Dev Misra

177003058

NVIDIA Issues and Recommendations

Overview

Almost 30 years ago, there existed three individuals that all believed that the next wave

of computing was going to exist within graphics-based visual processing. In a world where 3D

personal computer graphics was in its infancy, NVIDIA founders Jensen Huang, Curtis Priem, and

Chris Malachowsky saw an opportunity to capture the market by storm. Banding together, they

managed to secure $20 million of venture capital funding to build upon their ideas and work to

turn their dreams into a reality. On April 5, 1993, the American multinational technology

company NVIDIA was born. Six years later, in 1999, a graphics processor called the GeForce 256

was released, heralded by NVIDIA as "the world's first Graphics Processing Unit (GPU)."

Introducing an on-board Transformation and Lighting (T&L) engine to consumer-level 3D

hardware, the GeForce 256 outperformed other computer graphics visualization products

available at the time by a wide margin. According to one source, "This engine allowed the

graphics chip to undertake the heavily floating-point intensive calculations of transforming the

3D objects and scenes – and their associated lighting – into the 2D representation of the

rendered image. Previously, this computation was undertaken by the CPU, which could easily

bottleneck with the workload, and tended to limit available detail." Essentially, the release of

this graphics card revolutionized the visual computing market.

Due to the success of the GeForce 256, NVIDIA was contracted by technology giant

Microsoft to create and develop the graphics hardware for the Microsoft Xbox gaming console,

earning both a bright outlook for its future within the field of graphics hardware and an advance of $200 million, along with a secondary $400 million bond and stock offering a few months later. This ushered in the NVIDIA vs. ATI (acquired in 2006 by Advanced Micro Devices) era of graphics processors, where fierce competition between the two companies and advancement of personal computing technologies lead to exponential growth within the field of visual computing. Year by year, NVIDIA continued to invest heavily into R&D towards GPUs and has slowly but surely achieved a massive 80% market share in Q2 of 2020 (for discrete GPUs). While continuing to develop graphics cards, NVIDIA has leveraged its profits to branch out and expand onwards to multitudes of other projects, including: supercomputing, data centers, machine learning and artificial intelligence, robotics, ray tracing, virtual reality, and much more. NVIDIA has also acquired other companies and competitors along the way.

Despite the irrefutable success and massive growth of NVIDIA in the past years, the launch of the new RTX 3000 Series of graphics cards has highlighted a major concern within NVIDIA – Inbound and Outbound Logistics. Upon release, the graphics cards sold out almost instantaneously. Months after release, retailers and consumers alike continue to face heavy stock shortages for these graphics cards, an issue that is expected to remain for many months more. CEO Jensen Huang commented, "Our growth, in the near term, is more affected by the cycle time of manufacturing and flexibility of supply… We are in a good shape and all of our supply informs our guidance, but we would appreciate shorter cycle times—we would appreciate more agile supply chains. But, the world is constrained at the moment. And so, we just have to make the best of it. But, even in that condition, all of that is building for our guidance, and we expect to grow."

Issues

The RTX 3000 Series debuted with three separate GPUs capturing the mid-to-high end portions of the visual processing market: the GeForce RTX 3090 at $1,499; the GeForce RTX 3080 at $699, and the GeForce RTX 3070 at $499. Powered by Ampere – NVIDIA's 2nd generation RTX architecture – this highly anticipated new series of GPUs (touted to be the biggest generational leap ever made by NVIDIA) delivered all that consumers wanted and more. That is, except for one crucial component – the actual delivery of the product itself. Employing a staggered release for the graphics cards, the RTX 3080 was the first to release on September 17th, followed by the RTX 3090 on September 24th. Both of these products sold out instantly, and consumers have been hard pressed to locate the cards ever since, as they continue to sell out instantly after every restock. In a press conference in October, Huang expanded on the issue, stating, "I believe that demand will outstrip all of our supply through the year. Remember, we're also going into the double-whammy. The double-whammy is the holiday season. Even before the holiday season, we were doing incredibly well, and then you add on top of it the 'Ampere factor,' and then you add on top of that the 'Ampere holiday factor,' and we're going to have a really big Q4 season. The 3080 and 3090 have a demand issue… The demand issue is that it is much much greater than we expected — and we expected really a lot. Retailers will tell you they haven't seen a phenomenon like this in over a decade of computing. It hearkens back to the old days of Windows 95 and Pentium when people were just out of their minds to buy this stuff. So this is a phenomenon like we've not seen in a long time, and we just weren't prepared for it."

When looking at this situation, it is also important to consider the factor of the global COVID-19 pandemic. In a commentary after the release of Q3 financials, the CFO of NVIDIA Colette Kress stated, "Our gaming and data center market platforms have benefited from stronger demand as people continue to work, learn, and play from home… We are supply-constrained. One, by the overwhelming demand [for] our gaming desktop [RTX] 30 series. And we're going to work really hard over the next couple months to fine-tune that overall supply to meet that overall demand. But additionally, we are also supply-constrained in our overall data center [business]." Coronavirus and the related measures intended to slow its spread have caused a large spike in demand for video gaming and rapid growth of the work-from-home economy. This correlates directly with the launch of the RTX 3000 series, and demand for the graphics cards is at an unparalleled high. The RTX 3070, the most affordable option in the new series of graphics cards, was initially set to release on October 15th; however, due to postulated supply shortages in relation to the unexpected spike in demand, the release was delayed by two weeks to October 29th. Despite the delayed release, the RTX 3070 was fated to sell out within seconds after launch, and just like its co-products, it did.

If the lack of supply to meet unprecedented demand was not already a large enough source of grief for prospective consumers of the RTX 3000 series graphics cards, another derivative issue exists within the multitudes of bots that swarm websites upon launch. One huge reason that everyday consumers are unable to secure a product on release day is that humans are simply unable to purchase the products at the same speed that bots programmed specifically for the task do. To make matters worse, many of these bots are operated by what are referred to as "scalpers" – individuals who purchase the already limited supply of graphics

cards en masse and then relist them on vendors such as Amazon and eBay for exorbitant prices. This discourages consumers from even attempting to purchase products from online vendors, even months after the initial release of the graphics cards.

Recommendations

After the Q3 earnings release of NVIDIA on November 18th, CEO Jensen Huang said, "NVIDIA is firing on all cylinders, achieving record revenues in gaming, data center and overall. The new NVIDIA GeForce RTX GPU provides our largest-ever generational leap and demand is overwhelming. NVIDIA RTX has made ray tracing the new standard in gaming. We are positioning NVIDIA for the age of AI, when computing will extend from the cloud to trillions of devices." Because the supply shortage for the RTX 3000 Series graphics cards is primarily due to astronomic demand, NVIDIA has a few options in order to capitalize on the situation. One recommendation I would give to NVIDIA is to expand its product offerings while demand remains high. This is something that the company seems to already be in the process of doing, as recent leaks have surfaced of a new, even more affordable offering within the RTX 3000 Series – The RTX 3060 Ti. Leaked benchmarks of this new graphics card "show the RTX 3060 Ti outperforming the last-gen RTX 2080 Super [a MSRP $700 graphics card] in all scenarios, even in ray tracing, despite the fact that it reportedly features fewer RT cores (38 vs 48)." The RTX 3060 Ti is rumored to make its debut in early December, and is expected to be the first NVIDIA Ampere GPU to cost under $400. This new addition to the RTX 3000 Series product line is sure to give consumers a new purchase to look forward to during the holiday season, and will

hopefully shift some demand from the higher priced RTX 3080 and RTX 3070 graphics cards to the more affordable RTX 3060 Ti.

This new product offering by NVIDIA in a time of high demand for GPUs is also extremely likely to go out of stock minutes after release; however, there are a few things that NVIDIA can do in order to deter bots and scalpers from denying stock to aspiring human consumers. I recommend the implementation of a reCAPTCHA system to the checkout process for the new graphics cards. Although this is a low wall that many advanced bots are able to detect and get through, it is a deterrent to the majority of simple bots designed to speed up the checkout process, and will likely increase the amount of time that these bots take to purchase an item. Along with a reCAPTCHA system, NVIDIA can implement an initial limit to the number of graphics cards ordered online per IP address. Again, although this is a hurdle that advanced bots will be able to circumvent, this constraint should theoretically limit the amount of graphics cards that can be purchased by a single individual or entity. When the demand for the graphics cards settles down, this constraint can be removed. Together, these implementations would go a long way towards slowing down bots and scalpers, and if nothing else, show consumers that the company is making an active effort towards attempting to get customers the products that they desire.

Conclusions

Due to a variety of factors, be it the release of new technological architecture, the holiday season, or a global pandemic, there exists an extremely high demand for the newly launched NVIDIA graphics cards. Despite this high demand, capacity limits and long lead times

across the chipmaking industry globally cannot be ignored. NVIDIA executives have acknowledged the fact that it will take a few more months for product availability to catch up with demand, and the company is doing all that is within their power to ensure that customers can receive the graphics cards that they desire. Because of this, I believe that the most important thing that NVIDIA can currently do is to continue assuring customers that supply shortages are due to various external factors within the global environment, and that the company is operating with the interests of its customers in mind. It is no secret that NVIDIA as a company is primed for huge growth in upcoming years, and thus relatively smaller tokens of appreciation in times such as these would go a long way in building up brand name and reliability. It is logical to assume that if NVIDIA follows recommendations such as implementing bot-reducing measures for online purchases of graphics cards, the resulting satisfaction built within the consumer base will have a largely positive impact on future sales and word-of-mouth advertising. In my external analysis paper, I mentioned that NVIDIA's greatest strength exists in its product differentiation. If this same tactic of differentiation can be applied to consumer and community relations during these strained times, NVIDIA and all of its clientele will be sure to flourish during better days yet to come.

References

Alcorn, P. (2020, October 05). Nvidia CEO: GeForce RTX 3080 and 3090 Shortages To Last Until 2021. Retrieved November 18, 2020, from

https://www.tomshardware.com/news/nvidia-ceo-geforce-rtx-3080-and-3090-shortages-to-last-until-2021

Broekhuijsen, N. (2020, October 02). Nvidia Delays RTX 3070 By Two Weeks to Avoid Another Disappointing Launch. Retrieved November 18, 2020, from

https://www.tomshardware.com/news/nvidia-delays-rtx-3070-by-2-weeks-to-avoid-another-disappointing-launch

Jhonsa, E. (2020, November 19). Nvidia's CFO: We're Working Hard to Keep Up with Record Demand. Retrieved November 20, 2020, from

https://www.thestreet.com/investing/nvidias-cfo-were-working-hard-to-keep-up-with-record-demand

Johnston, M. (2020, November 16). Nvidia Earnings: What to Look For. Retrieved November 18, 2020, from

https://www.investopedia.com/nvidia-q3-2021-earnings-5087601

McKenna, B. (2020, November 19). NVIDIA's Earnings Report Was Stellar -- Yet Not Good Enough for All Investors. Retrieved November 19, 2020, from

https://www.fool.com/investing/2020/11/18/nvidias-earnings-report-was-stellar-yet-not-good-e/

Page, C. (2020, November 17). Nvidia RTX 3060 Ti outperforms the RTX 2080 Ti in leaked Nvidia benchmarks. Retrieved November 19, 2020, from

https://www.techradar.com/news/nvidia-rtx-3060-ti-outperforms-the-rtx-2080-ti-in-leaked-nvidia-benchmarks

Research at NVIDIA. (n.d.). Retrieved November 18, 2020, from

https://www.nvidia.com/en-us/research/research-areas/

Ridley, J. (2020, November 19). Nvidia is as frustrated with the lack of RTX 30-series stock as you are. Retrieved November 20, 2020, from

https://www.pcgamer.com/nvidia-is-as-frustrated-with-the-lack-of-rtx-30-series-stock-as-you-are/

Singer, G. (2013, April 10). History of the Modern Graphics Processor, Part 3. Retrieved November 18, 2020, from https://www.techspot.com/article/657-history-of-the-gpu-part-3/

Singer, G. (2013, April 3). History of the Modern Graphics Processor, Part 2. Retrieved November 18, 2020,

from https://www.techspot.com/article/653-history-of-the-gpu-part-2/