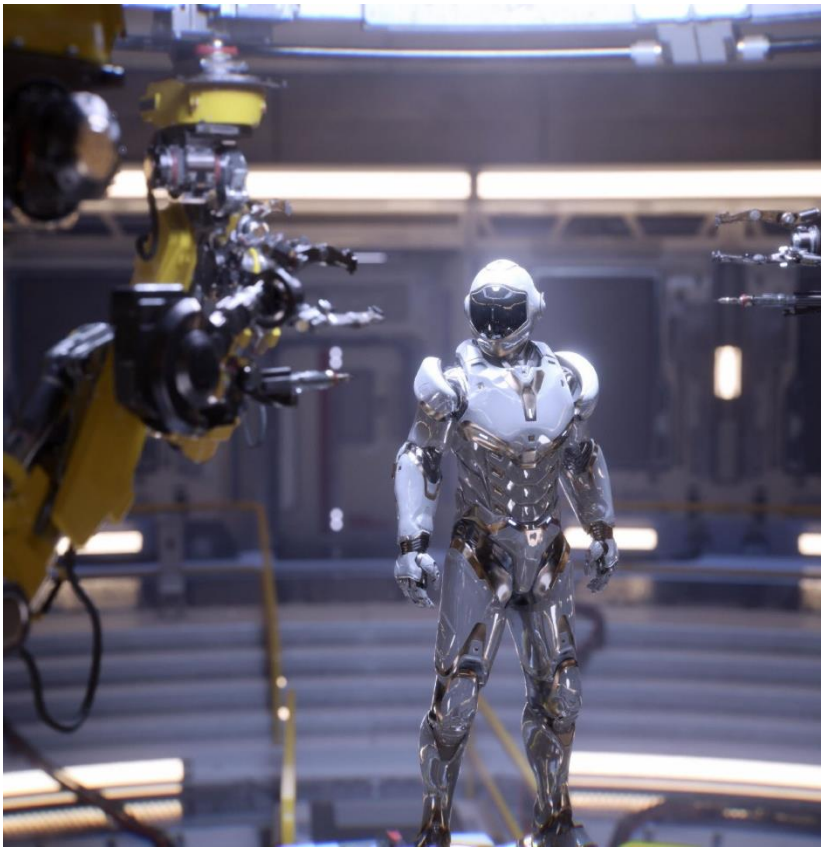

Source: https://nvidianews.nvidia.com/news/nvidia-unveils-quadro-rtx-worlds-first-ray-tracing-gpu

Drivers of The Modern Video Card

Joe Linogao

Introduction, “RTX-Off, RTX-On”

On August 13th, 2018, Nvidia announced the world’s first ray tracing General Processing Unit (GPU) the “Quadro RTX” with the new “Turing architecture.” [1] The card would allow artists, designers and other content creators to harness the power of real-time ray tracing, something thought to be unachievable only a decade ago. Ray tracing allows for the accurate simulation of light rays as oppose to using artificial light interactions. Simply put, this development in the video card industry allows creators (especially VFX editors), to see what they are creating, without having to wait however many hours to render a piece. Video cards have always aimed to accelerate graphical computations, and Nvidia’s new card and architecture advance this aim exponentially.

There are two primary aims which drive graphic/video card development: how much data can be offloaded onto the card, and how quickly can this be achieved? Over the last two decades, two companies, Nvidia and AMD, have dominated the market with their solutions to these questions.

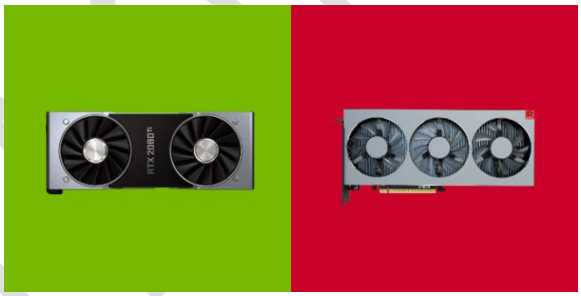

Figure 1: Nvidia RTX 2080 Ti and AMD Radeon VII

However, as COVID-19 increases isolationism, social distancing, and lack of real-time interaction, people are looking towards desktop computers as a means of escape. [2] Whether it is used to play games, create funny videos, or learn how to use 3D art programs, computers require some form of graphical computation. As such, new technological drivers for the video card market have developed, even before the times we are in.

What is the Purpose of a Video Card? Why Does it Exist?

A video card (GPU) is essentially a small computer inside a computer. The card is optimized for accelerating any computation that involves graphical data. [3] Having a component that is heavily engineered to perform a specific set of tasks puts less strain on the other parts of the computer. One way to visualize this concept is by imagining a chef making a cake with an automatic mixer. Here, the mixer is the GPU and the chef, the rest of the computer. The chef offloads ingredients into the mixer and allows it to automatically mix them together. This allows the chef to perform other tasks without worrying about mixing ingredients giving a sense of decentralized command.

IBM created the first prototype, the Professional Graphics Controller in 1984. The aim was to free up the central processing unit (CPU) by handling most of the video processing tasks. [3] Although the $5500 price, and lack of compatibility didn't receive much market appeal, it was the foundation needed for the GPUs we have today.

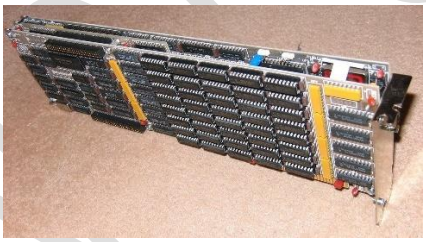

Figure 2: IBM's Prototype GPU

Today, we have GPUs that outweigh the processing power of technology from a decade ago. Manufacturing advancements drastically lowered the price of GPUs, which increased their production and use by consumers. [4] Much like most products, there are many companies that control the market. In the case of GPUs, Nvidia and AMD have created numerous "families" of graphics cards, with each year iterating from the last. But with the technology for offloading graphical data already in place, what else drives their development of GPUs?

Competition & Affordability; The Modern-Day Drivers

Competition

One driver of GPU technology mentioned indirectly is market competition. The competitive nature between Nvidia and AMD forces both sides to innovate each iteration of their cards. Moreover, Nvidia's developments in ray tracing and deep learning architecture puts pressure on AMD to keep up with them. Although both families of GPUs have their pros and cons the consumers ultimately decide the better product. Having healthy competition means that users have a fresh line of new GPUs every year, with every iteration topping the last.

Let us take the most recent line of Nvidia GPUs as an example. Nvidia are currently on their "3000" series line, with massive improvements to graphical computation and ray tracing technology. Ray tracing has allowed artists to have the ability to visualize how light rays interact with their projects without sacrificing performance. However, artists are not the only ones who benefit from this technology. Gamers now have the joy of playing computer games at a cinematic quality with the push of a button. After fixing many of the problems relating to in-game performance [5], the world of ray tracing has never been more prominent in desktop computers.

Figure 3: Ray tracing on vs. off

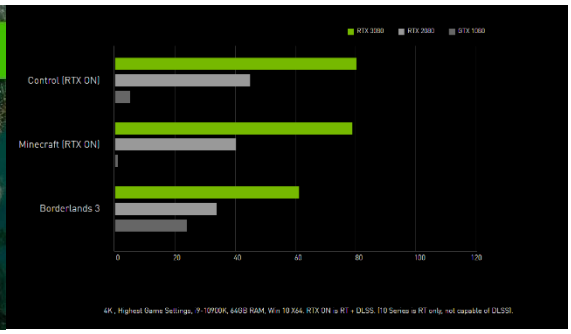

Figure 4: Nvidia RTX 3080 (Green) Performance vs. 2080 (Light Grey) and 1080 (Dark Grey)

AMD responded by creating a more powerful card, the RX 6900XT. When put against Nvidia's top tier cards, AMD manages to pull slightly ahead when it comes to gaming performance, as you can see in the image below.

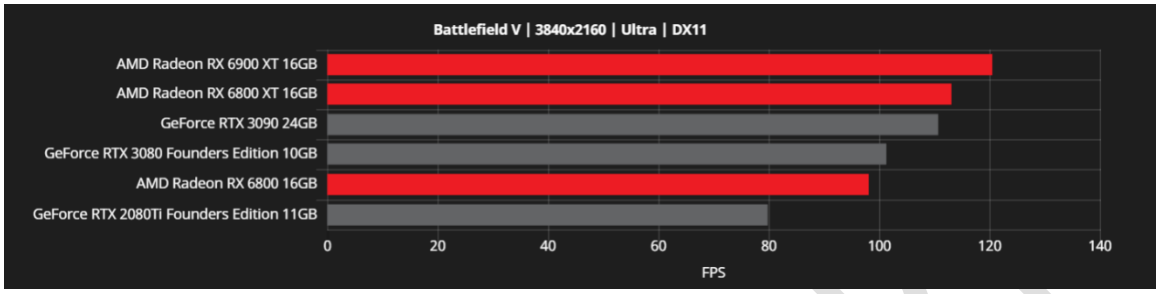

Figure 5: AMD (Radeon) vs. Nvidia (GeForce, Ray Tracing Off) Cards in Battlefield V

Although their ray tracing technology itself may not be as developed as Nvidia's, AMD have created GPUs that make people have to consider their options. In other words, consumers have a choice of either the best raw performance or, the best cinematic experience at the cost of performance. This choice hinges solely on the user's day-to-day life. The consumer could be a VFX artist or a gamer (or even both!). In the future, perhaps either company develops a card capable of ray tracing without having to sacrifice on overall performance.

Affordability

However, the main downside of the cards that I have mentioned is their price, which can range from $600 all the way to $1400. Although it is affordable in the eyes of a production studio, it is a heavy expense for many independent creators.

Luckily, there is a solution for those with a lower budget who want a GPU. Alongside the development of cards that focus on power, Nvidia and AMD both develop cards that will satisfy the needs of most people without breaking the bank. Since there is a demand for cheaper alternatives, companies must create cards that outperform others at certain price ranges. Therefore, the technological drivers change depending on the price range. At lower price ranges, the driver goes from who can create the most powerful card, to who can produce the cheapest card with the least compromises.

On Nvidia's side, they have the GTX 1660 Super, which is also the card I personally use. The card is currently priced at around $250 [6] and sometimes even lower than that. While it may not have any of the fancy ray tracing technology, it does offer a high-quality experience when playing most video games and provides enough power for independent creators. Looking at the figure below, we can see that every iteration of Nvidia's cheap alternatives improves from the last every year.

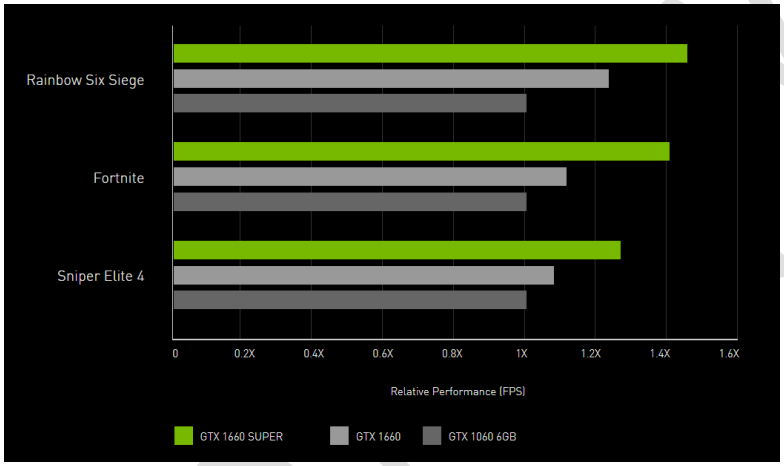

Figure 6: Comparison between Nvidia's "Budget" Cards, GTX 1660 Super (Green), 1660 (Light Grey), 1060, (Dark Grey)

On AMDs side, the Radeon RX 5500 XT is a budget gamer's dream. While it may not be as powerful as the GTX 1660, it can be found at almost $100 less [7]. Considering its performance relative to the price, it keeps up with the 1660. With the price of the RX 5500 XT, it makes consumers side with AMD despite the minor difference in performance. Winning the hearts of all consumers, regardless of budget ranges, just gives companies more motivation to create the best video cards that everyone can enjoy.

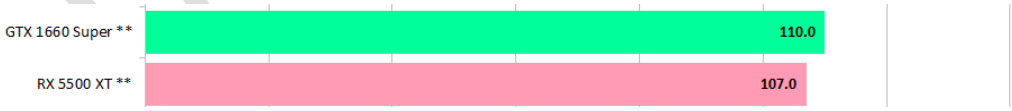

Figure 7: Comparison between Nvidia GTX 1660 Super and AMD RX 5500 XT

Conclusion; The Future of Video Cards

Video cards have come a long way since the first IBM prototype. The shift from simply processing graphical data faster to creating photorealistic experiences with ray tracing is fascinating to see. No matter what the use case of video cards is, the technology is only going to get better. Perhaps AMDs response to ray tracing will shift the balance [8], or Nvidia might dominate in other sectors, such as deep learning. [9] In any case, it is exciting to imagine what is there to come for the future of video cards. There could be a chance that a “no compromises” video card will be created. Only time will tell.

References

[1] G. Laguna, "NVIDIA Unveils Quadro RTX, World's First Ray-Tracing GPU, NVIDIA NEWSROOM," 13 August 2018. [Online].

Available: https://nvidianews.nvidia.com/news/nvidia-unveils-quadro-rtx-worlds-first-ray-tracing-gpu. [Accessed 8 November 2020].

[2] J. Novet, "The PC is suddenly cool again ... for now.," CNBC, 2020. [Online].

Available: https://www.cnbc.com/2020/05/04/pc-sales-usage-rise-during-coronavirus-lockdown.html. [Accessed 8 November 2020].

[3] C. McClanahan, "History and evolution of gpu architecture," A Survey Pa-per, vol. 9, 2010.

[4] N.M.Xi, "TheDuopolyAnalysisofGraphicsCardMar-ket," arXiv.org, Papers 2010.09074, October 2020. [Online]. Available: https://ideas.repec.org/p/arx/papers/2010.09074.html.

[5] A. S. O. M. M. S. a. O. A. A. Gamal, "A leap in theworld of graphics cards," *Evolution,* vol. 1, p. 2.

[6] "Gigabyte GeForce GTX 1660 SUPER OC 6G," Amazon, [Online]. Available: shorturl.at/rxJVY. [Accessed 11 November 2020].

[7] "MSI Radeon RX 5500 XT MECH 8G OC," Amazon, [Online]. Available: shorturl.at/gxzJS. [Accessed 11 November 2020].

[8] "AMD unveils three Radeon 6000 graphics cards with ray tracing and RTX-beating performance," EUROGAMER, 28 October 2020. [Online].

Available: https://www.eurogamer.net/articles/digitalfoundry-2020-10-28-amd-unveils-three-radeon-6000-graphics-cards-with-ray-tracing-and-impressive-performance. [Accessed 11 November 2020].

[9] J. B. L. J. A. D. N. R. T. A. V. D. J. K. N. A. Gawande, "Scaling Deep Learning Workloads: NVIDIA DGX-1/Pascal and Intel Knights Landing," *2017 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW),* pp. 399-408, 2017.

Appendix

CPU – Central Processing Unit, a main component of computers that acts as the brain.

Gamer – A person who plays video games.

GPU – Graphical Processing Unit, another way of referring to a video card.

GTX – Codename included in Nvidia video cards that don't have ray-tracing technology directly built in.

Quadro/ GeForce – Codenames for the different "families" of GPUs Nvidia has.

Radeon – AMD's family of GPUs.

Ray tracing – Rendering technique for generating an image by tracing the path of light as pixels in an image plane.

RTX – Codename included in Nvidia video cards that have ray-tracing technology.

Turing Architecture – The name Nvidia uses to describe the ray-tracing cores inside their RTX video cards.

VFX Artist – A media creator who specializes in art involving 3D video effects, visual effects etc.