



1998.00 EUR
incl. 19% VAT, plus [shipping](#)

Special offer. New/unused item.

INFERENCE ACCELERATOR

In the new era of AI and intelligent machines, deep learning is shaping our world like no other computing model in history. GPUs powered by the revolutionary NVIDIA Pascal™ architecture provide the computational engine for the new era of artificial intelligence, enabling amazing user experiences by accelerating deep learning applications at scale.

The NVIDIA Tesla P40 is purpose-built to deliver maximum throughput for deep learning deployment. With 47 TOPS (Tera-Operations Per Second) of inference performance and INT8 operations per GPU, a single server with 8 Tesla P40s delivers the performance of over 140 CPU servers. As models increase in accuracy and complexity, CPUs are no longer capable of delivering interactive user experience. The Tesla P40 delivers over 30X lower latency than a CPU for real-time responsiveness in even the most complex models.

REAL-TIME INFERENCE

The Tesla P40 delivers up to 30X faster inference performance with INT8 operations for real-time responsiveness for even the most complex deep learning models.

Name / Brand / Architecture

Manufacturer:	NVIDIA
Model:	Tesla P40
Reference card?:	Yes
Target market segment:	Server / High Performance Computing
Die name:	GP102
Architecture:	Pascal
Fabrication process:	16 nm
Transistors:	12 billion
Bus interface:	PCI-E 3.0 x 16
Launch date:	September 2016

Frequency

Base clock:	1303 MHz
Boost clock:	1531 MHz

Memory specifications

Memory size:	24 GB
Memory type:	GDDR5X
Memory clock:	1251 MHz
Memory clock (effective):	10008 MHz
Memory interface width:	384-bit
Memory bandwidth:	480.38 GB/s

Cores / Texture

CUDA:	6.1
CUDA cores:	3840
ROPs:	96
Texture units:	240

Electric characteristics

Maximum power draw:	250 W
---------------------	-------

Performance

Pixel fill rate:	125.09 Gigapixels/s
Texture fill rate:	312.72 Gigatexels/s
Single precision compute power:	11758.08 GFLOPS
Double precision compute power:	367.44 GFLOPS

External connectors

Power connectors:	1 x 8-pin
-------------------	-----------

Dimensions

Length:	10.51 inches (267 mm)
Width:	Dual-Slot

Other features / Support

Open CL support:	1.2
OpenGL support:	4.5
DirectX support:	12.0
Shader model:	5.0