# **Nvidia vs AMD: A historical market analysis and technical review of the GPU (Graphics Processing Unit)**

# 

# Nvidia vs AMD: the great GPU war

# 

# **ABSTRACT**

Throughout this project, I establish the history of GPUs; how they first came about, what different companies have done and why some haven't worked out which resulted in only two remaining competitors left. I also provide data of real companies in the charts I produced to give you a clear and realistic understanding of the market side of GPUs.

I believe that we don't realize the importance of GPUs and how it has impacted us on the things we do every day. Therefore, this project will get straight into what is truly important in today’s world. I have decided to write this project as I plan to build a computer, so completing this essay will help me gain knowledge as to which company produces better GPUs. As one of the components to build a computer is a GPU. Researching about GPUs will give a much clearer understanding of what GPUs are, and this will also benefit me as I want to pursue a career in technology.

GPUs are used to power the world's biggest supercomputer. Supercomputers are used for scientific and engineering applications which need a great amount of computation. Such as quantum mechanics, weather forecasting, and climate research. Studying the history of GPUs is a very important knowledge to have as it has allowed me to have a better sense of how they work as well as their impact on the current world. In addition, learning about the history of GPUs allows one to predict what the future of technology could hold for us. For example, the improvements in GPU performance has been so drastic that the “World’s top 300 internet services now have Nvidia GPUs in their data centers to do inference” (Jensen Huang)

There have been several articles that have focused on one particular subject such as the performance of GPUs or the current Market state of GPUs. So I am writing this project as an attempt to tie all these things together. And even better, having all the information in one essay makes it easier for the reader to access the information.

# **INTRODUCTION**

It's hard to believe that only three decades ago GPUs were a rare thing and very few people were lucky enough (or rich enough) to actually have one at home. However, GPU technology has rapidly evolved and nowadays it is less common to find a single home without a computer or a smart device that uses a GPU. GPUs have gone from being used from Games to phones and AI technology.

Computers are made up of hundreds of individual components to make them do the things you expect, such as running applications or sending data to other places. The CPU (Central processing unit) and GPU are just some of the parts that make the computer work. A CPU has a few cores optimized for sequential serial processing, it performs basic processes such as web surfing or running Office 365. On the other hand, GPUs have a massively parallel architecture consisting of thousands of smaller, more efficient cores for handling multiple tasks simultaneously [2]. It performs more intensive tasks such as gaming or video editing. GPUs implement a number of 2D and 3D graphics processing in hardware, making them much faster than a general-purpose CPU [3].

In a computer, a GPU can be present on the CPU (as known as integrated GPU) or embedded on the motherboard (as known as Discrete GPU).

GPUs are a very fundamental component in today’s day and age. Most people think of them as tools for playing games but they actually are powering the world all around us, with everything from self-driving cars to cancer research. For example, as of March 2020, Nvidia is doing their part to fight Covid-19 as they are working with Oxford Nanopore to help find a vaccine against the coronavirus. GPUs are now finding more generalized uses in computers for applications such as machine learning, oil exploration, image processing, and financial transactions [6].

# **LITERATURE REVIEW**

To help me explain what a GPU is and how it works in the introduction, I used some information from a computer science book: <https://www.amazon.co.uk/OCR-AS-Level-Computer-Science/dp/1910523054/ref=sr_1_1?keywords=a-level+computer+science+ocr&qid=1581779367&sr=8-1>)

This book is used in the A-level computer science course and it is a reliable source as it is from an exam board and is used in all schools. I also used the whitepapers and Keynotes from the companies to help me. These sources are reliable because their purpose is to educate on what the product is and what it is created by the companies.

To explain the market history of GPUs in Findings and discussions I used multiple websites and compared the information to ensure that it is correct

I found it difficult to find websites that included the Discrete market share between Nvidia and AMD as most websites include the GPU market share (Discrete and integrated GPUs). However, there were a couple of websites I used that just included the market share of Nvidia and AMD. Like I did before, I compared the websites to make sure the data I used was correct. These websites used the charts and graphs provided by Jon Peddie Research (PDR). PDR Jon Peddie Research offers the latest statistics, trends, and reports on computer graphics, gaming, content creation, design, and related technology. I find this source very useful as it is the only source I could find that showed the market share of the companies in 2009 whereas other sources began in 2015.

I found the Steam hardware survey most useful as data is collected from consumers, so the data is reliable. And the survey is optional and anonymous. Steam is a gaming platform that has peaked at a total of 18.5 million concurrent users in 2019 and with total registered users close to 100 million. The more people do the survey, the more reliable it is. Steam also runs a system check to verify that the information that they are getting from the consumers is correct. Hence why I included this in my article.

To compare the GPUs and produce the charts in Findings and discussion, I used a software called Pass mark: <https://www.videocardbenchmark.net/gpu_list.php>

Passmark is a software company that creates software utilities for performing benchmark tests on a computer system. It tests multiple components such as the GPU. Anyone can use it through a subscription to test their computers. This makes their data more varied, hence more reliable. The software then fetches the results and creates an overall benchmark score to compare other GPUs. The higher the score, the more performance the GPU has. This makes it easier to compare them without looking at specifications. especially for people who won’t understand what the specification means. Another reason this is useful is that you can see what products offer the best value of money by looking for a perfect combination of a high benchmark score paired with a low price. The score also helps people to understand the GPU’s capabilities, whether it is for everyday activities such as browsing websites, for games, or for video editing.

The rest of the data I used to compare the GPUs was collected from Wikipedia: <https://en.wikipedia.org/wiki/List_of_Nvidia_graphics_processing_units> <https://en.wikipedia.org/wiki/List_of_AMD_graphics_processing_units>

Some people say that Wikipedia is not a reliable source as it can be edited by anyone at any time. But Wikipedia cites its sources which you can confirm the information. But to make sure my data was reliable, I went on the official websites of Nvidia and AMD other websites to make sure all the information was correct: <https://www.nvidia.com/en-gb/about-nvidia/corporate-timeline> <https://ir.amd.com/investor-overview>

# **METHODOLOGY**

I found most of my research online by searching specific topics that I wanted to find. I think the advantages of using the internet are that there are so many results to each topic, meaning I had many resources to read through and discover new information. However, there were also disadvantages due to this, because some of the things that came up were irrelevant information, so I looked at many websites to decide what to use. Another disadvantage to finding the resources online is that some of the facts and statistics can be false or biased. So, I made sure that I checked with other reliable websites to validate my resources.

I found some of my other research by buying books related to how companies and GPUs work. (<https://www.amazon.co.uk/Competitive-Strategy-Dummies-Richard-Pettinger/dp/0470779306>). This book was very useful as it showed me how to structure my EPQ. For instance, the introduction, briefly explains what the book is about and how it is organized: It gave a brief summary of what each chapter is about so the reader can choose what to read. I implemented this in my introduction as I think it is a great way to interest the reader. I also used AMD and Nvidia whitepapers to help understand how a GPU works:

<https://www.nvidia.com/content/dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf>

<https://www.amd.com/system/files/documents/rdna-whitepaper.pdf>

The second reason why the book was useful in that it showed me how to compare the companies. So, in my case, it will be between AMD and Nvidia. For example, businesses have various strategies to help increase profits. One of those strategies is either focusing your business to have a cost advantage or brand advantage. Cost advantage (AMD) set out to deliver their products both cheaper and more cost-effective than the competitors. Brand advantage (Nvidia) set out to deliver their top-range products and command the highest prices in the competition.

The second book I bought gave me a better understanding of how GPUs work. (<https://www.amazon.co.uk/OCR-AS-Level-Computer-Science/dp/1910523054/ref=sr_1_1?keywords=a-level+computer+science+ocr&qid=1581779367&sr=8-1>). I’m glad that I picked this specific book as it was easy to understand as it used a simpler language compared to other ones I’ve read online.

When I was writing my project proposal about planning how I was going to research, I said that I was going to attend some lectures and keynotes from the companies, but later during my research, I found out that the AMD and Nvidia Keynotes occur three times per year. And, these conferences happen mostly in the USA. For example, the Nvidia conference (March 2019) happened in California, USA. In the end, I found all the Keynotes on YouTube. I found out that it was much better to watch the keynotes online as I could pause the videos and freely make notes instead of going to the conferences and missing out on some vital information. It had many interesting points and helped me broaden my knowledge on the topic.

After doing my first set of research near the end of 2019, I did some more in February 2020 because I was interested to see if there were any changes to GPUs. And there was. There was another conference for AMD in January.

In this video, I found that the new AMD mid-range GPUs (5600 series Graphics cards) turned out to be far more impressive as they outpaced Nvidia’s RTX 2060 and GTX 1660 ti cards. This amazing product meant that the large percentage of people who plan to buy a mid-range computer had a great choice, especially as the retail price was $279 compared to the RTX 2060, $350. This new set of information has affected my research as the new graphics cards produced by AMD might alter the GPU market share, as of 2020. This slightly changed my conclusion.

In the market history section, I compared the profits and revenues of the companies using the American currency. I used this because it is an international currency. As of the first quarter of 2019, American currency makes up 61% of all known central bank [foreign exchange reserves](https://www.thebalance.com/foreign-exchange-reserves-3306258). So, I thought it would benefit many researchers who are reading my article as they can compare the prices easier. I have seen this feature on other websites during my research, such as TechRadar:<https://www.techradar.com/uk/news/computing-components/graphics-cards/amd-vs-nvidia-who-makes-the-best-graphics-cards-699480>.

In findings and discussions, I used visual charts such as bar and line graphs as it would be easier to summarize large data sets in a visual form such as GPU benchmark score. Otherwise, it would make it difficult for the reader to understand the data. Furthermore, a graph summarizes how one quantity changes if another quantity that is related to it also changes, and in this dissertation, I will be comparing factors, such as performance between the GPUs of AMD and Nvidia so it would be very useful to use graphs.

In order to make my charts, I collected data from various websites and put them into a table in Excel. I used specific information for my charts and I will explain why I did this

Nvidia and AMD have produced hundreds of GPUs since the industry began. Because there were so many GPUs, I couldn’t use all of them for my charts as it would be hard to process all the data visually. I fixed this problem by picking out specific GPUs

GPUs are split into two sections, consumer and professional GPUs. Consumer cards are mainly aimed at-home use as they are built for everyday activities and gaming, whereas professional cards are for business use as they are built to run professional graphics software applications. They are also more expensive. Professional GPUs are not built for home use, so the average consumer doesn’t buy them, therefore I have used consumer GPUs for my charts.

However, there are still hundreds of consumer Graphics cards that were produced. To make my set of data even smaller, I found out the best Graphics card Nvidia and AMD have made every generation. I chose to do this, as every generation the architecture of the GPU has changed. So I can show how different architectures alter the performance of GPUs. I picked the correct graphics cards by finding which generation they were introduced on Wikipedia and found the best GPU made that generation by using the benchmark score on a website called Passmark. I ended up with 37 GPUs that had the highest score every generation from both companies. This has helped to make more charts more readable.

Other data I used for my charts: Manufacturing process node (MPN), the Release date of GPUs, and Generation it was produced. I got this from the same website on Wikipedia.

I wrote all of my data into a table in excel and produced my own graphs. For my first chart, I did MPN against Benchmarks score. I did this to show how the size of MPN affects the performance of GPUs.

The second chart is the Release year of the GPU against the benchmark score. From the data I collected, I found the average score of the GPUs produced each year and used this data against the release year. I did this to make it easier to compare the company's Graphics cards each year. The chart shows how performance has improved dramatically over the years.

The Final chart I made was comparing the GPU in the same generation of both companies against their benchmark score. Similar to the 2nd chart, I found the average score of the GPUs produced in the same generation. This will help the reader to understand which company was better for every generation.

# **FINDINGS**

Market history of GPUs

The 1970s

As early as 1951, MIT (A private research university in Massachusetts, USA) built a flight simulator for the Navy. Although it may be considered the first 3D graphics system, the base of today’s GPUs was formed in the mid-70s. [10].

In 1976, companies such as Motorola, RCA, and LSI began to produce video clips with highly advanced capabilities. RCA, In 1976 built the first “PIXIE” video chip, which was able to output a video signal at 62x128 resolution.

This video chip was quickly followed a year later with the release of the Atari 2600 game which included a Television Interface Adapter (TIA) [11].

In 1979, the Namco Galaxian arcade system took it to the next level. Its graphics supported RGB color and custom video hardware. This arcade is now known as the “golden age of arcade video games” [12].

The 1980s

In 1981, Motorola produced an MC6845 video address generator which became the basis for IBM (a computer hardware company), and they started using monochrome and a color display adapter in their computers [10].

Shortly after this, in 1983, Intel released the ISBX 275 video graphics controller. It was able to display eight colors at a resolution of 256x256 or monochrome at 512x512 [12].

In 1985, three Hong Kong immigrants formed Array Technology Inc, soon renamed as ATI technologies (later known as AMD). This company would lead the market for years with its wonderful line of graphics boards and chips [12].

ATI produced a color emulation card in 1986 which was responsible for $10 million in sales in the first year

The 1990s

Overall, the early 1990s was the time when a lot of graphics hardware companies were founded, such as Nvidia, ATI, ST Micro, and Matrox, the market followed a classic competition cycle with many companies being founded [11], followed by rounds of acquisitions and dwindling numbers of competitors. Among the winners founded during this time was Nvidia (the company was founded by Jensen Huang in April 1993) [10].

In 1991, ATI combined their Wonder and Mach8 series to build a GPU which pushed ATI to a milestone of $100 million in revenue in the same year [11].

In 1995, the first 3D graphics cards were introduced. 3DFx (One of the first companies that manufactured 3D graphics cards) produced voodoo graphics cards that were launched in November 1996, they took around 85% of the discrete market [12]. Other cards that only rendered 2D graphics became obsolete overnight. In the same year, Nvidia released their first product, NV1.

Two years later, Nvidia released the Riva 128 card which was the first card to combine 2D and 3D acceleration [13]. In the same year in March, ATI launched their Rage series graphics cards which outperformed the competitors which helped ATI to make them $47.7 million in profit and sales exceeding $600 million in revenue, whereas Nvidia had $13.3 million in revenue [11]. By the end of 1997, this company had nearly 25% of the discrete market share, behind 3dfx and ATI respectfully.

In August 1998, ATI improved their Rage series cards which helped AMD to become the top graphics supplier with 27% of the market, net income of $168.4 million, and sales of $1.15 billion. On the other hand, Nvidia had $4.1 million in profit and $158.2 million in revenue[11].

The term “GPU” would not be introduced until 1999 by NVIDIA [8]. This company shaped the future of modern graphics by producing GeForce 256. Along with the ATI’s Radeon 7500, these were the first true GPUs that were available at the consumer level [7]. Nvidia defined the term GPU as “a single-chip processor with integrated transform, lighting, triangle setup/clipping, and rendering engines that is capable of processing a minimum of 10 million polygons per second.” [13].

In 1999, Nvidia partnered with Microsoft for Nvidia’s GPU for the Xbox console. This resulted in Nvidia’s profit for the year is $41 million and revenue of $374.5 million. At the End of 1999, ATI made a revenue of $1.2 billion and $160 million in profit with a 32% share of the discrete market [11]. All the other competitors, such as Matrox, that were founded in the 1990s, became more insignificant as they couldn’t keep up with the competition. For instance, the Matrox Mystique graphics card produced by Matrox that was released in 1996 produced poor 3D images compared to the voodoo1 graphics card produced by voodoo [1]

The 2000s

In August 2000, ATI introduced their Radeon DDR card and Nvidia countered with the Geforce 2 GTS in September. ATI’s card beat Nvidia’s by 10-20% in benchmarks which helped ATI to maintain their first position in the discrete market [11]. At the end of 2000, Nvidia had revenue of $735.3 million.

In 2001, Nvidia released the GeForce 3 series which gave programmers the ability to program the pipelines [4] of the GPUs which has never been done before. They also upgraded the GPUs in the Xbox consoles which increased Nvidia’s market to 31% to ATI’s 17% [11]. This was because of the popularity of gaming consoles getting larger. This resulted in ATI's sales dropping to $1.04 billion as the company recorded a net loss of $54.2 million[11].

A year later, the first programmable GPUs hit the market: Nvidia GeForce FX and ATI Radeon 9700. And these two companies were competing against each other for features to achieve ever-increasing levels of performance and image quality. [7]. In the same year, Nvidia and ATI battled it out again, but in the end, ATI’s 9000 series outperformed Nvidia’s 5000 series. ATI bounced back with a $35.2 million profit in 2003 after losing $47.5 million in 2002 [11].

The battle between Nvidia and ATI lasted from 2000 to 2006. At this time 3dfx and ST Micro became increasingly irrelevant as it’s cards couldn’t keep up with Nvidia and ATI’s cards [10]. During this time, ATI still outcompeted as profit peaked at $204.8 million at the end of 2005 from nearly $2 billion in revenue [11]. This was the highest in the company’s history.

Between 2000 and 2006, these two companies would go head to head and delivered features that are now used today, such as shading, volumetric explosion, refraction, waves, vertex blending, shadow volumes, bump mapping, elevation mapping, and rasterization [10]

In 2006, ATI was acquired by AMD [11].

AMD cards were outperforming those of Nvidias but not until November 2007 when Nvidia produced the 8800 GTX card. This card demolished every single and dual-graphics competitor in outright performance. Despite that success, the company dropped three percent in discrete graphics market share [11].

In October 2008, Nvidia improved the 8800 series card which still beat its competition. This helped to push Nvidia to a 71% market share by year's end [11].

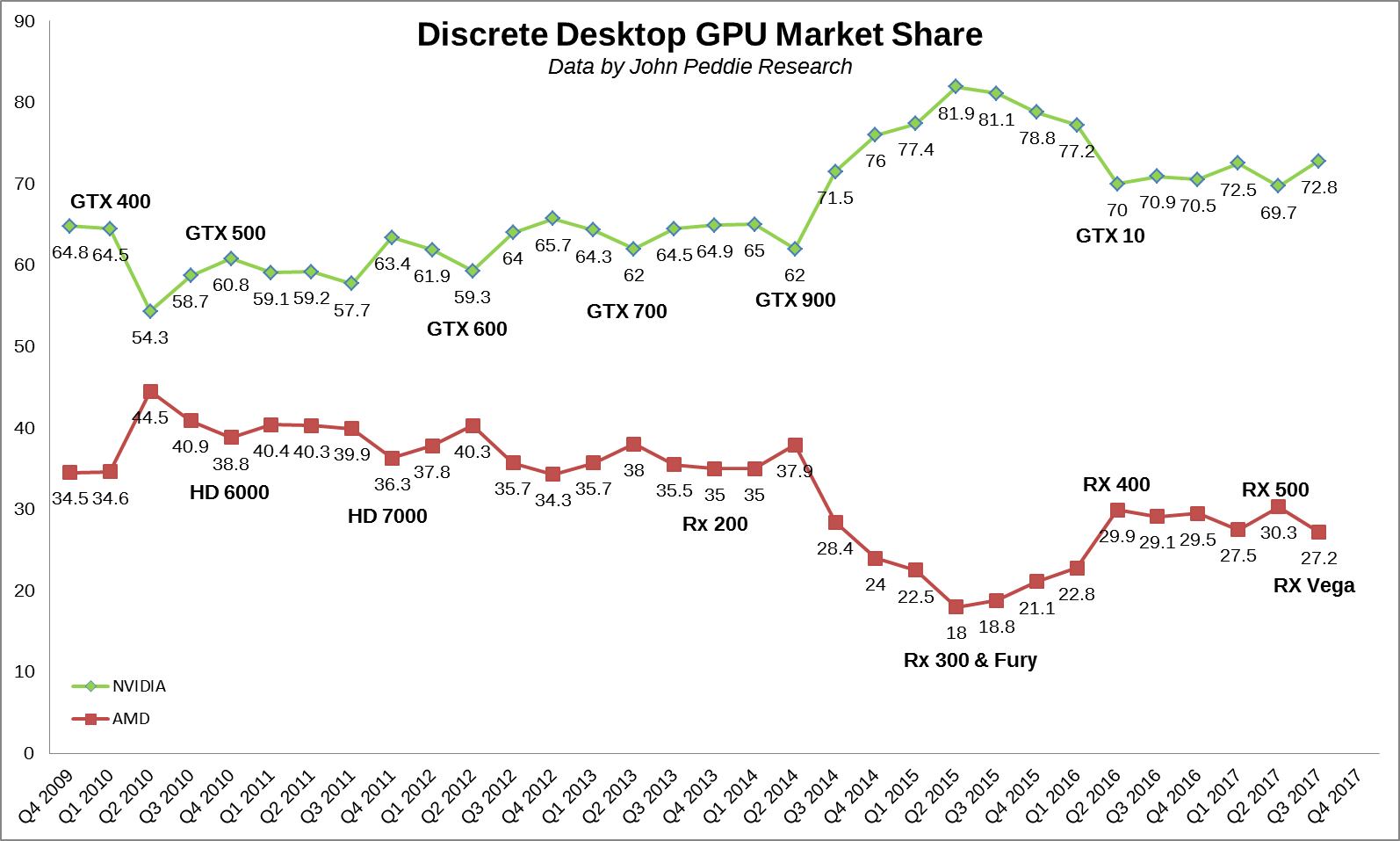
The 2010s

In 2010, Nvidia collaborated with many companies such as Audi. They helped to power the cars’ dashboards and increase navigation and entertainment systems [5].

In 2013, Nvidia produced the GeForce GTX Titan which was released in February. As of 2013, it was the fastest single GPU on the planet [13]. Nvidia later produced the GeForce GTX Titan in December 2018 which has now become the forerunner of graphics technology. Also, in 2013, Nvidia released their G-sync technology along with AMD’s free-sync technology that was released in 2015. The result of these features was that they helped to eliminate stutter and screen tearing which benefited consumers and gamers. [12].

In 2016, AMD released their GPUs that had a robust increase in performance per watt, but as of today, GPUs are not only used for graphics. They have found their way into machine learning, oil exploration, scientific image processing, statistics, linear algebra, 3D reconstruction, medical research [6].

Since 2009, the market share between Nvidia and AMD has been the same. With Nvidia holding still holding 70% of the market share to AMD’s 30%.

****

**Figure 4: Discrete market share from 2009 to 2017**

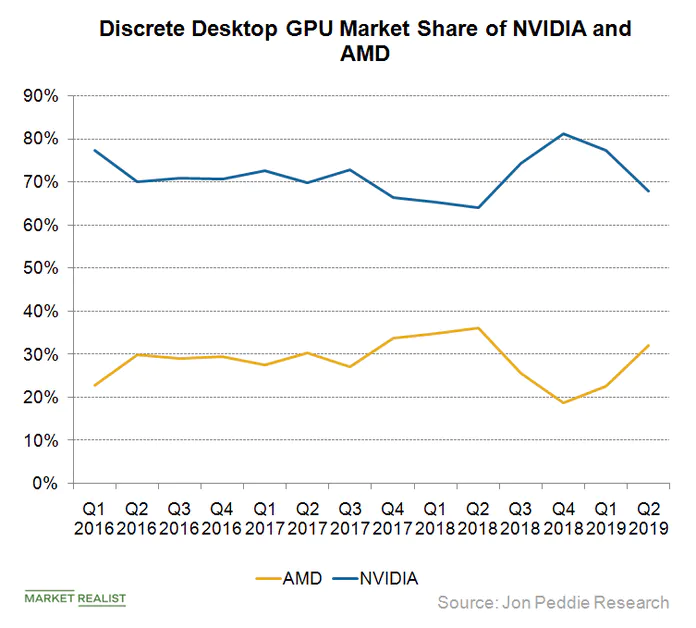
Figure 4 shows how much Nvidia holds in the GPU market share, with Nvidia’s lowest being in Q1 2010 at 54% and their highest being in Q2 2015 with 82%.

However, AMD had its best years from Q2 2010 to Q2 2012. With their HD family cards having an advantage over Nvidia's GTX series at that time [14]. The demand for cryptocurrency also resulted in AMD’s share increasing.

But in 2013, AMD’s market share started to slip badly after the first cryptocurrency boom. The boom meant that the prices of GPUs skyrocketed because of the demand of cryptocurrency. This meant that consumers couldn’t afford the high prices which made AMD lose half its graphics market in less than two years [14].

AMD’s market share increased in 2016 and 2017 due to crypto-related demand and their new architecture of GPUs called Polaris [15].

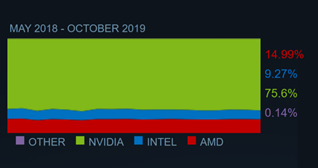
.

****

**Figure 5: Discrete market share from 2016 to 2018**

In Q1 2018, AMD hit a four-year record for market share, at 14.9 percent. But with its current market share in Q2 2018 set at 13 percent according to JPR, it’s clear those gains were temporary

In Q1 2018, AMD hit their highest market share since 2014 with 38% compared to Nvidia’s 62%. But this only lasted temporarily as AMD began to lose their share again in the next few quarters [14]. Nvidia gained this share because of their RTX titan released in December 2018. The raw performance of this card is why Nvidia gained back the share in 2018 as seen in figure 5 [13].



**Figure 6: Market share of GPUs between MAY 2018 – October 2019**

In figure 6, Nvidia GPUs contribute to a 75.6% market share, while AMD GPUs contribute to a 15% market share.

In Q4 2019, Nvidia launched their mid-range GPU, the GTX 16 super series. It was a smart play by Nvidia to release their GPUs before AMD could as they officially launched their RX 5500 series in mid-December. This meant that Nvidia increased its market share more than AMD as it had early sale advantages. This trend also be seen on the Steam Hardware survey

**Figure 7: GPU usage of a consumer survey, January 2020**

Steam is an online gaming platform where you can play, discuss and create games. Steam conducts a monthly survey, called Steam Hardware survey, which collects data about what kinds of computer hardware and software their customers are using. And the participation is optional and anonymous. In this case, we will be looking at GPU consumer use. This survey shows the percentage of people on steam using different types of GPUs, with the most consumed at the top.

In figure 7, in September 2019, the GTX 1060 and GTX 1050 ti reign supreme with just under a quarter of the consumption of GPUs. This is because many gamers can buy them at cheap prices in the used market. That’s why Nvidia’s share has seen a consistent rise. Nvidia takes the first 10 spots which just shows how popular their cards are, which further emphasizes why Nvidia reigns in the market share. The most consumed AMD card is the Radeon RX 580, which takes up the 11th spot. This highlights the 15% market share AMD has compared to the 75% Nvidia has.

In the next few pages, I will be comparing the performance of GPUs between Nvidia and AMD using a series of charts and graphs.

Benchmark is a test that helps to compare similar products, from mobile devices to computers. In this case, I will be comparing the GPUs produced by Nvidia and AMD. Each test on the GPU produces a score called benchmark score. This metric will be used to compare the important features such as performance. The higher the score, the higher the performance. I have obtained my data on Pass mark which has the scores of all the GPUs

The first two graphs will be comparing the Manufacturing process node (MPN) against the benchmark score of both companies

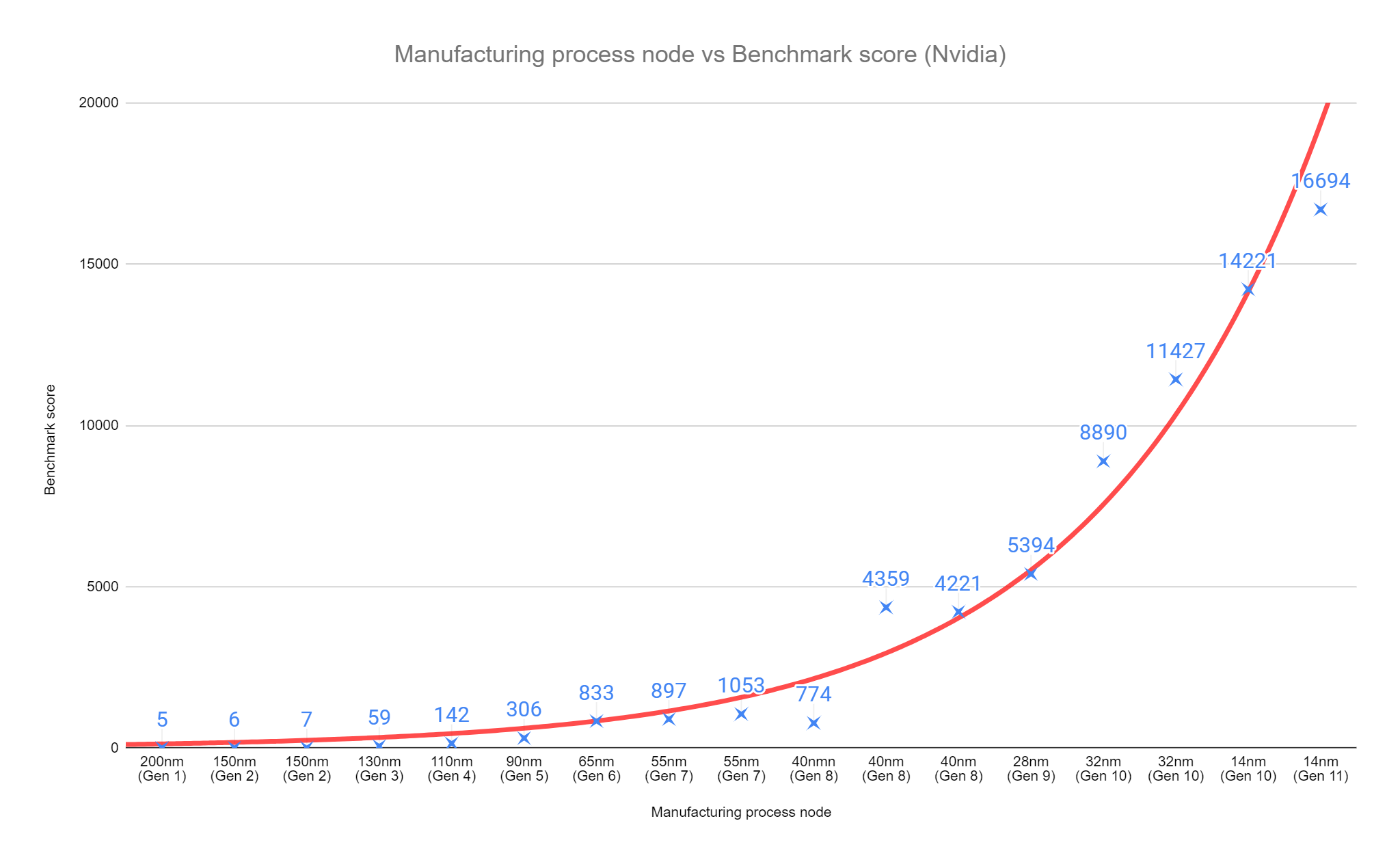


Figure 8: A chart showing MPN against Benchmark score of GPUs every generation (Nvidia)

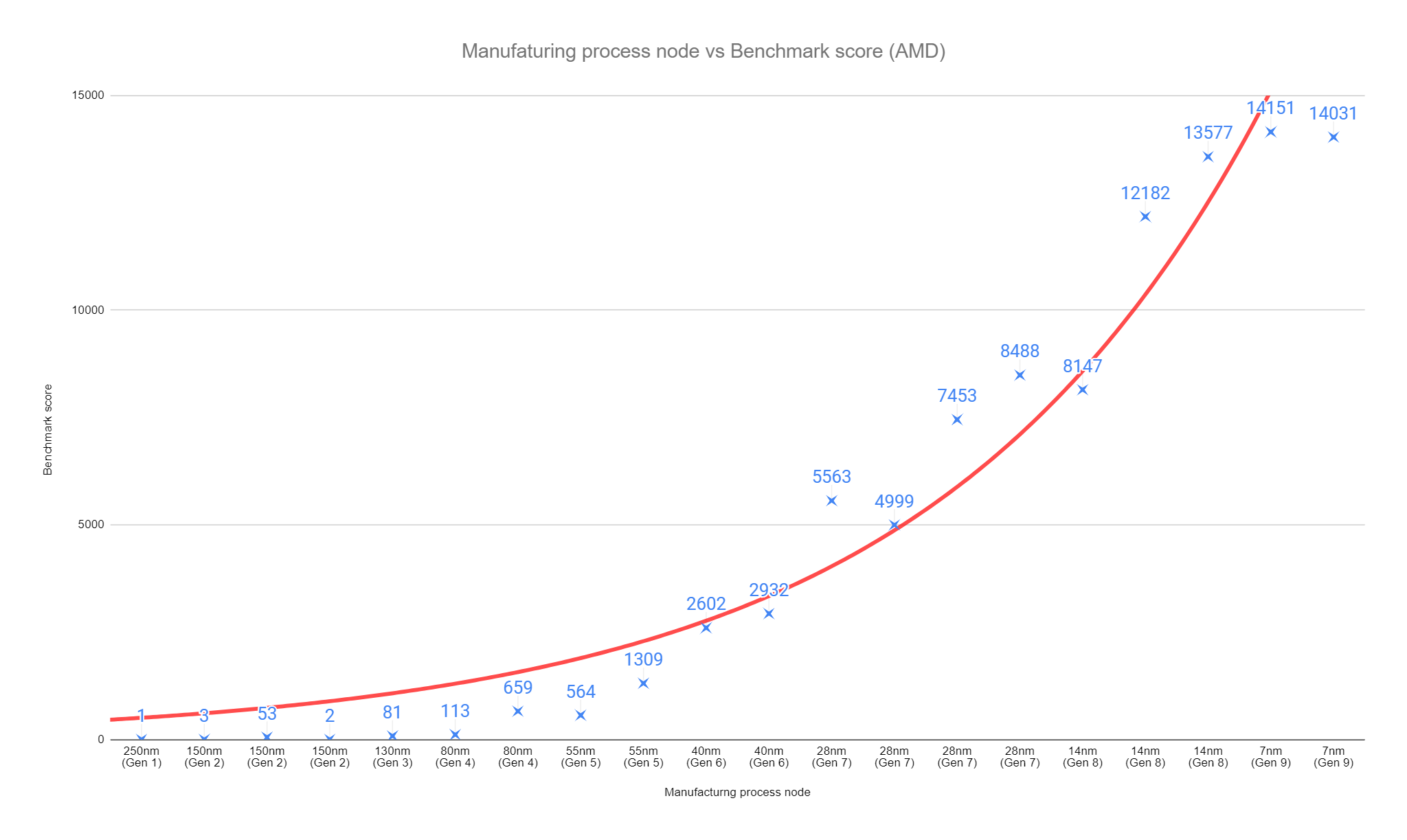


Figure 9: A chart showing MPN against Benchmark score of GPUs every generation (AMD)

Figures 8 and 9 show that the benchmark score of GPUs has increased exponentially over the generations. The trend line shows this. This is because the size of the MPN has decreased. The smaller the MPN, the higher the benchmark score which means higher performance.

Since the first generation, AMD has always been ahead of Nvidia in terms of the MPN. For instance, during the 7th generation, AMD was on 28nm while Nvidia was on 55nm. The benchmark score also provides evidence for this that AMD was ahead. For example, during the same generation, In figure 9, AMD’s latest GPU has a score of 8488, while in figure 8 Nvidia’s latest GPU has a score of 1053.

But since the last two generations, Nvidia’s GPUs have had a higher score than AMDs. Even though AMD is on 7nm their latest GPU with 14031, is lower than Nvidia’s latest with 16694 and they are on 14nm. You would think that AMD’s GPUs will have more performance since they have a smaller MPN but in this case, they aren’t. This is because Nvidia spent its R&D on optimizing the 14nm architecture instead of developing to a smaller MPN.

Optimization is key in higher performance and I’m going to show how optimization affects it.

In figure 9, during the 7th generation, AMD initially got a score on their GPU of 5563, and their second GPU got 4999. The newer GPU got a lower score as AMD didn’t optimize their graphics card properly, hence the lower score. This pattern can also be seen during the 8th generation in figure 1 where the first GPU got 4359 and the second got 4221.

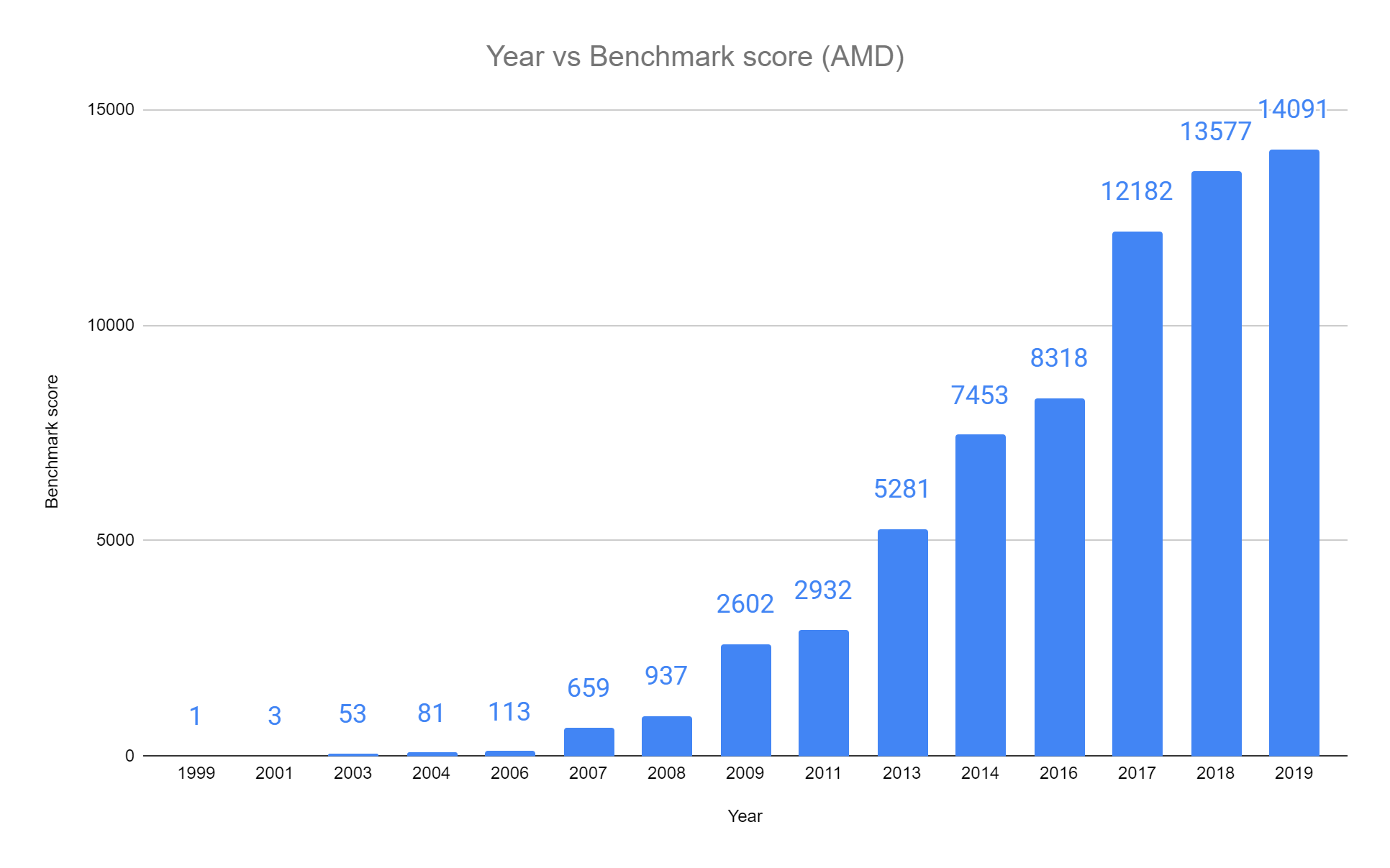
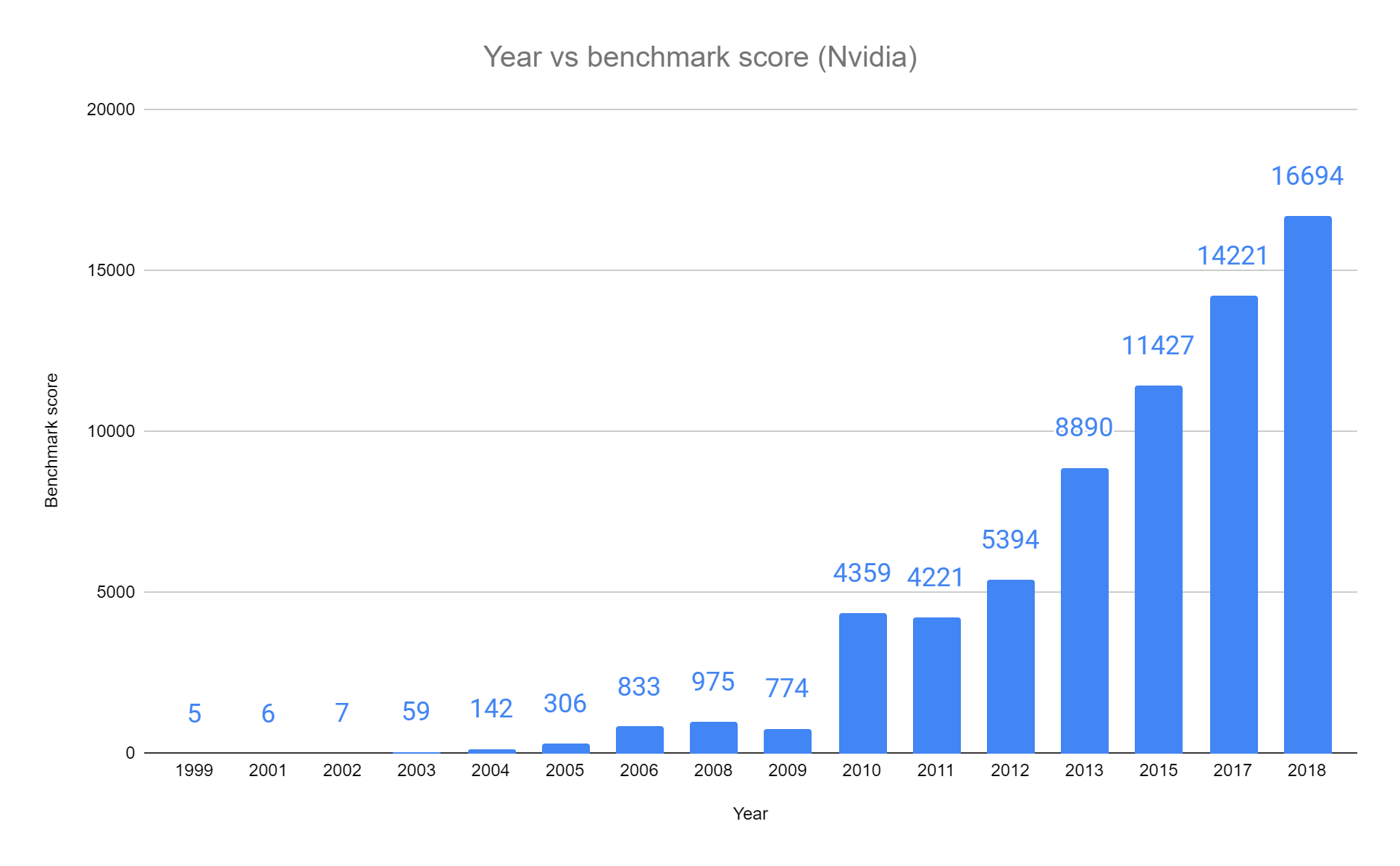
Figure 10: A chart showing Year against Benchmark score of GPUs every year (Nvidia)

Figure 11: A chart showing Year against Benchmark score of GPUs every year (AMD)

Figures 10 and 11 show the increase in scores over the years. Between 1999 and 2008, both companies produced GPUs that had very similar scores against their counterparts in the same year. But in 2009 onwards, in figure 11, AMD began to pick up the pace. In 2009, AMD’s average benchmark score for all their GPUs was 2602, whereas Nvidia’s in figure 10 was at 774. This trend carries on until 2012.

In 2013, Nvidia's average score was 3609 more than AMD’s. From 2013 onwards, Nvidia began to take over AMD in terms of score and this pattern has stayed the same ever since.

2017 onwards, AMD has been slowly catching up to Nvidia, with both the companies’ latest Graphics cards in 2019 having a difference of 2603 compared to 3609 in 2013.

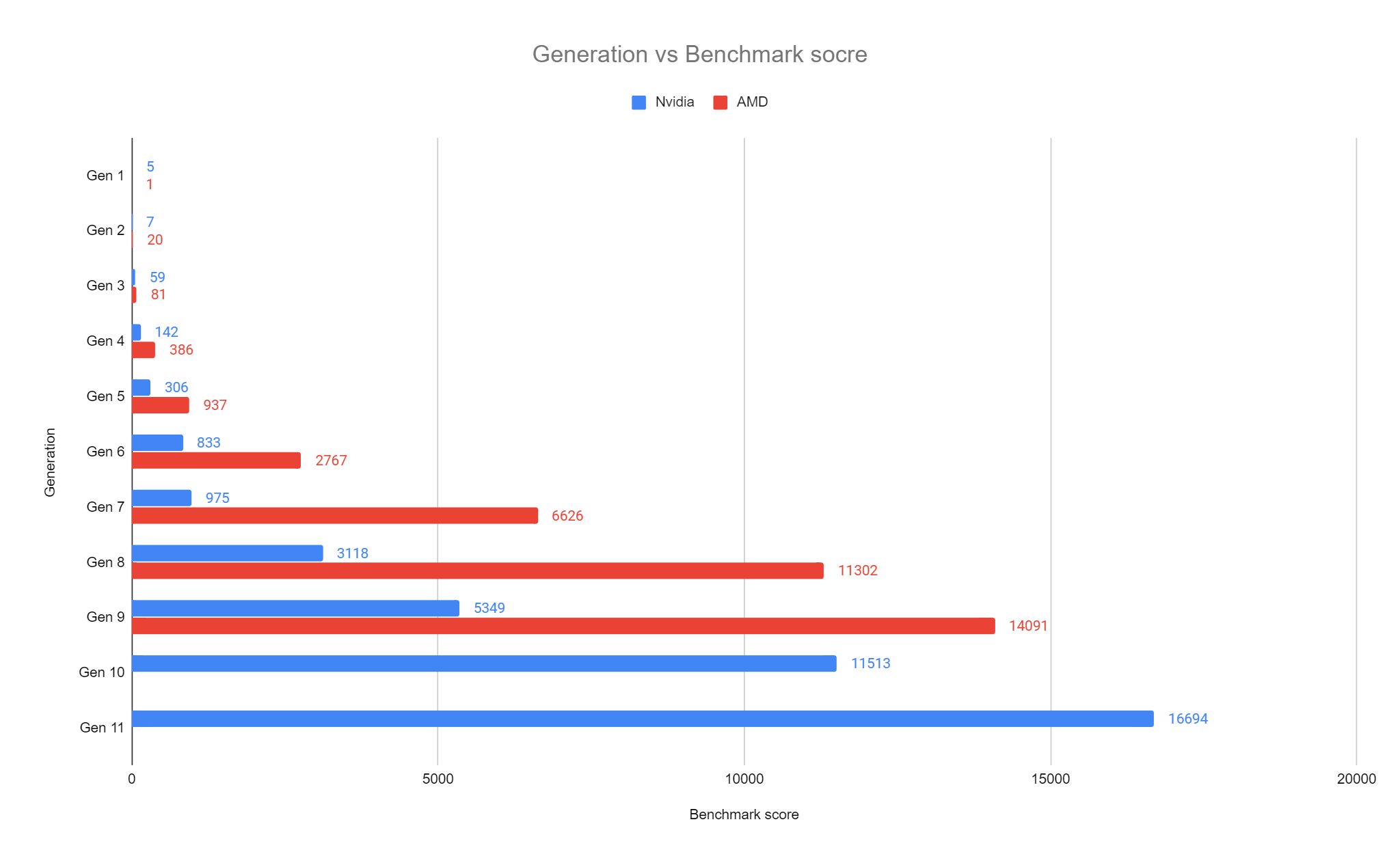


Figure 12: A chart showing Generation against Average Benchmark score of GPUs

In figure 12, Both companies started at a similar pace with AMD getting slightly ahead in performance in the first three generations. Between the 4th and 9th generations, AMD began to pull form Nvidia massively with AMD having nearly 3x more performance than Nvidia's counterpart.

However, Nvidia improved their architecture dramatically over the next two generations with their latest generation of GPUs of 16694 average scores compared to AMD’s latest with 14091. This shows that Nvidia’s architecture is far better than AMD’s.

# **DISCUSSIONS**

The race between AMD and Nvidia

In the late 1980s, AMD was the first to enter the GPU industry before Nvidia was founded in the early 1990s. AMD had an early start which made me believe that they had more time to produce better GPUs as they had more time to R&D. AMD already had sales of $100 million in 1991 compared to Nvidia which was founded in 1993. We can also see AMD being ahead in Figures 8 and 9 when each generation, AMD had a better MPN which led to better performance of GPUs. Because of this, I think that AMD was already winning the race against Nvidia.

In my opinion, Nvidia began to close in on AMD during the late 1990s. Their Graphics cards were improving at a much faster rate than AMD’s and began partnering up with big tech companies such as Microsoft. However. AMD still remained ahead in the race as they had sales of $600 million compared to Nvidia’s $13.3 million (in 1997), mainly because they were ahead in R&D. AMD’s market share was increasing, from 27% in 1998 to 32% in 1999. This had led me to believe that AMD was still winning the race

In the early 2000s, Gaming was becoming more popular which helped Nvidia partner with gaming companies like Microsoft (Xbox). From my understanding of the gaming industry, this helped Nvidia dramatically as they had a 31% market share compared to AMD’s 17%. AMD’s GPUs however between 2000 and 2006 still outcompeted Nvidia’s which helped to increase their revenue to $2 billion in 2005 from $1 billion in 2001. Despite Nvidia having a bigger market share, AMD was far ahead in sales and profits. So I still think AMD was ahead during this time.

This was the case until the second half of the 20s decade. During this time Nvidia had enough time to catch up with AMD in terms of R&D, such as architecture and MPN. These improvements can be seen in Figures 8 and 12. Nvidia produced the 8800 GTX card in 2007 which outperformed AMD’s cards which have never been the case before. This card helped Nvidia to become ahead in the battle. From my point of view, I think that this accomplishment and along with other partnerships with companies such as Audi helped them to win the battle against AMD. In 2008, the card and the partnerships helped to push Nvidia to 71% market share. In Figures 10 and 11, we begin to see Nvidia takeover AMD in 2010 and onwards.

Since 2008, Nvidia has held around the same market share with their lowest and highest being 54% and 82% respectively. Whereas AMD’s lowest and highest were 12% and 45% respectively. This can be seen in figure 4. The reflection of Nvidia’s market share can also be seen in Figure 7, the consumer survey. Where the top 10 consumer GPUs are all Nvidia as of January 2020

# **CONCLUSION**

In conclusion, I think that AMD had the lead over Nvidia since the GPU industry began and they kept this lead for nearly 3 decades until 2008. I think this is because AMD was founded at the start of the GPU industry whereas Nvidia joined a few years later. Furthermore, they held a big percentage of the market share, and their revenue and profits were much more than Nvidia. Between 2005 and 2008, Nvidia began to take more of the market share as they began partnerships with other companies and their GPUs saw an exponential increase in performance. Such as the 8800 GTX card. In 2008, Nvidia had a very large percentage of the discrete GPU market share, 71% compared to AMD’s 28%. Since 2008, the market share has been relatively the same. From this, I can say that Nvidia is winning the race.

Because the market share has been relatively the same since 2008, I can say that it will stay the same for the next few years. In figure 10 and 11, I mention how the difference in benchmark score of the competing GPUs every year have been getting closer. This shows that there will still be intense competition as both of the companies are keeping up with each other, unlike the other competitors who became insignificant.

However, with Intel planning to join the industry in late 2020, it will make a significant impact as they already contribute to 75% of the total graphics market (Including integrated and discrete cards). Intel already has the R&D to make GPUs as they have produced them before (i740 in 1998). If Intel follows through with its discrete GPU plans, it could mean more competition in the GPU space, which is always better for those who use GPUs. It could lead to faster innovation, as well as more competitive pricing when it comes to graphics cards.

People say that using the performance of the GPUs doesn’t show who is ahead in the race as the companies produce cards for different audiences. AMD is a cost leadership company which means that they deliver products that are more cost-effective, and therefore produce mid-range GPUs. Whereas Nvidia is a brand leadership company that delivers the best products but demands the highest prices, so, therefore, produces high-range GPUs. Hence why we see in figure 12, Nvidia recently has had higher benchmark scores than AMD. So using the performance of GPUs to compare the companies won’t be as useful.

However, AMD has decided to aim for mid-range GPUs as most customers want cheaper GPUs for the same performance. This will help AMD to get a bigger market share which will help them to get closer to Nvidia in the race.

I also took into account the percentage of market share the companies hold and the profits/revenues they have made. I took all of this into account when comparing the companies to help me decide who's ahead.

# **Evaluation**

I think that all the research I found was very useful in one way or another, whether I just used it to improve my understanding of the topic or if I was going to use it as a focal point in this dissertation. However, if I were to do this again or had more time and scientific help to go deeper into the project, I would have really liked to carry out some more of my own primary research. I don’t think that I would be able to do that on my own, but I could have asked to do one with someone who specializes in computing. For example, with someone’s help, I would build my own computer or use someone else's, so I can benchmark GPUs. I would make my test similar to the one I used for this project so the results that are produced are fair and reliable. But this would be expensive and time-consuming. Perhaps I could interview people on their views about the battle between Nvidia and AMD to see who they think is winning the race.

If I had more words for the dissertation, I think that I would’ve written more points to discuss. I would have researched other specific areas such as the new entrant, Intel, joining the GPU industry. I would have gone into more detail of the current state they are in now and what the impact will be if they do join the industry. Another area I would have researched on is explaining why there aren’t any other companies such as Samsung or ARM holdings entering the GPU industry. For instance, Nvidia and AMD have many years of expensive research and development, hence why there haven’t been any new entrants. Also, near the end of my market history, part of my Findings and discussions, I write about how the companies produce features such as “shading, volumetric explosion…” If I was writing this as my job, I would go into full detail in explaining what the features do and how they work together to produce the graphics we use today.

Most of the research I enjoyed typing up because the facts were interesting, but I feel that I could have used different sets of data for my primary research. One way to make this better is to collect the data of retail prices of the GPUs and to standardize my data all in one go. For instance, Price over benchmark score to show if it's cheaper now to get a graphics card that is really good. This new set of metrics/data can help people who want to buy GPUs to save money while not cutting performance.

Furthermore, I would have looked for better counterarguments. That is one of the things that I don’t think I perform best in because it was very hard to disprove that Nvidia produced better GPUs than AMD.

In my opinion, I think that I performed best in the findings and discussions part to my dissertation. I got to discuss research I had found and since I am really interested in this subject, I found that part quite easy to write. Particularly the information I found as to why Nvidia holds such a big percentage of the market share compared to AMD. I enjoyed writing all the findings and discussions because I got to share my opinion on the research that I had written up.

I found writing this dissertation very useful for me because it helped me learn how to do a long piece of writing and it improved my writing skills. It has helped prepare me for when I will need to write longer essays or dissertations in the future. It has also broadened my knowledge on GPUs, which was a topic I found intriguing and now I know many facts about it.

# 

# **Bibliography**

[1] Dutches, Texture Mappers (1995-1999),

<https://www.hattix.co.uk/hardware/index.php?page=video1.html>

[2] Christopher Barnatt, A guide to computing, last updated 8 may 2018

<https://www.explainingcomputers.com/hardware.html>

[3] Peter Bjorn Perlso, Graphics hardware history, September 1999

<http://titancity.com/articles/gfxcards.html>

[4] Crow Thomas. Evolution of the Graphical processing unit, 2004

<https://www.igi-global.com/chapter/history-and-evolution-of-gpu-architecture/139841>

[5] Nvidia, Nvidia history

<https://www.nvidia.com/en-gb/about-nvidia/corporate-timeline/>

[6] Nvidia, Research AI playground

<https://www.nvidia.com/en-us/research/>

[7] Venkat, Understanding the graphics pipeline, 2005

<https://www.cis.upenn.edu/~suvenkat/700/lectures/2/Lecture2.ppt>

[8] Dominik Goddeke, GPU computing with CUDA, 2009

<http://www.mathematik.uni-dortmund.de/~goeddeke/gpgpu/cuda-2009/1-History.pdf>

[10] Alex Woodie, The modern GPU: A graphic history, 2013

<https://www.hpcwire.com/2013/08/21/the_modern_gpu_a_graphic_history/>

[11] Graham Singer, The history of the modern graphics processor, 2019

<https://www.techspot.com/article/650-history-of-the-gpu/>

[12] The history of gaming: the history of GPUs

<https://community.shadow.tech/gben/blog/gaming/history-of-gaming-gpus>

[13] History of GPUs, 2019

<https://xoticpc.com/blogs/news/history-of-gpus>

[14] Joel Hruska, Charting 9 years of GPU markets shifts between Intel, AMD and Nvidia, September , 5, 2018

<https://www.extremetech.com/gaming/276425-charting-9-years-of-gpu-market-shifts-between-intel-amd-and-nvidia>

[15] Puja Tayal, AMD Gains Significant GPU market share from Nvidia, August 20, 2019

<https://marketrealist.com/2019/08/amd-gains-significant-gpu-market-share-from-nvidia/>