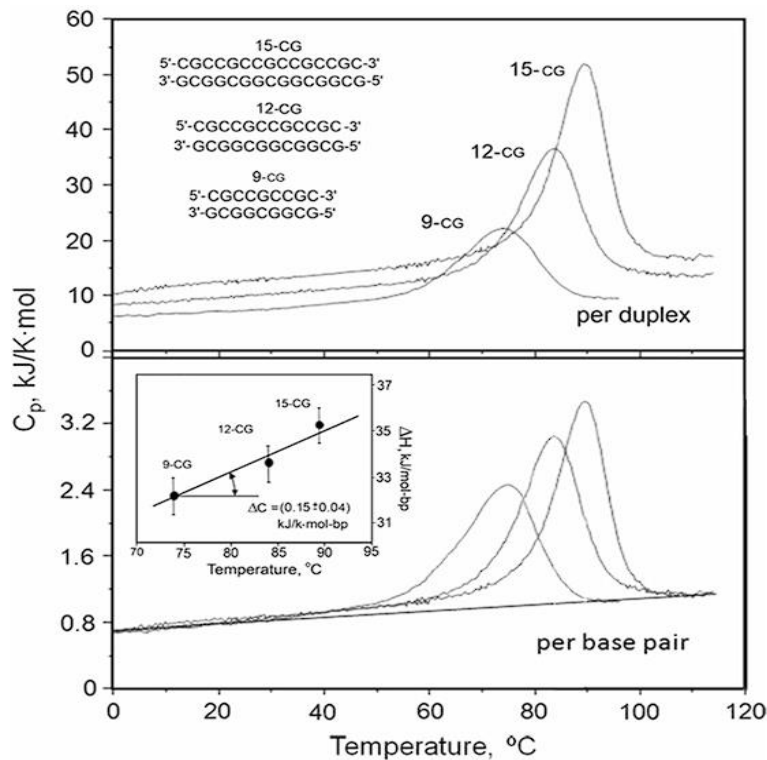


Figure 5 Partial Heat capacity functions for different base pair (bp) lengths and temperatures.

Methodology:

Conventional thermocycler systems cannot directly measure the temperature from the samples themselves, instead it uses several Digital or Analog temperature sensors or thermocouples, commercially available machines work on this principle, so the sample temperature are not accurately measured rather they are estimated only, in this research my propose is to design and simulate a new PCR thermocycler controller based on the thermal properties of the sample mixture, if the thermal properties of the samples are known, by using COMSOL Multiphysics one can simulate a new model that describes exactly how the temperature diffuses in the samples, in fact with proper material definition you can even measure the temperature in any domain, boundary, line, or area in the simulation, and by using this temperature data you can design of a thermal controller that follows the temperature profile of the samples rather than a normal feedback loop and a temperature tuner.

Practical parts followed in the design:

The total design of the project consisted of several steps, including:

- 1- Geometry design using BLENDER.
- 2- Modelling using COMSOL.
- 3- System identification using MATLAB.
- 4- Cosimulation using Livelink for Simulink.

1- Geometry Design:

Using Blender the heating block was made using a predefined geometry for a PCR plate, first ($114.3 \times 76.2 \times 21.6 \text{ mm}$ L/W/H) rectangular prism was formed, then the plate model was imported (96-well-plate manufactured by applied biosystems), and by using a subdivision modifier on subtract mode the geometry was made, the final product is a rectangular-prism shaped heating block with sample holes or openings to hold the PCR tubes, the complete process is shown in the images below.

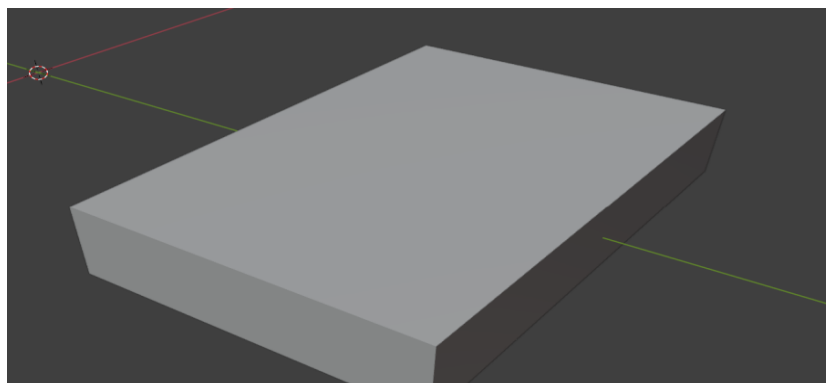


Figure 6 designed PCR block.

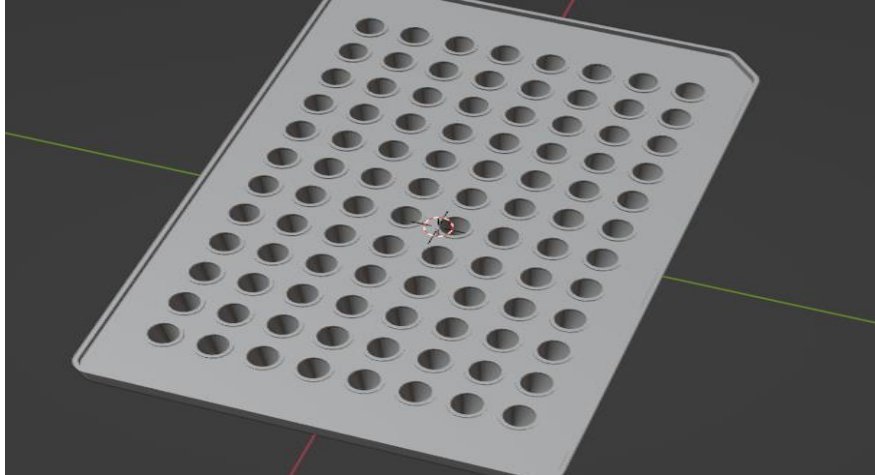


Figure 7 an upper projection of the sample plate.

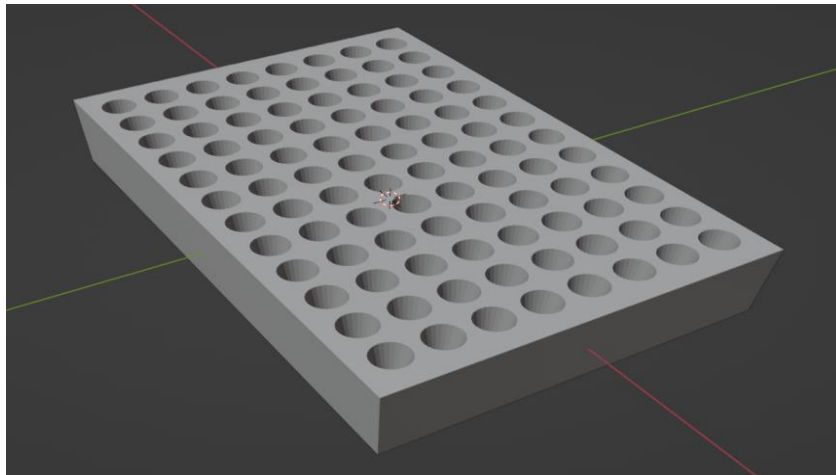


Figure 8 The heat block after applying the subdivision modifier.

2- Modelling using COMSOL Multiphysics:

The next step was to import the models to COMSOL, COMSOL is a simulation environment that is developed for user-friendly modelling, COMSOL is a powerful software for solving ordinary differential equations (ODEs) and partial differential equations (PDEs) for complex physics and engineering geometry, it also allows the study of several different physics phenomena at the same time (hence the name Multiphysics).

Modelling using COMSOL can follow some basic static steps:

a- Geometry Design/Import:

COMSOL has a CAD section that you can use to build the geometry, in this project the geometry was already built using Blender and the file was just imported to COMSOL.

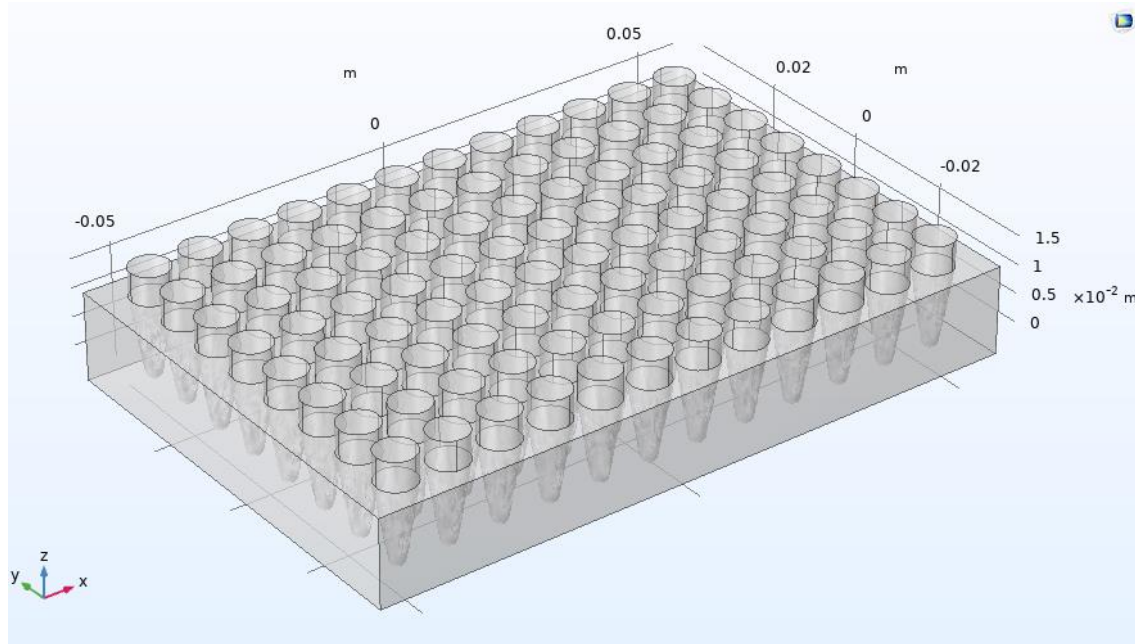


Figure 9 Final designed geometry in COMSOL.

- b- Material definition, in this section you start by adding proper material to each domain or boundary depending on your design, in the PCR case the choice was Aluminum6063-t83 for its good thermal conductivity and cheap price. It was applied to the heating block domain, also a sample material was defined using estimated values of the DNA mixture, air for the vacancy in the tube, and polypropylene for the tube housing.

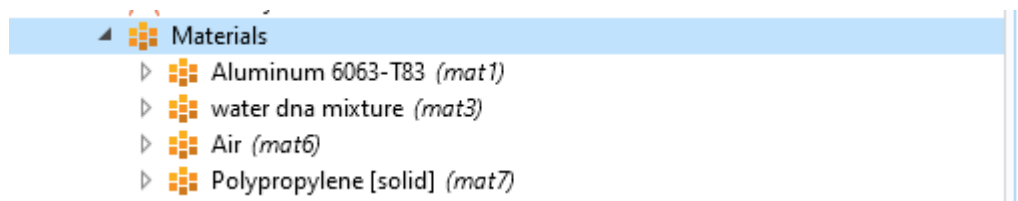


Figure 10 Material definition in COMSOL.

- c- Mesh: a mesh is an active segmentation of the 3D or 2D geometry that you have created, it is crucial for the software, as the solver relies on the segments to achieve an estimation or a solution for the PDE at hand, for complex geometries building a proper mesh can be difficult, and the solver may not reach convergence if the mesh is not correct, in this case the 3D meshes included a combination of free Triangular (which is a boundary mesh) followed by a Tetrahedral mesh (which is a domain mesh) this strategy was implemented for the samples As otherwise the sample overall shape will be heavily misshaped and the results will not be accurate, other parts of the project were only in the form of free Tetrahedral the mesh size was set to extremely coarse to reduce simulation time as much as possible.

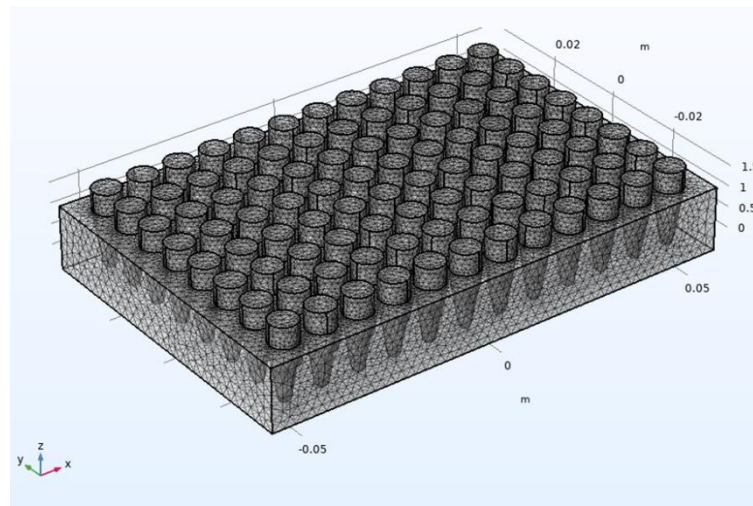


Figure 11 Final Geometry mesh.

- d- Physics definition: in this section proper physics is defined according to your geometry and the phenomena being studied, in this the Heat Transfer in Solids branch was used as it already predefines everything required to solve the PDE of the heat transfer equation [12].

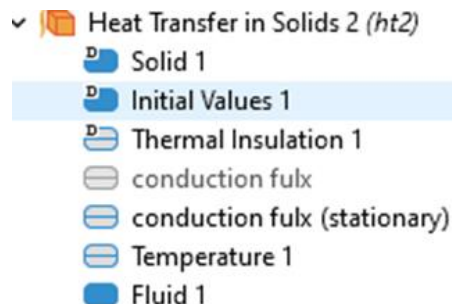


Figure 12 Physics definition in COMSOL.

- e- Running the solver: After the previous steps the simulation is run to get a solution, two solvers are available to choose from, a stationary and a time-dependant solver, depending on your intended study method whether being a stationary or a transient study.
- f- Study types:
There are two different study types used in this study:
 - stationary study: in stationary study, the software outputs the result with disregard of time, in other words, it outputs the result when the system stabilizes, or (time goes to infinity). In the initial model of the project, several stationary studies were used to get a general understanding of how heat distributes in the geometry regarding different boundary conditions and heat flux powers.

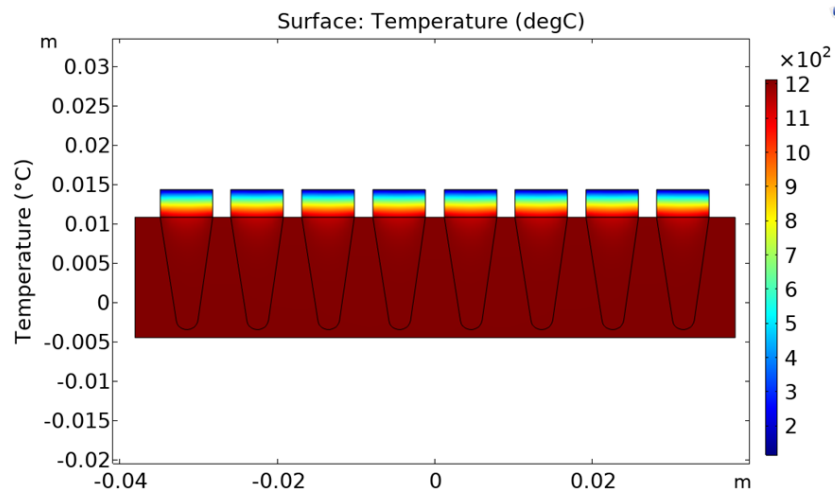


Figure 13 Visual result of the steady state study.

- Time-dependent study: this type of study is used when the user is interested in viewing the simulation over time or at different specified intervals, in this study it is crucial for viewing how heat is distributed over time. To generate a time-dependent solution, a piecewise function was created, the piecewise function used was at a time $t = 0$ would start from 2500 Watts then at $t = 60$ it would become -2500 Watts then at $t = 120$ it would rise back to 2500 Watts.

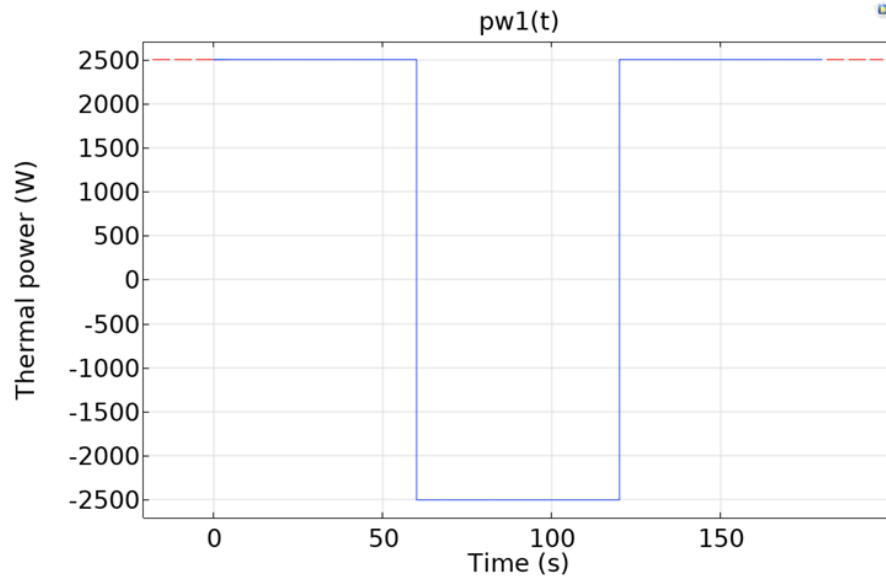


Figure 14 Piecewise function for time dependent study.

By using an integration over a surface (or a domain probe) a preliminary data was acquired for the block and the samples (the blue curve represents the data for the samples, and the green for the block).

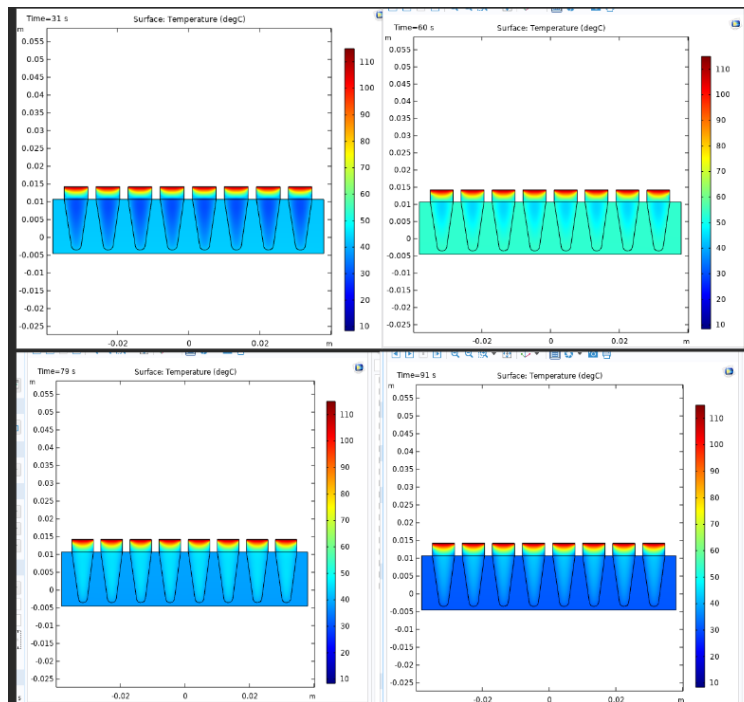


Figure 15 Visual representation on the geometry for the temperature profiles in all domains for time dependent study.

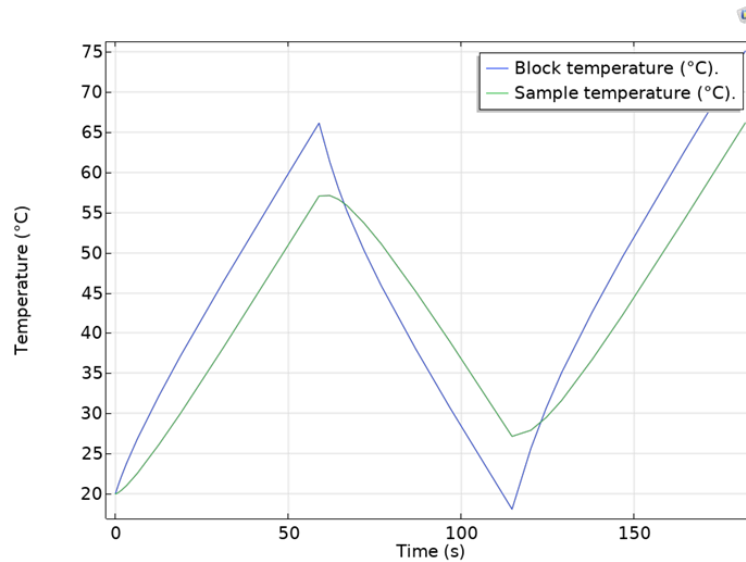


Figure 16 Average temperature profile for block and samples in degrees Celsius.

- 3- Obtaining the transfer functions for samples and block using system identification application:

By importing the generated temperature file to MATLAB and obtaining the individual data matrices a transfer function can be estimated by using the “system identification toolbox”, a transfer function for each domain can be estimated.

```

0.001708 s^3 + 0.0006876 s^2 + 1.352e-05 s + 2.964e-06
-----
s^3 + 0.009884 s^2 + 0.00604 s + 2.707e-06
Name: block_3_3
Continuous-time identified transfer function.

Parameterization:
  Number of poles: 3   Number of zeros: 3
  Number of free coefficients: 7
  Use "tfdata", "getpvec", "getcov" for parameters and their uncertainties.

Status:
Estimated using TFEST on time domain data "block".
Fit to estimation data: 82.43% (stability enforced)
FPE: 6.907, MSE: 6.184

```

Figure 17 the transfer function obtained using the 2D data.

Reference temperature profile:

The next step was to try and find a way to approximate the real input and output for the thermocyclers, after thorough search on the internet on datasets (thermal power as input and temperature as output), a supposed approach to reverse approximate the input power based on the transfer function and the desired output (temperature profile) was made, therefore a parametric MATLAB function was created, the function generates (plots) any required thermal graph with relevance of several considerations.

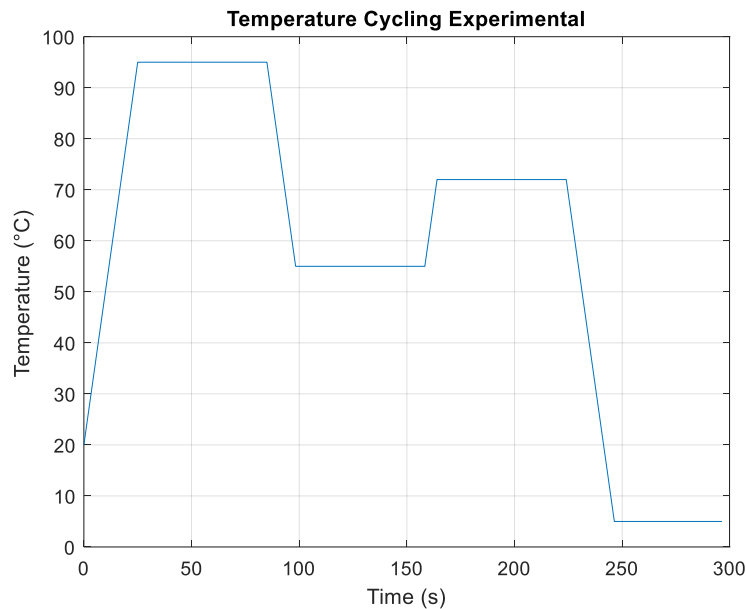


Figure 18 generated reference signal using MATLAB.

Estimating the input using a Cosimulation between COMSOL and Simulink using Livelink:

COMSOL Multiphysics offers a wide range of supplementary applications that can connect the simulation software to other processing software via an internal (Local) server for sending and receiving Data, these applications are indicated by using the word “Livelink,” examples are Livelink for “Simulink, MATLAB, Word, Excell.... etc.”, by using such software extended control capabilities can be utilized, in this project Livelink for Simulink was used to estimate the input signal using a Realtime PID controller, the process starts by creating an interpolation function with an arbitrary value (for example 500 watts) then linking that function to a parameter, after that creating a Cosimulation node and defining the inputs (interpolation function connected to the heat flux) and the outputs (block and samples probes) and creating the accordant FMU file, a new Simulink file is created and the FMU file imported to the COMSOL Cosimulation block.

The Analysis of the Cosimulation Diagram:

The Cosimulation block has one input and two outputs, the input is thermal power in watts, and the outputs are the sample and block average temperatures in degrees kelvin. The PID was the main control Method used in this simulation, it is important to note that two individual feedback methods are used a normal block temperature feedback and sample temperature feedback, hence depending on the type of feedback being used, two individual PID tunings have to be made because of do two different transfer functions defined for the sample and block domains. Also, the automatic PID tuning provided by the tuner app in MATLAB **cannot** be used in this core simulation because the tuner cannot linearize the system as it is not defined in a direct mathematical formula rather it is linked to the COMSOL model, also, To verify the mathematical model of the project a Kelman filter can be applied using a state space representation of the system, the Cosimulation file was further to include the Kalman filter. By using System Identification Toolbox to the output and input data after applying the PID for both domains the sample and block temperature domains, two new transfer functions were generated, also, a first-order state space representation can be obtained for the block temperature to be used in the Kalman filter.

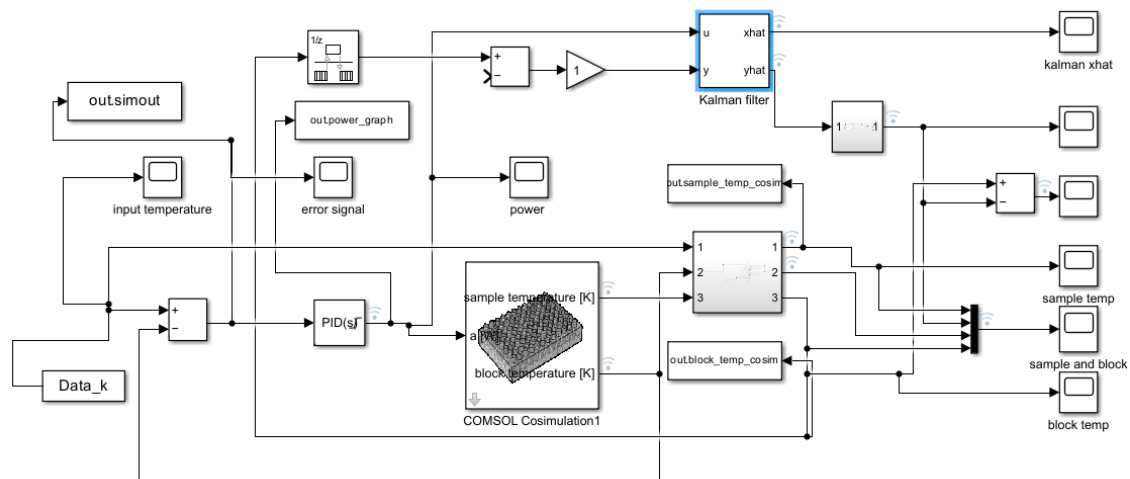


Figure 19 Cosimulation diagram.

Cosimulation Results and Discussion:

The system has several scopes to visually represent everything that is being computed, the scopes from left to right are:

- Input temperature scope (this scope plots the reference input graph).
- Error signal (this scope plots the difference between the input and the output).
- Power signal (this scope plots the power scope in watts that is being delivered to the COMSOL model, this is the most important signal as it describes the required power signal needed to get the desired graph).
- Sample and block scopes (these scopes describe the sample and block temperatures in degrees Celsius as the output of the system, also there is a two-channel scope to visualize both signals simultaneously).
- Kalman filter output.

Generated graphs:

block and sample temperature profiles vs time:

The following graph shows the difference between sample temperature and block Temperature vs time in block temperature feedback mode, it shows how the sample temperature profile is lagging from the block temperature indicating that a better more stable temperature control can be designed.

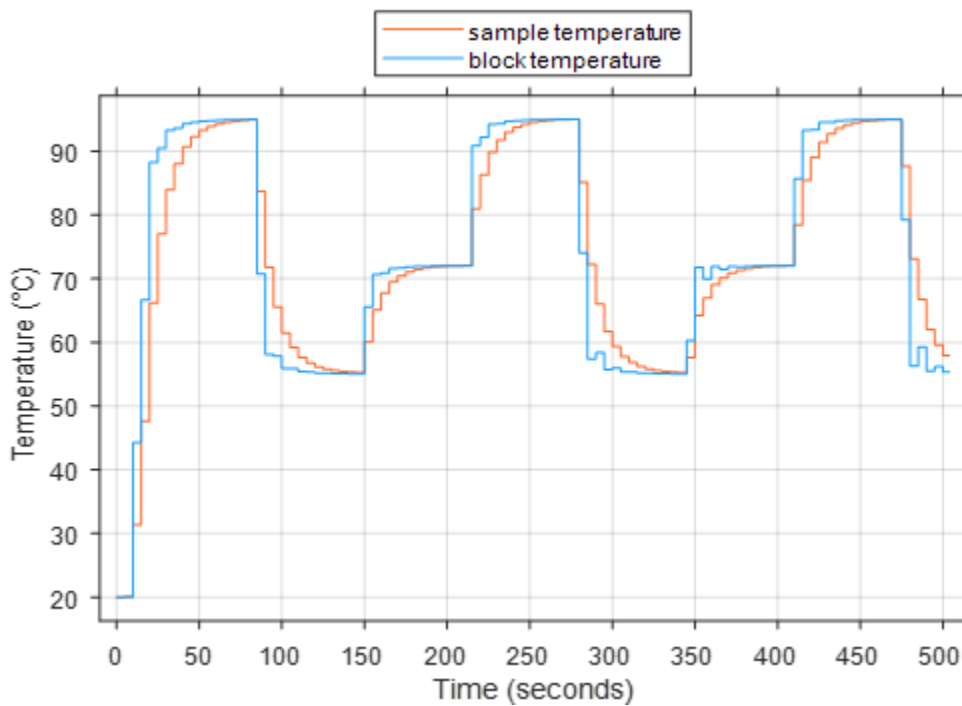


Figure 20 Sample and block average temperatures.

Thermal power vs time:

The thermal power signal is extremely important for this simulation as it represents the input signal to the plant, its value depends on the saturation threshold set in the PID box.

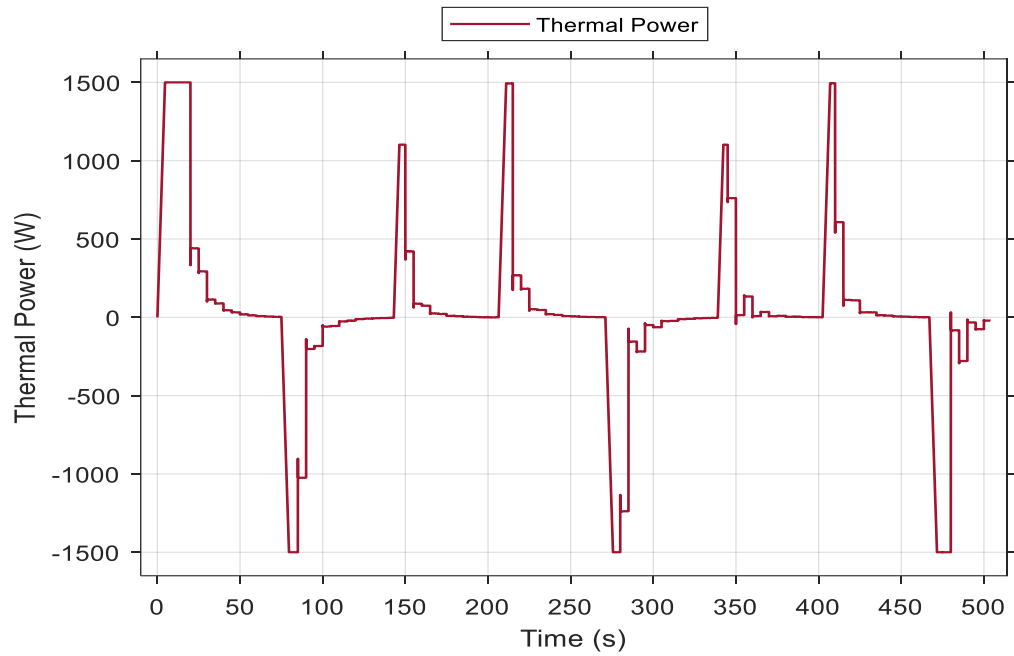


Figure 21 Thermal power (input signal).

Error signal:

The error signal was the input signal for the PID controller.

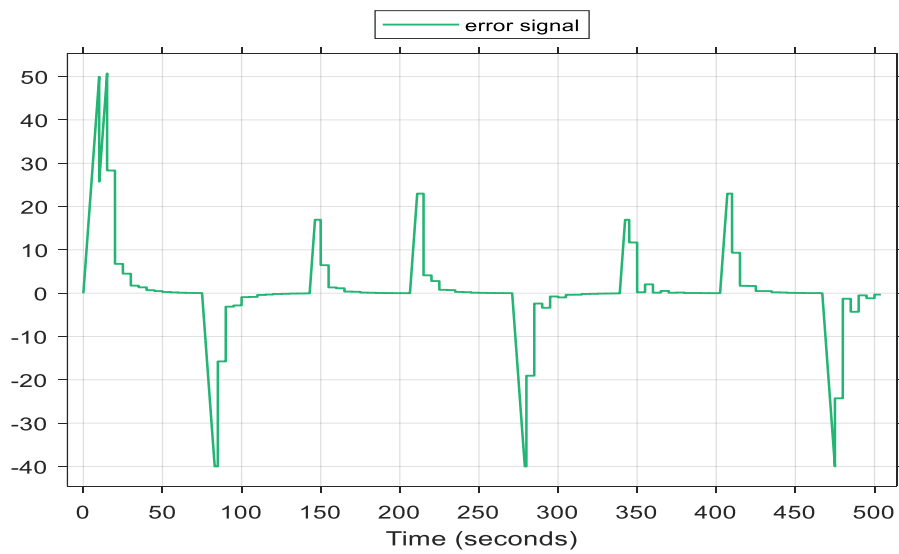


Figure 22 Error signal graph.

Required temperature (Reference graph):

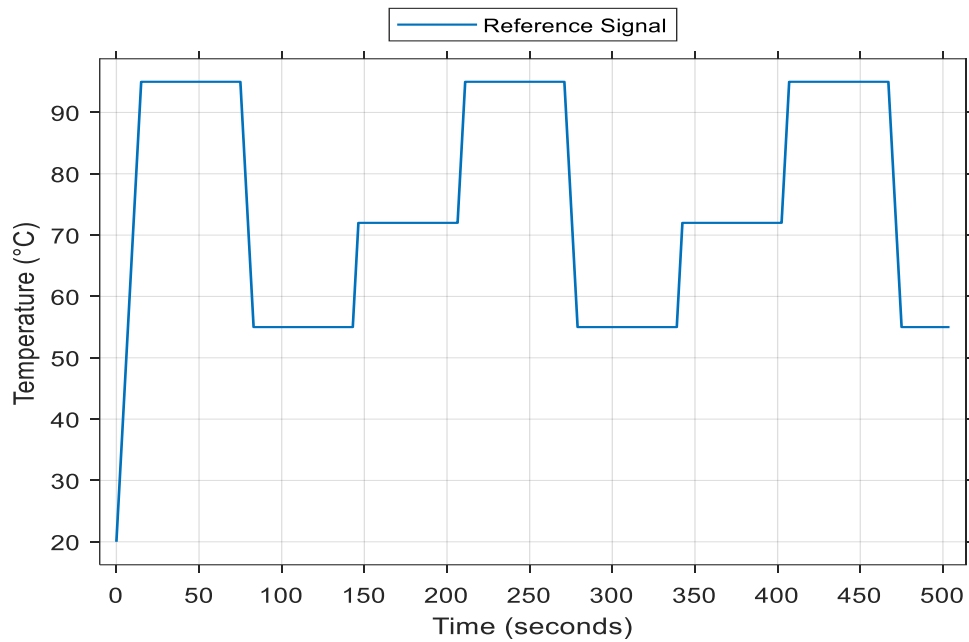


Figure 23 Reference signal graph.

Kalman filter output:

The graph shows the output estimation for the Kalman filter and how it estimates the output based on the state space model and the input and outputs of the system.

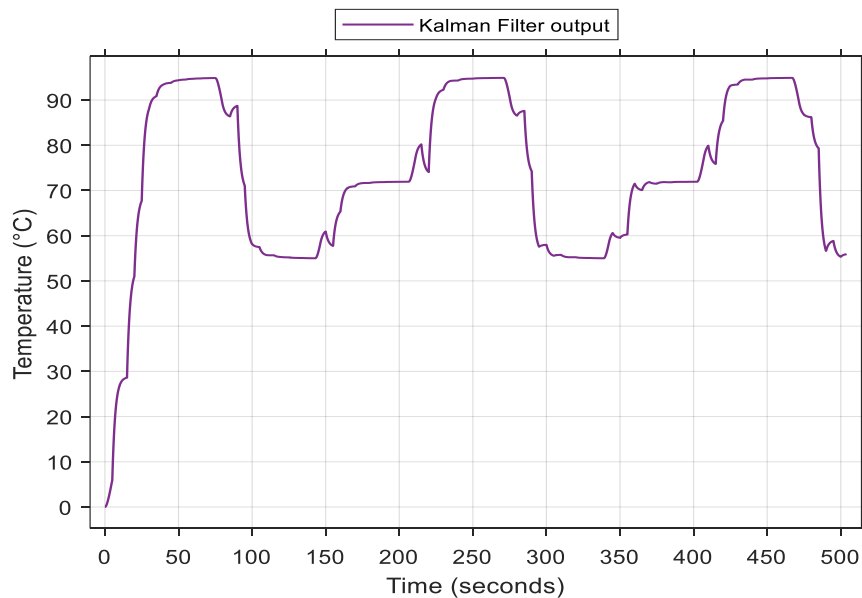


Figure 24 Kalman filter output.

Multiplexed graphs for different feedback modes:

The following results are the final results represented by the simulation; they are meant to visually compare the different outputs for different modes in the simulation, Where:

The yellow graph is the input signal, the purple graph is the block temperature, the blue graph is the sample temperature, and the orange graph is the graph for the Kalman filter.

Final Results for block feedback:

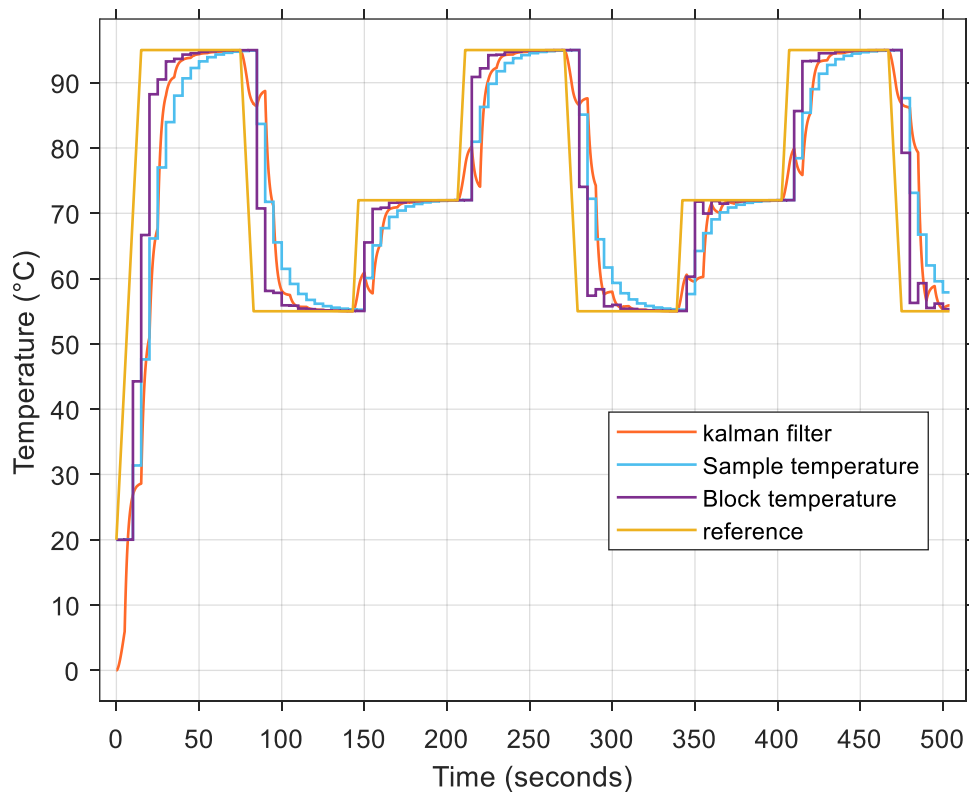


Figure 25 Final multiplexed results for block feedback.

Final Results for sample feedback:

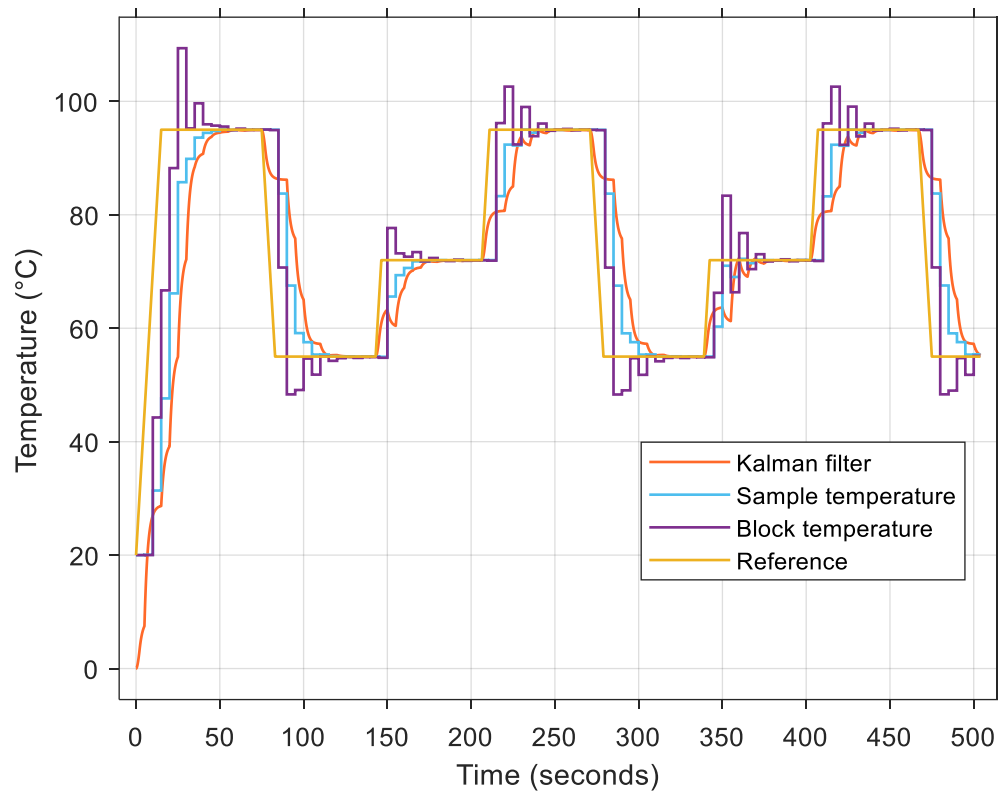


Figure 26 Final multiplexed results for sample feedback.

An important observation is that for sample feedback control mode, the temperature of the block is increasing/decreasing above/below the set point, which can be interpreted since there is a difference between the thermal properties of both domains and the samples are included inside the block's domain, thermal compensation has to be made to increase/decrease the temperature of the samples more quickly, meaning more thermal energy has to be provided/removed from the block to increase/decrease the temperature of the samples more rapidly. It is worth noting that a 5 sample communication step was used to reduce the time of the simulations, also there is a difference between the two feedbacks where the temperature graph of the samples are significantly increased at the sample feedback mode, ensuring more stability and higher accuracy which is the main aim of the project.

Final estimation of the transfer functions of the sample and block domains:

A final estimation of the new transfer functions was made using the system identification app from the data acquired after Cosimulation model:

Block transfer function:

```
From input "ul" to output "yl":  
0.001536 s^3 + 0.003011 s^2 + 0.0001557 s + 2.816e-06  
-----  
s^3 + 0.05589 s^2 + 0.001071 s + 9.587e-08  
Name: tf_3_3_block_cosimulation  
Continuous-time identified transfer function.  
  
Parameterization:  
Number of poles: 3    Number of zeros: 3  
Number of free coefficients: 7
```

Figure 27 final block transfer function estimation.

Sample transfer function:

```
From input "ul" to output "yl":  
-4.063e-05 s^3 + 0.0008289 s^2 + 0.0004603 s + 2.652e-06  
-----  
s^3 + 0.1722 s^2 + 0.0009429 s + 5.066e-08  
Name: tf_3_3_sample_cosimulation  
Continuous-time identified transfer function.  
  
Parameterization:  
Number of poles: 3    Number of zeros: 3  
Number of free coefficients: 7
```

Figure 28 final sample transfer function estimation.

Summary:

The science of designing and simulating a polymerase chain reaction thermocycler is an engaging field, it is a highly interdisciplinary project that requires a deep understanding of several fields of engineering, biology, and biotechnology. Since this innovative technology is always in rapid advancement, this project resembles an embodiment of the capabilities of further development and improvements, proving that increasing the stability and accuracy of the PCR test is possible. By combining the functionality of Blender, COMSOL, MATLAB, and Simulink, developing such a complicated control mechanism is achieved. The results of this project required deep and thorough research of the problem at hand, to create a state-of-the-art innovative technology to reduce sample degradation, and non-specific binding of primers to DNA and RNA, contributing to the development of human knowledge and PCR technology.

References:

- [1] Alberts, Bruce. *Molecular biology of the cell*. Garland science, 2017.
- [2] Erlich, Henry A. *PCR technology*. Vol. 246. New York: Stockton press, 1989.
- [3] Mullis, Kary B. "The Polymerase Chain Reaction." Nobel Lecture, December 8, 1993. NobelPrize.org, Nobel Media AB 2022.
<https://www.nobelprize.org/prizes/chemistry/1993/mullis/lecture/>. //accessed on 15/4/2024.
- [4] "PCR Troubleshooting 101: How to Address Non-Specific Amplification." Genetic Education, 20 May 2020,
<https://geneticeducation.co.in/pcr-troubleshooting-101-how-to-address-non-specific-amplification>.
//accessed on //15/4/2024.
- [5] Garibyan, Lilit, and Nidhi Avashia. "Research techniques made simple: polymerase chain reaction (PCR)." *The Journal of investigative dermatology* 133.3 (2013): e6.
- [6] Brown, Terry A. *Genomes 5*. CRC Press, 2023.
- [7] Hughes, Simon, and Adrian Moody, eds. *PCR: methods express*. Scion Publishing Ltd, 2007.
- [8] Wu, Peng, Shu-ichi Nakano, and Naoki Sugimoto. "Temperature dependence of thermodynamic properties for DNA/DNA and RNA/DNA duplex formation." *European journal of biochemistry* 269.12 (2002): 2821-2830.
- [9] Dragan, Anatoliy, Peter Privalov, and Colyn Crane-Robinson. "Thermodynamics of DNA: heat capacity changes on duplex unfolding." *European Biophysics Journal* 48 (2019): 773-779.
- [10] Spink, Charles H. "Differential scanning calorimetry." *Methods in cell biology* 84 (2008): 115-141.
- [11] Höhne, Günter Walter Heinrich, Wolfgang Hemminger, and Hans-Jürgen Flammersheim. *Differential scanning calorimetry*. Vol. 2. Berlin: Springer, 2003.
- [12] COMOL documentation/user's-guide/the-heat-transfer-features/Domain-features/Solid.
[COMSOL Documentation](#) // accessed on 4/7/2024.