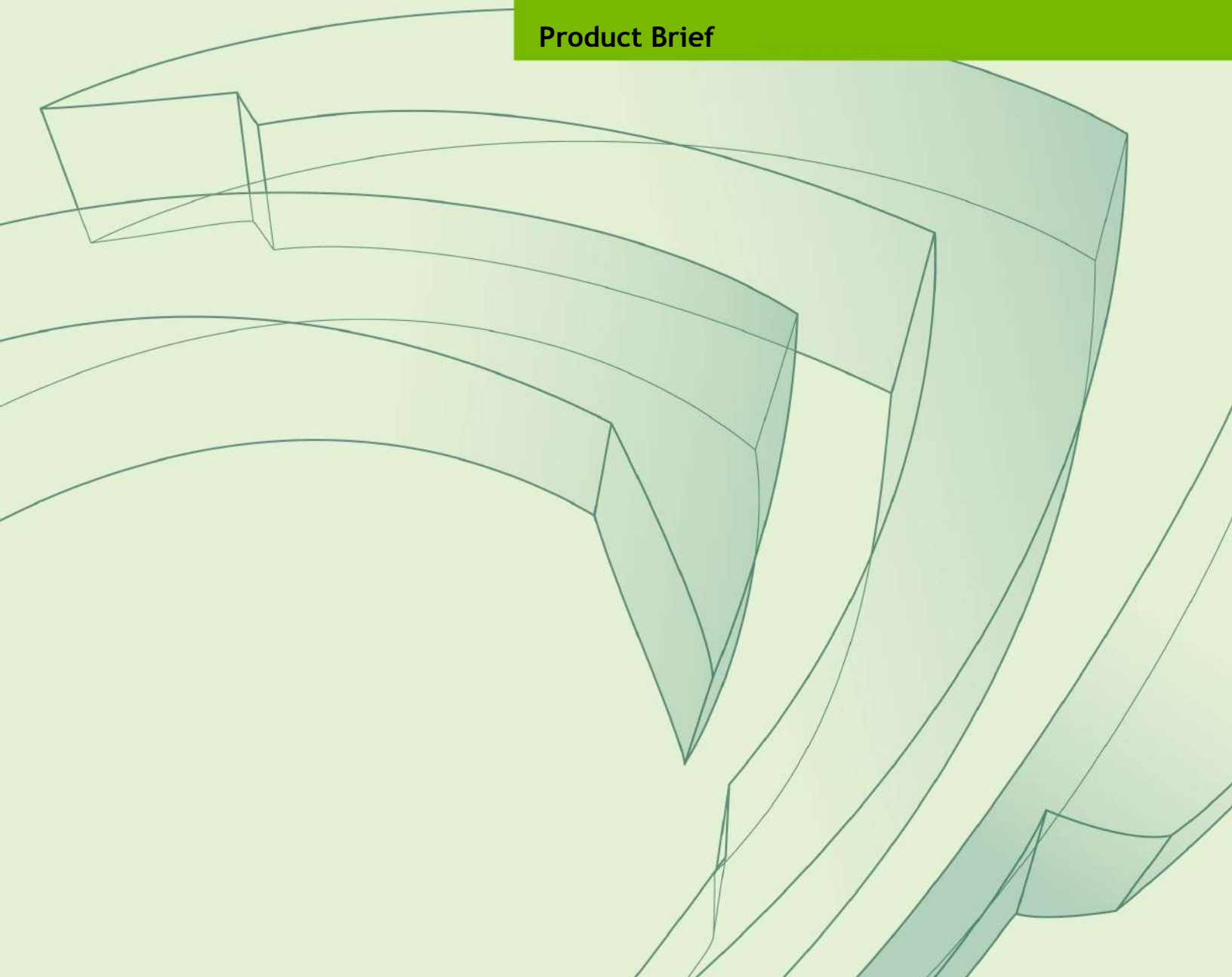




TESLA P40 GPU ACCELERATOR

PB-08338-001_v01 | November 2016

Product Brief



DOCUMENT CHANGE HISTORY

PB-08338-001_v01

| Version | Date | Authors | Description of Change |
|---------|-------------------|---------|-----------------------|
| 01 | November 29, 2016 | WT, SM | Initial Release |

TABLE OF CONTENTS

- Overview** 1
 - Airflow Direction Support 2
- Specifications**..... 4
 - Product Specifications 4
 - Form Factor 6
 - Power Connector Placement 7
 - CPU 8-Pin to PCIe 8-Pin Dongle 8
- Support Information**..... 9
 - Certificates and Agencies 9
 - Certifications 9
 - Agencies 9
 - Languages 10

LIST OF FIGURES

| | | |
|-----------|-------------------------------------------------------|---|
| Figure 1. | NVIDIA Tesla P40 Board..... | 2 |
| Figure 2. | Tesla P40 (PG610 SKU 200): Left-to-Right Airflow..... | 2 |
| Figure 3. | Tesla P40 (PG610 SKU 200): Right-to-Left Airflow..... | 3 |
| Figure 4. | Tesla P40 Board Dimensions | 6 |
| Figure 5. | CPU 8-Pin Power Connector..... | 7 |
| Figure 6. | CPU 8-Pin to PCIe 8-Pin Dongle..... | 8 |

LIST OF TABLES

| | | |
|----------|---------------------------------------------------------|----|
| Table 1. | Product Specifications | 4 |
| Table 2. | Memory Specifications | 5 |
| Table 3. | Software Specifications..... | 5 |
| Table 4. | Board Environmental and Reliability Specifications..... | 6 |
| Table 5. | Supported Auxiliary Power Connections | 7 |
| Table 6. | Languages Supported | 10 |

OVERVIEW

The NVIDIA® Tesla® P40 GPU Accelerator is a dual-slot 10.5 inch PCI Express Gen3 graphics card based on a high-end NVIDIA® Pascal™ graphics processing unit (GPU). The Tesla P40 GPU Accelerator has 24 GB GDDR5 memory and a 250 W maximum power limit.

The Tesla P40 GPU Accelerator is offered as a 250 W passively cooled board that requires system air flow to properly operate the card within its thermal limits. It is designed for single precision GPU compute tasks as well as to accelerate graphics in virtual remote workstation environments. A new feature of the Tesla P40 GPU Accelerator is the support of the “INT8” instruction which is optimized for deep learning inference. As a result, Tesla P40 delivers 47 TOPS (Tera-operations per second) of inference performance which enable high throughput and real-time inference services.

In addition, deep learning models today are trained on GPU servers but typically deployed in separate cluster for inference. The Tesla P40 dramatically simplifies operations such that the same server can now be used for both high performance training and inference.

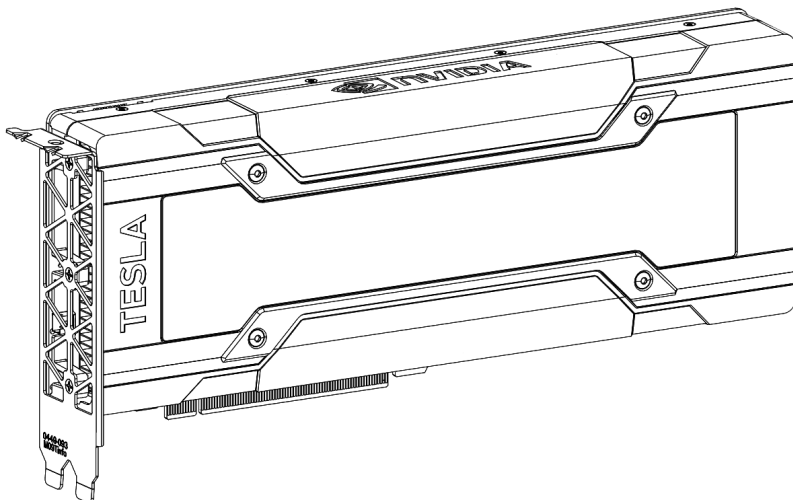


Figure 1. NVIDIA Tesla P40 Board

AIRFLOW DIRECTION SUPPORT

The Tesla P40 board supports bidirectional airflow. The following diagrams illustrate the airflow options supported.

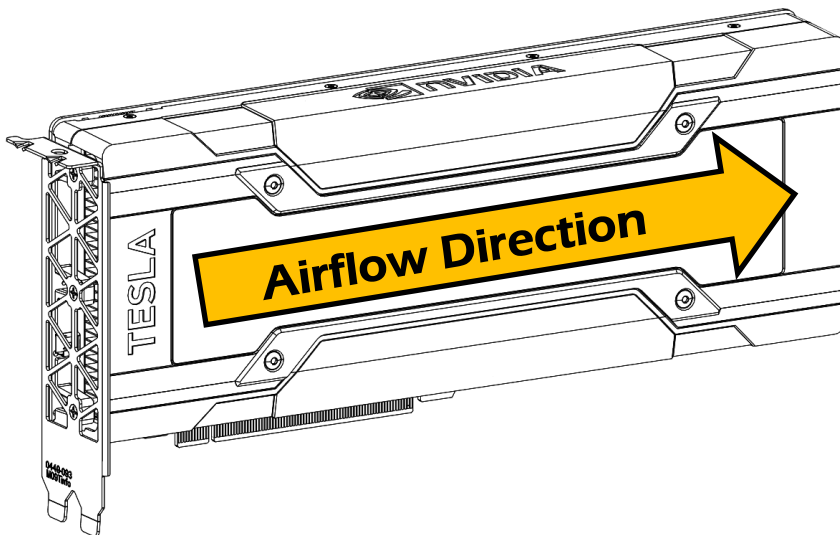


Figure 2. Tesla P40 (PG610 SKU 200): Left-to-Right Airflow

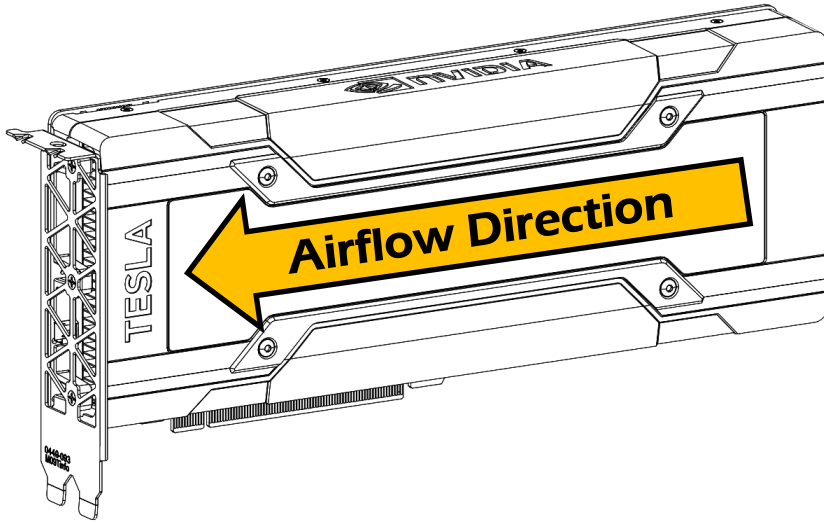


Figure 3. Tesla P40 (PG610 SKU 200): Right-to-Left Airflow

SPECIFICATIONS

PRODUCT SPECIFICATIONS

Table 1 provides the product specifications for the Tesla P40 board.

Table 1. Product Specifications

| Specification | | Description |
|------------------------------|---------------------|------------------------------------------------------------------------------------------|
| Product SKU | | PG610 SKU 200 NVPN: 699-2G610-0200-100 |
| Total board power | | 250 W |
| GPU SKU | | GP102-895-A1 |
| PCI identifiers | | Device ID: 0x1B38 Vendor ID: 0x10DE Sub-Vendor ID: 0x10DE Sub-System ID: 0x11D9 |
| NVIDIA® CUDA® cores | | 3840 |
| GPU clocks | Base | 1303 MHz |
| | Boost | 1531 MHz |
| VBIOS | EEPROM size | 4 Mbit |
| | UEFI | Supported |
| PCI Express interface | | PCI Express 3.0 ×16 Lane and polarity reversal supported |
| Thermal cooling solution | | Passive |
| Physical dimensions | | NVIDIA Form Factor 3.0 compliant 4.376 inches × 10.5 inches, dual-slot |
| Power connectors and headers | | One CPU 8-pin auxiliary power connector |
| Weight | Board | 968 Grams (excluding bracket and extenders) |
| | Bracket with screws | 20 Grams |

| Specification | | Description |
|---------------|----------------------|-------------|
| | Long offset extender | 52 Grams |
| | Straight extender | 42 Grams |

Table 2 provides the memory specifications for the Tesla P40 board.

Table 2. Memory Specifications

| Specification | | Description |
|----------------------|-------------|------------------------|
| Memory clocks | Performance | 3615 MHz |
| | Idle | 405 MHz |
| Memory size | | 24 GB |
| Memory I/O | | 384-bit |
| Memory configuration | | 24 pcs 512M × 16 GDDR5 |
| Memory bandwidth | | Up to 347 GB/s |

Table 3 provides the software specifications for the Tesla P40 board.

Table 3. Software Specifications

| Specification | Description |
|-----------------------------------|-----------------------------------------------------------------------------------------------------------|
| Compatibility modes supported | Compute (default) Graphics |
| Base address | BAR0: 16 MB BAR1: 32 GB (Compute) or 256 MB (Graphics) BAR3: 32 MB I/O BAR: 4 KB (Graphics only) |
| PCI class code | 0x03 - Display controller |
| PCI sub class codes | 0x02 - 3D Controller (Compute) 0x00 - VGA-Compatible Controller (Graphics) |
| ECC support | Support configurable (enabled by default) |
| SMBus (8-bit address) | 0x9E (write), 0x9F (read) |
| SMBus direct access | Supported |
| SMBPBI (SMBus Post Box Interface) | Supported |

Table 4 provides the environment conditions specifications for the Tesla P40 board.

Table 4. Board Environmental and Reliability Specifications

| Specification | Condition |
|-----------------------------------|-------------------------------------------------------------------------------------------------------|
| Operating temperature | 0 °C to 45 °C |
| Storage temperature | -40 °C to 75 °C |
| Operating humidity | 5% to 95% relative humidity |
| Storage humidity | 5% to 95% relative humidity |
| Mean time between failures (MTBF) | Uncontrolled environment: 703379.3 hours at 35 °C Controlled environment: 9913208.1 hours at 35 °C |

FORM FACTOR

The NVIDIA Tesla P40 board conforms to the NVIDIA Form Factor 3.0 specification.

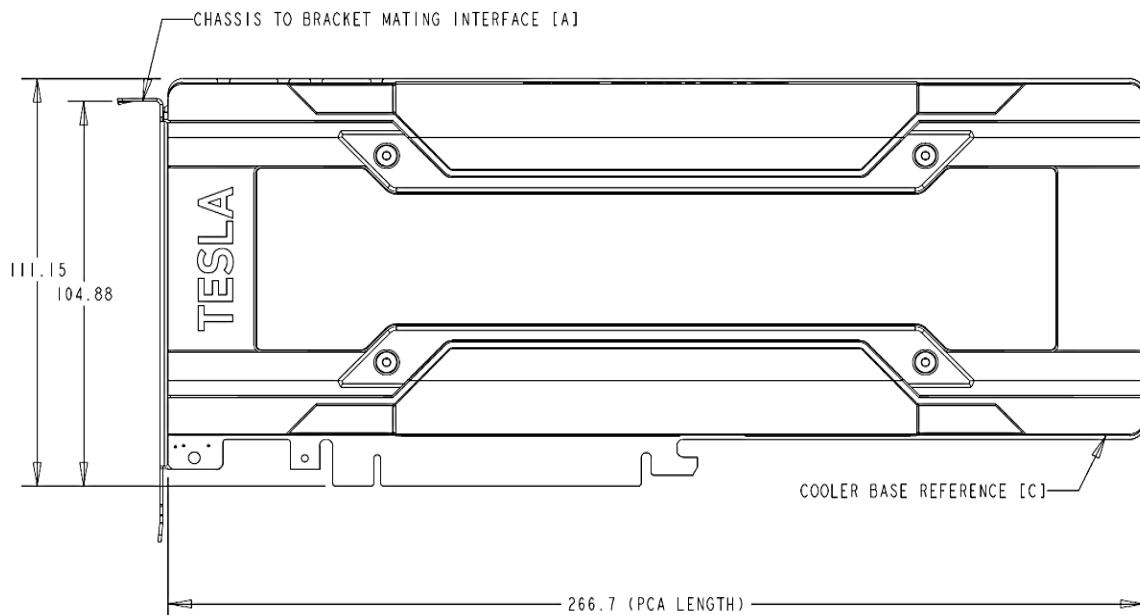


Figure 4. Tesla P40 Board Dimensions

POWER CONNECTOR PLACEMENT

The board provides a CPU 8-pin power connector on the East edge of the board.

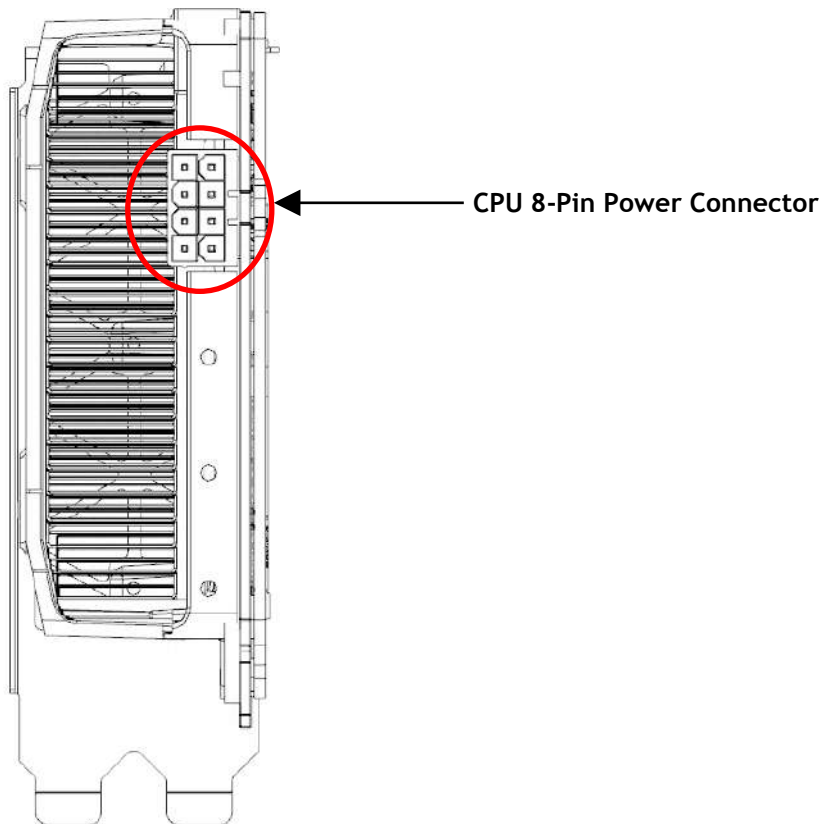


Figure 5. CPU 8-Pin Power Connector

Table 5 lists supported auxiliary power connections for the Tesla P40 board.

Table 5. Supported Auxiliary Power Connections

| Board Connector | PSU Cable |
|--------------------------|---------------------------------------------------------------------------------------------------------------------|
| CPU 8-pin | 1x CPU 8-pin cable |
| CPU to PCIe 8-pin dongle | 2x PCIe 8-pin cable 2x PCIe 6-pin cable ¹ 1x PCIe 8-pin cable and 1x PCIe 6-pin cable ¹ |

Notes:

¹Each PCIe 6-pin cable must be capable of carrying up to 50% of the average power and tolerate 50% of the peak currents.

CPU 8-Pin to PCIe 8-Pin Dongle

Figure 6 lists the pin assignments of the dongle.

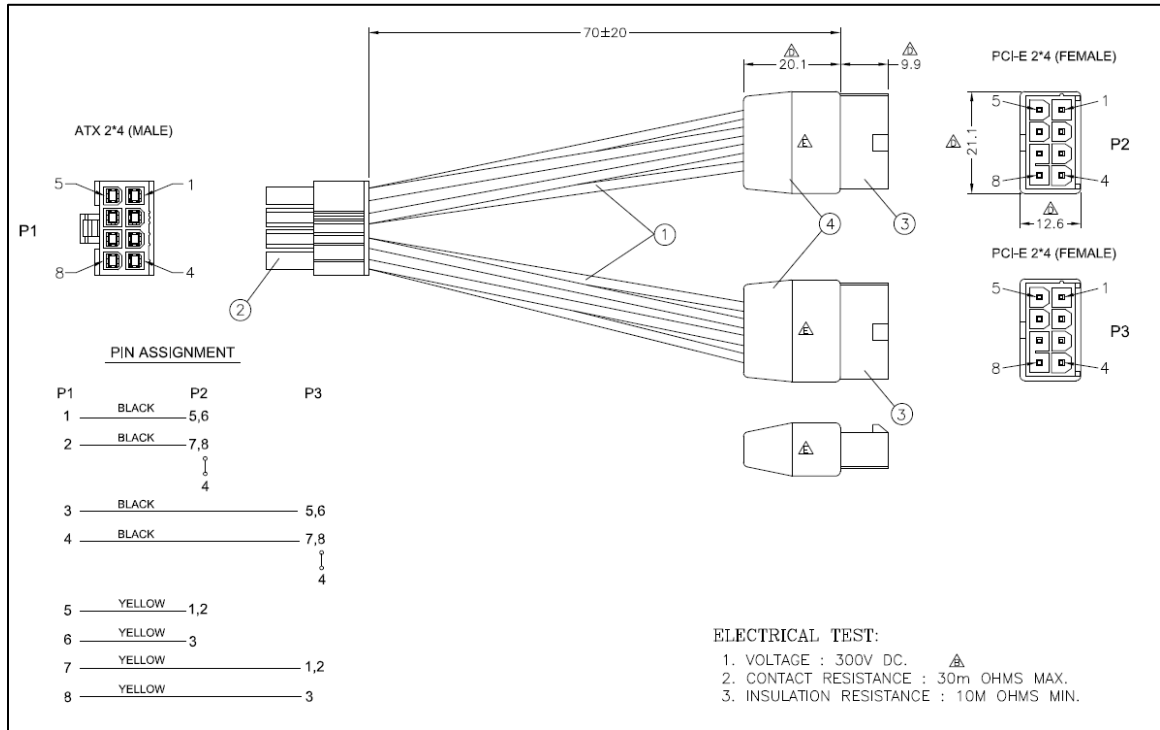


Figure 6. CPU 8-Pin to PCIe 8-Pin Dongle

SUPPORT INFORMATION

CERTIFICATES AND AGENCIES

Certifications

- ▶ Windows Hardware Quality Lab (WHQL):
 - Certified Windows 7, Windows 8.1, Windows 10
 - Certified Windows Server 2008 R2, Windows Server 2012 R2
- ▶ Ergonomic requirements for office work W/VDTs (ISO 9241)
- ▶ EU Reduction of Hazardous Substances (EU RoHS)
- ▶ Joint Industry guide (J-STD) / Registration, Evaluation, Authorization, and Restriction of Chemical Substance (EU) – (JIG / REACH)
- ▶ Halogen Free (HF)
- ▶ EU Waste Electrical and Electronic Equipment (WEEE)

Agencies

- ▶ Australian Communications and Media Authority and New Zealand Radio Spectrum Management (RCM)
- ▶ Bureau of Standards, Metrology, and Inspection (BSMI)
- ▶ Conformité Européenne (CE)
- ▶ Federal Communications Commission (FCC)
- ▶ Industry Canada - Interference-Causing Equipment Standard (ICES)
- ▶ Korean Communications Commission (KCC)
- ▶ Underwriters Laboratories (cUL, UL)
- ▶ Voluntary Control Council for Interference (VCCI)

LANGUAGES

Table 6. Languages Supported

| Languages | Windows ¹ | Linux |
|-------------------------------|----------------------|-------|
| English (US) | Yes | Yes |
| English (UK) | Yes | Yes |
| Arabic | Yes | |
| Chinese, Simplified | Yes | |
| Chinese, Traditional | Yes | |
| Czech | Yes | |
| Danish | Yes | |
| Dutch | Yes | |
| Finnish | Yes | |
| French (European) | Yes | |
| German | Yes | |
| Greek | Yes | |
| Hebrew | Yes | |
| Hungarian | Yes | |
| Italian | Yes | |
| Japanese | Yes | |
| Korean | Yes | |
| Norwegian | Yes | |
| Polish | Yes | |
| Portuguese (Brazil) | Yes | |
| Portuguese (European/Iberian) | Yes | |
| Russian | Yes | |
| Slovak | Yes | |
| Slovenian | Yes | |
| Spanish (European) | Yes | |
| Spanish (Latin America) | Yes | |
| Swedish | Yes | |
| Thai | Yes | |
| Turkish | Yes | |

Note:

¹Windows 7, Windows 8, Windows 8.1, and Windows 10 are supported.

Notice

The information provided in this specification is believed to be accurate and reliable as of the date provided. However, NVIDIA Corporation (“NVIDIA”) does not give any representations or warranties, expressed or implied, as to the accuracy or completeness of such information. NVIDIA shall have no liability for the consequences or use of such information or for any infringement of patents or other rights of third parties that may result from its use. This publication supersedes and replaces all other specifications for the product that may have been previously supplied.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and other changes to this specification, at any time and/or to discontinue any product or service without notice. Customer should obtain the latest relevant specification before placing orders and should verify that such information is current and complete.

NVIDIA products are sold subject to the NVIDIA standard terms and conditions of sale supplied at the time of order acknowledgement, unless otherwise agreed in an individual sales agreement signed by authorized representatives of NVIDIA and customer. NVIDIA hereby expressly objects to applying any customer general terms and conditions with regard to the purchase of the NVIDIA product referenced in this specification.

NVIDIA products are not designed, authorized or warranted to be suitable for use in medical, military, aircraft, space or life support equipment, nor in applications where failure or malfunction of the NVIDIA product can reasonably be expected to result in personal injury, death or property or environmental damage. NVIDIA accepts no liability for inclusion and/or use of NVIDIA products in such equipment or applications and therefore such inclusion and/or use is at customer’s own risk.

NVIDIA makes no representation or warranty that products based on these specifications will be suitable for any specified use without further testing or modification. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer’s sole responsibility to ensure the product is suitable and fit for the application planned by customer and to do the necessary testing for the application in order to avoid a default of the application or the product. Weaknesses in customer’s product designs may affect the quality and reliability of the NVIDIA product and may result in additional or different conditions and/or requirements beyond those contained in this specification. NVIDIA does not accept any liability related to any default, damage, costs or problem which may be based on or attributable to: (i) the use of the NVIDIA product in any manner that is contrary to this specification, or (ii) customer product designs.

No license, either expressed or implied, is granted under any NVIDIA patent right, copyright, or other NVIDIA intellectual property right under this specification. Information published by NVIDIA regarding third-party products or services does not constitute a license from NVIDIA to use such products or services or a warranty or endorsement thereof. Use of such information may require a license from a third party under the patents or other intellectual property rights of the third party, or a license from NVIDIA under the patents or other intellectual property rights of NVIDIA. Reproduction of information in this specification is permissible only if reproduction is approved by NVIDIA in writing, is reproduced without alteration, and is accompanied by all associated conditions, limitations, and notices.

ALL NVIDIA DESIGN SPECIFICATIONS, REFERENCE BOARDS, FILES, DRAWINGS, DIAGNOSTICS, LISTS, AND OTHER DOCUMENTS (TOGETHER AND SEPARATELY, “MATERIALS”) ARE BEING PROVIDED “AS IS.” NVIDIA MAKES NO WARRANTIES, EXPRESSED, IMPLIED, STATUTORY, OR OTHERWISE WITH RESPECT TO THE MATERIALS, AND EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES OF NONINFRINGEMENT, MERCHANTABILITY, AND FITNESS FOR A PARTICULAR PURPOSE. Notwithstanding any damages that customer might incur for any reason whatsoever, NVIDIA’s aggregate and cumulative liability towards customer for the products described herein shall be limited in accordance with the NVIDIA terms and conditions of sale for the product.

Trademarks

NVIDIA, the NVIDIA logo, CUDA, Pascal, and Tesla are trademarks and/or registered trademarks of NVIDIA Corporation in the U.S. and other countries. Other company and product names may be trademarks of the respective companies with which they are associated.

Copyright

© 2016 NVIDIA Corporation. All rights reserved.