



# NVIDIA A2 Tensor Core GPU

Product Brief

# Document History

PB-10727-001\_v02

Version	Date	Authors	Description of Change
01	November 8, 2021	AV, SM	Initial Release
02	March 18, 2022	VK, SM	<ul style="list-style-type: none"><li>• Removed Secure Boot from Table 1</li><li>• Updated Secure Boot and CEC description in Table 3</li><li>• Updated “Root of Trust” section</li><li>• Updated Table 4 “Board Environmental and Reliability Specifications”</li><li>• Added Table 5 “Root of Trust Feature Set”</li></ul>

# Table of Contents

<b>Overview</b> .....	<b>1</b>
<b>Specifications</b> .....	<b>2</b>
Product Specifications .....	2
Environmental and Reliability Specifications .....	4
<b>Airflow Direction Support</b> .....	<b>5</b>
<b>Product Features</b> .....	<b>6</b>
PCI Express Interface Specifications .....	6
PCIe Support .....	6
Polarity Inversion and Lane Reversal Support .....	6
Root of Trust .....	7
Form Factor .....	8
Hockey Stick Board Retention .....	9
<b>Support Information</b> .....	<b>10</b>
Certifications .....	10
Agencies.....	10
Languages .....	11

## List of Figures

Figure 1.	NVIDIA A2 NVFF 5.0 HHL with Full Height Bracket.....	1
Figure 2.	NVIDIA A2 Airflow Direction.....	5
Figure 3.	NVIDIA A2 PCIe Card Dimensions with Full Height Bracket .....	8
Figure 4.	NVIDIA A2 PCIe Card Dimensions with Low Profile Bracket.....	8
Figure 5.	NVIDIA A2 Removable Hockey Stick Tab.....	9
Figure 6.	NVIDIA A2 Removable Hockey Stick Tab Assembly.....	9

## List of Tables

Table 1.	Product Specifications .....	2
Table 2.	Memory Specifications .....	3
Table 3.	Software Specifications.....	3
Table 4.	Board Environmental and Reliability Specifications .....	4
Table 5.	Root of Trust Feature Set.....	7
Table 6.	Languages Supported .....	11

---

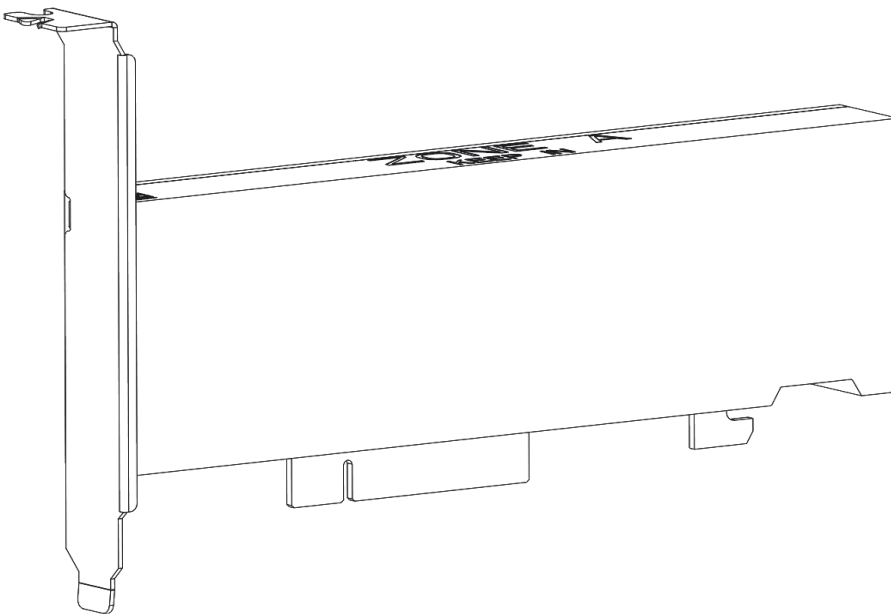
# Overview

The NVIDIA A2 Tensor Core GPU is a compact, lower power product, that delivers entry-level acceleration for Deep Learning, Graphics and Video processing in any server. It is a half-height (low profile), half-length, single slot card featuring 16 GB of GDDR6 memory and a 60 W maximum power limit. The A2 supports x8 PCIe Gen4 connectivity. It is a passively cooled card with a superior thermal design that requires system airflow to operate and handles challenging ambient environments with ease (NEBS-3 capable).

The NVIDIA A2 is powered by the NVIDIA Ampere Architecture. It provides revolutionary multi-precision performance to accelerate deep learning and machine learning training, as well as inference, video transcoding, AI audio and video effects, rendering, data analytics, virtual workstation, virtual desktop, and many other workloads.

As part of NVIDIA AI, the A2 supports all AI frameworks and network models, delivering dramatic performance and efficiency that maximizes the utility of at-scale deployments.

**Figure 1. NVIDIA A2 NVFF 5.0 HHL with Full Height Bracket**



---

# Specifications

## Product Specifications

Table 1 through Table 3 provide the product, memory, and software specifications for the NVIDIA A2 PCIe card.

Table 1. Product Specifications

Specification	NVIDIA A2
Product SKU	PG179 SKU 220 NVPN: 699-2G179-0220-xxx
Total board power	60 W default 60 W maximum 40 W minimum
Thermal solution	Passive
Mechanical Form Factor	HHHL-SS (half-height, half-length, single-slot)
PCI Device IDs	Device ID: 0x25B6 Vendor ID: 0x10DE Sub-Vendor ID: 0x10DE Sub-System ID: 0x157E
Four part ID (VID:DEVID:SVID:SSID) <sup>1</sup>	10DE:25B6:10DE:157E
GPU clocks	Base: 1440 MHz Maximum boost: 1770 MHz
VBIOS	EEPROM size: 16 Mbit UEFI: Supported
PCI Express interface	Physical x8 PCIe lanes PCIe Gen4 x8, x4; Gen3 x8 Lane and polarity reversal supported
Performance States	P0, P8
Zero Power	Not supported
NEBS readiness	Supported

Specification	NVIDIA A2
Weight	Board: 260 Grams (excluding bracket) Bracket (Full height) with screws: 14 Grams Bracket (Half height) with screws: 9 Grams
Note: <sup>1</sup> The NVIDIA A2 is uniquely identified by its complete four part ID.	

Table 2. Memory Specifications

Specification	Description
Memory clock	6251 MHz
Memory type	GDDR6
Memory size	16 GB
Memory bus width	128 bits
Peak memory bandwidth	200 GB/sec

Table 3. Software Specifications

Specification	Description <sup>1</sup>
SR-IOV support	Supported: 16 VF (virtual functions)
BAR address (physical function)	BAR0: 16 MiB <sup>1</sup> BAR1: 16 GiB <sup>1</sup> BAR3: 32 MiB <sup>1</sup>
BAR address (virtual function)	BAR0: 4 MiB (256 KiB per VF) <sup>1</sup> BAR1: 32 GiB, 64-bit (2 GiB per VF) <sup>1</sup> BAR3: 512 MiB, 64-bit (32 MiB per VF) <sup>1</sup>
Message signaled interrupts	MSI-X: Supported MSI: Not supported
ARI Forwarding	Supported
Driver support	R470.82 or later
Secure Boot	Supported (See "Root of Trust" section)
CEC firmware	v6.07 or later (for CEC-enabled cards)
NVIDIA® CUDA® support	CUDA 11.1 or later
Virtual GPU software support	Supports vGPU 14.0 or later
NVIDIA-Certified Systems™ Program	NVIDIA-Certified Systems v2.5 or later
PCI class code	0x03 – Display Controller
PCI sub-class code	0x02 – 3D Controller
ECC support	Enabled (by default); can be disabled using software

Specification	Description <sup>1</sup>
SMBus (8-bit address)	0x9E (write), 0x9F (read)
SMBus direct access	Supported
SMBPBI (SMBus Post-Box Interface)	Supported

Note:

<sup>1</sup>The KiB, MiB, and GiB notation emphasizes the “power of two” nature of the values. Thus,

- 256 KiB = 256 x 1024
- 16 MiB = 16 x 1024<sup>2</sup>
- 64 GiB = 64 x 1024<sup>3</sup>

## Environmental and Reliability Specifications

Table 4 provides the environment conditions specifications for the A2 PCIe card.

Table 4. Board Environmental and Reliability Specifications

Specification	Description
Ambient operating temperature	0 °C to 50 °C
Ambient operating temperature (short term) <sup>1</sup>	-5 °C to 55 °C
Storage temperature	-40 °C to 75 °C
Operating humidity (short term) <sup>1</sup>	5% to 93% relative humidity
Operating humidity	5% to 85% relative humidity
Storage humidity	5% to 95% relative humidity
Mean time between failures (MTBF)	Uncontrolled environment: <sup>2</sup> 2,502,369 hours at 35 °C Controlled environment: <sup>3</sup> 3,270,359 hours at 35 °C

Notes: Specifications in this table are applicable up to 6000 feet.

<sup>1</sup>A period not more than 96 hours consecutive, not to exceed 15 days per year.

<sup>2</sup>Some environmental stress with limited maintenance (GF35).

<sup>3</sup>No environmental stress with optimum operation and maintenance (GB35).

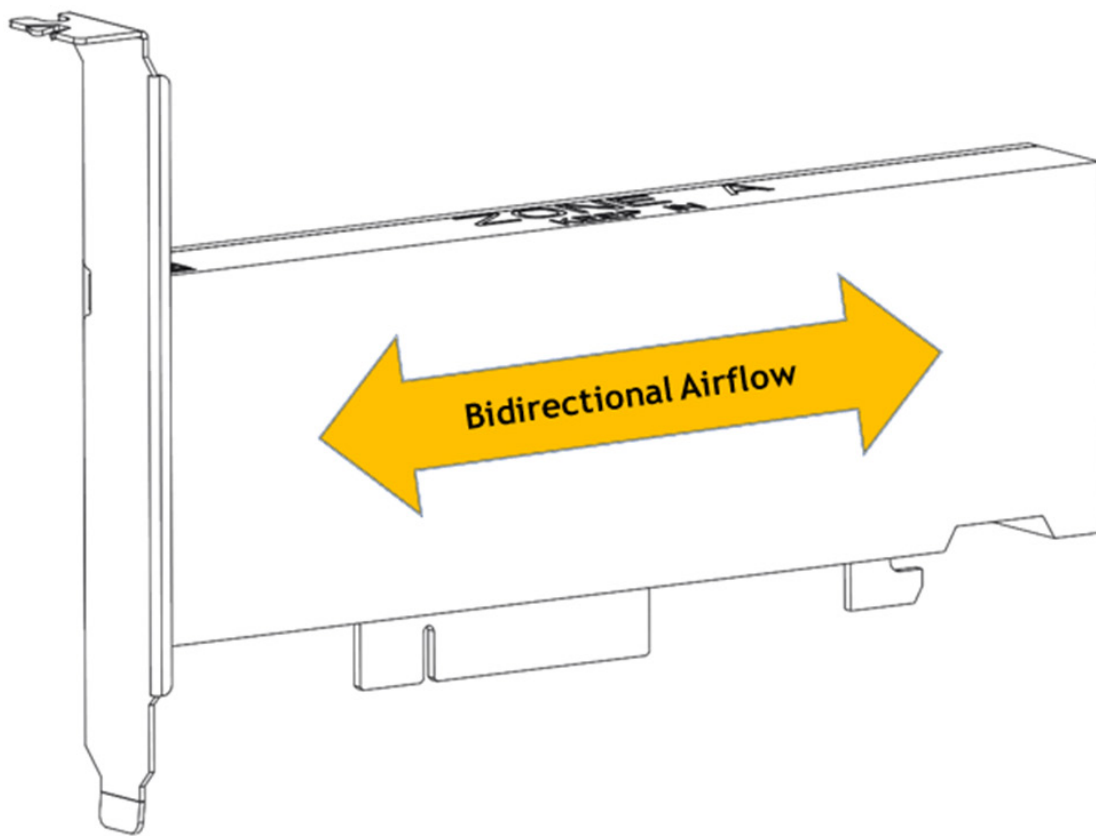


---

# Airflow Direction Support

The NVIDIA A2 PCIe card employs a bidirectional heat sink, which accepts airflow either left-to-right or right-to-left directions.

Figure 2. NVIDIA A2 Airflow Direction



---

# Product Features

## PCI Express Interface Specifications

The following subsections describe the PCIe interface specifications for the A2 PCIe card.

### PCIe Support

The A2 card supports PCIe Gen4. Gen4 x8 interface should be used when connecting to the A2 PCIe card.

### Polarity Inversion and Lane Reversal Support

Lane Polarity Inversion, as defined in the PCIe specification, is supported on the A2 PCIe card.

Lane Reversal, as defined in the PCIe specification, is supported on the A2 PCIe card. When reversing the order of the PCIe lanes, the order of both the Rx lanes and the Tx lanes must be reversed.

## Root of Trust

The NVIDIA A2 GPU has a primary root of trust within the GPU chip that provides the following:

- ▶ Secure boot
- ▶ Secure firmware upgrade
- ▶ Firmware rollback protection
- ▶ Ability to disable In-band firmware update (established after each GPU reset)
- ▶ Secure application processor recovery

In addition, NVIDIA offers some A2 boards with an onboard CEC1712 chip, acting as a secondary root of trust, extending the security capabilities allowing for firmware attestation, key revocation, and out-of-band firmware updates. The CEC1712 device authenticates the contents of the GPU firmware ROM before permitting the GPU to boot from its ROM. For CEC1712-enabled cards, the root of trust feature occupies up to two I2C addresses (in addition to the SMBus addresses). I2C addresses 0xAA and 0xAC should therefore be avoided for system use.

Identification of the two variants of A2 boards (with or without CEC1712) can be done using the 900-level part number on the back of the GPU or running the `nvidia-smi -q` command.

- ▶ 900-2G179-XXXX-1XX A2 GPUs without CEC1712 (secondary root of trust)
- ▶ 900-2G179-XXXX-0XX A2 GPUs with CEC1712 (secondary root of trust)

The following table shows the features that are available using the primary and secondary root of trust.

**Table 5. Root of Trust Feature Set**

Features	Primary Root of Trust within GPU Chip	Secondary Root of Trust Using External CEC Chip on Board
Secure Boot	Yes	Yes
Secure Firmware Upgrade	Yes	Yes
Firmware Rollback Protection	Yes	Yes
In-Band Firmware Update Disable	Yes <sup>1</sup>	Yes
Key Revocation	No	Yes
Firmware Attestation	No	Yes

Notes:

<sup>1</sup>In-Band Firmware Update Disable” feature must be established after every GPU reset.

# Form Factor

In this product brief, nominal dimensions are shown.

Figure 3. NVIDIA A2 PCIe Card Dimensions with Full Height Bracket

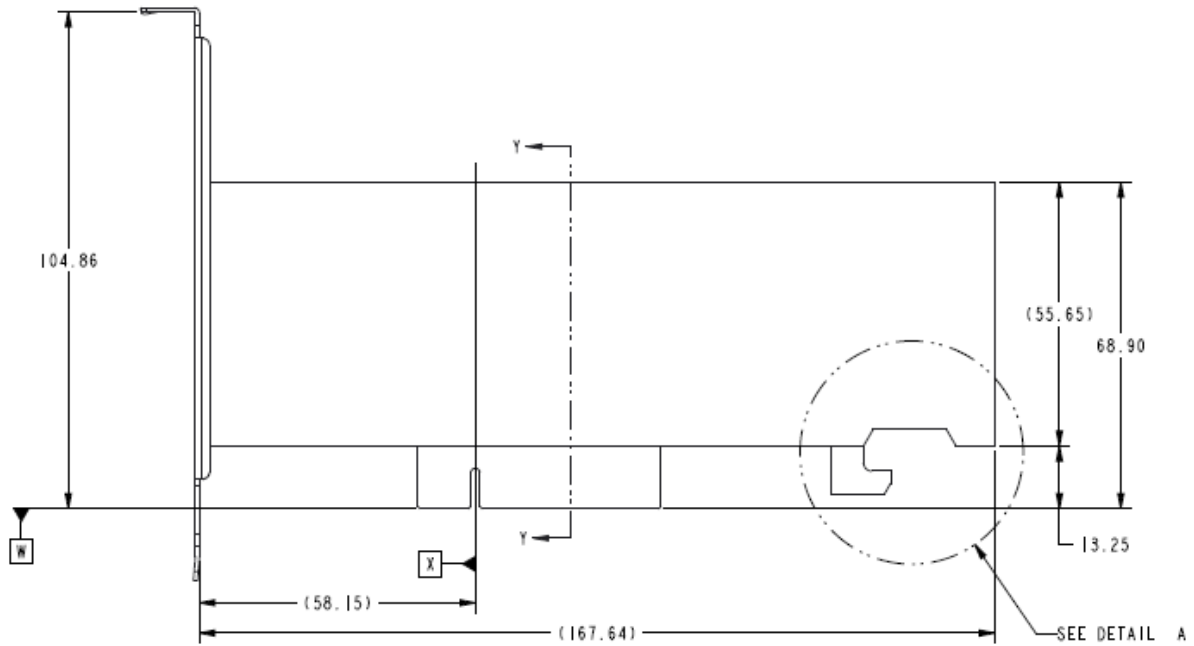
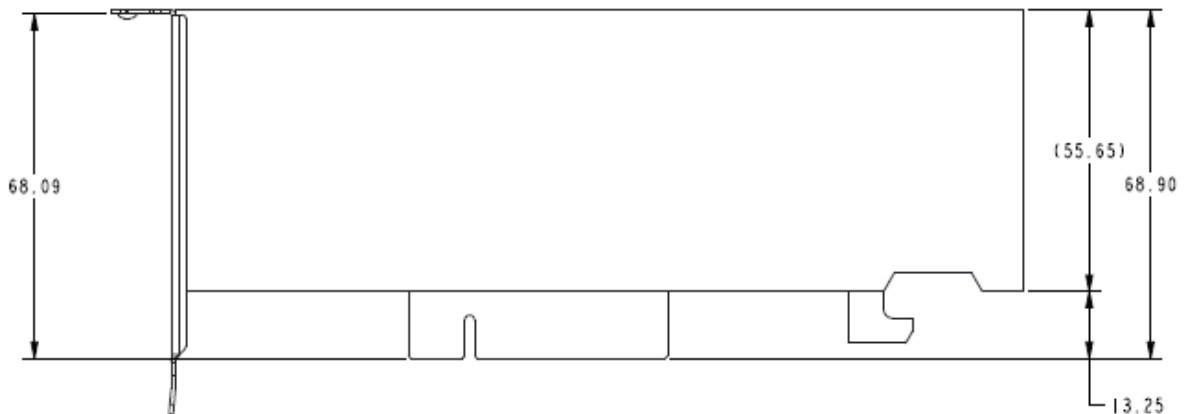


Figure 4. NVIDIA A2 PCIe Card Dimensions with Low Profile Bracket



## Hockey Stick Board Retention

The NVIDIA A2 enables south edge board retention using a “hockey stick” tab located to the east of the PCIe card fingers, as shown in Figure 5. The A2 hockey stick tab is optional. A2 hockey stick tab assembly is depicted in Figure 6.

Figure 5. NVIDIA A2 Removable Hockey Stick Tab

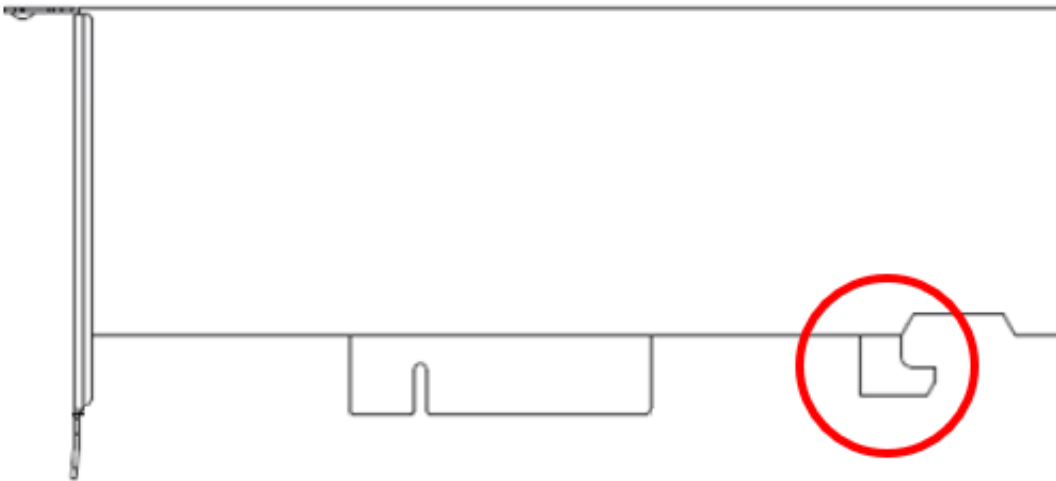
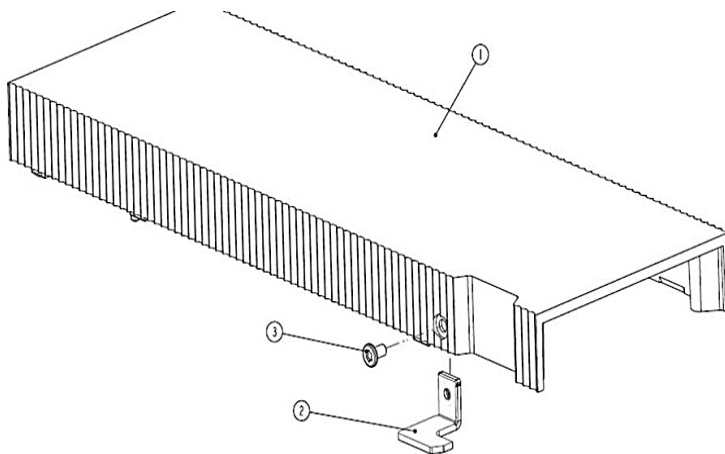


Figure 6. NVIDIA A2 Removable Hockey Stick Tab Assembly



NO.	NVPN	Q'TY	DESCRIPTION
1	367-0541-000	1	COVER-A
2	097-0299-000	1	HOCKEY STICK
3	151-1133-000	1	SCREW-HOCKEY STICK

1. ALIGNMENT FIXTURE FOR HOCKEY-STICK POSITIONING IS REQUIRED
2. TORQUE SCREW TO 2.0-2.5 in-lbf

---

# Support Information

## 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 <sup>1</sup>	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:

<sup>1</sup>Microsoft Windows 7, Windows 8, Windows 8.1, Windows 10, Windows Server 2008 R2, Windows Server 2012 R2, and Windows 2016 are supported.

## Notice

This document is provided for information purposes only and shall not be regarded as a warranty of a certain functionality, condition, or quality of a product. NVIDIA Corporation ("NVIDIA") makes no representations or warranties, expressed or implied, as to the accuracy or completeness of the information contained in this document and assumes no responsibility for any errors contained herein. 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 document is not a commitment to develop, release, or deliver any Material (defined below), code, or functionality.

NVIDIA reserves the right to make corrections, modifications, enhancements, improvements, and any other changes to this document, at any time without notice.

Customer should obtain the latest relevant information 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 ("Terms of Sale"). NVIDIA hereby expressly objects to applying any customer general terms and conditions with regards to the purchase of the NVIDIA product referenced in this document. No contractual obligations are formed either directly or indirectly by this document.

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 this document will be suitable for any specified use. Testing of all parameters of each product is not necessarily performed by NVIDIA. It is customer's sole responsibility to evaluate and determine the applicability of any information contained in this document, ensure the product is suitable and fit for the application planned by customer, and perform 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 document. NVIDIA accepts no 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 document 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 document. 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 document is permissible only if approved in advance by NVIDIA in writing, reproduced without alteration and in full compliance with all applicable export laws and regulations, and accompanied by all associated conditions, limitations, and notices.

THIS DOCUMENT AND 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. TO THE EXTENT NOT PROHIBITED BY LAW, IN NO EVENT WILL NVIDIA BE LIABLE FOR ANY DAMAGES, INCLUDING WITHOUT LIMITATION ANY DIRECT, INDIRECT, SPECIAL, INCIDENTAL, PUNITIVE, OR CONSEQUENTIAL DAMAGES, HOWEVER CAUSED AND REGARDLESS OF THE THEORY OF LIABILITY, ARISING OUT OF ANY USE OF THIS DOCUMENT, EVEN IF NVIDIA HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. 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 Terms of Sale for the product.

## Trademarks

NVIDIA, the NVIDIA logo, CUDA, and NVIDIA Certified Systems 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

© 2021, 2022 NVIDIA Corporation. All rights reserved.