

# Nvidia Titan RTX

Department: 電機所

Student ID: 110061635

Name: 王瑞賢

## ● OUTLINE

1. Introduction
2. Nvidia Titan RTX Spec
3. Applications
4. Conclusion
5. References

# 1. Introduction

Titan RTX is a powerful PC graphics processor designed for use by researchers, scientists, and web creators.

## ◆ History

Because there are many generations of the Titan series, this side will not be introduced one by one, only to introduce its origin and Titan's previous generation on the good.

## ● GTX Titan



This Titan without any suffixes is the first generation of the Titan series, the full name is NVIDIA GeForce GTX Titan, Titan means giant in Greek, and Huang's intention in naming this card Titan was obvious, he wanted to make a giant-like performance

card, and he did. It was released in early 2013, the core architecture for the Kepler architecture, a generation of Titan released after the GTX6 series cards, but earlier than the GTX 7 series cards, so it can be considered an upgraded version of the GTX 680, that is, as the GTX 680Ti to see. Performance higher than GTX 780 and lower than GTX 780Ti, At that time, it was a well-deserved card king.

- GTX Titan B



This graphics card is also called Black Titan, the second generation of the Titan series, the full name is NVIDIA GeForce GTX Titan Black. AMD released its flagship single-core graphics card R9 290X at the end of FY13, which outperformed Titan by

20 points but debuted at less than half the price. So NVIDIA felt it was time for this 20<sup>th</sup> anniversary graphics card to take a back seat, and in early 2014, the Titan II GeForce GTX Titan Black officially took over! The Black Titan is the same as the generation of Titan, this graphics card is still GK110 Kepler core.

- GTX Titan Z



This graphics card is also called Z Titan, is the third generation of the Titan series, the full name is NVIDIA GeForce GTX Titan Z. Although the previous Titan graphics card performance

explosion, in the single-core card hard to find rivals, but can not become a real card king, because and dual-core card compared to the performance gap is still a lot. So in the middle of 2014, Titan Z made its grand debut, just 4 months after the release of the Black Titan, the new flagship made in such a short period of time is naturally nothing innovative, in fact Titan Z is just two Titan B got on a PCB board.

- Titan V

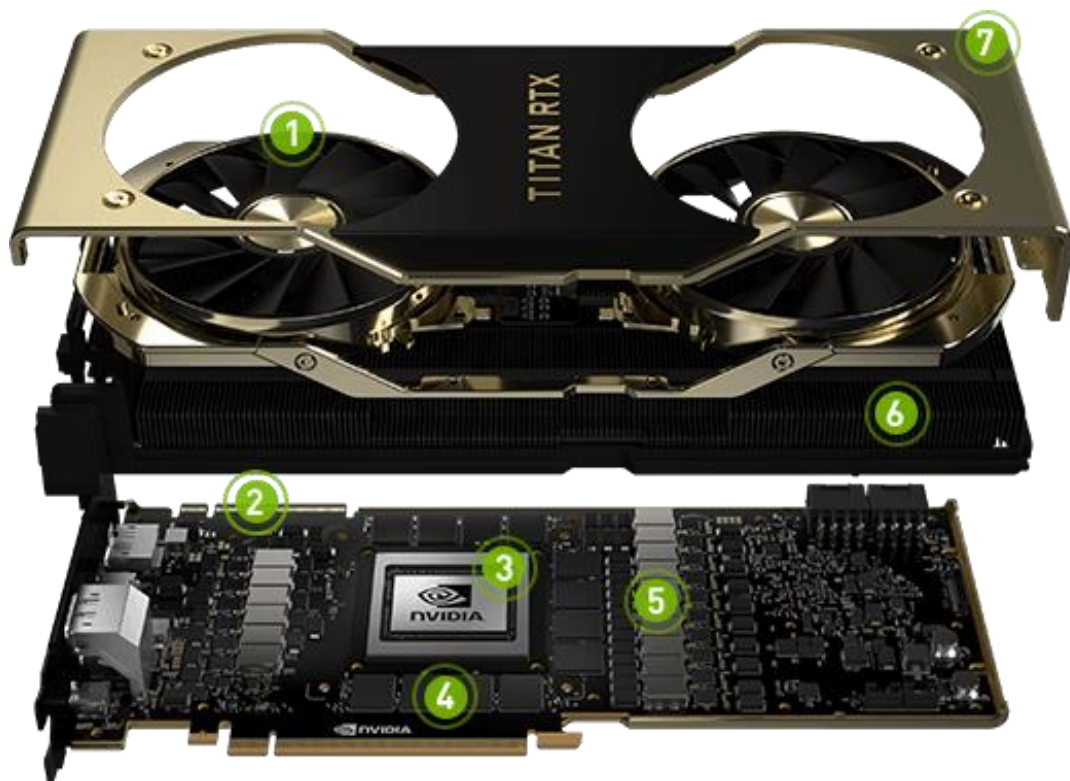


Titan V graphics card, it is the seventh generation of the Titan graphics card series, it follows the then latest Volt architecture core GV100 stream processor, frequency and other parameters

beyond the GP102 core Titan XP, from the runtime finally exceeded AMD's ultimate flagship PRO DUO, after two years finally again took the crown of graphics card king, but this graphics card does not seem to be very hot.

## 2. Nvidia Titan RTX Spec

The next part of the Titan RTX specifications, as indicated in the diagram, will be broken down into nine main parts to do the introduction.



(1)Fan

The first is the fan, which is made up of two sets of 13 fan blades that allow for quieter cooling and higher airflow.



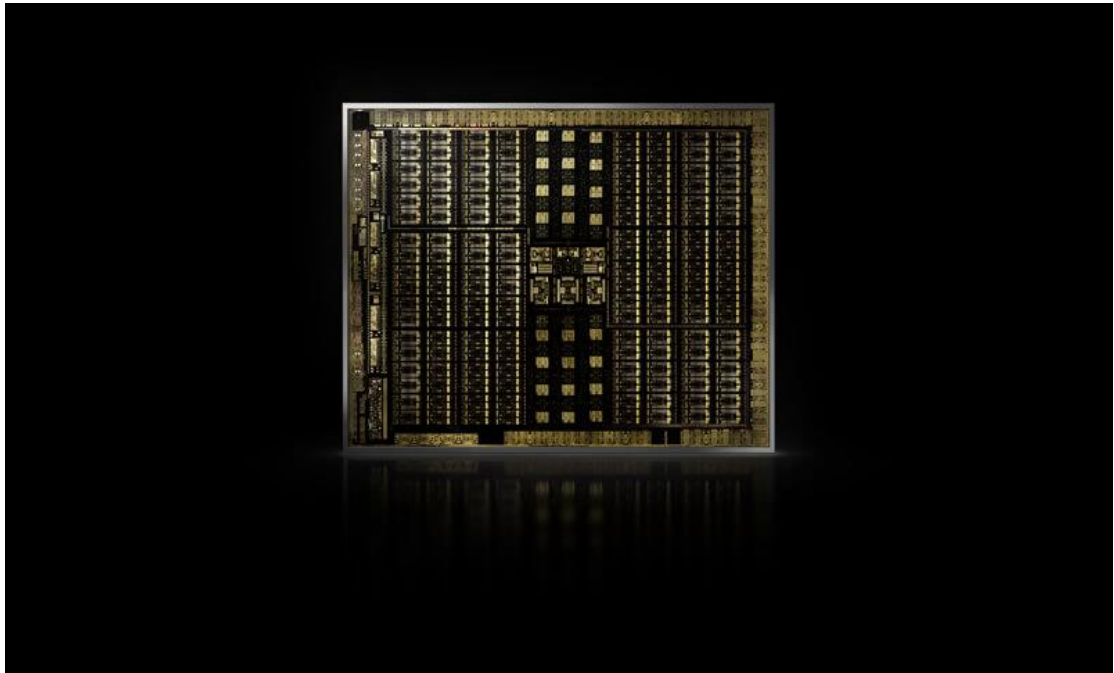
## (2) TITAN RTX NVLINK™ Bridge



As you can see in the picture, an accessory connects the graphics cards together, and as explained earlier, it allows the memory of the graphics cards to be stacked on top of each other. Its main use is in deep learning of AI, because deep learning consumes a lot of graphics card memory, and if it encounters a memory bottleneck, even if the chip performance is high, the effect of speed enhancement is very limited, so this NVlink bridge is needed.

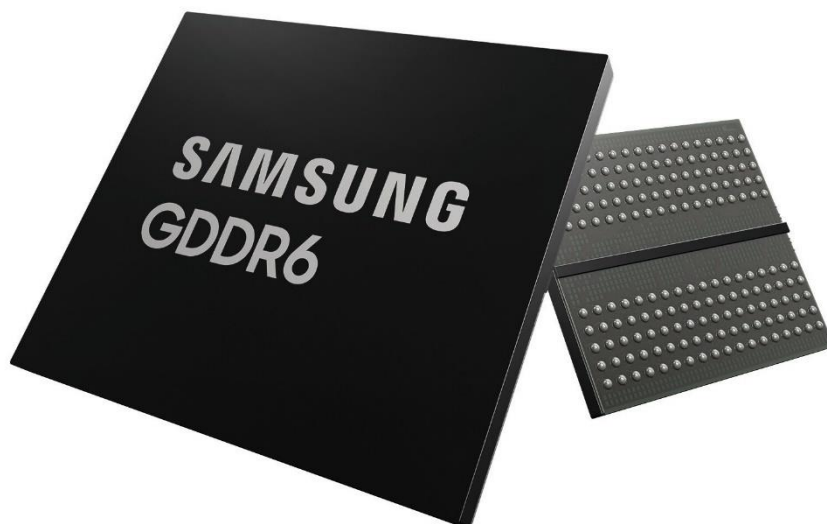


### (3) NVIDIA TURING GPU



The third is Nvidia's latest Turing architecture GPU, which contains 72 RT cores, 576 Tensor cores, and 4608 CUDA cores for parallel computing.

### (4) GDDR6 MEMORY



The fourth is GDDR6 memory, which can provide up to 672GB

of memory bandwidth per second, greatly enhancing its throughput and processing of some very large files.

#### (5)POWER SUPPLY



The fifth is the power supply, which provides 13 phases of power supply to ensure proper operation and overclocking.

#### (6)VAPOR CHAMBER

This full-card vapor chamber is 2X larger to maximize heat spreading and heat transfer to the fin stack.

#### (7)COVER

A forged and machine-finished diecast aluminum cover with diamond-cut edge detailing provides a rigid, lightweight frame



for an open design with beautifully smooth, continuous curves.

## (8) VIRTUALLINK



Next is Virtual Link, which integrates the power, video, audio, and data signals needed for VR devices into a single USB-C cable and connects directly from the graphics card, significantly simplifying the number of cables needed for head-mounted devices. However, other than NVIDIA, almost all of the VR devices or graphics cards that came out later did not use the Virtual Link connection standard because there was no one else to use it, so NVIDIA no longer included the Virtual Link port with the RTX20 super series of graphics cards that came out later.

(9) DISPLAY PORT 1.4 8K @60 Hz

			
DisplayPort	HDMI		
	bandwidth	resolution	
DisplayPort 1.4	32.4 Gbps	up to 8K@60Hz	
HDMI 2.0	18 Gbps	up to 8K@30Hz	

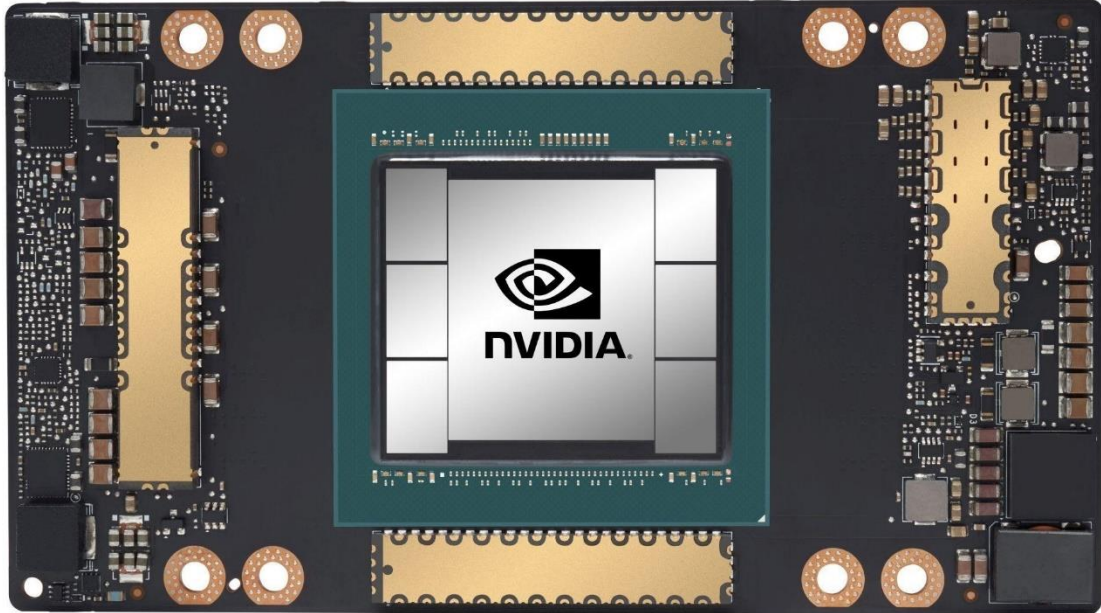
resolution	refresh rate (Hz)	HDMI 2.0	DisplayPort 1.4
1440p	165	Yes	Yes
1440p	240	No	Yes
4k	60	Yes	Yes
4k	120	no	Yes

Display port, which may be less known than HDMI. Compared to HDMI, DisplayPort does not require certification and licensing fees, so any manufacturer can make it. Like HDMI, it can transmit video and audio signals at the same time, but there are still some slight differences.

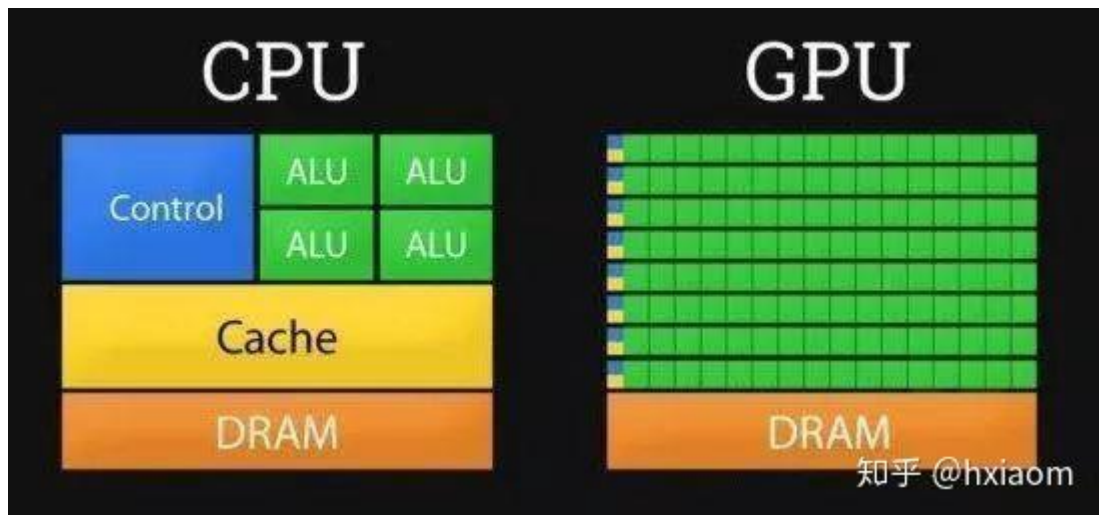
In the latest HDMI 2.1 and DisplayPort 2.0, the highest picture quality can reach 8k, but the bandwidth of DP is even higher than HDMI, and the data transfer volume is larger, meaning that DP can provide a better screen update rate with the same resolution, so DP is relatively more mainstream in the gaming market now.

### 3. Applications

- Deep Learning



For deep learning training, GPUs have become the best choice for gas pedals. Most computations are essentially parallel floating-point computations, i.e., large matrix multiplications that require significant memory bandwidth and size for optimal performance, and these requirements are very much in line with those of HPC.



The evolution of CUDA and NVIDIA's computing business coincided with the advancement of research in machine learning, which only became "deep learning" again around 2006, and GPU acceleration of neural network models provided orders of magnitude of acceleration compared to CPUs, which in turn re-promoted deep learning to its current buzzword. Meanwhile, NVIDIA's graphics competitor ATI was acquired by AMD in 2006; OpenCL 1.0 was only released in 2009, the same year AMD divested its GlobalFoundries fab.

As Deep Learning researchers and academics successfully used CUDA to train neural network models faster, NVIDIA only released their cuDNN library of optimized deep learning primitives, with many HPC-centric BLAS (Basic Linear Algebra

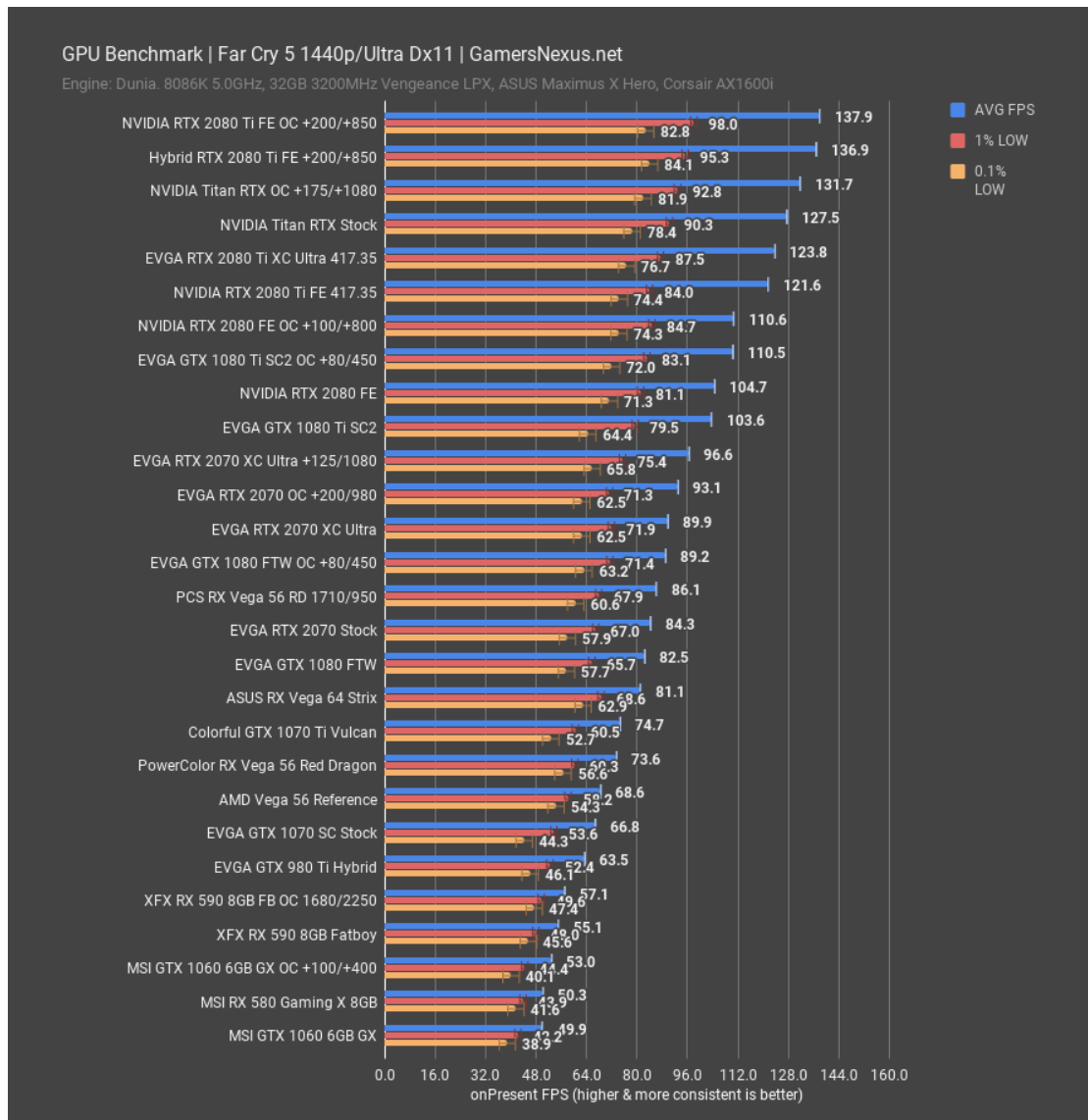
subroutines) and corresponding cuBLAS precedents. cuDNN brings researchers to create and optimize CUDA code to abstract away the needs of researchers to improve DL performance. As for AMD's counterpart, MIOpen, which was only released under the ROCm umbrella in 2017, the current basis for support is not comprehensive.

So in that sense, while both NVIDIA and AMD's underlying hardware are suitable for DL acceleration, NVIDIA GPUs end up being the reference implementation for deep learning.



## ● Gaming

In terms of gaming performance, the Titan RTX was also the top graphics card at the time, with the most powerful scores in all games or benchmarks.



(At that time, the 20 series had not yet been released)

Although he is still the strongest among all graphics cards, but his price is also very high. The performance in games is not

much better than other graphics cards. Although it can improve the performance of the game by 10%, it may cost more than 100% of the price. Unless you are already wealthy, if you want to play games, it is more attractive to choose a general gaming graphics card.

#### **4. Conclusion**

When compared to the Titan RTX benchmark, the GeForce RTX 2080 Ti performs at about 89% of the performance of the Titan RTX, while the GeForce RTX 2080 performs at about 78% of the Titan RTX, and the previously launched GeForce RTX 1080 performs at only 60% of the Titan RTX.

As mentioned in the previous chapter, the Titan RTX is not suitable for gamers because of its low price.

In fact, after the launch of 2080ti, even in the field of deep learning, few people would choose to use Titan RTX. many 2080Ti can be installed in the server to replace Titan because it is not the best choice in terms of price/performance ratio. This card was only brilliant for a while and was almost replaced by the newly released 2080Ti after only a few months.

## 5. Reference

<https://www.nvidia.com/en-us/deep-learning-ai/products/titan-rtx/>

<https://www.notebookcheck.net/NVIDIA-TITAN-RTX-Desktop-GPU-Review.400241.0.html>

<https://www.tomshardware.com/reviews/nvidia-titan-rtx-deep-learning-gaming-tensor,5971.html>

<https://www.anandtech.com/show/13668/nvidia-unveils-rtx-titan-2500-top-turing>

<https://www.gamersnexus.net/hwreviews/3418-nvidia-titan-rtx-review-overclocking-gaming-power-thermals>

<https://kknews.cc/zh-tw/game/2z2bv69.html>

<https://zhuanlan.zhihu.com/p/355146807>