Cambricon

寒 武 纪 科 技

MLU100-D Series Intelligent Processing Card User Manual

Revision History

Document Name		MLU100-D Series Intelligent Processing Card User Manual	
Revision		V1.0.1	
Created Date		2018-09-01	
Revision No.			
	Draft Date	Revision No.	Notes:
0	2018.09.01	V1.0.0	First Release
1	2018.10.23	V1.0.2	Peak Processing Capacity and PCIe BAR updated
2			
3			
4			

Technical supports contact: service@cambricon.com

www.cambricon.com

Content

1. Product Brief

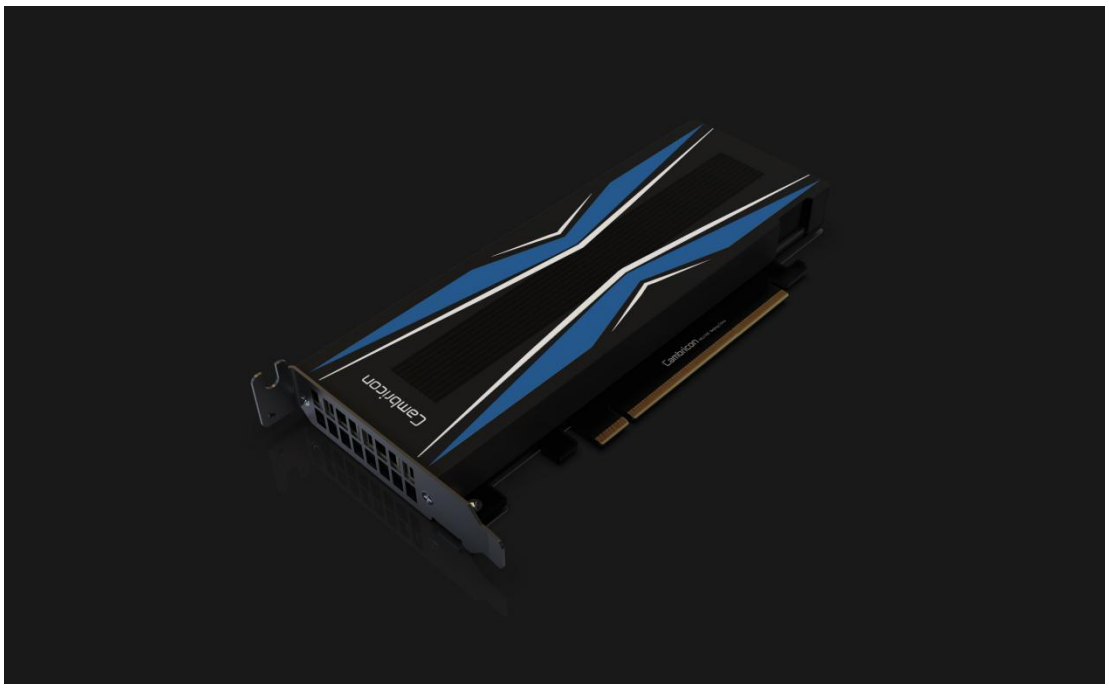

Figure. 1 MLU100-D Series Intelligent Processing Card appearance

Cambricon MLU100-D Series Intelligent Processing Card

The MLU100-D Series is the fourth generation of Intelligent Processing Cards provided by Cambricon Technologies, on which populated one piece of Cambricon MLU100 chip. It can provide extremely high calculating capacity to the Cloud Inference. Compared with classical general-purpose CPU, MLU100-D can achieve a much bigger Performance/Power-consumption ratio when handling the AI (Artificial Intelligence) related works, which makes it a real AI-ready processor.

MLU100-D Series Intelligent Processing Card has a theoretical equivalent Peak Processing Capacity of about 128TOPS(INT8), and it can be integrated into a server or workstation by a PCIe Gen3 x16 single slot. The Card's heat dissipation solution is passive so the server (workstation) system should be in charge of providing cooling airflow. The typical power consumption of the card is about 40-70W. MLU100-D Series Intelligent Processing Card can support up to 16GB DDR4 with ECC capacity.

Cambricon MLU100-D Series Intelligent Processing Card's key features:

1. Versatility —— As a universal intelligent processor, it can support many kinds of Deep Learning technologies and multimodal intelligent processing such as vision, voice and nature language, and it can be used in a wide range.
2. Sparsification —— By bringing sparsity technology into AI chip, it can achieve a theoretical peak performance of 128TOPS (INT8).
3. High efficiency —— Compared with GPU, it has featured special Instruction Set and processor architecture dedicated to deep learning and AI, so it has much higher Performance/Power-consumption ratio.
4. Perfect software development environment —— Cambricon Technologies provides a thorough and mature development environment Cambricon NeuWare for MLU100-D series processors, and it has included a full set of tools such as Application Development, Function Debugging and Performance Optimizations.

5. CE, FCC, UL certificates.

2. Product specifications

2.1 Performance specifications

Table 1 describes the performance specifications of Cambricon MLU100-D Series Intelligent Processing Card.

Card Type	MLU100-D3/D4
Processor Architecture	Cambricon MLUv01
Core Clock	1 GHz
Peak FP16	16 TFLOPS (Without Sparsity) 64 TFLOPS (With Sparsity)
Peak INT8	32 TOPS (Without Sparsity) 128 TOPS (With Sparsity)
Video Coding/Decoding Support	Yes
Total Memory Size	8GB/16GB
Memory Bus Width	256-bit
Memory Bandwidth	102.4 GB/s
System Interfaces	PCI Express Gen3, x16, Supporting Lane Reversal
PCIe ID	PCIe Vendor ID 0xCABC PCIe Device ID 0x0100 PCIe Sub-Vendor ID0xCABC PCIe Sub-System ID0x00C3/0x00C4
Outline	167mm(L)*69.2mm(W), Single-Slot, Half-Height Half-Length
Typical Power	40-70W
TDP	75W
ECC Supporting	Yes

Table 1 PCIe Card hardware Specifications

2.2 Software specifications

Table 2 Describes the software specifications of MLU100-D Series Card:

PCIe Base address（MLU100）	PF （1，64bit）： BAR0: 256MB prefetchable BAR2: 16MB prefetchable BAR4: 64MB prefetchable VF（4，64bit）: BAR0: 16MB non-prefetchable BAR2: 16MB non- prefetchable BAR4: 16MB non-prefetchable
ECC Protect	Yes (Enabled by default)
SMBus (8bit Address)	0x8E (write) 0x8F (read)

Table 2 PCIE Card Software Specifications

SMBUS Register is 32-bit wide, and Table 3 describes how to read a register:

DIRECTION	M->S	M->S	S->M	M->S	S->M
BITS	1	8	1	8	1
CONTENT	S	SLAVE ADDRESS（Write）	ACK	REGISTER ADDRESS	ACK

M->S	M->S	S->M	S->M	M->S	S->M	M->S	S->M	M->S	S->M	M->S	M->S
1	8	1	8	1	8	1	8	1	8	1	1
Sr	SLAVE ADDRESS（Read）	ACK	DATA[7:0]	ACK	DATA[15:8]	ACK	DATA[23:16]	ACK	DATA[31:24]	N	P

Table 3 SMBUS register reading

Table 4 describes the definition and address of SMBUS registers:

Registers	Address	Access	Description
Total Card Power	0x01	RO	Card Power consumption, Float Data, Unit W
Card Temperature	0x02	RO	Card Temperature, Float Data, Unit ℃
Chip Temperature	0x03	RO	Chip Temperature, Float Data, Unit ℃
PCIE Vendor ID and Device ID	0xA0	RO	[15:0] Vendor ID [31:16] Device ID
PCIE Sub-Vendor ID and Sub-System ID	0xA1	RO	[15:0] Sub-Vendor ID [31:16] Sub-System ID
Device Name	0xF0	RO	

Vendor Name	0xF1	RO	
Hardware Revision	0xF2	RO	
Firmware Revision	0xF3	RO	
Manufacturing Time	0xF4	RO	
Device ID	0xF5	RO	

Table 4 SMBUS Registers Description

2.3 Working Environment Specifications

Operating Temperature	0℃ ~ 45℃
Storage Temperature	-40℃ ~ 75℃
Operating Humidity	5%—95% Relative Humidity
Storage Humidity	5%—95% Relative Humidity

Table 5 PCIE Card Working Environment

2.4 Outline Dimension Specifications

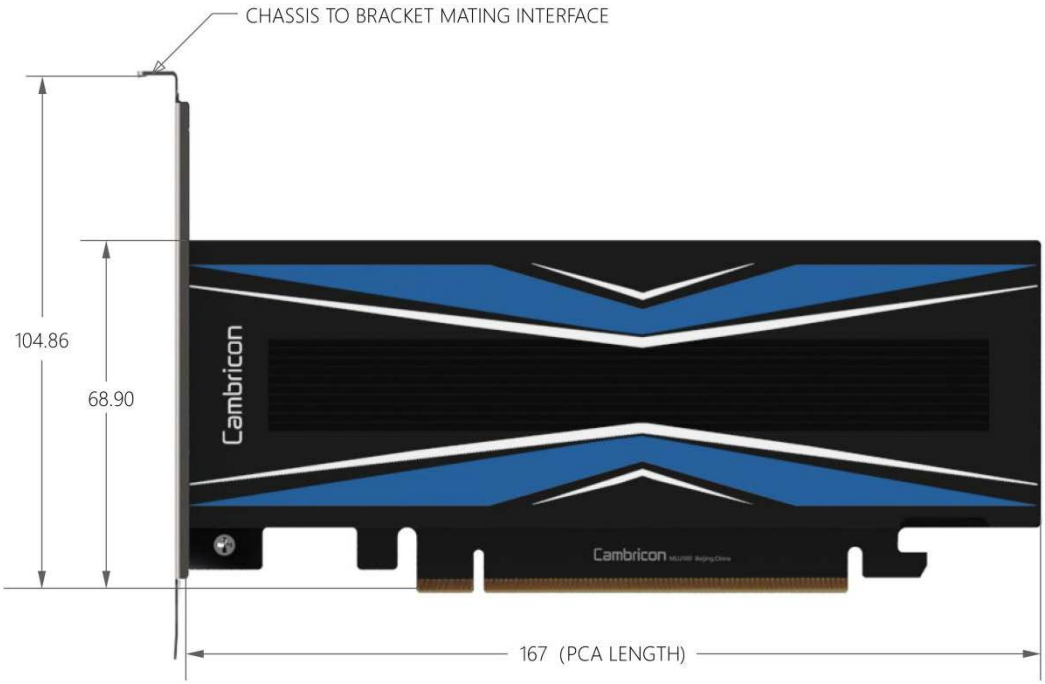

Figure 2 MLU100-D outline dimension（high chassis）

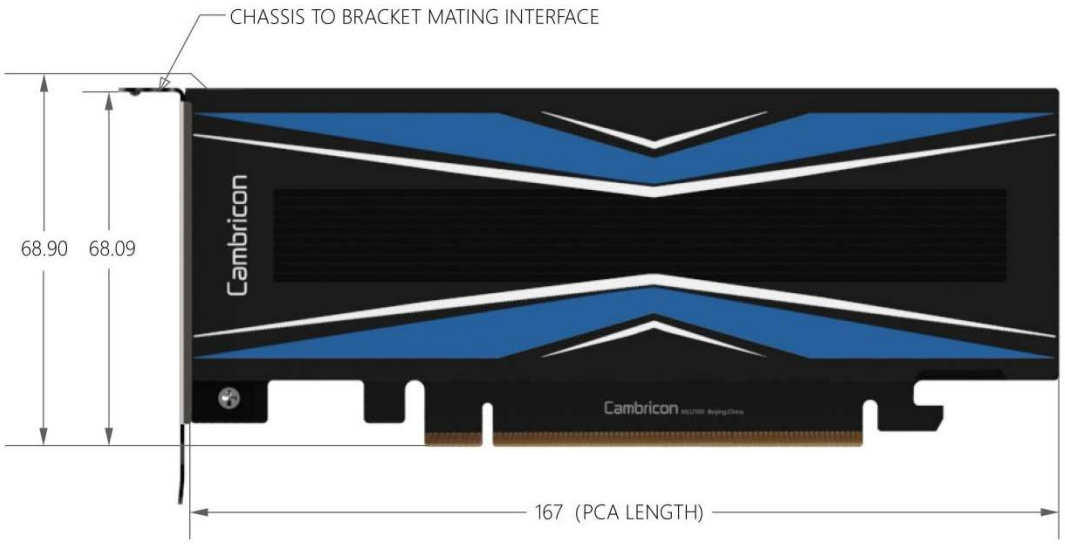

Figure 3 MLU100-D outline dimension（low chassis）

2.5 Heat Dissipation Specifications

MLU100-D Series Intelligent Processing Card is using a passive heat dissipation method. The system designers should pay attentions to below for the Airflow Volume at the entrance and Temperature of Ambient relationship, so as to design a cooling system for the card and guarantee the card is operating within the green recommendation area. The core clock frequency of the chip will be decreased to half when the temperature of chip is over 87℃, and shut down when over 90℃.

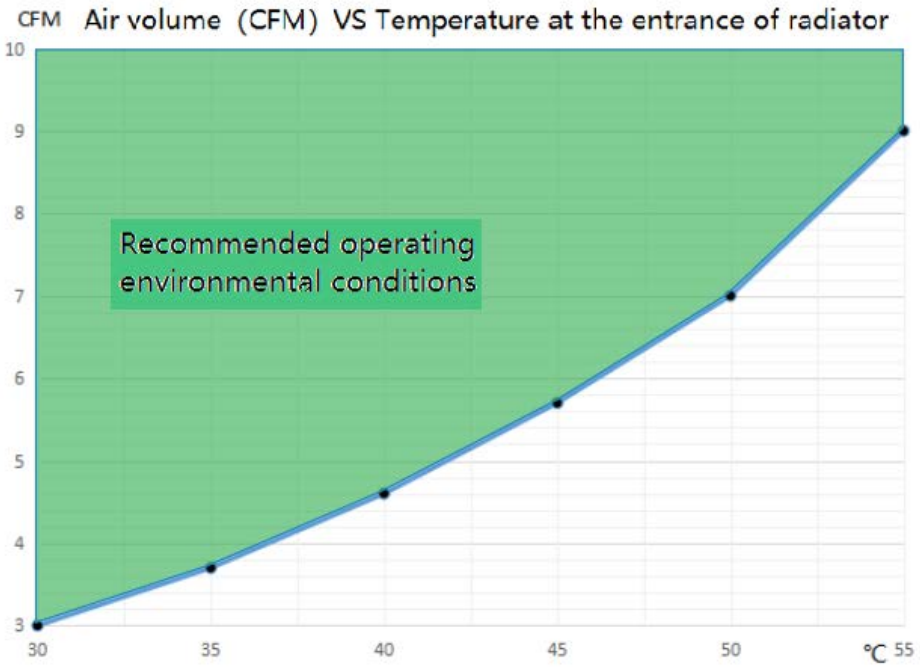

Figure 4 Temperature of Ambient vs Airflow Volume

Below is the supported direction of airflow: It supports either of directions.

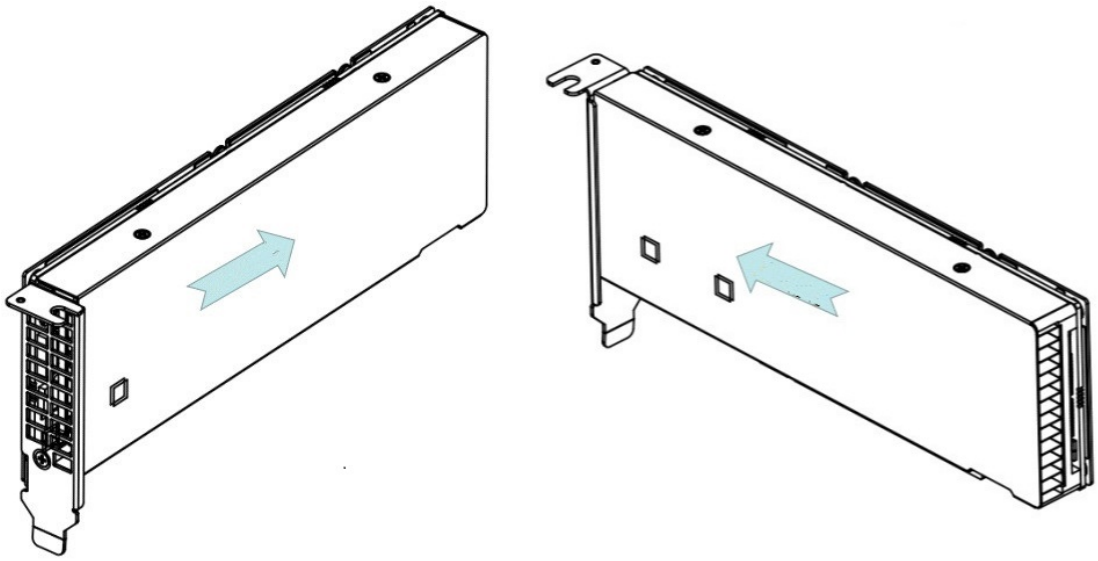

Figure 5 Supported direction of airflow

2.6 Power Supplies Specifications

Power Supply	Min (V)	Normal(V)	Max(V)	Current Avg (A)	Current Peak (A)
PCIe edge connecter（12V）	10.8V	12V	13.2V	5.5A	11A
PCIe edge connecter（3V3）	3.0V	3.3V	3.6V	1.5A	2.25A

Table 7 Power Supplies - Voltages and Currents Specifications

3. Software Development Environment

Cambricon Technologies provides a software development environment: Cambricon NeuWare

NeuWare can fully support all kinds of mainstream programming framework, such as TensorFlow, Caffe, Caffe2, MXNet and ONNX. With above mentioned Programming Frameworks, users can easily and conveniently develop and deploy their Deep Learning Applications on Cambricon MLU100-D Series Intelligent Processing Cards. At the same time, NeuWare provides complete runtime system and driver software to speed up the system integration procedure.

NeuWare also provides a full set of software tools such as Application Development, Function Debugging and Performance Optimizations. The Application Development Tools include Machine Learning Library, Runtime Library, Compiler, Model retraining tools and Domain-Specific (for example Video Analysis) SDK; The Function Debugging Tools can fulfill all the requirements from different levels of Programming Framework and Function Library; The Performance Optimization Tools include tools for performances analysis and system monitoring.

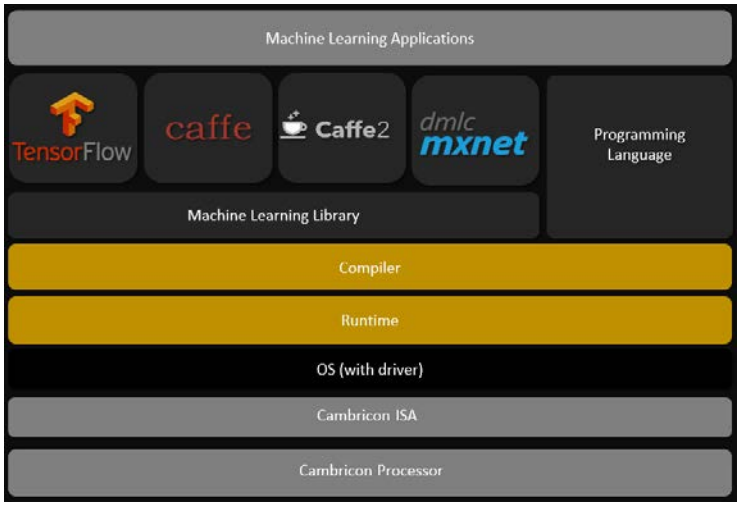

Figure 6 Cambricon NeuWare

For more information, please refer to: www.cambricon.com
Tel: +86-10-83030003
Email: business@cambricon.com
Address: 11 Floors, Block D, Zhizhen Building, 7 Zhichun Road, Haidian District, Beijing

4. FCC notes

This device complies with Part 15 of the FCC Rules.

Operation is subject to the following two conditions: (1) This device may not cause harmful interference, and (2) this device must accept any interference received, including interference that may cause undesired operation.

NOTE: This equipment has been tested and found to comply with the limits for a Class B digital device, pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference in a residential installation. This equipment generates, uses and can radiate radio frequency energy and, if not installed in accordance with the instructions, may cause harmful

interference to radio communications. However, there is no guarantee that interference will not occur in a particular installation. If this equipment does cause interference to radio or television reception, which can be determined by turning the equipment off and on, the user is encouraged to try to correct the interference by one or more of the following measures:

- Reorient or relocate the receiving antenna
- Increase the separation between the equipment and receiver
- Connect the equipment into an outlet on a circuit different from that to which the receiver is connected
- Consult the dealer or an experienced radio/TV technician for help

Information User:
Changes or modifications not expressly approved by the party responsible for compliance could void the user's authority to operate the equipment