

PART NUMBER:
TCSK80M-PB

NVIDIA TESLA K80 by PNY
THE WORLD'S FASTEST ACCELERATOR
FOR DATA ANALYTICS AND SCIENTIFIC



Accelerate your most demanding single and double precision workloads in scientific computing, seismic processing, and data analytics applications by upgrading to the NVIDIA Tesla K80 dual-GPU accelerator. It delivers up to 2.2x faster performance than the Tesla K20X, up to 2.5x faster performance than the Tesla K10, and up to 10x faster performance than CPUs on real-world applications.

The Tesla K80 accelerator delivers more than 2x application speed-up compared to the previous generation of accelerators, and up to 10x faster performance compared to CPUs. With exclusive features like 24 GB of GDDR5 memory, 480 GB/s memory bandwidth, and improved GPU Boost technology, the Tesla K80 delivers the computational horsepower that allows you to crunch through petabytes of data and run simulations faster than ever before.

Tesla GPU Accelerators are built on the NVIDIA Kepler™ compute architecture and powered by CUDA,® the world's most pervasive parallel-computing model. This makes them ideal for delivering record acceleration and compute performance efficiency for applications in fields including: Machine Learning and Data Analytics, Seismic Processing, Computational Biology and Chemistry, Weather and Climate Modeling, Image, Video, and Signal Processing, Computational Finance/Physics, CAE and CFD.

The Kepler-based Tesla family of GPUs is part of the innovative Tesla Accelerated Computing Platform. As the leading platform for accelerating data analytics and scientific computing, it combines the world's fastest GPU accelerators, the widely used CUDA parallel computing model, and a comprehensive ecosystem of software developers and software vendors.

The innovative design of the TESLA K80 compute architecture includes:

Zero-power Idle

Increases data center energy efficiency by powering down idle GPUs when running legacy nonaccelerated workloads.

2x Shared Memory and 2x Register File

Increases effective throughput and bandwidth with 2x shared memory and 2x register file compared to the K40.

GPU Boost

Enables the end-user to convert power headroom to higher clocks and achieve even greater acceleration for various HPC workloads. Dynamically scales GPU clocks for maximum application performance and improved energy efficiency

TESLA K80¹ - PRODUCT SPECIFICATIONS

MEMORY SIZE PER BOARD	24 GB GDDR5 (12 GB per GPU)
MEMORY INTERFACE	384-bit
MEMORY BANDWIDTH	480 Gb/s
CUDA CORES	4992
PEAK DOUBLE PRECISION FLOATING POINT PERFORMANCE	2.91 Tflops (GPU Boost Clocks) 1.87 Tflops (Base Clocks)
PEAK SINGLE PRECISION FLOATING POINT PERFORMANCE	8.74 Tflops (GPU Boost Clocks) 5.6 Tflops (Base Clocks)
SYSTEM INTERFACE	PCI Express 3.0 x16
MAX POWER CONSUMPTION	300 W
THERMAL SOLUTION	passive heat sink
FORM FACTOR	111.15 mm (H) x 267 mm (L) Dual Slot, Full Height
DISPLAY CONNECTORS	None
POWER CONNECTORS	8-pin CPU power connector
PACKAGE CONTENT	1x Power Adapter (2 x PCIe 8-pin to single CPU 8-pin)
PART NUMBER	TCSK80M-PB

TESLA K80 - Features

ECC MEMORY ERROR PROTECTION	Meets a critical requirement for computing accuracy and reliability in datacenters and supercomputing centers. Both external and internal memories are ECC protected in Tesla K80.
SYSTEM MONITORING FEATURES	Integrates the GPU subsystem with the host system's monitoring and management capabilities such as IPMI or OEM-proprietary tools. IT staff can thus manage the GPU processors in the computing system using widely used cluster/grid management solutions.
L1 AND L2 CACHES	Accelerates algorithms such as physics solvers, ray-tracing, and sparse matrix multiplication where data addresses are not known beforehand
ASYNCHRONOUS TRANSFER WITH DUAL DMA ENGINES	Turbocharges system performance by transferring data over the PCIe bus while the computing cores are crunching other data.
FLEXIBLE PROGRAMMING ENVIRONMENT WITH BROAD SUPPORT OF PROGRAMMING LANGUAGES AND APIS	Choose OpenACC, CUDA toolkits for C, C++, or Fortran to express application parallelism and take advantage of the innovative Kepler architecture.
TESLA GPUBOOST	End-user can convert power headroom to higher clocks and achieve even greater acceleration for various HPC workloads on Tesla K80. Dynamically scales GPU clocks for maximum application performance and improved energy efficiency
2X SHARED MEMORY AND 2X REGISTER FILE	Increases effective throughput and bandwidth with 2x shared memory and 2x register file compared to the K40
ZERO-POWER IDLE	Increases data center energy efficiency by powering down idle GPUs when running legacy nonaccelerated workloads

Software and Drivers

SOFTWARE APPLICATION PAGE	www.nvidia.com/teslaapps
TESLA GPU COMPUTING ACCELERATORS SUPPORTED OS	Linux (64-bit) and Windows (64-bit)
DRIVERS	www.pny.eu/drivers
LEARN MORE ABOUT TESLA DATA CENTER MANAGMENT TOOLS	www.nvidia.com/softwarefortesla

TESLA in Comparison

	TESLA K80 ¹	TESLA K40	TESLA K20X	TESLA K20	TESLA K10 ¹
Peak double-precision floating point performance (board, boost clocks)	2.91 Tflops	1.66 Tflops	1.31 Tflops	1.17 Tflops	0.19 Tflops
Peak single-precision floating point performance (board, boost clocks)	8.74 Tflops	5 Tflops	3.95 Tflops	3.52 Tflops	4.58 Tflops
Number of GPUs	2 x GK210	1 x GK110B	1 x GK110		2 x GK104
Number of CUDA cores	4992	2880	2688	2496	3072
Memory size per board (GDDR5)	24 GB	12 GB	6 GB	5 GB	8 GB
Memory bandwidth for board (ECC off)TM	480 Gb/s	288 Gb/s	250 Gb/s	208 Gb/s	320 Gb/s
Architecture features	SMX, Dynamic Parallelism, Hyper-Q				SMX
System	Servers	Servers & Workstations	Servers	Servers & Workstations	Servers

¹ Specifications are shown as aggregate of two GPU's

