THERMAL LOAD OF THREE-FAN COOLING SYSTEM OF A GRAPHIC PROCESSING UNIT(GPU)

AIZUDDIN AZRI BIN CHE AZIZ

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

THERMAL LOAD OF THREE-FAN COOLING SYSTEM OF A GRAPHIC PROCESSING UNIT(GPU)

AIZUDDIN AZRI BIN CHE AZIZ

This report is submitted
in fulfillment of the requirement for the degree of
Bachelor of Mechanical Engineering (Thermal Fluid)

Faculty of Mechanical Engineering
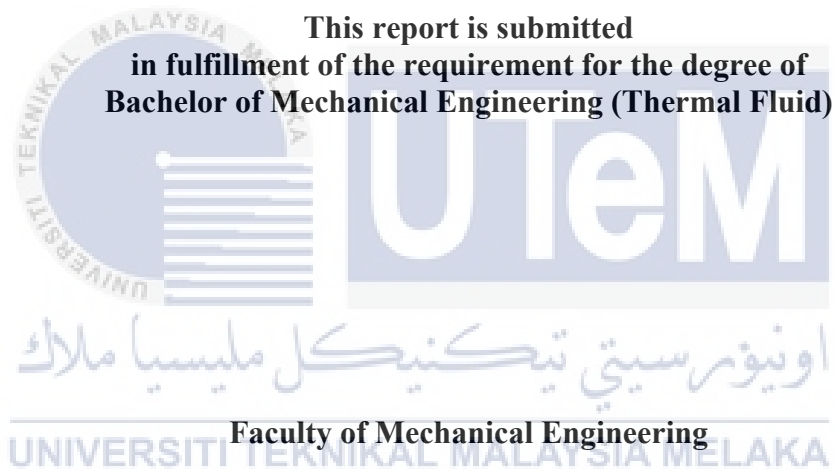
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

JANUARY 2022

DECLARATION

I declare that this project report entitled "Thermal Load of Three-Fan Cooling System of a Graphic Processing Unit" is the result of my own work except as cited in the references

Signature	:	....
Name	:	....
Date	:	....

APPROVAL

I hereby declare that I have read this project report and in my opinion this report is sufficient in terms of scope and quality for the award of the degree of Bachelor of Mechanical Engineering (Thermal Fluid).

Signature : ....

Name of Supervisor : ....

Date : ....

DEDICATION

To my dear beloved father, mother, friends and my supportive supervisor that

endlessly encouraging me to keep going forward.

ABSTRACT

Graphics Processing Unit (GPU) has gone through many revolutionary changes throughout the decade. GPU's processing power can challenge the existing Central Processing Unit (CPU) in the task of running high-profile software. However, any electronics that flow current into it will generate heat. The thermal problem is the major problem for GPUs. The optimal operating GPU requires minimum fan speed and core temperature with the highest efficiency (hashing power). This research focused on obtaining the optimal response at a certain core clock and memory clock via the optimization tool of Design-Expert software. The response was recorded while applying constant amount of power supplied to the GPU. The GPU used is in the category of three-fan which is ASUS ROG Strix GTX1070. The relationship between the clocking (core and memory) and GPU responses (fan speed, core temperature, hash rate) were observed. By that, Central Composite Design (CCD) generated one equation for each fan speed, core temperature, and hash rate. Next, the optimization process suggests several new clocks settings that give the best performance than the current setting. For the validation and confirmation process, the best one core and memory clock was selected. The results from the validation process prove that the predicted response was precise with less than 2% deviation.

ABSTRAK

Unit Pemprosesan Grafik (GPU) telah melalui banyak perubahan yang besar sepanjang dekad. Kuasa pemprosesan GPU mampu mencabar Unit Pemprosesan Pusat (CPU) yang sedia ada dalam tugas menjalankan perisian berprofil tinggi. Walau bagaimanapun, apabila komponen elektronik mengalirkan arus elektrik ke dalamnya ia akan menghasilkan haba. Masalah haba adalah masalah utama untuk GPU. GPU yang beroperasi secara optimum memerlukan kelajuan kipas dan suhu teras yang minimum dengan kecekapan tertinggi (kuasa hashing). Penyelidikan ini akan memberi tumpuan untuk mendapatkan tindak balas optimum pada jam teras dan jam memori yang tertentu menggunakan alat pengoptimuman perisian Design-Expert. Tindak balas GPU telah direkodkan sepanjang kuasa yang malar dibekalkan kepada GPU. GPU yang digunakan dalam penyelidikan ini adalah GPU tiga kipas yang bernama ASUS ROG Strix GTX1070. Perkaitan antara jam (teras dan memori) dan tindak balas GPU (kelajuan kipas, suhu teras, kadar hash) direkodkan. Hasil daripada ini, Reka Bentuk Komposit Pusat (CCD) telah menjana satu persamaan untuk setiap kelajuan kipas, suhu teras dan kadar hash. Seterusnya, proses pengoptimuman telah mencadangkan beberapa tetapan jam baharu yang memberikan prestasi yang lebih baik daripada tetapan semasa. Bagi proses pengesahan pula, satu tetapan jam (teras dan memori) terbaik baharu telah dipilih. Hasil proses validasi membuktikan bahawa respons yang diramalkan adalah tepat dengan sisihan kurang daripada 2%.

ACKNOWLEDGEMENT

First and foremost, I'd like to express my gratitude and praise to Allah the Almighty, my Creator and Sustainer, for the opportunity to finish this project despite of all the hardship and circumstances throughout the journey. Additionally, I'd like to express my gratitude to Universiti Teknikal Malaysia Melaka (UTeM) for providing the best supervisor for me. My dear supervisor, Dr Muhammad Zulfattah bin Zakaria has given me continuous support, encouragement, counsel and inspiration that has been a great help for me. His unwavering patience in guiding will always be cherished through the eternity.

Additionally, sincere thanks to my friends, Mr Hazim and Mr Sufi that always helps and cooperate with me during discussion and studies. They have provided a huge amount of helps since the starting of this project. Lastly, to my father, Mr. Che Aziz and my mother, Mrs. Faridah for their prayers, love and support that are priceless and heavily appreciated. They have been accompanying me through the ups and downs of this semester.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF ABBREVIATION

GPU	Graphic Processing Unit
CPU	Central Processing Unit
IBM	International Business Machines
HPC	High Performance Computing
R&D	Research & Development
RSM	Response Surface Methodology
VRAM	Video Random Access Memory
VRM	Voltage Regulator Module
RAM	Random Access Memory
CCD	Central Composite Design
BBD	Box-Behnken Design
ANOVA	Analysis of Varience
DOE	Design of Experiment
OC	Overclocking
C.V	Coefficient of Variation
LOF	Lack of Fit
PC	Personal Computer

LIST OF SYMBOLS

W = Watts

Q_{max} = Maximum heat transfer capacity

Q_{cond} = Rate of heat conduction

k = Thermal conductivity

A = Cross sectional area

$T_1 - T_2$ = Temperature difference

Δx = Thickness

Q_{conv} = Rate of heat convection

h = Convection heat transfer coefficient

A_s = Surface area where heat convection takes place

T_s = Surface temperature

T_∞ = Temperature of the fluid sufficiently far from the surface

Q_{rad} = Rate of heat radiation

ε = Emissivity

σ = Stefan-Boltzmann constant

T_s = Surface temperature

T_{surr} = Absolute temperature of surrounding

CFM = Cubic feet per minute

rpm = revolutions per minute

ºC = Degree Celcius

MH/s = Mega hash per second

CHAPTER 1

INTRODUCTION

1.1 Background

Graphic Processing Unit (GPU) was originally more like a video display card that was made by International Business Machines (IBM) in 1981. Monochrome Display Adapter (MDA) was originally in single monochrome to allow high-resolution text and symbol display at 80 x 25 characters that was useful for drawing forms. Years later, IBM came out with first graphic card with full-color display (PC, 2019). After that, the technology developed and evolved into having 8nm chip and having memory of 24GB that able to support hardware-raytracing, variable-rate shading and more. Mainly GPU is used to improve the performance of 3D rendering. Time after time, affected by modern technology, it able to produce stunning visual effects and natural scenes with advanced lighting and shadowing methods. Developers has also started to greatly increase the capability of GPU to boost its performance so that it able to do deep learning, high performance computing (HPC) and more in a shorter period of time (Intel, 2020).

Graphic cards are at the mercy of names like Nvidia and AMD, which exert great control over the market. EVGA, ASUS, and MSI are a couple of the many third-party graphic card manufacturers that purchase their chips from these two companies (Crider, 2018). Following that, these third-party companies will come out with their

own package with hardware such as cooler, lighting, plus additional video ports or maybe a bit more attractive lighting option. At this particular point is where single, dual even three-fan GPU were invented. In the absence of anything that could be done to help reduce the heating issues, these various GPU producers realized the value of R&D in figuring out the best method to deal with it.

GPUs significantly outperformed CPUs in terms of speed, making them a valuable medium for computing (Breitbart, 1999). GPUs consist of chips and RAM (random access memory) which chips are used to process the data received from the input port and RAM are used to store the data of the image processed (S. Gillis, 2020). Now with the latest GPU that can generate almost 1695MHz at boost clock, it must be a challenging task to keep maintaining the GPU's temperature at its best to keep on having the best performance.

1.2 Problem Statement

Electronic devices such as central processing unit (CPU) or graphic processing unit (GPU) have a high heat generating capacity which equivalent to nuclear reactors, which makes it difficult to control their heat output on an individual component basis (Hirasawa et al.,2005). When the GPU becomes too hot, its performance degrades rapidly. However, most GPUs are equipped with thermal insulation which will lessen the amount of heat generated to maintain full efficiency of it. Normally, there are a few factors contributing to the overheating issue such as heatsink mount, fan speed or maybe the fans are clogged with dust. Moreover, it is a difficult task to get rid of the

heat and also to monitor the temperature around the chip since GPU is close-packed (Feng & Li, 2013).

It is obvious that thermal management is the main issue if it is about GPU. However, with the nonstop research that has been done to resolve the issue, little by little discoveries of ways to maintain the GPU at its best temperature have been figured out. There is a strong relationship between the clocking (core and memory) to the GPU response. Hence, there is a need to examine how clocking will affect the GPU responses such as fan speed, core temperature and hashing power.

1.3 Objectives

The core objectives of the project are as follow:

1. To analyze the relationship between clocking (core and memory) and GPU responses (fan speed, core temperature, power consumption, and hash rate).

2. To locate the optimal fan speed, core temperature and the highest possible efficiency (hashing power) at certain core clock and memory clock via optimization tool.

1.4 Scope of Project

The scope of the project is:

1. Collecting literature review for GPU cooling system.

2. Get the Three-Fan Graphic Processing Unit GTX 1070 response (fan speed, core temperature and hash rate) at certain core and memory clock using Afterburner and Phoenixminer software.

3. Finding solutions using numerical optimization using Design-Expert software

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

This chapter will explain in depth of how the GPU really works. It will briefly explain which components that contribute to the heat released and which components that helps to counter the problem of heat throttling that affecting the performance of the GPU. Also included in this section is the fundamental knowledge and equations related to this project. In this chapter, detailed information to be applied onto this project will be provided.

Figure 2.1: AORUS GeForce RTX™ 3090 XTREME 24G

4

2.2 Component Contributing to Heat

Pushing better performance meaning higher heat will be dissipated from the GPU itself. Then when its temperature rises to a certain degree, the GPU's performance will be dropped to prevent the GPU from overheating and damaging the GPU permanently. The normal allowable temperature of a normal operating processor is below 70°C but the increment of 2°C will make it 10% less reliable than its normal working condition (Choi et al., 2012). Moreover, temperature is another critical consideration because high processor temperatures can damage performance, shorten the life of the processor and lead to reliability issues (Kang et al., 2011).

The main contributor of heat on a GPU is the processor itself. To differentiate the processor from the GPU, it usually located in the center of the GPU having a square shape. All processors are Integrated Circuits (IC) which means that it will generate heat as soon as it is switched on. There are two company in control of GPU market, Nvidia and AMD with their GPU architecture of Pascal and Polaris respectively (Verma, 2017). The processor can be considered as the brain of the GPU that will handle all the calculation, simple arithmetic, mathematical, input/output and control operations based on the instruction given to it. As electricity is running into a GPU processor, it can get hot in the same way as a CPU does. The addition of Watts would raise the temperature much further, in addition to the potential to overclock the GPU.

Second component that contribute to the heat is the Video Random Access Memory (VRAM) of the GPU. This component can be considered as the second most important component on a GPU. In this component is the location where both graphics data and game textures are placed for GPU processing. RAM is made up out of

hundreds of billions of transistors that are packed onto the chipset to transform with an order of command. These transistors act like a switch that turns on and off according to commands given to them. As we know these devices have certain current resistance that will heats up, so the amount of heat produced is determined by the command it is executing such as running graphics-intensive software with additional background software running at the same time.

The third one would be Voltage Regulator Module (VRM). Living up to its name, it is used to transform higher voltage from the power supply into lower voltage ranges for GPU usage. It usually transforms 12V to about 1V to 1.5V approximately, which is the voltage range at which GPUs typically run. Each graphics card has different numbers of VRM (Evanson, 2020). GPU with higher power will requires more VRM to control the voltage input so it able to turn down the voltage when idling to keep the heat down and to save power (Verma, 2017). VRMs may get extremely hot, much hotter than the GPU, so they need adequate cooling to prevent the graphics card from shutting down.
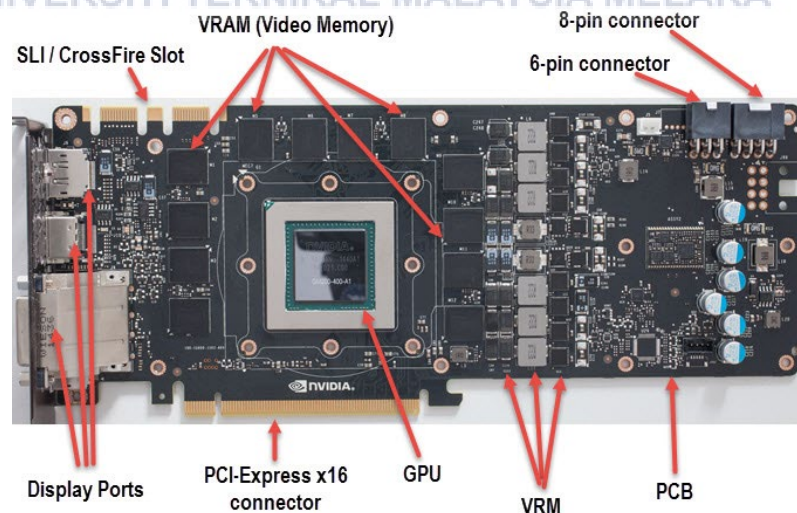
Figure 2.2: Components on a GPU

2.3 Component to Counter Heat Problem

When power is delivered to any semiconductor, heat will be generated through the process. Since the 90s, cooling systems for graphic card have gone through evolutions because in the early days, GPU needs no cooling system since the chips were bare. Then in 2001, heatsink was introduced to GPUs attached by using thermal paste. Graphics cards have become more power hungry as their computational power has increased. Heatsinks and the fans that cooled them had grown in size since then, but they were still small and light. Larger heatsinks and fans that also cool RAM were introduced in 2003. In the same year, Nvidia released the FX 5800 Ultra, a GPU with a blow-out cooling system that turned out to be the loudest in history. Then, there was also new model introduced with heat pipes later that year. Hence, there are main components which are heatsink, fans, thermal paste and heat pipes.
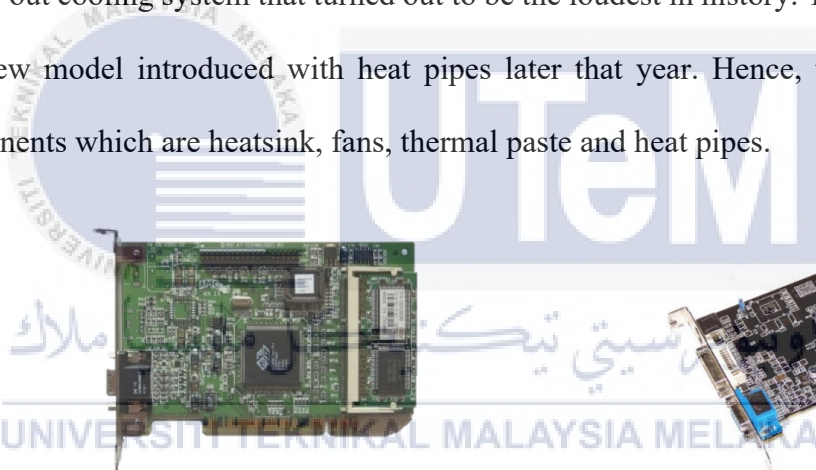
Figure 2.3: Rage Pro with no cooling, 1997   Figure 2.4: Radeon 7000 with heatsink, 2001

Figure 2.5: Radeon 9800 XT with larger heatsink plus fans, 2003   Figure 2.6: FX 5800 Ultra with blow-out style fan, 2003

2.3.1 Heatsink

Heatsink can be considered as the most important component for cooling the GPU processor. It usually located on top of any component that generate heat to reject most of heat generated to maintain the performance of GPU. The performance of active GPU processor cooling heatsinks primarily depends on the forced air convection created by computer fans. These components are used in electronics as well as high-power electrical components and regarded to be the best and most cost-effective cooling option. It comes in a variety of forms and base materials, each having its own flowing flow (Khattak & Ali, 2019). In most cases, heat created by processors is transmitted to a heat sink by conduction and subsequently to the environment by natural, mixed, or forced convection. Moreover, heatsink plays an important role when it comes to overclocking the GPU since heatsink is the closest component to the GPU processor. Overclocking is a process where the input Watts is increased to get more performance from the GPU. As the temperature increases, the heat sink's low heat removal efficiency may cause damage to the electronic component. As a result, if the excessive heat created by the electronic element is not evacuated in a timely manner, chip reliability and service life are jeopardized. (Hamdi et al., 2018). Best way to enhance the heat dissipation is by adding heat pipes or by increasing the fin area of the heatsink. However, both options would result in a directly increase the cost and system size (Choi et al., 2012).

2.3.2 Heat Pipes

Heat pipes are metal pipes which filled with a liquid such as ammonia, acetone, or water. It began to be used on a larger scale around year of 2003. The liquid evaporates as it heats up and travels through the pipe. As it travels, it loses heat and cools down, returning to a liquid state. The heat pipe is an excellent way to transfer heat from one point to another because this loop is infinite. The pipes are usually connected to a cooling plate above the GPU and then transfer the heat to a heatsink that is located away from the GPU. A heat pipe can be thought of as a passive heat pump that removes heat due to physical laws. (Badalan & Svasta, 2017). A heat pipe is a device with excellent thermal conductivity that allows heat to be transported while keeping a nearly constant temperature across its heated and cooled parts (Jouhara & Hussam, 2018). The cooling system's performance can be improved by adding heat pipes (Choi et al., 2012).

According to Jouhara & Hussam (2018), a heat pipe is a passive thermal transfer device that uses phase-change processes and vapour diffusion to transmit huge quantities of heat over relatively long distances with no moving elements. Heat pipes are made out of an evacuated tube that is partly filled with a working fluid that can transform in both liquid and vapour phases. The evaporator is on the left side of the heat pipe, while the condenser is on the right. When the fluid enters the evaporator with high temperature, it will evaporate and turn into vapour that can flow to the condenser with high velocity. The vapour condenses and releases its latent heat as soon as it reaches the condenser part. After that, it flows back to the evaporator either by

gravity or via a capillary wicking mechanism. All heat pipe uses the same operating principle, only materials and geometry changes with each heat pipes.

2.3.3 Cooling Fan

The term "air cooling" is really common to cooling fan system. It is a basic, standard cooling technique, through which heat generated by electrical components is transmitted to a fan that blows or inhales the heat (Nugroho et al.,2019). Heat pipes dissipate heat better when it is combined with a heat sink with fan making them share the cooling function together (Hirasawa et al., 2005). Fan can be used either to release the heat generated from the heatsink or absorbing the cooler air outside GPU to be distributed inside the GPU. Fans for GPU cooling system varies in size, shape or the spinning direction. Each of them has different functionality that contributes to dissipating the heat generated from the GPU. According to Choi et al. (2012), active air cooling heatsinks have encountered some development challenges, including manufacturing costs and sizing the system to fit within the confines of a desktop PC chassis. Nowadays, high-end GPU will require cooling that can remove at least 2-3 times the amount of heat generated since the over-clocking trend started for GPU. However, these cooling fans has one major problem that been bothering any of the GPU users. That particular problem is with the noise created when the fans are turned on. It became a major problem that the high-end user demanding the fans to be produced to a lesser noise generation (Choi et al.,2012).

2.4 Design of Fan

2.4.1 Angle of Blade

The angle between the chord of the blade and the plane of rotation, also known as the angular position, is measured at a particular point along the blade's length. It is a good option to use the blade angle to change the fan's angle of attack. Whether the fan is operating at a constant speed or at variable speeds, adjusting the fan blade angle is needed to provide the most effective angle of attack. According to Huang & Gau (2012), a revised optimum fan blade may improve fan airflow, thus optimizing the axial-flow fan's performance. The setting angle of the blade, the blade root chord and the blade end chord have all been redesigned to provide an optimum fan blade. During the experimental setup, the number of blades, fan speed, and fan diameter are all maintained constant. In the Figure 2.7, the system variables for the original fan blade and the redesign fan blade are displayed.

	Original fan blade	Optimal fan blade	Designed fan blade #1	Designed fan blade #2	
NACA airfoil (mpta)	4609	5608	4506	4609	
Blade root chord (Lr, mm)	18.0	19.38	17.59	18.17	
Blade end chord (Le, mm)	33.0	37.86	30.64	34.18	
Setting angle (θ,°)	47	59	33.18	52.22	
Number of blades	7	7	7	7	
Fan speed (ω, rpm)	2500	2500	2500	2500	
Hub diameter (D_{Hub}, mm)	40.0	40.0	40.0	40.0	
Fan diameter (D, mm)	110.0	110.0	110.0	110.0	
Hub height (H_u+H_l=H, mm)	15.0	15.0	15.0	15.0	
Fan gap (mm)	2.0	2.0	2.0	2.0	
Air volume flow rate (Q, CFM)	78.19	88.84	63.0	83.0	

Figure 2.7: The system variables for various fan blades and fabricated original and optimal fan

11

From Figure 2.7 that the blade setting angle has an effect on the air volume and flow rate. The air volume flow rate was greatest at 88.84 CFM when the setting angle was set at 59 degrees, whereas the lower the setting angle, the lower the air volume flow rate, which was about 63 CFM. From the research of Pirunkaset et al., (2008), blade angles have an impact on a tiny cooling tower's performance. The fan blade angles in the cooling tower of 5 tonnes of refrigeration were set at 59°, 67°, 75°, and 83°, respectively, in the study. According to the graph Figure 2.8, a blade angle of 59° produces the most CFM, followed by 67°, 75°, and 83°. The absence of performance curves at different blade angles may be attributed to manufacturers not properly adjusting the fan blades after installation. As a result, determining the optimal blade angle is critical since it impacts the cooling system's performance and efficiency.
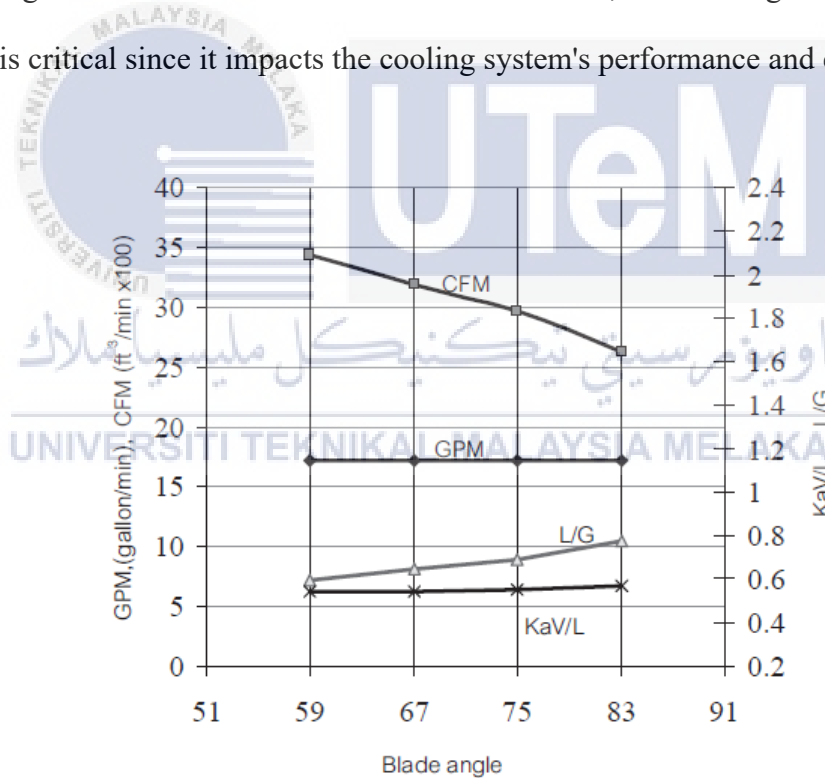
Figure 2.8: The effect of blade angle on CFM

2.4.2 Design of Blade

An axial fan is a kind of fan that draws in fresh air by spinning an axial airflow around an axis. This fan has a circular flow entrance and exit. The flow-producing function of the fan may be assisted by generating a pressure differential and, as a consequence, a force that propels air through the fan. Optimizing the chord length and installation angle of the blade along the blade height, according to Wu & Huang (2019), may help to improve the lift-drag ratio. Using orthogonal optimization, three design alternatives (straight blades, C-type blades, and forward swept blades) were investigated in this study as shown in Figure 2.9.

Straight blade fan   Fan with a 16° forward-sweep blade   C-type blade fan

Figure 2.9: Design of blade fan

From Figure 2.10, a quantitative analysis of the three fan designs revealed that the forward-swept blade fan was the most efficient overall, while the C-type blade fan was just marginally less efficient. The straight blade fan, on the other hand, was the least efficient overall. Forward-swept blade and C-type blade aerodynamics are better than straight blade aerodynamics while operating under non-design conditions. The primary benefits of this model are the forward swept blade and the C-type, which is low noise.

13

	Straight blade	Forward-sweep blade	C-type blade	Units
Speed	985	985	985	[r min^{-1}]
Volume flow rate	50.9628	50.6329	51.1943	[m^3 s^{-1}]
Pressure rise	1496.05	1520.47	1405.6	[Pa]
Total pressure efficiency	93.5621	94.8116	94.0066	[%]

Figure 2.10: Blade fans simulation results
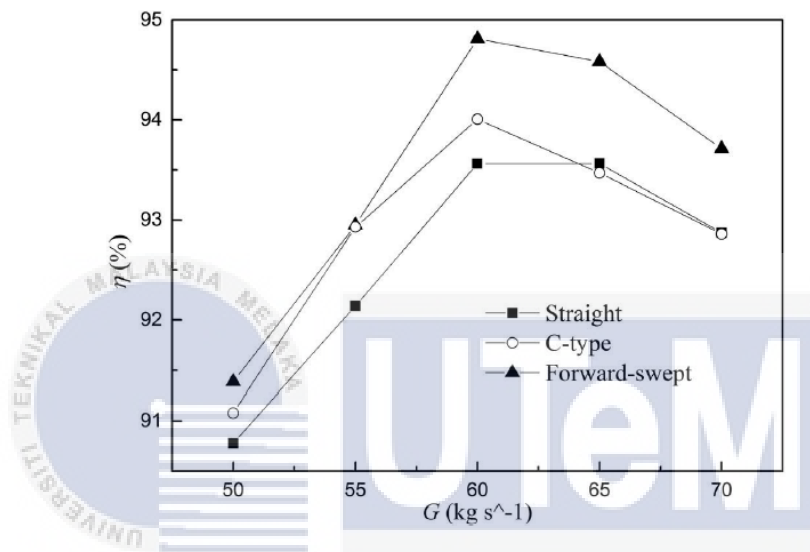
Figure 2.11: Mass flow rates using three kinds of fan total pressure efficiency curve.

2.4.3 Numbers of Blade

For the fan, the outside part of the blade performs the most work. As a result, the number of blades is crucial. According to Zhang and Jin (2011), blade number, hub ratio, blade angle, and blade number all have an impact on the aerodynamic and noise performance of small axial fans.

14
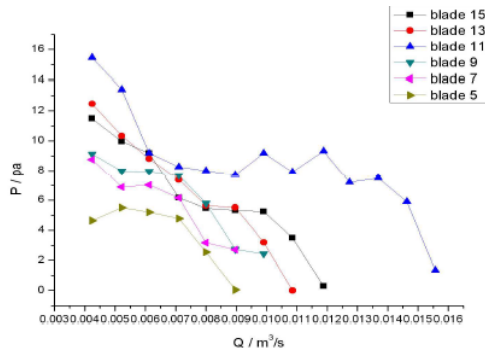
Figure 2.12: Flow rate-static pressure curve
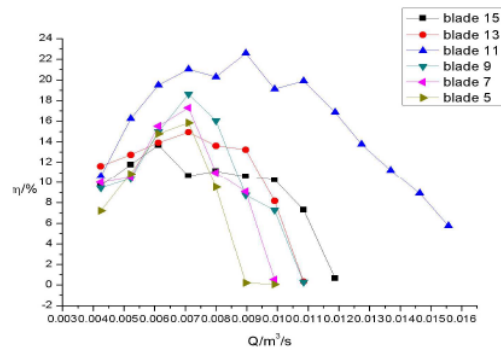
Figure 2.13: Flow rate-efficiency curve

Based on the Figure 2.14, the total pressure and efficiency of a fan system increase as the number of blades increases. When the number of blades is 11, and the flow range is high, the total pressure and efficiency are at their peak. Because the static pressure gradient is the smallest, the flow loss is at its lowest point, and the efficiency is at its highest.

(a) blade number=5 (b) blade number=7 (c) blade number=9

(d) blade number=11 (e) blade number=13 (f) blade number=15
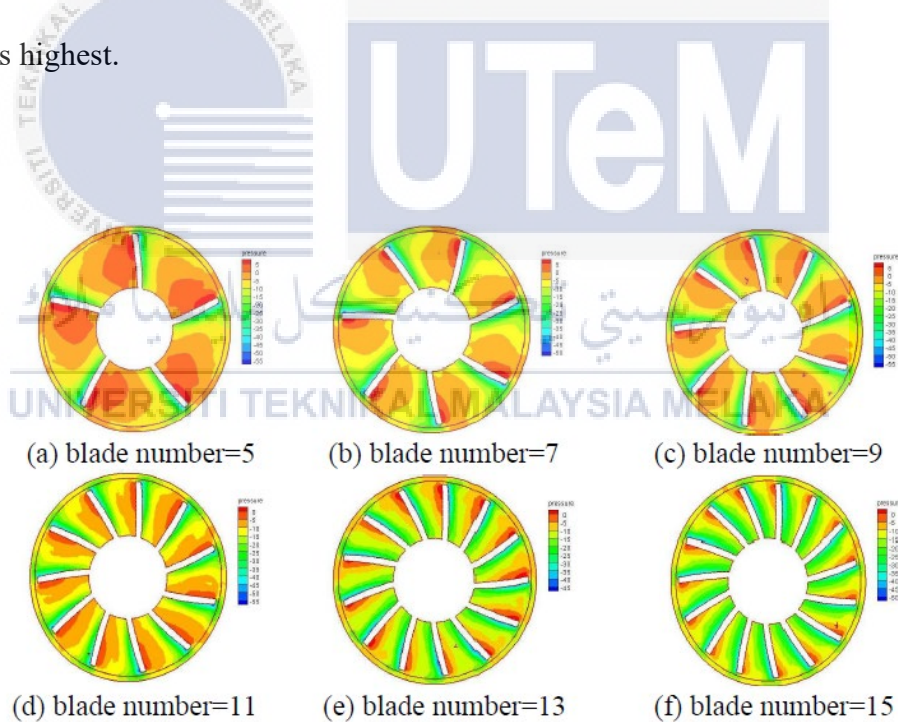
Figure 2.14: The static pressure field distribution

In general, the number of blades increases overall pressure and efficiency. When there are 11 blades, overall pressure and efficiency are at their peak as shown in

the Figure 2.14. Because of the static pressure gradient is at its greatest at the lowest flow point, flow loss is minimal and efficiency is at its highest.
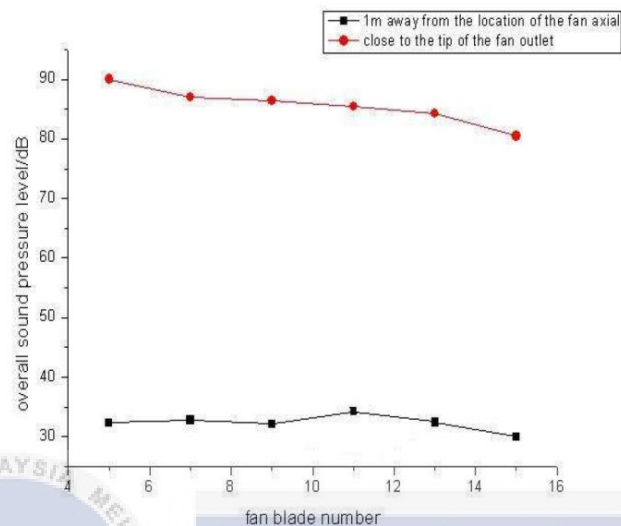
Figure 2.15: Fan blade number against overall sound pressure level

In a fan constant flow field, eddy current is one of the key factor that generates broadband noise, and the effective eddy currents are mostly located in the tip clearance, which has a direct influence on the strength of broadband noise. As the number of blades increases, the sound pressure level of aerodynamic noise at the tip of the fan blades decreases. Figure 2.15 proven that the sound pressure level rises again at 1 metre away from the impeller centre point. This may be due to static electricity produced by the fan blade clearance.

According to Rajabi et al., (2017), the maximum flow rate rises accordingly to the number of blades rises. Although certain blades have been changed, the rotational speed has remained constant at 2900 revolutions per minute. In this case, the research divides fans into three types: four, five, and six blades.

Figure 2.16: Changes in number of blades

The characteristic curves in the Figure 2.17 indicate that when the flow rate is raised from 200 CFM to 400 CFM, the pressure will rise. The static pressure is significantly decreased when the flow rate is raised from 400 CFM to 800 CFM. When six bladed propellers are used, the maximum pressure at 400 CFM is increased by 32%. The surface of the blades becomes bigger as the number of blades increases, increasing the amount of energy transmitted when they in contact with fluid. Outflow pressure rises as a consequence of this. When there are high flow rates, the most efficient separations are used, and increasing the number of blades has little effect on the output pressure. The static pressure generated by the different number of blades is almost equal to 42 Pa at the maximum flow rate of 800 CFM, as seen in the graph in Figure 2.17. Increasing the number of blades has no effect on the flow rate range but it will increase the inlet flow pressure while increasing the outlet static pressure.
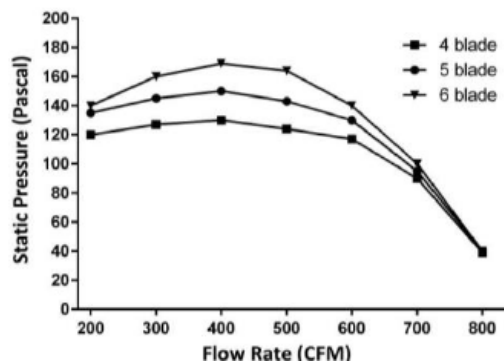
Figure 2.17: Characteristics curves of the fans with different number of blades

2.5 Heat Transfer

Heat transfer is a process where the thermal energy is move from matter to another. The three mechanisms of heat transfer are conduction, convection, and radiation. Conduction is occurs between particles of two material that are in contact with each other that result to energy transfer of more energetic particles to the less energetic particles (Mjallal et al., 2018). Heat transfer of conduction for GPU usually happened in between the heat sink by the GPU's chipset. Equation involved for conduction is:

$$Q = kA\frac{T_1 - T_2}{\Delta x} \tag{2.1}$$

Where,

Q_{cond} = Rate of heat conduction, W

k = Thermal conductivity, $\frac{W}{mK}$ / $\frac{W}{m°C}$

A = Cross sectional area, m^2

$T_1 - T_2$ = Temperature difference, K

Δx = Thickness, m

Convection happened when there is temperature difference between surface of a solid and a fluid in motion. The molecular vibration that occurs within the solid and the large-scale fluid particle movement of energy are two elements that cause convection to occur due to the energy flow. The equation for heat convection is:

$$Q = hA_s(T_s - T_\infty) \tag{2.2}$$

Where.

   Q_{conv} = Rate of heat convection, W

   h = Convection heat transfer coefficient, $\frac{W}{m^2} \cdot °C$

   A_s = Surface area where heat convection takes place, m^2

   T_s = Surface temperature, $°C$

   T_∞ = Fluid temperature, $°C$

Conduction and convection both need a medium to transport heat unlike thermal radiation. Radiation is a process where the source of thermal energy is transported by electromagnetic waves (Mjallal et al., 2018). Equation involved for radiation is:

$$Q = \varepsilon \sigma AS(T_s{}^4 - T_{surr}{}^4) \qquad (2.3)$$

Where.

   Q_{rad} = Rate of heat radiation, W

   ε = Emissivity,

   σ = Stefan-Boltzmann constant, $5.67 \times 10^{-8} W/m^2 K^4$

   T_s = Surface temperature, $°C$

   T_{surr} = Absolute temperature of surrounding, $°C$

2.6 Equipments, Tools and Software Used

2.6.1 Graphic Processing Unit (GPU)

For this research, GPU from ASUS will be used for the test to meet the need of triple fans GPU. The specification of the GPU as shown in Table 2.1.

Table 2.1: GPU specification

Product Specifications	ASUS ROG Strix GTX1070
Video Memory	8GB
Clock Speeds (Boost / Base)	1835MHz/1632MHz
Memory Type	GDDR5
CUDA Core	1920
Bandwidth	256 Gb/s
Memory Speed	8 Gbps
Memory Interface	256-bit
Resolution	Digital Max Resolution 7680 x 4320
Dimension	11.73 " x 5.28 " x 1.57 "
Recommended PSU	500W
Picture	

2.6.2 Response Surface Methodology (RSM)

Response Surface Method (RSM) is a method for designing experiments, creating models, analysing the impacts of multiple factors, and reaching the best circumstances for desired results using a small number of planned experiments. It can be used to show how a set of input variables can affect a specific response over a particular region of interest as well as which input values will result to maximum or

minimum response (Trinh & Kang, 2010). Mujtaba et al. (2020) stated that the effect of independent input process parameters on output variables was visualized using RSM. The tool managed to help researchers lessen the number of experiments to get the optimum results. RSM also proven to be more efficient and cost effective since it requires minimum amount of testing and time consumed. It was also used to examine the impact of a large independent and dependent factors (Hooda et al., 2012). According to Zhang et al. (2017), RSM has been a simple yet efficient strategy for balancing accuracy and efficiency in a variety of general reliability problems. Plus, RSM is a more accurate predictor of the effect of factors on response and a more effective tool for optimization. RSM is an approach that combines mathematical and statistical tools for modelling and optimising response variables. It incorporates quantitative independent variables (Gopalakannan & Senthilvelan, 2013).

According to Azcarate, Pinto, & Goicoechea (2020), RSM is applied with a series of consecutive steps starting from screening, building response surface, modelling and simultaneous optimization. Screening is a method where variables which may have an effect to the process are examined. The objective is to begin by considering many variables using a screening design and statistical analysis of the results by employing graphical tools and analysis of variance (ANOVA). Then, it is determined which factors have a substantial effect on the reaction or responses. After that, response surface design is built to provide the experimental data. Commonly used central composite design (CCD) is implemented which built with two-level (-1 and +1) factorial design points, axial or star points and centre points. Next, RSM helps develop a mathematical model that describes the link between each response and its associated factors. The model can be a first-, second-, or higher-order polynomial function that is fitted using data acquired during experimentation using the

experimental design used for that reason. Notably, the fitted model must accurately describe the data relationship in order to produce accurate predictions within the experimental region. Last step which is optimization which the process is by applying optimization criteria such as maximization, minimization, reaching a target value or a range of-values. To meet the latter criteria, several functions are constructed to ensure that the desirability values fall within an acceptable range of response values.

2.6.3 Central Composite Design (CCD)

There are a few experimental designs that can be used in RSM. However, the most commonly used are Box-Behnken Design (BBD) and central composite design (CCD). Among alternative experimental designs, the CCD was chosen due to its capacity to construct response surfaces with fewer runs necessary. The CCD is a two-level factorial design that has been supplemented with centre and axial points to allow for the fitting of quadratic models (Laid et al., 2021). According to Wagner, Mount, & Giles (2014), the central composite design is a two-level factorial with the addition of 2k (k is the number of independent variables) points (star points) between the axes and repeat points at the centroid. For some case, due to the fact that star points outside of the plane may not be feasible processing conditions, it is necessary to locate the star points within the plane with the other coordinate points as shown in Figure 2.16.
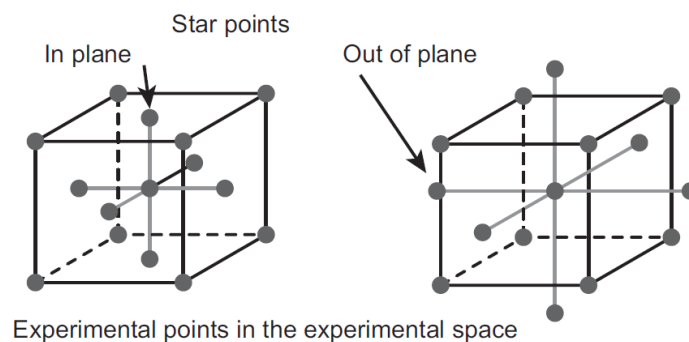

Figure 2.18: Central composite design with star points

CCD is frequently used in response surface methodology to construct a second-order polynomial for the response variables without requiring a full factorial design of experiments. Three different points exist in CCD: factorial points, central points, and axial points. The factorial points are vertices of the n-dimensional cube derived from a full or fractional factorial design with factor levels coded as –1, +1. The central point is the point in the design space that is in the centre. Axial points are located symmetrically with respect to the central point on the coordinate system's axes at a distance a from the design centre (Sahoo & Barman, 2012). The CCD design is an efficient method for sequential experimentation because it provides a sufficient amount of information for testing lack of fit without requiring an unusually large number of design points. On the basis of the CCD, there are four critical stages for optimising experiments: (1) conducting statistically designed experiments for the experimental plan, (2) recommending a mathematical model for the experimental data and focusing on analysis of variance data, (3) directly controlling the model's efficiency with diagnostic plots, and (4) estimating the response and verifying the model (Demirel & Kayan, 2012). Moreover, the central composite design (CCD), a technique that falls under the category of experimental design, is used for multivariable optimization and development, as well as multivariable analysis and method condition. There are numerous advantages to using a CCD over conventional "one factor at a time" experiments. Since the factors involved in an experiment are being changed at the same time, it generally requires fewer experiments than a full factorial design (Topkafa & Ayyildiz, 2016).
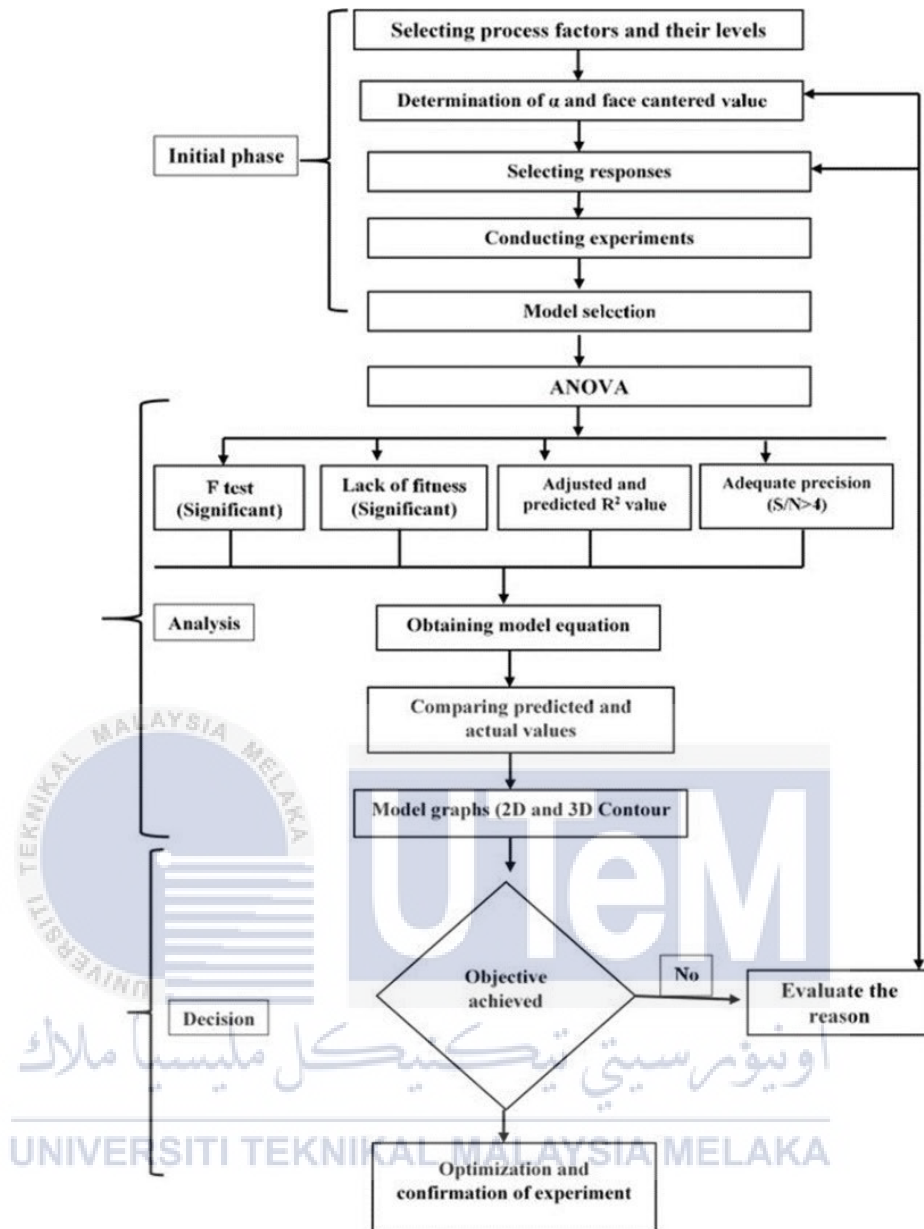
Figure 2.19: CCD flow diagram

2.6.4 Design-Expert 13

Design-Expert is a statistical software solely use for design of experiments (DOE). It offers screening, characterisation, optimization, resilient parameter design, mixture designs, and combination designs. This software helps users to examine the results of each solution one at a time. At the final result of the optimization and prediction processes, a decision is made about the design to use and the number of

levels to set for known factors (Iva Rezić, 2011). It also able to screen as many possible variables as possible by using Design-Expert test matrices. Analysis of variance (ANOVA) is used by Design-Expert to prove these variables' statistical significance. It also includes a number of graphs that aid in identifying and visualising standout effects and illustrates the impact of each element to the desired results.

Experiments is a step in the engineering method for determining the research outcome. There are a number of factors that must be addressed in order for the experiment to be completed correctly. Design-Expert could be useful to determine the number of variables in a single experiment. Design-Expert is a piece of software that assists with experiment design and optimization. A numerical optimizer, based on the validated predictive models, assists the user in determining the ideal values for each of the factors in the experiment. It is designed to make experiments simpler with the best manipulated variables figured out before the real experiment begins, getting the best result will take lesser time. Figure 2.17 shows Design-Expert interface.
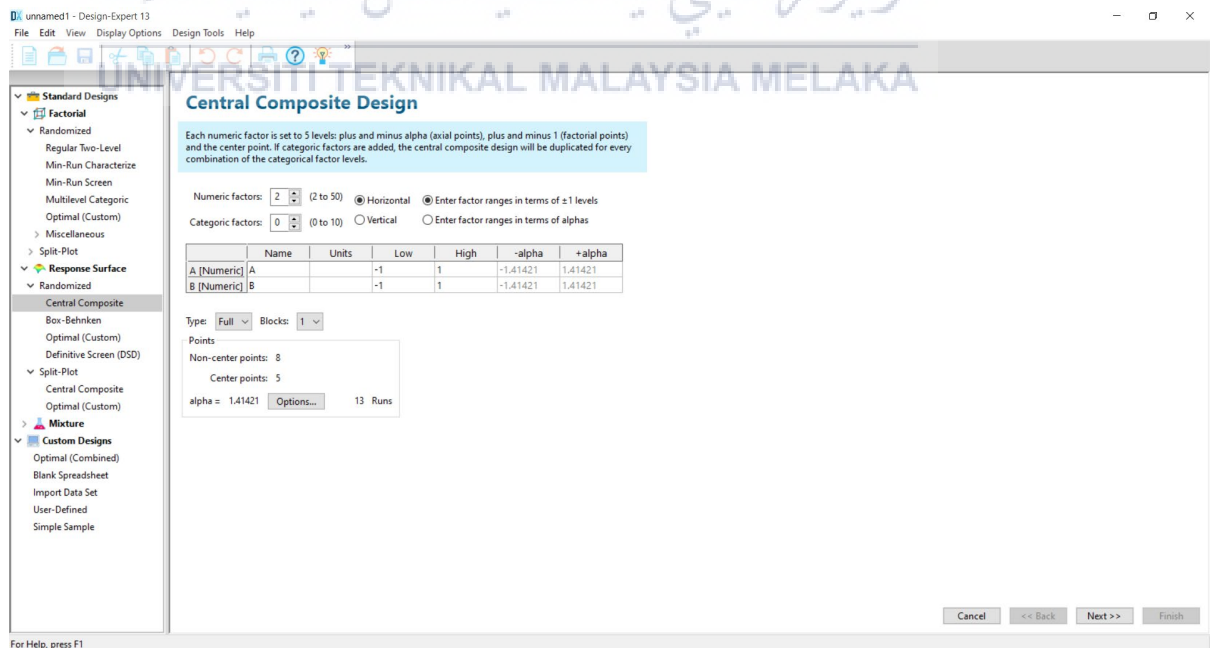
Figure 2.20: Design-Expert interface

2.6.5 MSI Afterburner

GPUs consists of processor that can be manipulated by software unlike other hardware such as mouse, keyboard or monitor. MSI or Micro Star International has built a software specifically for this exact functionality called MSI Afterburner. This software can manipulate overall GPU's performance such as memory clock, core clock, fan speed, core voltage, temperature limit, power limit and others. The speciality of this software is that it can be used with other GPU manufacturer such as ASUS, Palit, EVGA, Zotac or either Nvidia or AMD. The rule of thumb is the higher the voltage supplied to the GPU, the better the performance will be but the GPU's temperature needs to be taken into consideration so that it will not damage the GPU. This software utilizes the OC scanner to determine which card the system is using and from there, it lets user to tune their card according to their likings so it can maximize the performance of users GPU. Figure 2.18 shows the interface of MSI Afterburner.
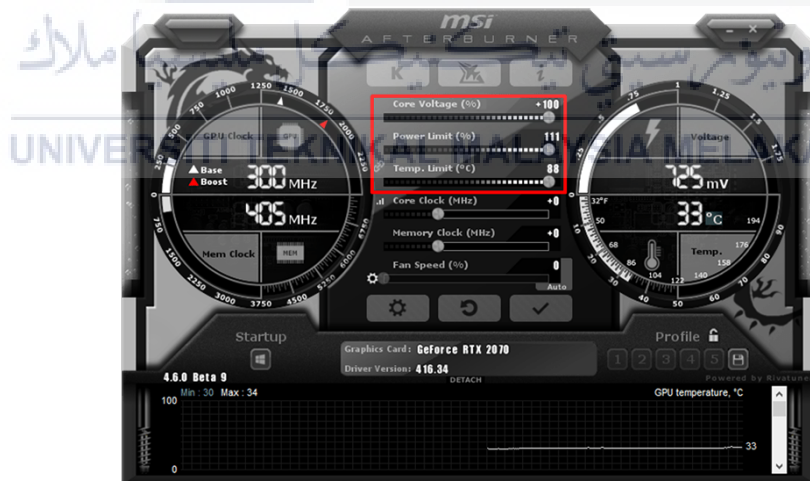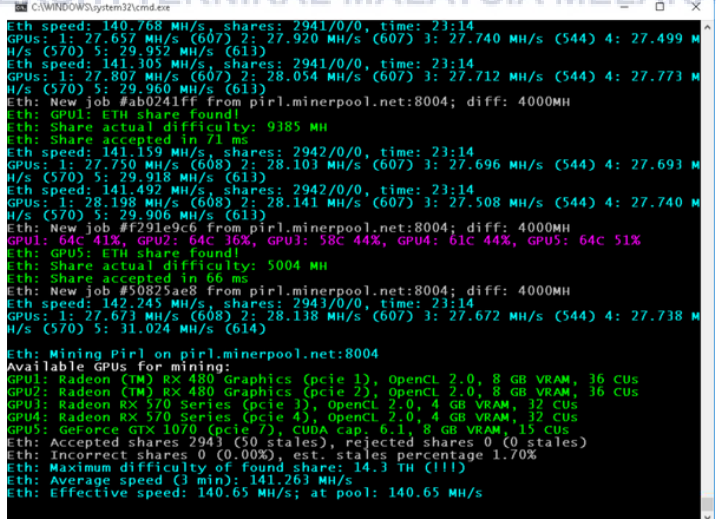
Figure 2.21: MSI Afterburner user interface

2.6.6 PhoenixMiner

PhoenixMiner is a cryptocurrency mining software which is used to create new cryptocurrencies and add new blocks to an existing blockchain. The mining party

26

receives newly mined cryptocurrency as compensation for their contribution to the blockchain. A computer's graphics processing unit (GPU) is used to assist with cryptocurrency mining. Today, the majority of mining is carried out via a mining pool, which distributes the reward among its members via a network of computers.

PhoenixMiner, a cryptocurrency miner, generates cryptocurrency using the Dagger Hashimoto (Ethash) algorithm. This list includes several well-known cryptocurrencies, including Ethereum, Ethereum Classic, and MOAC. The software is compatible with graphics cards from both Nvidia and AMD. It runs on Windows x64 and Linux x64. According to the creators, it is currently the fastest and most cost-effective Ethereum/Ethash miner available. There is, however, a 0.65% service charge, which means that the user will "work" for developers for 35 seconds out of every 90 minutes of GPU mining. By comparison, Claymore's miner development fee is 1%. In this research, the software is utilizes to collect the GPU's hashing power in hash rate (MH/s). Hash rate refers to the amount of computational power used to process the transactions and mine cryptocurrency.

Figure 2.22: Phoenixminer's interface

27

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter will discuss the strategy that will be used to achieve the goals of the research project in a successful manner. In order to achieve the objectives, a flowchart will be used to demonstrate the process step-by-step. This chapter will cover every aspect of the research, including the methods, software, processes and procedures that were implemented. The purpose of this chapter is to ensure that all plans and activities are properly managed to avoid any errors that may occur as a result of a poor approach to planning. Also, all the information from literature review is utilized to conduct the experiment.
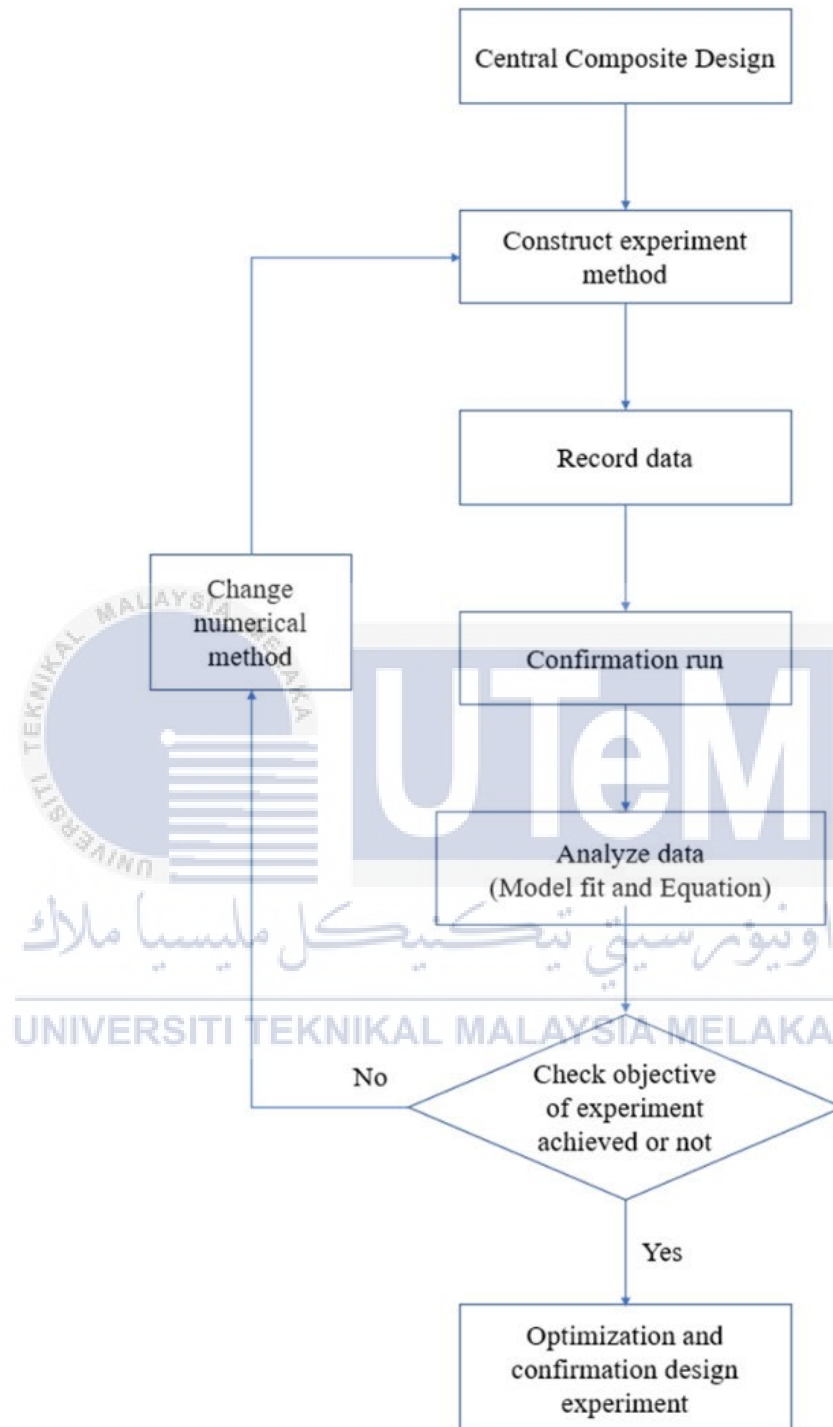
3.2 Flowchart

Figure 3.1: Flowchart of methodology

3.3 Schematic Diagram
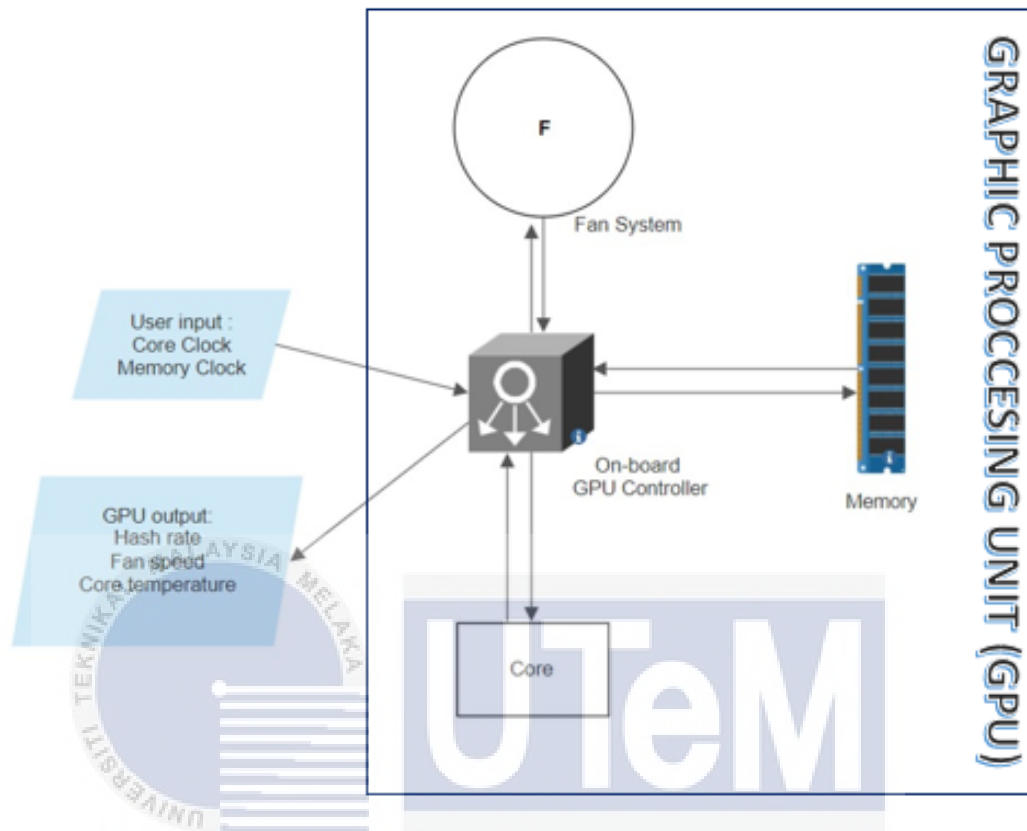
Figure 3.2: Schematic diagram of the GPU setup

3.4 Design of Experiment (DOE)

Design of Experiment (DOE) is a set of mathematical and statistical techniques for decreasing the number of experiments required to determine the effect of parameters on a process' response (Hatami, 2015). The numeric factor in the Central Composite Design (CCD) can be filled with 2 to 50 variables, whereas the Box-Behnken Design (BBD) can be filled with 3 to 21 variables. The experiment's manipulated variables are represented by numerical factors. There are only two numeric factors in this study which is the memory clock and the core clock of the GPU, BBD was chosen to optimise the data from the experiments. The experiments were designed using Response Surface Methodology (RSM) based on the CCD in Design-

Expert Version 13.0 to examine the relationship between clocking (core clock, memory clock) and GPU response (fan speed, core temperature, hash rate).

The experimental data were analysed statistically using Design-Expert Version 13.0. Different statistical parameters of various polynomial models such as lack-of-fit, predicted and adjusted multiple correlation coefficients and coefficients of variation were applied to determine the right clocking with the highest GPU efficiency. A significant difference was discovered after analysing variance and computing f-value at probabilities of 0.5, 0.1 and 0.01. From the data collected, Design-Expert made a response plot to display the relationship between emulsifying conditions and response variables. When variables and responses are optimised, the CCD design helps in determining the number of tests to run. Each variable's minimum, intermediate, and maximum values are denoted by the symbols -a, 1, 0, +1, and +a, respectively, as shown in Table 3.1

Table 3.1: Independent variables and their corresponding levels for GPU clocking

Independent variable / Coded levels / Symbol	-a	-1	0	+1	+a
Core Clock (MHz) — A	-200	-141.421	0	141.421	200
Memory Clock (MHz) — B	300	387.868	600	812.132	900

According to the Table 3.1, the core clock is set between -200MHz and 200MHz, while the memory clock is set between 300MHz and 900MHz. The value were obtained from Minerstat website (minerstat.com). The factorial +1 and -1 design points are depicted in Figure 3.2 by the square's four corners. The parameters +1 and -1 are used to define the study region's boundaries, which are believed to contain the

optimal situation. On the other hand, axial points frequently fall outside of this range. The axial +alpha and -alpha design points are represented in the Figure 3.2 by four-star points. Even for the most severe axial runs, there are alphas to ensure that everything continues to work correctly. The emphasis must be on something that is operable.
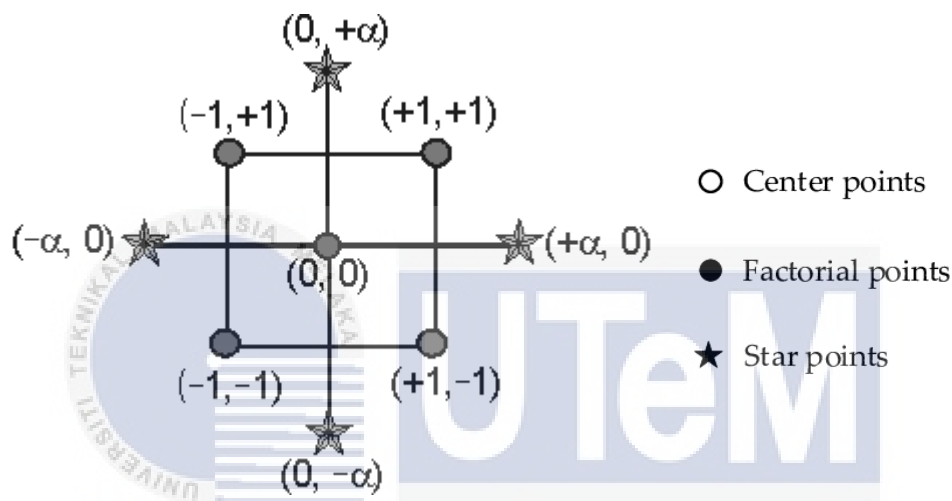
Figure 3.3: Classic CCD for 2 factors square's four corners

3.5 Experimental Setup

Values of core clock and memory clock that were specified by Design-Expert software will be applied to MSI Afterburner software. From the software, GPU response of the fan speed (rpm), core temperature of GPU processor (ºC) and the hash rate (MH/s) can be observed through Phoenixminer. These responses will be analysed and optimised in Design-Expert to determine the optimal clocking settings. According to Minerstat (minerstat.com), the optimal core clock and memory clock value for GTX 1070 are 100 MHz and 300 MHz respectively. Thus, for the purpose of conducting this research, the core clock and memory clock are set to be (-200 to 200 MHz) and (300 to 900 MHz) in the numerical factors in terms of alphas since the limitations of

maximum and minimum has been set. 13 different run values were given by Design-Expert consist of five central points, four fractional points and four axial points by applying CCD design. Figure 3.2 explains the setup for this research. The experiment was then repeated continuously multiple time to get constant response (fan speed, core temperature and hash rate) from GPU.

Central Composite Design

Each numeric factor is set to 5 levels: plus and minus alpha (axial points), plus and minus 1 (factorial points) and the center point. If categoric factors are added, the central composite design will be duplicated for every combination of the categorical factor levels.

Numeric factors: 2 (2 to 50) ● Horizontal   ○ Enter factor ranges in terms of ±1 levels

Categoric factors: 0 (0 to 10) ○ Vertical   ● Enter factor ranges in terms of alphas

	Name	Units	Low	High	-alpha	+alpha
A [Numeric]	Core Clock	MHz	-141.421	141.421	-200	200
B [Numeric]	Memory Cloc	MHz	387.868	812.132	300	900

Type: Full ∨ Blocks: 1 ∨

Points

Non-center points: 8

Center points: 5

alpha = 1.41421   Options...   13 Runs

Figure 3.4: Design-Expert interface for CCD

Std	Run	Space Type	Factor 1 A:Core Clock MHz	Factor 2 B:Memory Clock MHz	Response 1 Fan speed rpm	Response 2 Core Temperatu… C	Response 3 Hashrate MH/s
12	1	Center	0	600			
13	2	Center	0	600			
7	3	Axial	0	300			
4	4	Factorial	141.421	812.132			
1	5	Factorial	-141.421	387.868			
10	6	Center	0	600			
11	7	Center	0	600			
6	8	Axial	200	600			
3	9	Factorial	-141.421	812.132			
2	10	Factorial	141.421	387.868			
8	11	Axial	0	900			
5	12	Axial	-200	600			
9	13	Center	0	600			

Figure 3.5: Design-Expert response interface

3.6 Statistical Analysis and Optimization of Clocking and GPU Response

Design-Expert software was used to conduct a regression analysis of the GPU response (fan speed, hash rate, and core temperature). The software will perform an analysis of variance (ANOVA) on the built models. ANOVA is frequently used to determine the degree of dissimilarity between sets of data. It establishes the model's significance through an analysis of the model's statistical fit. Then, using the optimization constraint option, the numerical setting for the ideal fan speed and temperature can be established. CCD is an effective method for fitting a quadratic surface model sequentially. Numerous design alternatives that do not fit are eliminated from CCD's evaluation process.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Result

The core clock and memory clock are put to the test by using the Afterburner software to overclock the GPU as suggested by Design-Expert software as Figure 3.5. The result of the response is recorded in Table 4.1 presented the likes of fan speed, core temperature and hash rate. This is done to get the optimal fan speed with the highest efficiency at a certain core clock and memory clock.

Table 4.1: Experimental design and response value obtained by the GPU

Standard order	Space type	Independent input variable (MHz)		Response (experiment)		
		A: Core clock	B: Memory clock	Fan speed (rpm)	Core temperature (ºC)	Hash rate (MH/s)
1	Factorial	-141.421	387.868	1934	53.3	21.6
2	Factorial	141.421	387.868	1925	53.1	24.5
3	Factorial	-141.421	812.132	1923	53.1	21.5
4	Factorial	141.421	812.132	1908	53	24.3
5	Axial	-200	600	1934	53.2	21.1
6	Axial	200	600	1914	53	24.8
7	Axial	0	300	1930	53.1	23.1
8	Axial	0	900	1921	53	22.9
9	Center	0	600	1922	53	23
10	Center	0	600	1930	53.1	23
11	Center	0	600	1930	53.1	23
12	Center	0	600	1927	53.1	23
13	Center	0	600	1922	53	23

4.2 CCD and ANOVA

Response 1: Fan speed

Table 4.2: ANOVA for fan speed response

Source	Sum of squares	df	Mean square	F-value	p-value	
Model	549.05	2	274.53	19.89	0.0003	significant
A-Core Clock	341.71	1	341.71	24.76	0.0006	
B-Memory Clock	207.35	1	207.35	15.02	0.0031	
Residual	138.03	10	13.80			
Lack of Fit	73.23	6	12.20	0.7534	0.6404	not significant
Pure Error	64.80	4	16.20			
Cor Total	687.08	12				

Table 4.3: Fit Statistics for fan speed response

Std. Dev.	3.72	R²	0.7991	Adeq. Precision	13.0290
Mean	1924.62	Adjusted R²	0.7589		
C.V. %	0.1930	Predicted R²	0.6730		

Response 2: Core Temperature

Table 4.4: ANOVA for core temperature response

Source	Sum of squares	df	Mean square	F-value	p-value	
Model	0.0668	2	0.0334	11.10	0.0029	significant
A-Core Clock	0.0425	1	0.0425	14.11	0.0037	
B-Memory Clock	0.0244	1	0.0244	8.09	0.0174	
Residual	0.0301	10	0.0030			
Lack of Fit	0.0181	6	0.0030	1.01	0.5224	not significant
Pure Error	0.0120	4	0.0030			
Cor Total	0.0969	12				

Table 4.5: Fit Statistics for core temperature response

Std. Dev.	0.0549	R²	0.6894	Adeq. Precision	9.7153
Mean	53.08	Adjusted R²	0.6273		
C.V. %	0.1034	Predicted R²	0.4747		

Result obtained from ANOVA for both fan speed and core temperature suggested that the model constructed are in satisfactory condition and significant. The lack of fit for both responses is not significant which indicates the model is adequate. Table 4.2 and Table 4.4 shown that both of p-value has a value of less than 0.05 which signify a statistically accepted model. Having a p-value larger than 0.10 indicates that the input variables is insignificant in relation to the responses recorded.

In addition, the correlation between R^2 and adjusted R^2 is the second factor to verify the constructed model by Design-Expert. The value of R^2 for both responses are high in response to the independent input factors of core clock and memory clock indicating that the model generated is reliable. The predicted R^2 of fan speed which is 0.6730 compared to the adjusted R^2 which equals to 0.7589 is reasonably close with the difference of less than 0.2, indicates that the model fits the data and can be used to interpolate successfully. Similar to core temperature predicted R^2 of 0.4747 and adjusted R^2 of 0.6273 also having less than 0.2 difference meaning the model provides good predictions. Plus, both of the responses have an adequate precision of more than 4 signalling for highly reliable model.

Then, visual diagnostics were carried out by comparing the experimentally measured values to the predicted values derived by the software using modelled equations. An example of the CCD model's ability to produce a good correlation between the input variable and the expected response values is shown by Figure 4.1 (a) and (b). Both of graphs show that there are no obvious outliers in the predicted and actual values. It is possible to use both the models in the experiment's design.
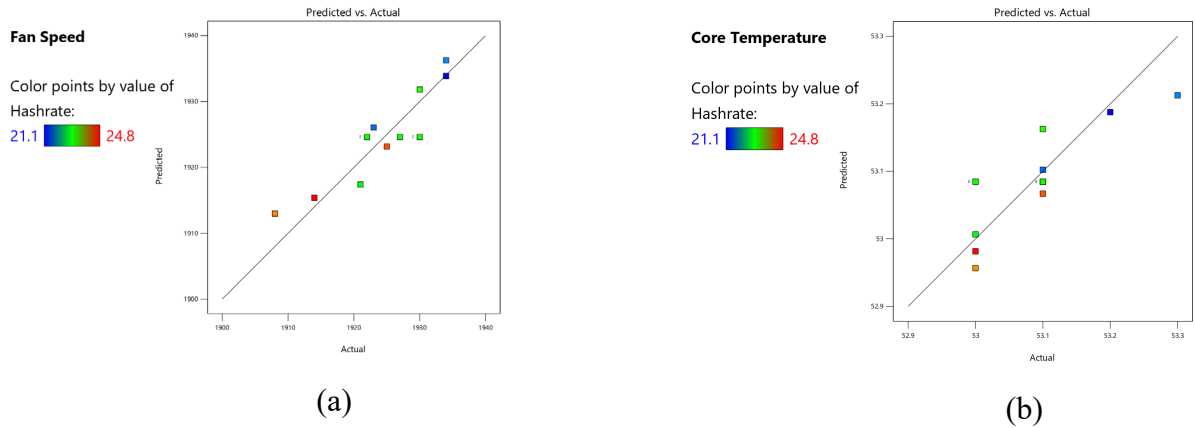
Figure 4.1: Graph of predicted (calculated) vs. actual (experimental) for:
(a) Fan speed (b) Core temperature

Response 3: Hash rate

Table 4.6: ANOVA for hash rate response

Source	Sum of squares	df	Mean square	F-value	p-value	
Model	14.98	2	7.49	2186.04	< 0.0001	significant
A-Core Clock	14.94	1	14.94	4359.70	< 0.0001	
B-Memory Clock	0.0425	1	0.0425	12.39	0.0055	
Residual	0.0343	10	0.0034			
Lack of Fit	0.0343	6	0.0057			
Pure Error	0.0000	4	0.0000			
Cor Total	15.02	12				

Table 4.7: Fit Statistics for hash rate response

Std. Dev.	0.0585	R²	0.9977	Adeq. Precision	137.4483
Mean	22.98	Adjusted R²	0.9973		
C.V. %	0.2547	Predicted R²	0.9950		

Validation of the results for this project is by monitoring the R^2, adjusted R^2, predicted R^2 and the adequate precision. The results shown in Table 4.6 is done by proposed linear model. When p-value shows value that were less than 1 and high amount of f-value, it considered significant in ANOVA. There is a 0.01% chance that

such a large f-value occurs as a result of noise. The ANOVA table shows that both A and B, as well as all of their interactions, are statistically significant.

After that, R^2 and adjusted R^2 is evaluated to consider the constructed model. The value of R^2 and adjusted R^2 for response that close to 1 implying that changes for response (hash rate) depends almost 100% to the input variable (core clock and memory clock). In other words, the developed models are highly dependent on the input variables, and all variables used in this experiment are significant. Although there is a small difference (less than 0.2) between the adjusted and predicted R2 values, this model fits the data well and can be used for interpolation. Due to the software does not provide the p-value for Lack of Fit, a diagnostic plot comparing actual values to predicted values could be used to validate results. The lack of outliers in the graph of Figure 4.2 validates the hash rate models, allowing them to be used in the experimental design.
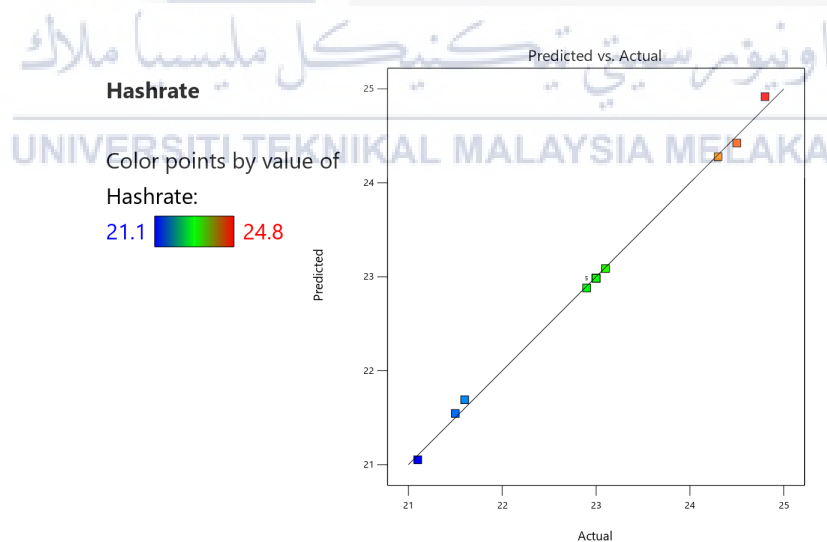
Figure 4.2: Graph of predicted (calculated) vs. actual (experimental) for hash rate

4.3 Effect of Independent Variables on Response Variables (Graph And Equation)

4.3.1 Fan Speed
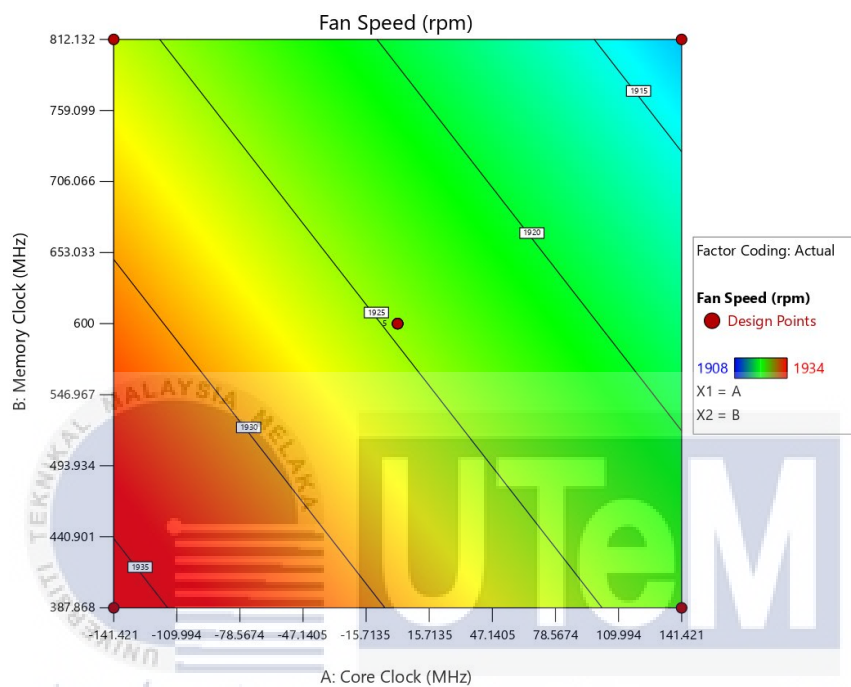
Figure 4.3: Contour plot for the combined effect of core clock (A), memory clock (B) and fan speed

$$\text{Fan speed} = 1924.62 + -6.53553*A + -5.09099*B \qquad (4.1)$$

Figure 4.3 represents the relationship between fan speed and the memory clock and core clocks. It can be seen that with the core clock maximum at 141.421 MHz and memory clock at 812.132MHz, the fan speed will decrease to lower than 1915 rpm. The red colour of the contour plot represents higher fan speed and blue represents the lower fan speed. Core clock values between -141.421 MHz to 47.1405 MHz and memory clock values between 387.868 MHz to 759.099 MHz are within the yellow and red region indicating a higher fan speed. Meaning that when the core clock and memory clock increase, the fan speed will decrease.

From the formula given by the software, A represents core clock and B represent memory clock. By referring that, it stated that the fan speed will be further reduced by A component with multiplier of -6.53553 and B component with multiplier of -5.09099 due to negative (-) sign before the multiplier components. Hence, it clearly shows that A has more influence to lower the fan speed compared to B. Additionally, the software has validated that this component because the p-value for core clock is lesser than p-value for memory clock.

4.3.2 Temperature
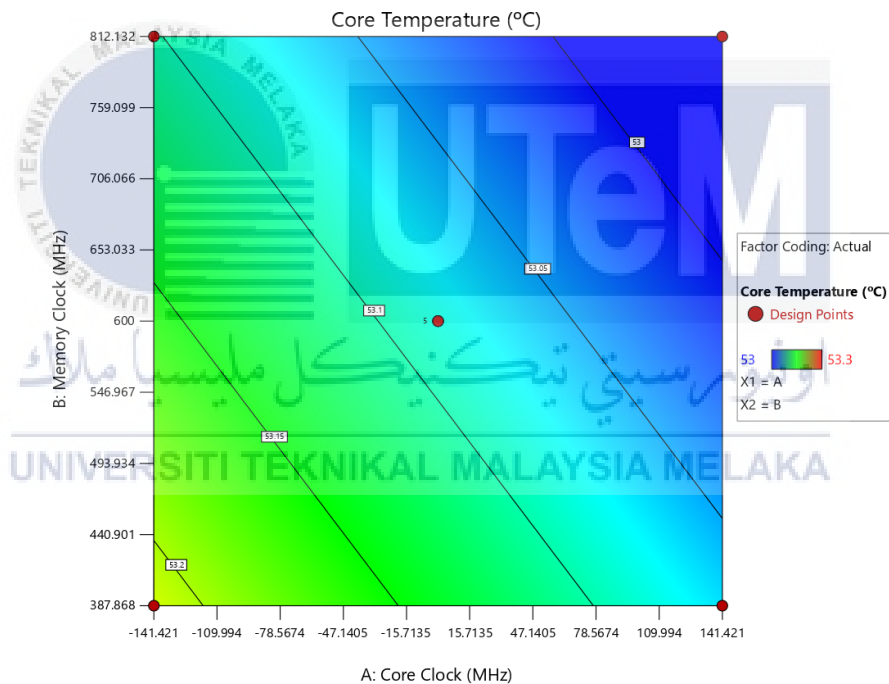
Figure 4.4: Contour plot for the combined effect of core clock (A), memory clock (B) and core temperature

$$\text{Core temperature} = 53.0846 + -0.0728553*A + -0.0551777*B \qquad (4.2)$$

Figure 4.4 shows the correlation between core temperature of the GPU and the memory and core clock. As can be seen, when the core clock is at maximum of 141.421 MHz and the memory clock is at 812.132 MHz, the core temperature dropped to 53ºC. The contour plot's yellowish green colour indicates a higher core

41

temperature, while the blue colour indicates a lower core temperature. Between -141.421 MHz and 78.5674 MHz for the core clock shows a higher temperature, and 387.868 MHz to 812.132 MHz for the memory clock also indicates a higher temperature because both are in the yellowish green region. That is, as the core and memory clocks increase, the core temperature decreases.

According to the formula, A indicates the core clock and B indicates the memory clock. As a result of the negative (-) sign preceding the multiplier components, it was stated that the core temperature would be further reduced by A component with a multiplier of -0.0728553 and B component with a multiplier of -0.0551777. As a result, it is clear that A (core clock) has a bigger effect on lowering the core temperature than B (memory clock). Additionally, the software validated that this component exists due to the fact that the p-value for the core clock is less than the p-value for the memory clock.

4.3.3 Hash Rate
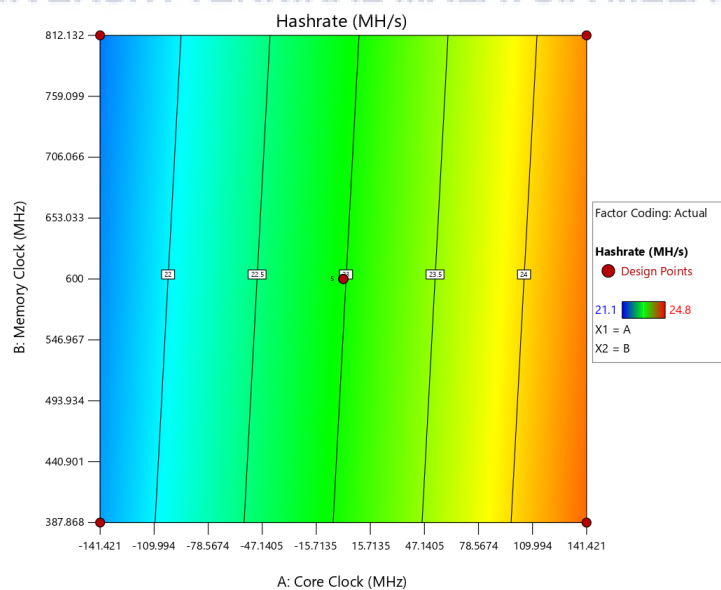
Figure 4.5: Contour plot for the combined effect of core clock (A), memory clock (B) and hash rate

$$\text{Hash rate} = 22.9846 + 1.36657*A + -0.0728553*B \qquad (4.3)$$

In Figure 4.5 it shows how hash rate and the speed of the memory and core clock are linked. When the core clock is at 141.421 MHz and the memory clock is at 812.132 MHz, the hash rate increases up to 24 MH/s. The contour plot's red colour means that the hash rate is higher, and the blue colour means that the hash rate is lower. The highest value of hash rate was recorded in the core clock range between 109.994 MHz and 141.421 MHz. It means that as the core clocks increases, the hash rate will also increase. By contrast, the memory clock has little effect on the hash rate value.

It is stated in the formula that A (core clock) has a larger multiplier of +1.36657, implying that A (core clock) has a significant impact effect on the value of hash rate. In contrast, with a multiplier of 0.0728553, B (memory clock) has a smaller effect on the hash rate value and, as a result of the negative multiplier, will actually decrease the value. Furthermore, the software confirms the existence of this component because the p-value of the core clock is less than the p-value of the memory clock.

4.4 Data Optimization

Table 4.8: Numerical optimization setting

Name	Goal	Lower Limit	Upper Limit	Lower Weight	Upper Weight	Importance
A: Core Clock	is in range	-141.421	141.421	1	1	3
B: Memory Clock	is in range	387.868	812.132	1	1	3
Fan speed	minimize	1908	1934	1	1	3
Core Temperature	minimize	53	53.3	1	1	3
Hash rate	maximize	21.1	24.8	1	1	3
Ambient Temperature	none	30	30	1	1	3

Hot Spot Temperature	none	64	64.3	1	1	3
Power Consumption	none	112.6	113.5	1	1	3

Optimization is the process of determining the optimal values for an objective function within a specified domain (Iva Rezić, 2011). Following the completion of all the necessary analysis, numerical optimization was carried out using the desirability function in Design-Expert Software. By that, 5 different solutions were discovered, each with a different level of independent variables. The goals selected for the optimization of the GPU settings were core clock and memory clock are 'in range' because the values for core clock and memory clock obtained from the minerstat.com website are within the optimum range of-values for these parameters. To achieve these limits in performance, the core clock is set to 141.421 MHz at the lowest possible frequency and 141.421 MHz at the highest possible frequency. Meanwhile, the lower limit of the memory clock is 387.868 MHz and the upper limit is 812.132 MHz. Additionally, the goals for the response variables include the lowest possible fan speed, the lowest possible core temperature, and the highest possible hash rate. Those settings have produced 5 possible solutions, with only one of them being selected as the optimal setting for the GPU to operate at its best. Table 4.9 lists the experimental fan speed, core temperature, and hash rate for each of the suggested solutions that were chosen for testing with the predicted and experimental fan speed, core temperature, and hash rate.

44

Table 4.9: Numerical optimization solution

Number	Core Clock	Memory Clock	Fan speed	Core Temperature	Hash rate	Hot Spot Temperature	Power Consumption	Desirability	
1	141.421	812.132	1912.989	52.957	24.278	63.957	113.361	0.885	
2	131.761	812.130	1913.435	52.962	24.185	63.962	113.342	0.870	
3	141.421	746.131	1914.573	52.974	24.301	63.974	113.268	0.865	
4	141.421	738.554	1914.755	52.976	24.304	63.976	113.258	0.862	
5	**141.421**	**724.663**	**1915.088**	**52.979**	**24.308**	**63.979**	**113.238**	**0.858**	**Selected**

For confirmation and validation, the experiment will be repeated using one solution provided by the software and the constructed equation. The optimal solution is the one with the highest hash rate for a specified memory and core clock. The experiment is then run with the selected core and memory clocks and GPU response to compare them to the software-predicted values. At the optimization condition, the predicted and experimental value deviations were less than 2%. This discrepancy is due to technological limitations, as the sensor is not sensitive to small changes in GPU response and the software (Phoenixminer and Afterburner) can only show results up to one decimal place. These small deviations can be related to a reliable testing device and a mathematical model capable of predicting the value of the GPU response. Without the need to manually find each core clock and memory clock, numerical optimization based on CCD can predict the desired lowest fan speed, lowest core temperature, and highest possible hash rate at constant power consumption. The constructed equation can be used to predict the GPU response using any core clock and memory clock setting. On the basis of the validation process, it is possible to conclude that the accuracy of CCD has been experimentally demonstrated. In conclusion, the CCD method is the correct method for this experiment.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Core clock and memory clock are two main factors for an optimal performance of a graphic processing unit (GPU). Optimal performance of a GPU is not only reflected by its maximum processing power (hash rate) but more importantly - highest performance at a stable and continuous operation. This is vital especially for a GPU to be utilized under non-stop, high-power-density operation. Result shows that a non-harmonious core and memory clock setting effects the overall performance adversely, leads to lower hash rate and an increase in core temperature. This situation leads to even more energy is required to cool down the GPU. This situation fatigues the cooling system consists of fan bearing, rotor, electrical connection, and other parts. The harmonious clocking of both parameters was obtained from literature and varies from on-board chips to another. However, for GTX1070 used in this project, it was shown that the core memory clock can range from -200 to 200 MHz, while memory clock can be set from 300 to 900 MHz.

Acting upon the above-mentioned data set, a design on experiment was built. This is vital to minimize the amount of experiment as the setup and experiment running for each clock setting is time-consuming. Optimization tool namely Design Expert was used to accommodate the process. The chosen design was Central Composite Design or CCD as literature shows that it was previously used by many researchers for an almost identical data set and CCD was shown to produce an equation which predicts

the results with an accuracy of over 95 percent. Using CCD for this project, only 13 experiments were required to be conducted, and all experiments were done in less than seven weeks. The generated equation of hash rate, fan speed and core temperature were obtained and it suggested a new clock setting for even better (the best actually) performance than the current settings. To prove it experimentally, the core and memory of the suggested setting was again tested, and results show that the predicted hash rate and fan speed were precise with less than 2 percent deviation.

5.2 Recommendation

Based form our literature review done during this project, it was clear that there are many factors contributing to performance of a graphic processing unit (GPU) for examples: number of fans on the GPU, rotation direction of fans either in the same direction or different, fan type (axial or radial), heat transfer fluid (air or liquid) and many more. This project only covers three-fan GPU and only on the ASUS ROG Strix GTX1070 model. There are plenty of other models which yet to be tested for an optimal operating parameter. That is on the side of the hardware.

Considering the software, the design of experiment via Central Composite Design (CCD) used in this project is also only one out of many designs available for optimization process. Response Surface Method or RSM, Box Behnken and many other optimization designs are still not yet tested. Even some literature shows that these designs were also capable to obtain the desired parameters at high accuracy. But due to time constraint, this was not done. It is highly advisable to execute the project again via these designs and compare them with CCD so that a more in-depth discussion can be produced.

REFERENCES

Ahmadian-Elmi, M., Mohammadifar, M., Rasouli, E., & Hajmohammadi, M. R. (2021). Optimal design and placement of heat sink elements attached on a cylindrical heat-generating body for maximum cooling performance. *Thermochimica Acta*, *700*, 178941. https://doi.org/10.1016/j.tca.2021.178941

Ahmed, H. E., Salman, B. H., Kherbeet, A. S., & Ahmed, M. I. (2018). Optimization of thermal design of heat sinks: A review. *International Journal of Heat and Mass Transfer*, *118*, 129–153. https://doi.org/10.1016/j.ijheatmasstransfer.2017.10.099

Anish, M., & Kanimozhi, B. (2018). Experimental investigation and heat transfer process on longitudinal fins with different notch configuration. *International Journal of Ambient Energy*, *39*(1), 34–37. https://doi.org/10.1080/01430750.2016.1222965

Asirvatham, L. G., Nimmagadda, R., & Wongwises, S. (2013). Heat transfer performance of screen mesh wick heat pipes using silver-water nanofluid. *International Journal of Heat and Mass Transfer*, *60*(1), 201–209. https://doi.org/10.1016/j.ijheatmasstransfer.2012.11.037

Azcarate, S. M., Pinto, L., & Goicoechea, H. C. (2020). Applications of mixture
experiments for response surface methodology implementation in analytical
methods development. *Journal of Chemometrics*, *34*(12).
https://doi.org/10.1002/cem.3246

Bădălan, N., & Svasta, P. (2017). Fan vs. passive heat sink with heat pipe in cooling
of high power LED. *2017 IEEE 23rd International Symposium for Design and
Technology in Electronic Packaging, SIITME 2017 - Proceedings*, *2018-
Janua*, 296–299. https://doi.org/10.1109/SIITME.2017.8259911

Breitbart, J. (1999). Case studies on GPU usage and data structure design. *Diplomová
Práce, Universit\\"at Kassel, Germany*.
http://www.geeks3d.com/downloads/200808/Jens_Breitbart_thesis.pdf

Chen, J. S., & Chou, J. H. (2015). The length and bending angle effects on the cooling
performance of flat plate heat pipes. *International Journal of Heat and Mass
Transfer*, *90*, 848–856.
https://doi.org/10.1016/j.ijheatmasstransfer.2015.06.032

Choi, J., Jeong, M., Yoo, J., & Seo, M. (2012). A new CPU cooler design based on an
active cooling heatsink combined with heat pipes. *Applied Thermal
Engineering*, *44*, 50–56. https://doi.org/10.1016/j.applthermaleng.2012.03.027

Crider, M. (2018, October 4). Does It Matter Which Graphics Card Manufacturer You
Choose? – Review Geek. Retrieved April 25, 2021, from

www.reviewgeek.com website: https://www.reviewgeek.com/7683/does-it-matter-which-graphics-card-manufacturer-you-choose/

Crypto Mining Monitor and Management Software. (n.d.). Retrieved January 9, 2022, from minerstat website: https://minerstat.com

Culham, J. R., Khan, W. A., Yovanovich, M. M., & Muzychka, Y. S. (2007). The influence of material properties and spreading resistance in the thermal design of plate fin heat sinks. *Journal of Electronic Packaging, Transactions of the ASME*, *129*(1), 76–81. https://doi.org/10.1115/1.2429713

De Schampheleire, S., De Kerpel, K., Deruyter, T., De Jaeger, P., & De Paepe, M. (2015). Experimental study of small diameter fibres as wick material for capillary-driven heat pipes. *Applied Thermal Engineering*, *78*, 258–267. https://doi.org/10.1016/j.applthermaleng.2014.12.027

Demirel, M., & Kayan, B. (2012). *Application of response surface methodology and central composite design for the optimization of textile dye degradation by wet air oxidation*. 1–10.

Evanson, N. (2020, March 16). Anatomy of a Graphics Card. Retrieved from TechSpot website: https://www.techspot.com/article/1988-anatomy-graphics-card/

Feng, S., Shi, M., Yan, H., Sun, S., Li, F., & Lu, T. J. (2018). Natural convection in a cross-fin heat sink. *Applied Thermal Engineering*, *132*(December), 30–37. https://doi.org/10.1016/j.applthermaleng.2017.12.049

Feng, Z., & Li, P. (2013). Fast thermal analysis on GPU for 3D ICs with integrated microchannel cooling. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, *21*(8), 1526–1539. https://doi.org/10.1109/TVLSI.2012.2211050

Gopalakannan, S., & Senthilvelan, T. (2013). Application of response surface method on machining of Al–SiC nano-composites. *Measurement*, *46*(8), 2705–2715. https://doi.org/10.1016/j.measurement.2013.04.036

Hatami, M., Cuijpers, M. C. M., & Boot, M. D. (2015). Experimental optimization of the vanes geometry for a variable geometry turbocharger (VGT) using a Design of Experiment (DoE) approach. *Energy Conversion and Management*, *106*, 1057–1070. https://doi.org/10.1016/j.enconman.2015.10.040

Hirasawa, T., Kawabata, K., & Oomi, M. (2005). Evolution of the heatsink technology. *Furukawa Review*, *27*, 25–29.

Hooda, A., Nanda, A., Jain, M., Kumar, V., & Rathee, P. (2012). Optimization and evaluation of gastroretentive ranitidine HCl microspheres by using design expert software. *International Journal of Biological Macromolecules*, *51*(5), 691–700. https://doi.org/10.1016/j.ijbiomac.2012.07.030

Huang, C.-H., & Gau, C.-W. (2012). An optimal design for axial-flow fan blade: theoretical and experimental studies. *Journal of Mechanical Science and Technology*, *26*(2), 427–436. https://doi.org/10.1007/s12206-011-1030-7

Hussain, A. A., Freegah, B., Khalaf, B. S., & Towsyfyan, H. (2019). Numerical investigation of heat transfer enhancement in plate-fin heat sinks: Effect of flow direction and fillet profile. *Case Studies in Thermal Engineering*, *13*(December 2018). https://doi.org/10.1016/j.csite.2018.100388

Intel. (2020). What Is a GPU? Graphics Processing Units Defined. Retrieved from Intel website: https://www.intel.com/content/www/us/en/products/docs/processors/what-is-a-gpu.html

Jayapragasan, C., & Reddy, K. (2017). DESIGN OPTIMIZATION AND EXPERIMENTAL STUDY ON THE BLOWER FOR FLUFFS COLLECTION SYSTEM. *Journal of Engineering Science and Technology*, *12*(5), 1318–1336.

Jouhara, H. (2018). Heat Pipes. *Comprehensive Energy Systems*, *4–5*(m), 70–97. https://doi.org/10.1016/B978-0-12-809597-3.00403-X

Kang, S., Choi, H. J., Kim, C. H., Chung, S. W., Kwon, D., & Na, J. C. (2011).
Exploration of CPU/GPU co-execution: From the perspective of performance,
energy, and temperature. *Proceedings of the 2011 ACM Research in Applied
Computation Symposium, RACS 2011*, 38–43.
https://doi.org/10.1145/2103380.2103388

Khattak, Z., & Ali, H. M. (2019). Air cooled heat sink geometries subjected to forced
flow: A critical review. *International Journal of Heat and Mass Transfer*, *130*,
141–161. https://doi.org/10.1016/j.ijheatmasstransfer.2018.08.048

Laid, T. M., Abdelhamid, K., Eddine, L. S., & Abderrhmane, B. (2021). Optimizing
the biosynthesis parameters of iron oxide nanoparticles using central composite
design. *Journal of Molecular Structure, 1229*, 129497.
https://doi.org/10.1016/j.molstruc.2020.129497

Mjallal, I., Farhat, H., Hammoud, M., Ali, S., Shaer, A. A. L., & Assi, A. (2018).
Electronic Devices. *Cooling Performance of Heat Sinks Used in Electronic
Devices*, *02003*, 1–4.
https://doi.org/https://doi.org/10.1051/matecconf/201817102003

Mujtaba, M. A., Masjuki, H. H., Kalam, M. A., Ong, H. C., Gul, M., Farooq, M.,
Soudagar, M. E. M., Ahmed, W., Harith, M. H., & Yusoff, M. N. A. M. (2020).
Ultrasound-assisted process optimization and tribological characteristics of

biodiesel from palm-sesame oil via response surface methodology and extreme learning machine - Cuckoo search. *Renewable Energy*, *158*, 202–214. https://doi.org/10.1016/j.renene.2020.05.158

Pambudi, N. A., Bugis, H., Kuncoro, I. W., Setiawan, N. D., Hijriawan, M., Rudiyanto, B., & Basori, B. (2019). Preliminary experimental of GPU immersion-cooling. *E3S Web of Conferences*, *93*, 1–4. https://doi.org/10.1051/e3sconf/20199303003

PC, X. (2019, September 11). History of GPUs. Retrieved April 28, 2021, from XOTIC PC website: https://xoticpc.com/blogs/news/history-of-gpus#:~:text=Back%20in%201999%2C%20NVIDIA%20popularized

Pirunkaset, M., & Laksitanonta, S. (2008). Study on the Effect of Blade Angle on the Performance of a Small Cooling Tower. *Nat. Sci.)*, *42*, 378–384.

Rajabi, N. & Rafee, Roohollah & Frazam-Alipour, S.. (2017). Effect of blade design parameters on air flow through an axial fan. International Journal of Engineering, Transactions A: Basics. 30. 1583-1591. 10.5829/ije.2017.30.10a.20.

Rezić, I. (2011). Prediction of the surface tension of surfactant mixtures for detergent formulation using Design Expert software. *Monatshefte Fur Chemie*, *142*(12), 1219–1225. https://doi.org/10.1007/s00706-011-0554-y\

Sahoo, P., & Barman, T. K. (n.d.). *ANN modelling of fractal dimension in machining.* 159–226. https://doi.org/10.1533/9780857095893.159

Singh, G. (2017). *Natural Convection Heat Transfer from Modified 1Degree 2 Degree And 3 Degree Outward Expansion of Pin Fins. 7*(8), 14567–14570.

Solomon, A. B., Ramachandran, K., & Pillai, B. C. (2012). Thermal performance of a heat pipe with nanoparticles coated wick. *Applied Thermal Engineering*, *36*(1), 106–112. https://doi.org/10.1016/j.applthermaleng.2011.12.004

Tharayil, T., Asirvatham, L. G., Cassie, C. F. M., & Wongwises, S. (2017). Performance of cylindrical and flattened heat pipes at various inclinations including repeatability in anti-gravity – A comparative study. *Applied Thermal Engineering,* *122*, 685–696. https://doi.org/10.1016/j.applthermaleng.2017.05.007

Topkafa, M., & Ayyildiz, H. F. (2017). An implementation of central composite design: Effect of microwave and conventional heating techniques on the triglyceride composition and trans isomer formation in corn oil. *International Journal of Food Properties*, *20*(1), 198–212. https://doi.org/10.1080/10942912.2016.1152481

Trinh, T. K., & Kang, L.-S. (2010). Application of Response Surface Method as an Experimental Design to Optimize Coagulation Tests. *Environmental*

Engineering   *Research*, *15*(2),   63–70.

https://doi.org/10.4491/eer.2010.15.2.063

Verma, A. (2017, March 29). Graphics Card Components & Connectors Explained.

Retrieved May 24, 2021, from Graphics Card Hub website:

https://graphicscardhub.com/graphics-card-component-connectors/

Wagner, J. R., Mount, E. M., & Giles, H. F. (2014). Design of

Experiments. *Extrusion*, 291–308. https://doi.org/10.1016/b978-1-4377-3481-
2.00025-9

What is a GPU (Graphics Processing Unit)? Definition from WhatIs.com, & S. Gillis,

A. (2020, December). What is a GPU (Graphics Processing Unit)? Definition

from WhatIs.com. Retrieved from SearchVirtualDesktop website:

https://searchvirtualdesktop.techtarget.com/definition/GPU-graphics-
processing-unit

Wu, Y., & Huang, D. (2019). Optimization design of axial fan blade. *Journal of the*

Chinese Institute of Engineers, *42*(6), 473–478.

https://doi.org/10.1080/02533839.2019.1611478

Zhang, L., & Jin, Y. Z. (2011). Effect of Blade Numbers on Aerodynamic Performance

and Noise of Small Axial Flow Fan. *Advanced Materials Research*, *199-200*,

796–800. https://doi.org/10.4028/www.scientific.net/amr.199-200.796

Zhang, X., Xiong, G., Shen, Z., Zhao, Y., Guo, C., & Dong, X. (2017). A GPU-based parallel slicer for 3D printing. *IEEE International Conference on Automation Science and Engineering*, *2017-Augus*, 55–60. https://doi.org/10.1109/COASE.2017.8256075

Zhou, S., Zhou, H., Yang, K., Dong, H., & Gao, Z. (2021). Research on blade design method of multi-blade centrifugal fan for building efficient ventilation based on Hicks-Henne function. *Sustainable Energy Technologies and Assessments*, *43*, 100971. https://doi.org/10.1016/j.seta.2020.100971
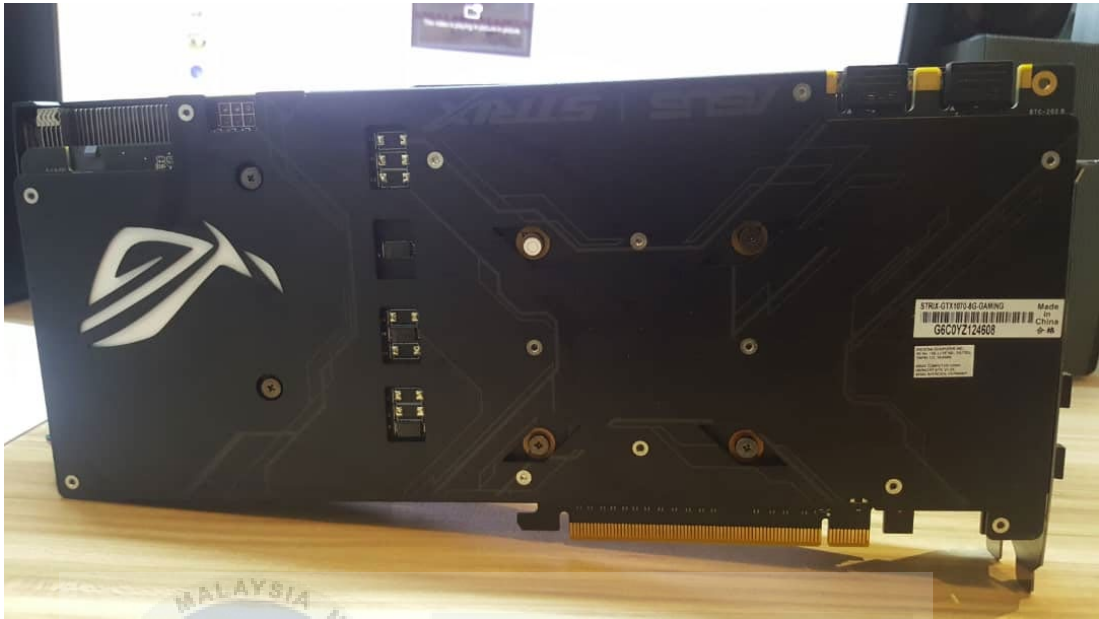
APPENDICES

Appendix A: Top view of GPU

Appendix A: Top view of GPU

Appendix B: Back view of GPU

Appendix C: GPU inside PC (View 1)

Appendix D: GPU inside PC (View 2)