

JBOG 0.1 Thermal Simulation

Phase 2



**Author:**

**Jean-Jacques Chanut**, Mechanical engineer Sesame

**Issue date**: 04/14/2020

Contents

[2. Revision History 3](#_Toc38446778)

[3. Scope 4](#_Toc38446779)

[4. Overview of the tested design 4](#_Toc38446780)

[5. NVIDIA Quadro and TESLA Thermal specification 6](#_Toc38446781)

[6. Challenge of the tested design 6](#_Toc38446782)

[7. GPU Modelization 6](#_Toc38446783)

[8. Ambient condition 6](#_Toc38446784)

[9. Simulations 7](#_Toc38446785)

[10.1. Closed casing with 2 GPU in line 7](#_Toc38446786)

[10.2. Closed casing with 4 GPU in line 13](#_Toc38446787)

[10. Conclusion 19](#_Toc38446788)

# Revision History

|  |  |  |
| --- | --- | --- |
| Date | Name | Description |
| 03/31/2019 | JJ Chanut | Creation |
| 04/14/2019 | JJ Chanut | Conclusion |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Scope

This document provides the results of thermal simulations on the Flotherm software of the JBOG solution #2 proposed in the document "JBOG specifications REV 0.1".

The purpose of these simulations is to test the validity of the concept. These results will also be used to define the thermal profile of the JBOG to guide the mechanical and electronic design of the JBOG.

# Overview of the tested design



Figure 1: JBOG concept 0.1

JBOG is a 2OU chassis integrating 4 full length double slot GPUs (NVIDIA Quadro and Tesla products).  
JBOD is masters by a compute node thanks to 4 mini SAS HD cables.

Une image contenant capture d’écran

Description générée automatiquement  
Figure 2: PCIE interface diagram

The PCIe boards are plugged on a vertical riser and are arranged two by two and aligned horizontally on the right side of the server.

On the left side, the main card which integrates the PCIe switch as well as the mini SAS HD ports, a USB port and an RJ45 port.

A close up of a computer

Description automatically generated  
Figure 3: Front view

# NVIDIA Quadro and TESLA Thermal specification

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Value** | **Units** |
| Max Total graphics power (TGP) | 260 | W |
| GPU shutdown temperature | 94 | °C |
| GPU slowdown temperature (50% clock slowdown) | 91 | °C |
| GPU target temperature | 84 | °C |
| GPU maximum operating temperature | 89 | °C |
| Maximum fan inlet temperature | 45 | °C |

# Challenge of the tested design

The horizontal alignment of the GPUs can cause breakdowns at the blowers integrated level because they are in the same direction of air circulation. To avoid this problem, we decided to remove them and add fans at the back of the JBOG chassis. We need to create enough airflow in the chassis to cool the GPUs.

The horizontal alignment also induces that the GPU furthest from the air inlet will be warmer, it is also necessary to pay attention to the inlet temperature in the radiators which must not exceed the limits indicated in the specifications (45 ° C). This can be difficult because we carry out our simulations with an ambient temperature of 35 ° C.

# GPU Modelization

We reproduced a NVIDIA Quadro RTX 8000 card from the data provided by PNY and simulated its integration into a chassis.  
It consumes the most electrical power and requires the most cooling in the professional range of NVIDIA products.

This model contains the board itself and its standard heatsink. We have only modeled the larger components, as well as those that have a determining effect on temperature. This model doesn’t integrate the standard NVIDIA casing as well as the optional front bracket.

NVIDIA heatsink is composed with fins assembled on top of a vapor chamber in contact with the processor die. The flotherm version used in this simulation doesn’t support vapor chamber models, so we replace it with a a material with a thermal conductivity of 5000W/mK.

# Ambient condition

All the simulations will run with these ambient conditions.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Value** | **Units** |
| Ambient temperature | 35 | °C |
| Radiant temperature | 35 | °C |
| Pressure | 1 | ATM |
| Humidity | 50 | % |

# Simulations

## Closed casing with 2 GPU in line

* + 1. Test 1

**Configuration:** This simulation integrates 2 GPU in line, with **the casing open at the right side** of the chassis.  
**2 fans 40x40x56 mm with 37.1 CFM** airflow each are located at the back of the casing.

A screenshot of a cell phone

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Main measurement recorded** | | | |
| **Parameter** | **Value** | **Targeted value** | **Units** |
| GPU 1 temperature | 558 | 84 | °C |
| GPU 2 temperature | 281 | 84 | °C |
| Heatsink 2 air inlet temperature | 67 | 45 | °C |
| Heatsink 1 air inlet speed | 0.05 | NA | m/s |
| Heatsink 2 air inlet speed | 0.35 | NA | m/s |
| Air temperature casing outlet | 37.8 | NA | °C |

A picture containing stationary, implement, pencil

Description automatically generated

Temperature is higher at the front of the chassis

A screenshot of a computer screen

Description automatically generated

There is low speed air flow from the front

Air inlet at the side of the chassis

Most of the air is drawn in from the rear side of the chassis and does not pass through the heat sinks.  
There is not enough air flow in the heat sinks to perform reliable cooling.

We need to close the side to guide the air to pass through the heat sinks.

* + 1. Test2

**Configuration:** This simulation integrates 2 GPU in line, with a casing flushing the sides of the GPUs.  
2 fans 40x40x56 mm with 37.1 CFM airflow each are located at the rear of the casing.

A screenshot of a cell phone

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Main measurement recorded** | | | |
| **Parameter** | **Value** | **Targeted value** | **Units** |
| GPU 1 temperature | 85.3 | 84 | °C |
| GPU 2 temperature | 106 | 84 | °C |
| Heatsink 2 air inlet temperature | 63 | 45 | °C |
| Heatsink 1 air inlet speed | 4.75 | NA | m/s |
| Heatsink 2 air inlet speed | 3.7 | NA | m/s |
| Air temperature casing outlet | 52.2 | NA | °C |

Unlike the first simulation, the temperature is lower at the front of the case.  
The air passes first through the first heatsink and is therefore warmer when it reaches the second heatsink.

Figure 4: Main components temperatures

*A screenshot of a video game

Description automatically generated*

AIR Flow

*A picture containing screenshot

Description automatically generated*

AIR Flow

* ΔT between the 2 GPU is 23.8 °C
* heatsink 2 inlet temperature is too high

So we need to find a way to cool the air at the rear heatsink inlet.

* + 1. Test 3

**Configuration:** This model integrates **2 GPU** in line, with a casing flushing the sides of the GPUs.

2 fans **40x40x56 mm with 37.1 CFM** airflow each are located at the rear of the casing.  
  
Too cool the rear heatsink air inlet temperature.   
To cool the air temperature at the rear heatsink inlet, we make an on the side so that fresh air can be brought into the space between the two processors.

A screenshot of a cell phone

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Main measurement recorded** | | | |
| **Parameter** | **Value** | **Targeted value** | **Units** |
| GPU 1 temperature | 86.2 | 84 | °C |
| GPU 2 temperature | 99.7 | 84 | °C |
| Heatsink 2 air inlet temperature | 69.7 | 45 | °C |
| Heatsink 1 air inlet speed | 5.39 | NA | m/s |
| Heatsink 2 air inlet speed | 3.26 | NA | m/s |

A close up of a device

Description automatically generated

A picture containing implement, stationary, pencil

Description automatically generated

A picture containing pencil

Description automatically generated

* GPUs ΔT is quite lower
* Air flow is speed is higher at the rear heatsink inlet .
* Rear heatsinks inlet temperatures remains too hot but is closer to the target.
* Rear processor temperature remains too hot but is closer to the target

The opening allows the rear processor to cool more efficiently at the expense of the front processor.  
This is because the volume of air intake is now separated into two inlets.

* + 1. Test 4

**Configuration:** Same configuration as the previous simulation but with 2 fans with 50 CFM  
We expect that this higher air volume will reduce the temperatures.

|  |  |  |  |
| --- | --- | --- | --- |
| **Main measurement recorded** | | | |
| **Parameter** | **Value** | **Targeted value** | **Units** |
| GPU 1 temperature | 83.1 | 84 | °C |
| GPU 2 temperature | 88 | 84 | °C |
| Heatsink 2 air inlet temperature | 61 | 45 | °C |
| Heatsink 1 air inlet speed | 5 | NA | m/s |
| Heatsink 2 air inlet speed | 3.94 | NA | m/s |

We are getting closer to the target temperatures, by switching to a 4 GPU system, we will be able to provide more air into the server, so we hope to reach the targets.

* + 1. First serie observations
* The air must be properly guided to optimize the amount of air passing through the radiators.
* It is important to lower the inlet temperature of the second heatsink, this balances the temperature difference between the two GPUs.   
  For this we can provide an ambient air intake between the two heatsinks, and even if we do not reach the objective of 45 °C, we note a significant drop in the temperature of GPU # 2.   
  We expect to get closer to this target temp with larger fans during simulations with 4 GPUs, indeed the air flow will be much higher with 80 mm fans.
* The air flow must be very fast to allow efficient cooling. Indeed, we had to increase the volume flow of the fans.

## Closed casing with 4 GPU in line

* + 1. Test 2.1

**Configuration:** This model integrates **4 GPU in 2 lines**, with an air guide flushing the sides of the GPUs.  
**2 fans 80x80x38 mm with 141.9 CFM** airflow each are located at the back of the casing.

This model does not provide lateral opening at the air guide level, we wish to see the influence of a significant air flow on the behavior of the model.

A screenshot of a cell phone

Description automatically generated

|  |  |  |  |
| --- | --- | --- | --- |
| **Main measurement recorded** | | | |
| **Parameter** | **Value** | **Targeted value** | **Units** |
| GPU 1 temperature | 74 | 84 | °C |
| GPU 2 temperature | 82.5 | 84 | °C |
| GPU 3 temperature | 89.3 | 84 |  |
| GPU 4 temperature | 90.4 | 84 |  |
| Heatsink 2 air inlet temperature | 56.2 | 45 | °C |
| Heatsink 1 air inlet speed | 7.7 | NA | m/s |
| Heatsink 2 air inlet speed | 4.75 | NA | m/s |
| Air temperature casing outlet | 50 | NA | °C |

A picture containing toy

Description automatically generated

A circuit board

Description automatically generated

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

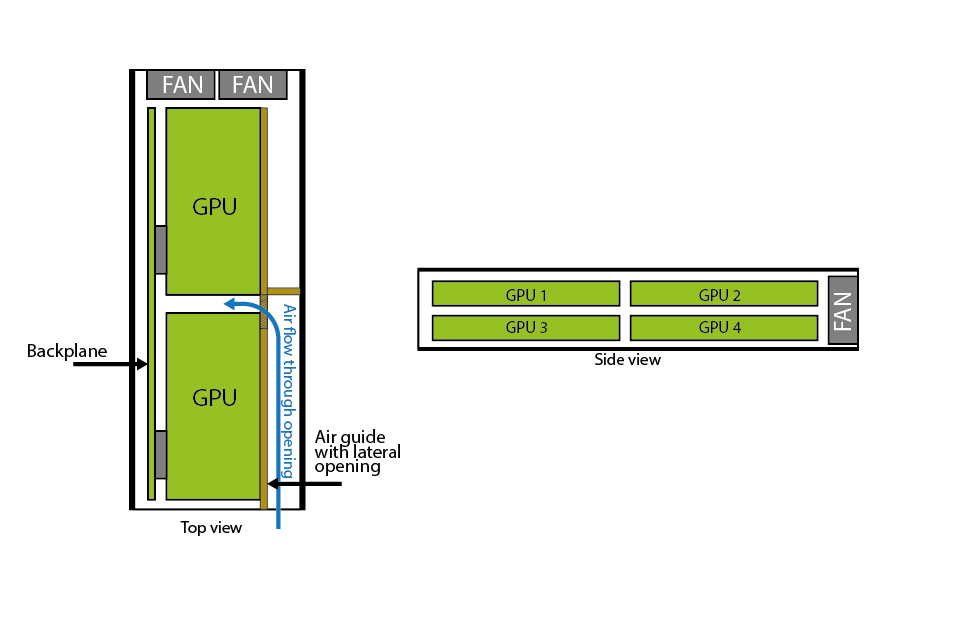
A picture containing toy, colorful

Description automatically generated

* GPU above and below temperatures (1 &3, 2 & 4are relatively similar).
* Air flow is constant in the chassis.
* Rear heatsinks inlet temperatures are too hot.
* GPU at the rear of the casing are too hot.
  + 1. **Test 2.2**

This model integrates **4 GPU in 2 lines**, with an air guide flushing the sides of the GPUs.  
**2 fans 80x80x38 mm with 141.9 CFM** airflow each are located at the back of the casing.

We add a lateral opening at the air guide level.



|  |  |  |  |
| --- | --- | --- | --- |
| **Main measurement recorded** | | | |
| **Parameter** | **Value** | **Targeted value** | **Units** |
| GPU 1 temperature | 78.4 | 84 | °C |
| GPU 2 temperature | 76.4 | 84 | °C |
| GPU 3 temperature | 72.7 | 84 | °C |
| GPU 4 temperature | 84.2 | 84 | °C |
| Heatsink 2 air inlet temperature | 53,6 | 45 | °C |
| Heatsink 1 air inlet speed | 7.7 | NA | m/s |
| Heatsink 2 air inlet speed | 4.75 | NA | m/s |
| Air temperature casing outlet | 42.6 | NA | °C |

**A close up of a device

Description automatically generated**

**A picture containing text

Description automatically generated**

**A close up of text on a black background

Description automatically generated**

**A picture containing toy, room, table

Description automatically generated**

We have relatively similar results to the previous simulation in terms of airflow behavior.

On the other hand, it is faster and allows a more efficient cooling of the processors.

The only concern is that air inlet temperatures of the rear heatsinks are higher than expected.

# Conclusion

The model presented in the specifications needs to be revised, an air guide for the GPUs needs to be installed, and a source of fresh air between the front and rear processors needs to be installed.

The airflow must be extremely fast and therefore requires powerful fans.

Nevertheless, we are unable to reach the target temperature at the rear heatsinks inlets, which is to be put into perspective as our simulations are done at an ambient temperature of 35°C.

It should also be noted that we performed these simulations without the official thermal profile of the GPU from NVIDIA, in fact we reproduced the processor heatsinks by estimating their cooling capacity. We have taken a margin by modeling a vapor chamber with relatively low heat conduction compared to what is possible.

It is necessary to confirm these simulations by physical tests.

These simulations allow us to provide a first thermal model of the JBOG and thus give specifications to start the mechanical design.