

CS/SE 2340 Computer Architecture Homework 6: Exploring a CPU/GPU

MY GPU: **NVIDIA RTX 2060 SUPER**



Section I: Overview

Since the first commercial GPU was developed by NVIDIA in 1999, The United States; and a large majority of the entire world has made the technology commonplace in any personal computer for gaming or other types of modern tasks that requires the extensive **concurrency** applications unique to GPUs.

While other major corporations like AMD and Intel market and expand on the technology developed by NVIDIA, NVIDIA remains a titan of innovation for next-generation GPU technology.

Whether it be their standard GeForce series of GPUs for the common gamer with the GTX technology or their more recent RTX technology including the Turing and Ampere architectures which introduced **real-time ray tracing**, rasterization, simulation, and A.I., their GPUs stand at the forefront to address any and all consumer needs

Such is the case with the very topic of this report: NVIDIA's RTX 2060 Super GPU.

Launched on July 9, 2019, as part of the **GeForce 20 series**, The RTX 2060 Super is based on NVIDIA's **Turing microarchitecture**.

With base specs of: 2176 **CUDA cores**, A **base clock rate** of 1470 Megahertz, a **boost clock rate** of 1650 Megahertz, 8 gigabytes of **GDDR6 SDRAM**, a memory bandwidth of 448 gigabytes per second, and 10.8 billion **transistors**, The RTX 2060 Super is a multi-purpose GPU fit and capable for almost any consumer needs.

However, the NVIDIA RTX 2060 Super is NOT the best GPU. Nor is it the cheapest GPU. While the Turing microarchitecture of the NVIDIA RTX 2060 Super is capable, NVIDIA's **Ampere microarchitecture** is much more advanced, and for all intents and purposes; better.

With an average eBay price of \$701 as of November, 2021, the GPU is cheaper than it's more technologically advanced cousins but still more costly than it's GTX predecessors like the GTX 1660 Super with a low cost of only \$229 as of December, 2021.

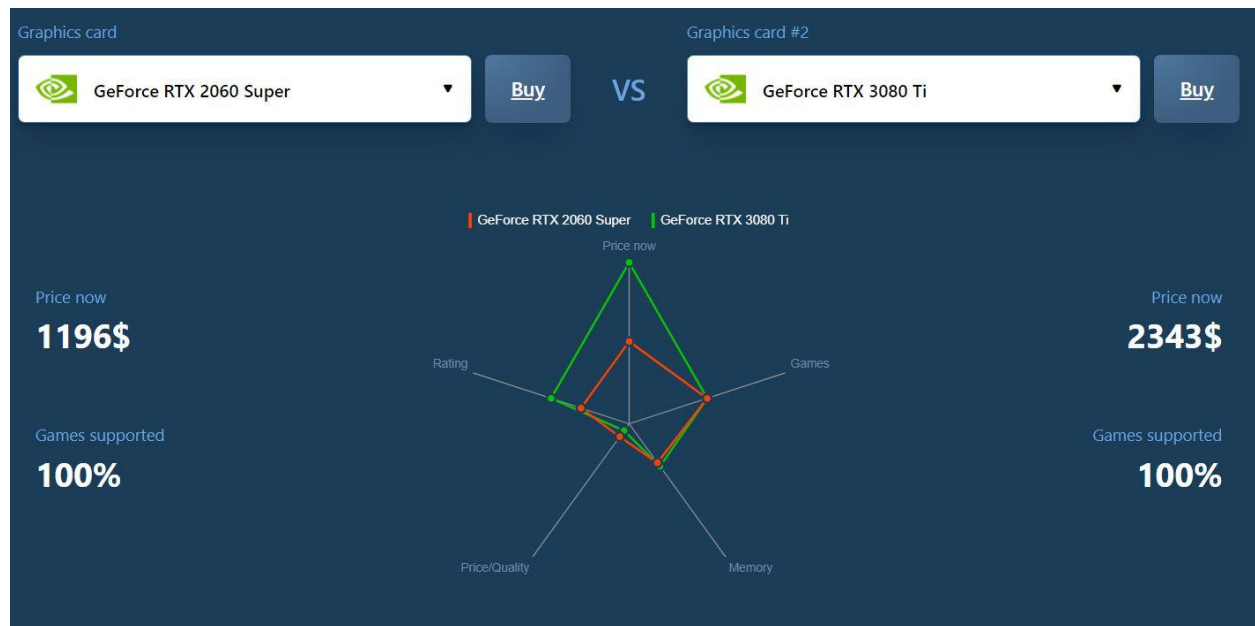
That being the case, why would anyone buy a RTX 2060 Super? The interesting answer is: niche.

Sitting in a "goldilocks zone" of sorts, the NVIDIA RTX 2060 Super is perfect for those who desire a high-performance GPU without needing the more advanced specs offered by the more expensive GPUs like the RTX 3080 Ti who use the Ampere microarchitecture.

With better performance than the RTX 2060 and nearly identical spec to the more advanced RTX 2070 with a cheaper price tag, the RTX 2060 Super fits in a perfect middle ground of cost vs. performance. It delivers excellent results for an affordable price.

The intended market of the RTX 2060 Super would be any gamer who wants a high-performing GPU for an affordable price range that would disqualify the more recently developed GPUs like the NVIDIA Titan. The targeted consumer would be someone who intends to play graphically intensive games or software at good performance, but not willing to spare the extra money to play those same games at unnecessarily higher graphical levels without any real purpose.

In summary, the NVIDIA RTX 2060 Super is a well rounded, multi-purpose GPU based on the Turing microarchitecture that is perfect for any consumer who wants high performance without an absurd cost.



Section II: Detailed Analysis

When examining NVIDIA's RTX 2060 Super on a more technical level, we must first be familiar with the RTX technology and the Turing microarchitecture the GPU utilizes.

RTX or “**Ray Tracing Texel Extreme**” are GPUs implementing **real-time ray tracing** to enhance graphic performance and make experiences like gaming more realistic.

The Turing microarchitecture, named after the famous Computer Scientist: Alan Turing, is said to be the biggest architectural leap in over a decade as claimed by NVIDIA. With the benefits of enabling major advances in efficiency and performance for PC Gaming, professional graphic design, and even **deep learning inferencing**, it's hard to deny their statement.

As for what causes the boost in graphical performance: NVIDIA states that their new **streaming multiprocessor architecture** with staggering shader execution efficiency, and a newly developed memory system which supports the latest **GDDR6** memory technology are what give the Turing microarchitecture its high-performance.

In combination with these technological advancements, the Turing microarchitecture boasts groundbreaking A.I. and ray-tracing applications, including but not limited to: advanced shading features, increased performance efficiency, image quality enhancement, **tensor core** innovation, and never before seen levels of geometric complexity.


Of course, that's not to mention the fact that the technology used in the Turing microarchitecture is cumulative; inheriting the enhancements to the **NVIDIA CUDA** platform that was introduced in the predecessor **Volta microarchitecture** that allowed **independent thread scheduling**, **hardware-accelerated Multi Process Service (MPS)** with **address space isolation** for multiple applications and even **Cooperative groups**.

Since the NVIDIA RTX 2060 Super is designed off of the Turing microarchitecture, it inherits these innovative features; boasting an additional 256 **CUDA Cores**, 32 **Tensor Cores** and 4 **RT Cores** compared to its 2060 predecessor.

As such, we can attribute most of the NVIDIA RTX 2060 Super's performance to the RTX platform design and Turing microarchitecture.

Section III: Reviews

First Review - Jarrod's Tech (Youtuber):

 Nvidia RTX 2060 Super vs RTX 2060 - Is Super Worth It?

In this review, Jarrod reviews the RTX 2060 Super and how it compares to the original predecessor: The RTX 2060.

Comparing and contrasting the differences in the base specs between the two GPUs, it can be observed that the RTX 2060 Super is superior in every field of comparison with the exception of its boost clock rate of 1,650 Megahertz which is slightly lower than the RTX 2060's boost clock rate of 1,680 Megahertz.

After comparing the base specs, Jarrod begins testing the performance of the two GPUs with his own PC.

After testing the two GPUs on various graphically intensive and famous games, He provides data regarding the average performance of each GPU for each game.

While the degrees of difference in performance varies depending on the game being run, it is clearly observed across all test trials that the RTX 2060 Super runs at a better performance; allowing a higher FPS (Frames Per Second) on average when RTX is disabled. The difference in performance becomes more magnified proportionally to the graphical resolution.

When RTX is enabled, an even larger improvement could be observed in the RTX 2060 Super compared to the RTX 2060. In other words, The RTX 2060 Super improved performance in making use of the RTX platform and by extension, it's Turing microarchitecture to allow better performance with RTX than its predecessor.


Across all tests, Jarrod asserts that on average, the RTX 2060 Super boasts a performance that is 13.16% faster than the RTX 2060 for lower graphical resolutions.

When switching to higher graphical resolutions, this performance boost jumps to 16.39%; proving the RTX 2060 Super's superiority in running advanced graphics with excellent performance.

In the final segments of the review Jarrod compares their respective Total Power Draw and pricing. For the Total Power Draw, it is observed that the RTX 2060 Super has a greater Total Power Draw, but makes up for it with a greater relative performance improvement.

Finally, the pricing. While the prices he discusses for the two have changed considerably since he released the video in 2019, it was originally \$400 for the RTX 2060 Super and \$350 for the RTX 2060. The \$50 minor price difference was well worth the upgrades provided to the RTX 2060 Super. (As a side note, I purchased my own RTX 2060 Super that I now use during the same time period for a similar price).

Second Review - Linus Tech Tips (Youtuber):

 Nvidia's New SUPER Cards! - RTX 2060 & 2070 SUPER Review

To begin with, the reviewer, Anthony, begins a breakdown of the "new" release, the RTX 2060 Super and other Super releases at the time such as the RTX 2070 Super.

Starting with an opinionated critique of the branding and product design of the "Super" GPUs, he moves on to assessment of what the "Super" GPUs provide compared to their predecessors.

It is here where the observations are made of the similarity between the RTX 2060 Super and the RTX 2070. Comparing the base specs of the "Super" GPUs compared to the "non-super" GPUs a noticeable observation can be made of their relative performances.

Despite being considered an “2060,” the base specs of the RTX 2060 Super are actually closer to the base specs of the RTX 2070, improving the cost-performance for the model while making the base RTX 2060 inferior in every sense.

This is seen in how the RTX 2060 Super has 2,176 CUDA Cores, 34 RT Cores, 272 Tensor Cores, and 8 GB GDDR6 Memory are only slightly inferior to the RTX 2070 with 2,304 CUDA Cores, 36 RT Cores, 288 Tensor Cores, and the same Memory and TDP (Thermal design power).

While the RTX 2070 Super also has an improvement to its base specs that make it more comparable to the base RTX 2080, it has noticeably less “bang-for-buck” than the RTX 2060 Super, further cementing its cost-performance worth.

The launch pricing at the time of video also supports the worth of the “Super” GPUS as they are slightly lower than the non-super versions of the next tier. With the RTX 2070 costing \$499 and the RTX 2060 Super costing \$399. A \$100 price difference for only a slight increase in performance between the two.

Anthony then praises the value of the “Super” GPUS albeit inferring the cause of their low pricing to be competition in the GPU market at the time of their release.

After comparing their base specs and prices, he then proceeds to testing the performance of the GPUs.

By comparing the performance of the GPUs by their average FPS in popular video games, it can be observed that the RTX 2060 Super displays impressive performance. Competing head-to-head with the more expensive RTX 2070 on average.

With RTX enabled, the performance of the RTX 2060 Super is even more pronounced, as it “jumps above its weight class” as a 60 series GPU.

After testing game performance, Anthony then tests the marketing claims of NVIDIA, verifying and confirming them. Moving onward, the GPUs are tested in other productivity software where no major difference in performance is observed.

With their performance confirmed, Anthony then tests the thermals of the GPUs to confirm their temperature thresholds. From these tests, it can be observed that the RTX 2060 Super has well-made cooling only getting 2°C higher than the RTX 2070 Super at 71 °C. Similar results can be seen when measuring power consumption of the GPUs.

As Anthony notes however, the most interesting test is regarding the “Super’s” boost clocks. Under normal performance tests, the GPUs barely go over their base clock rate, but when gaming, their sustained performance was well-above the boost clock’s advertised.

In his conclusion, Anthony notes how NVIDIA plans to discontinue the base RTX 2070 and RTX 2080 GPUs, and he affirms that the “Super” GPUs like the RTX 2060 Super are almost always a better buy when taking into account both cost and performance.

In his final notes, Anthony notes the causes for the release of the “Super” GPUs and their good values. Ending on his statement that NVIDIA has finally released a GPU that offers a fair price for ray-tracing.

Third Review - Techradar (Reviewer: Jackie Thomas):

<https://www.techradar.com/reviews/nvidia-geforce-rtx-2060-super>

To start the review, the reviewer, Jackie, elaborates on the timing of the RTX 2060 Super’s release. Discussing how it’s predecessor the RTX 2060 gave consumers the first real look at the capabilities of ray-tracing, yet still limited them to gaming in graphical resolutions of 1080p. Per his statement, the RTX 2060 Super hits a sweet spot between performance in value that other GPUs simply cannot match.

With his position on the GPU clarified, Jackie then moves on to discussing the RTX 2060 Super’s pricing and availability. At the time of the review, the selling price was around \$400.

When discussing the pricing, Jackie takes note of NVIDIA’s claims that the RTX 2060 Super’s performance is comparable to the original RTX 2070’s. Not only does the reviewer affirm these claims, he even states the RTX 2060 Super surpasses the RTX 2070 in some tests performed.

After discussing the RTX 2060 Super, Jackie then brings up the GPUs competitor, the AMD Radeon RX 5700 XT which was released at the same price and has similar performance. Comparing the base specs and performance of these two competing GPUs, Jackie notes that the RTX 2060 Super does have the advantage over the AMD Radeon RX 5700 XT due to its RT Cores that support ray tracing. Since the AMD Radeon RX 5700 XT doesn’t support ray tracing at all, the NVIDIA RTX 2060 Super clearly wins out.

After comparing performances, Jackie then goes deeper into the features and chipset of the RTX 2060 Super GPU: examining its Turing architecture and its varying improvements in core counts and VRAM.

The most notable feature is its 8 GB of GDDR6 VRAM which is a great increase from the 6 GB of its predecessor, the RTX 2060.

Next, the reviewer moves onto examining the GPU's performance. In this data, Jackie observes and claims that the RTX 2060 Super not only has a huge 21% performance increase over the predecessor RTX 2060, but actually performed slightly better than the RTX 2070 as well.

As the reviewer concludes, he notes that the RTX 2060 Super is able to handle almost any game at 1440p and claims the Nvidia GeForce RTX 2060 Super might be the GPU for most gamers to get. Expanding on how it's well-priced and with great relative performance, he ends by saying "We're just left wondering why Nvidia didn't release this card sooner – we finally have the value-oriented Nvidia Turing graphics card."

Section IV: Technical Terms

Concurrency:

Concurrency is the ability of different parts or units of a program, algorithm, or problem to be executed out-of-order or in partial order, without affecting the final outcome. This allows for parallel execution of the concurrent units, which can significantly improve overall speed of the execution in multi-processor and multi-core systems. In more technical terms, concurrency refers to the decomposability of a program, algorithm, or problem into order-independent or partially-ordered components or units of computation.

Ray Tracing:

Ray tracing is a method of graphics rendering that simulates the physical behavior of light. Thought to be decades away from reality, NVIDIA has made real-time ray tracing possible with NVIDIA RTX™ the first-ever real-time ray-tracing GPU—and has continued to pioneer the technology since. Powered by NVIDIA RT Cores, ray tracing adds unmatched beauty and realism to renders and fits readily into preexisting development pipelines.

GeForce 20 series:

A family of GPUs developed by NVIDIA. It is the successor of the GeForce 10 series.

Turing microarchitecture.

A GPU microarchitecture developed by NVIDIA and named in honor the famed Computer Scientist Alan Turing. Key elements of the Turing microarchitecture include dedicated artificial intelligence processors, (Tensor Cores), and dedicated ray tracing processors.

CUDA Cores:

CUDA Cores are parallel processors. These cores are responsible for: processing all data fed in and out of the GPU and performing graphics calculations. CUDA itself, is a parallel computing platform and programming model that enables a significant increase in the performance of computing by utilizing the power of the GPU.

Base Clock:

Base Clock, Core Clock or BCLK is the operating frequency of the graphic processing chip inside a GPU.

Boost Clock:

Boost clock is the boost of the processing speed done by a processor to itself automatically to enable high performance when under heavy load. Another term used in conjunction with this is overclocking, which means a manual increase in the speed of which pushes a processor to it's limits.

GDDR6 SDRAM:

Graphics Double Data Rate 6 Synchronous Dynamic Random-Access Memory is a type of synchronous graphics random-access memory known for its high bandwidth (Hence it's double data rate) interface specifically designed for the use in GPUs and general high performance computing. It is also the successor to GDDR5. GDDR6 is also associated with VRAM, (Virtual Random Access Memory.)

Transistor:

A semiconductor device commonly used in computing to amplify, generate, control, or switch electrical signals and power. A basic building block of nearly all modern electronics and computing and a major active component of microchips.

Ampere Microarchitecture:

A GPU microarchitecture design by NVIDIA and the successor to both the Volta and Turing architectures. Named after the Mathematician, Andre-Marie-Ampere, it was announced as part of the next generation GeForce 30 series. Designed for the age of elastic computing, the NVIDIA Ampere technology claims to provide unmatched

acceleration at every scale serving as the core of AI and HPC (High-Processing Computing).

Ray Tracing Texel Extreme (RTX):

An advanced professional visual computing platform developed by NVIDIA. RTX enables real time ray tracing, something that had been historically restricted to non-real time applications (CGI and renderings). NVIDIA RTX is also used as a platform for AI technologies and deep learning through cutting edge features like NVIDIA DLSS and Nvidia Broadcast. Per the company's claims, "RTX is the new standard."

Deep Learning Inferencing:

Contrary to Deep Learning Training, Deep Learning Inferencing, (DLI) refers to the use of fully trained deep neural networks (DNN) to make inferences or predictions on new data that a model may have never seen before.

Streaming Multiprocessor Architecture (SMs):

A part of the GPU used to run CUDA kernels. CUDA kernels are essentially functions executed on the GPU.

Tensor Core:

A technology featuring multi-precision computing designed for efficiency A.I. inference. In other words, they accelerate the matrix operations which are foundational to AI.

Volta Microarchitecture:

A GPU microarchitecture developed by NVIDIA and the predecessor to the Turing microarchitecture.

Independent Thread Scheduling:

A GPU feature which: enables interleaved execution of statements from divergent branches and enables execution of fine-grain parallel algorithms where threads within a warp may synchronize and communicate. This allows the GPU to yield execution of any thread to either make better use of its execution resources or to allow one thread to await data being produced by another thread.

Multi Process Service (MPS):

MPS is an alternative, binary compatible implementation of the CUDA API (Application Programming Interface). It allows multiple CUDA processors to share a single GPU context where each process receives some subset of the available connections.

Address Space Isolation:

A security idea that a system with virtual memory can create isolated address spaces which are only available to part of the given system or to certain specified processes.

Hardware Acceleration:

The use of computer hardware designed to perform functions more efficiently compared to software running on a general-purpose CPU.

Cooperative Groups:

A CUDA API used for defining and synchronizing groups of threads in a CUDA program.

RT Cores:

Cores in a GPU used to accelerate the math needed to use ray tracing by tracing virtual rays of light through a scene. These cores are an example of an application-specific integrated circuit.

References

Anthony. (n.d.). *Nvidia's new super cards! - RTX 2060 & 2070 super review ...*

Youtube. Retrieved December 6, 2021, from

<https://www.youtube.com/watch?v=bnEPBcCkM2Y>.

author Hi, A. G.- post. (2021, October 2). *What does RTX & GTX stand for in*

Nvidia gpus? CG Director. Retrieved December 6, 2021, from

<https://www.cgdirector.com/rtx-gtx-nvidia-gpus/>.

Cooperative groups: Flexible cuda thread programming. NVIDIA Developer Blog.

(2020, August 25). Retrieved December 6, 2021, from

[https://developer.nvidia.com/blog/cooperative-groups/#:~:text=At%20its%20simplest%2C%20Cooperative%20Groups,GPU%20\(Compute%20Capability%203.0%2B\)](https://developer.nvidia.com/blog/cooperative-groups/#:~:text=At%20its%20simplest%2C%20Cooperative%20Groups,GPU%20(Compute%20Capability%203.0%2B).).

Cuda Refresher: The Cuda Programming Model. NVIDIA Developer Blog. (2020, August 25). Retrieved December 6, 2021, from

<https://developer.nvidia.com/blog/cuda-refresher-cuda-programming-model/>.

Deep learning training vs deep learning inference (explained). Premio Inc. (n.d.).

Retrieved December 6, 2021, from

<https://premioinc.com/blogs/blog/deep-learning-training-vs-deep-learning-inference>.

Graphics cards. GeForce RTX 2060 Super vs RTX 3060 [in 7 benchmarks].

(n.d.). Retrieved December 6, 2021, from

<https://technical.city/en/video/GeForce-RTX-2060-Super-vs-GeForce-RTX-3060>.

Introducing geforce RTX super graphics cards: Best in class performance, plus

Ray Tracing. (n.d.). Retrieved December 6, 2021, from

<https://www.nvidia.com/en-us/geforce/news/gfecnt/geforce-rtx-20-series-super-gpus-out-now/#:~:text=The%20GeForce%20RTX%202060%20receives,Cores%20and%204%20RT%20Cores>.

Introducing the nvidia RTX ray tracing platform. NVIDIA Developer. (2021,

August 5). Retrieved December 6, 2021, from

<https://developer.nvidia.com/rtx/raytracing>.

Jarrold. (n.d.). *Nvidia RTX 2060 super vs RTX 2060 - is Super ... - youtube*.

Youtube. Retrieved December 6, 2021, from

<https://www.youtube.com/watch?v=SLkw-rdSfw4>.

Nvidia GeForce RTX 2060 super 8GB specs. GPUSpecs.com. (2020, July 13).

Retrieved December 6, 2021, from

<https://gpuspecs.com/card/nvidia-geforce-rtx-2060-super-8gb>.

Nvidia GeForce RTX 2060 super specs. TechPowerUp. (2021, November 9).

Retrieved December 6, 2021, from

<https://www.techpowerup.com/gpu-specs/geforce-rtx-2060-super.c3441>.

NVIDIA. (2021, November 10). *Nvidia Turing architecture in-depth.* NVIDIA

Developer Blog. Retrieved December 6, 2021, from

<https://developer.nvidia.com/blog/nvidia-turing-architecture-in-depth/>.

NVIDIA. (n.d.). *Nvidia Ampere Architecture: The heart of the modern data center.*

NVIDIA. Retrieved December 6, 2021, from

<https://www.nvidia.com/en-us/data-center/ampere-architecture/>.

NVIDIA. (n.d.). *Nvidia RTX Technology.* NVIDIA. Retrieved December 6, 2021,

from <https://www.nvidia.com/en-us/design-visualization/technologies/rtx/>.

NVIDIA. (n.d.). *Nvidia Turing: Reinventing Gaming Graphics.* NVIDIA. Retrieved

December 6, 2021, from <https://www.nvidia.com/en-us/geforce/turing/>.

NVIDIA. (n.d.). *Stream like a boss with Nvidia RTX graphics card.* NVIDIA.

Retrieved December 6, 2021, from <https://www.nvidia.com/en-us/geforce/rtx/>.

Specs dictionary - cuda cores: GamersNexus - gaming PC builds & hardware

benchmarks. GamersNexus. (n.d.). Retrieved December 6, 2021, from

<https://www.gamernexus.net/dictionary/2-cuda-cores>.

Thomas, J. (2020, September 21). *Nvidia Geforce RTX 2060 Super Review*.

TechRadar. Retrieved December 6, 2021, from

<https://www.techradar.com/reviews/nvidia-geforce-rtx-2060-super>.

The ultimate factsheet on Core Clock vs boost clock: 2 clocks are not the same.

HWIRIG. (n.d.). Retrieved December 6, 2021, from

<https://hwirig.com/ultimate-factsheet-core-clock-vs-boost-clock-2-clocks-2/amp/>.

Walton, J. (2021, December 1). *GPU prices, December 2021*. Tom's Hardware.

Retrieved December 6, 2021, from

<https://www.tomshardware.com/news/gpu-pricing-index>.

Whitelock, A. (2021, December 1). *The Best Cheap Graphics Card Prices and deals for December 2021*. TechRadar. Retrieved December 6, 2021, from

<https://www.techradar.com/news/best-graphics-card-deals#:~:text=Nvidia%20GeForce%20GTX%201660%20Super,-A%20great%20budget&text=Even%20in%202021%2C%20The%20GTX,%2F%20%C2%A3180%20%2F%20AU%24330>.

Wikimedia Foundation. (2021, December 1). *Geforce 20 series*. Wikipedia.

Retrieved December 6, 2021, from

https://en.wikipedia.org/wiki/GeForce_20_series.

Wikimedia Foundation. (2021, December 1). *Turing (microarchitecture)*.

Wikipedia. Retrieved December 6, 2021, from

[https://en.wikipedia.org/wiki/Turing_\(microarchitecture\)](https://en.wikipedia.org/wiki/Turing_(microarchitecture)).

Wikimedia Foundation. (2021, December 2). *Concurrency (computer science)*.

Wikipedia. Retrieved December 6, 2021, from

[https://en.wikipedia.org/wiki/Concurrency_\(computer_science\)](https://en.wikipedia.org/wiki/Concurrency_(computer_science)).

Wikimedia Foundation. (2021, December 5). *Transistor*. Wikipedia. Retrieved December 6, 2021, from

<https://en.wikipedia.org/wiki/Transistor#:~:text=A%20transistor%20is%20a%20semiconductor,building%20blocks%20of%20modern%20electronics.&text=A%20voltage%20or%20current%20applied,through%20another%20pair%20of%20terminals.>

Wikimedia Foundation. (2021, November 13). *Hardware acceleration*. Wikipedia. Retrieved December 6, 2021, from

https://en.wikipedia.org/wiki/Hardware_acceleration.

Wikimedia Foundation. (2021, November 2). *GDDR6 SDRAM*. Wikipedia. Retrieved December 6, 2021, from

https://en.wikipedia.org/wiki/GDDR6_SDRAM.

Wikimedia Foundation. (2021, November 29). *Nvidia RTX*. Wikipedia. Retrieved December 6, 2021, from https://en.wikipedia.org/wiki/Nvidia_RTX.

Your graphics, now with Super Powers. NVIDIA. (n.d.). Retrieved December 6, 2021, from <https://www.nvidia.com/en-us/geforce/graphics-cards/rtx-2060-super/>.

