

GPTshop.ai GH200 Silent 624GB Manual

Version 1.5 / September 2024

2024 GPTshop.ai UG (limited)

For the latest information and updates please see: <https://gptshop.ai>

Table of contents

Quickstart.....	2
System specifications.....	2
Package Contents	3
System Rear I/O.....	4
Power Sub-System.....	5
Fan speed adjustment.....	6
BMC.....	10
Firmware.....	10
Operating system.....	10
Drivers.....	11
Software.....	12
Nvidia Resources.....	13



Quickstart:

Username and password for BMC:

Username: root

Password: 0penBmc (zero!!!)

Username and password for preinstalled OS (optional):

Username: x

Password: xz

1.) Install OS

Ubuntu Server for ARM: <https://cdimage.ubuntu.com/releases/24.04/release/ubuntu-24.04.1-live-server-arm64+largemem.iso>

2.) Install Drivers

```
curl -O https://us.download.nvidia.com/tesla/560.35.03/NVIDIA-Linux-aarch64-560.35.03.run
```

```
sudo apt install build-essential
```

```
sudo sh NVIDIA-Linux-aarch64-560.35.03.run
```

System specifications:

Nvidia GH200 Grace Hopper Superchip

72-core Nvidia Grace CPU

Nvidia H200 Hopper Tensor Core GPU

480GB of LPDDR5X memory with EEC

144GB of HBM3e memory

624GB of total fast-access memory

NVlink-C2C: 900 GB/s of bandwidth

Programmable from 400W to 1000W TDP (CPU + GPU + memory)

2x High-efficiency 2000W PSU

2x PCIe gen4 M.2 slots on board

2x PCIe gen4 2.5" drive slots (NVMe) without Bluefield-3

4x PCIe gen4 2.5" drive slots (NVMe) with Bluefield-3

3x FHFL PCIe Gen5 x16

1x USB 3.0 port (mini USB hub included: 1x USB 3.0, 2x USB 2.0)

2x RJ45 10GbE ports

1x RJ45 IPMI port

1x Mini display port

1x Micro USB port

Halogen-free LSZH power cables

Stainless steel bolts

25 decibels (configurable)
281 x 553 x 539 mm (11.1 x 21.8 x 21.2")
27 kg (60 lbs)

TPM (optional, SPI mode)
ACPI compliance, S0, S5 support

System rating:
100-120Vac, 50/60Hz, 10A
200-240Vac, 50/60Hz, 10A

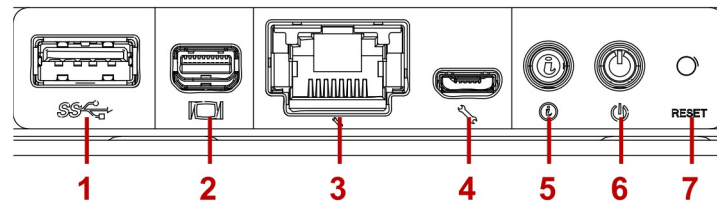
System management: IPMI v2.0 Compliant, on board "KVM over IP" support, Dedicated GbE management NIC port from PHY RTL8211FS to BMC

Operating environment:
Operating temperature: 0°C to 35°C (41°F to 95°F) at 1829M and up to 50°C at 0M. (No performance drop)
Non-operating temperature: -40°C to 70°C (-40°F to 158°F)
Operating relative humidity: 5% to 85%RH
Non-operating relative humidity: 50% to 93%RH

Package Contents







1x GH200 624GB system
2x Power cord
1x Alan key
1x MiniDP to DP adapter
1x Mini USB hub: 1x USB 3.0, 2x USB 2.0

System Rear I/O

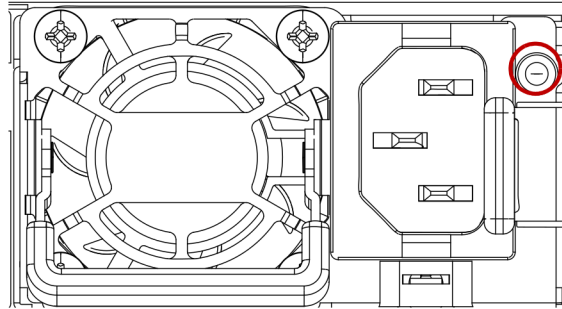


System Rear I/O

System Rear I/O Definition

No.	ICON	NAME	DESCRIPTION
1		USB 3.0 port	Connect to USB device Note: The USB device connected must be: No bigger than 17.7 mm (W) x 9.5 mm (H) to avoid interference with other ports.
2		Mini-DP connector	Maximum display resolution: 1920x1200 32bpp@60Hz (reduced blanking)
3		Dedicated NIC	Dedicated RJ45 connector
4		Micro USB port	Transmit in serial signal for debug or terminal concentrator
5		Location button/LED	Press to trigger on/off ID LED. Blue blinking - Identifier; Off - Normal.
6		Power button/LED	Press to power on the system. Press and hold for four seconds to power off the system. Blue blinking - System power off; On - System power on.
7		Reset button	Press to reset the system

Power Sub-System



PSU to Mainboard Module Description

Two power supply unit are supplied in the system. A secondary PSU is available for redundancy functionality.

Redundant AC Power Supply Units by Model

PSU	AC INPUT
(2) 2000W 73.5mm Platinum PSU	100-240VAC 50/60Hz, AC support

Power Supply Unit LED

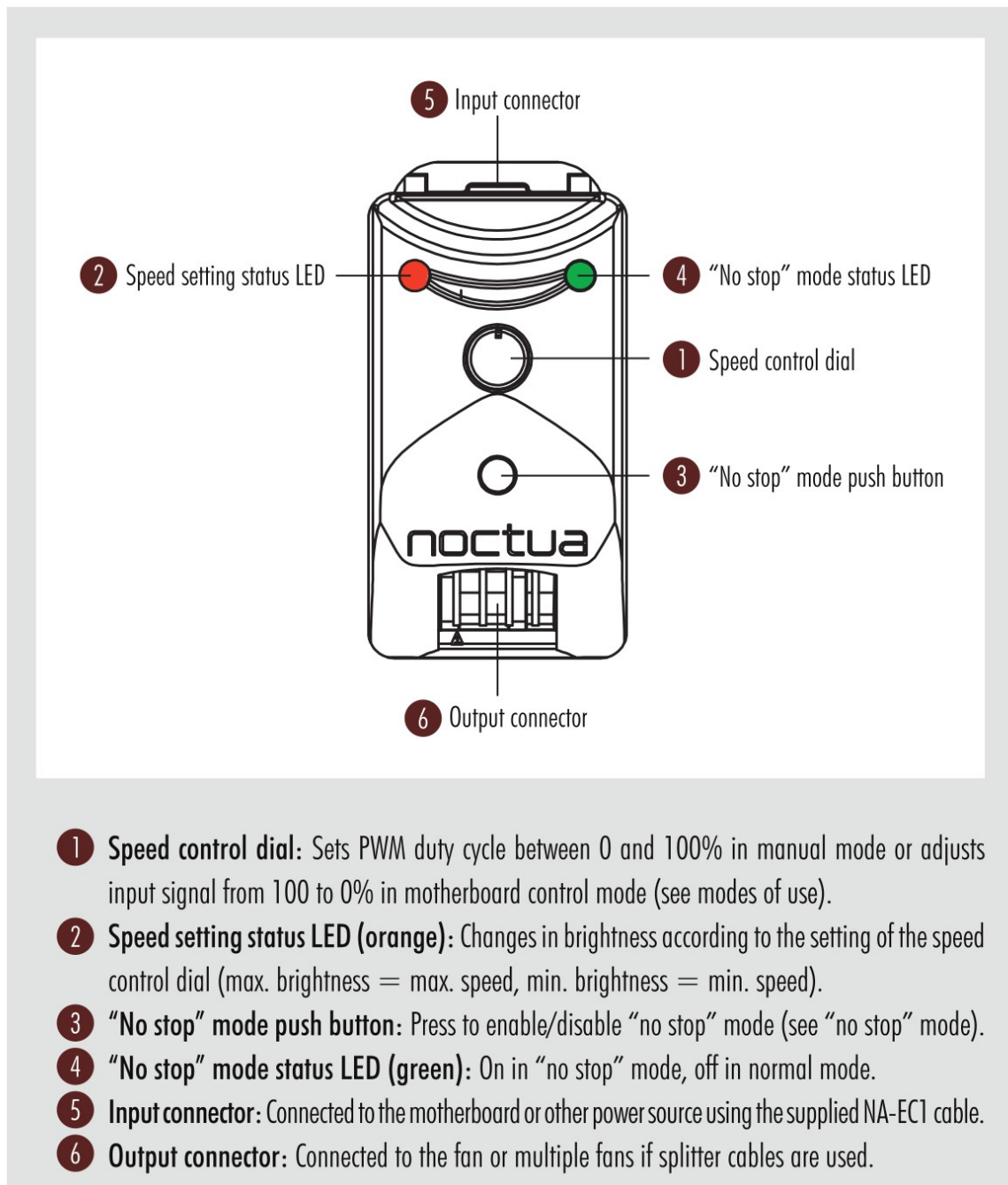
PSU LED COLOR	DESCRIPTION
Green On	Output ON and OK
Green Blinking	AC present / Only VSB ON
Amber On	AC Lost; with a second PSU in parallel still with AC input power. PSU critical event causing a shutdown; failure, OCP, OVP, Fan Fail.
Amber Blinking @1Hz	PSU warning events where the PSU continues to operate; high temp, high power, high current, slow fan, UV.

Fan speed adjustment

Some fan settings like „lower critical“ can be adjusted in the BMC or via IPMItool.

Manual adjustment of the PWM duty cycle:

Open the rear cover. Manually tune the fan PWM duty cycle of individual fans (see picture below).



Modes of use

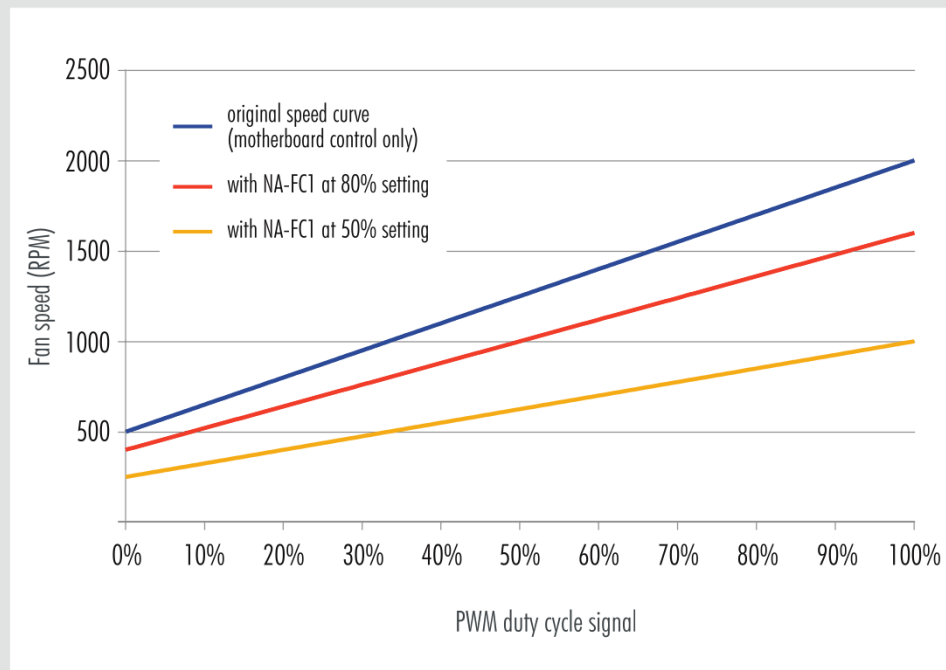
a) Manual control

If the NA-FC1 receives no PWM signal on the input side (from the motherboard), it will act as a manual speed controller. The NA-FC1 will then generate a PWM signal between 0 and 100% PWM duty cycle depending on the setting of the speed control dial ①. Turn the speed control dial clockwise to increase speed and anticlockwise to reduce speed. The orange speed setting status LED ② will change in brightness according to the speed setting.

Note that the NA-FC1 can be used in manual mode when powered both from motherboard fan headers or from other power sources. It just depends on whether there is an input PWM signal or not, so simply deactivate automatic fan control in the BIOS of your motherboard if you would like to use the NA-FC1 in manual mode.

b) Adjusting automatic motherboard fan control

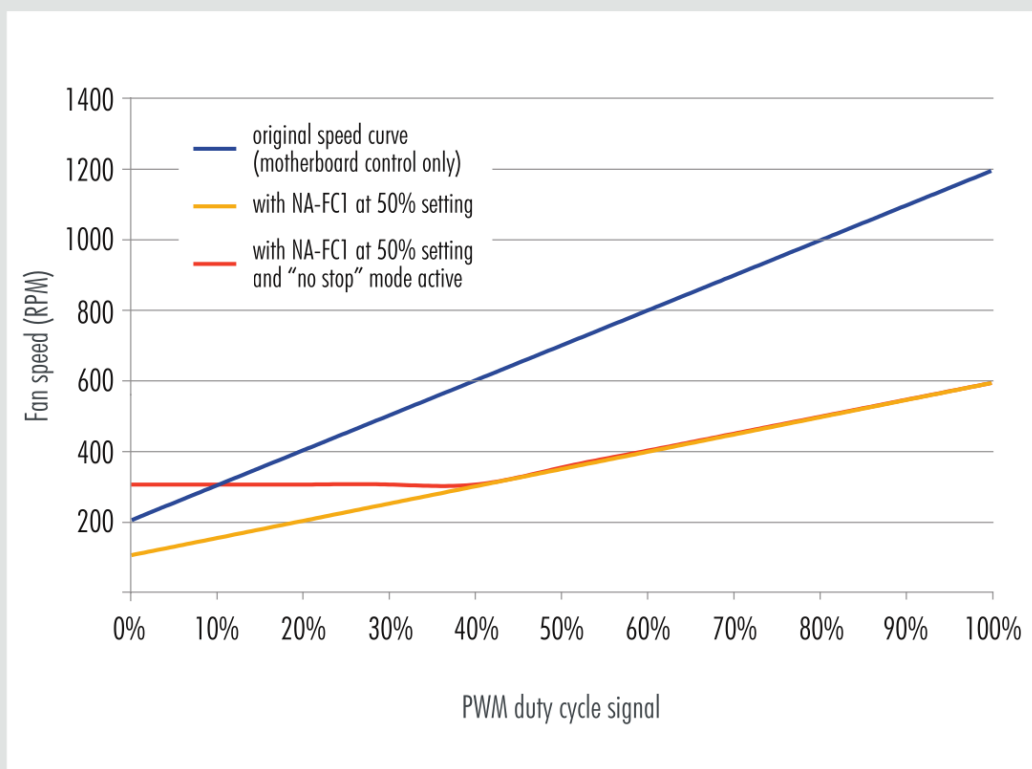
If the NA-FC1 receives an input PWM signal, it will adjust the incoming signal between 100% and 0%, depending on the setting of the speed control dial ①. For example, if the dial is set to 50% and the motherboard provides a 100% PWM duty cycle signal, the NA-FC1 will reduce the signal by 50%, thus giving an output PWM duty cycle of 50%. If the motherboard provides 80% and the NA-FC1 is set to 50%, it will put out 40% etc. This way, the NA-FC1 will work in tandem with the motherboard's automatic fan control, allowing you to lower the original control curve in order to achieve lower fan speeds and noise levels:



Turn the speed control dial clockwise to increase speed and anticlockwise to reduce speed. The orange speed setting status LED ② will change in brightness according to the speed setting.

“No stop” mode

Regardless of whether the NA-FC1 works as a manual controller or adjusts an incoming PWM duty cycle signal from the motherboard, turning the speed control dial to a very low setting (or using low-speed fans) may result in very low fan speed or the fan stopping altogether. If you would like to prevent the fan from stopping, press the “no stop” mode push button ③ in order to activate “no stop” mode. In “no stop” mode (indicated by the green “no stop” mode status LED ④ being on), the NA-FC1 will not allow the fan to fall below a minimum fan speed of 300rpm, regardless of how low the speed control dial setting or incoming PWM duty cycle signal is:



“No stop” mode can be used in order to avoid motherboard fan errors that can occur below certain speed levels or if the fan stops. To exit “no stop” mode, press the “no stop” mode push button ③ again, so that the green “no stop” mode status LED ④ goes off.

Please note that many PWM fans have a minimum speed higher than 300rpm and as the function of “no stop” mode is to not let fans go below 300rpm, it is only useful for fans which are generally able to go below this speed. For example, a fan with 500rpm minimum speed will act the same with “no stop” mode on and off.

Additional notes

PWM duty cycle and fan speed

Response to PWM input can vary from fan to fan and is not always linear, so for example some 2000rpm fans may run at 1000rpm at 50% PWM duty cycle while others may run at 1200rpm. The NA-FC1 can only set a specific duty cycle, not a specific fan speed. If you would like to set a specific fan speed, please either use hardware monitoring software provided by the motherboard manufacturer or 3rd-party tools like SpeedFan in order to check actual fan speeds during setup.

Minimum speeds and behaviour below 20% duty cycle

Note that some PWM fans will stop at duty cycles below 20% while others will keep running at the same speed as they have at 20%. Also, the minimum speed achieved at 20% varies from fan to fan. As this is determined by the fan's motor and internal electronics, the NA-FC1 cannot make fans go slower or stop, which are designed not to go below their 20% speed and/or have a high minimum speed setting. Please refer to the fan specifications for the minimum PWM speed. If you want to prevent the fan from stopping, please use the "no stop" mode.

Changing fans

When changing fans, please briefly disconnect the NA-FC1 from the power source first in order to make sure that the cache is cleared and the new fan can be tracked correctly.

Lags when switching into "no stop" mode at low speeds

Please note that it may take up to 30 seconds until the fan is brought up to 300rpm when switching into "no stop" mode at low speed settings.

BMC

How to connect to BMC:

Connect network cable to the dedicated IPMI port. Get the IP.

Open browser and enter IP: <https://192.168.178.x>

Standard username and password for BMC

username: root

password: OpenBmc (zero!!!)

Firmware Update

Download the firmware from <https://gptshop.ai>

Select update firmware in BMC to update (use tar file, do not extract).

Operating system

Standard username and password for preinstalled OS (optional):

username: x

password: xz

Install OS yourself:

Any ARM linux distribution with kernel > 6.5 should work just fine.

Example Ubuntu (standard):

Download the version you would like to have (64k kernel is highly recommended).

Ubuntu Server for ARM: <https://ubuntu.com/download/server/arm>

Intall with USB stick or over network via BMC (slower).

With older versions (not recommended) you might need to select the HWE kernel. Using the newest 64k kernel is highly recommended.

There are special Nvidia kernels for Ubuntu 22.04: [linux-nvidia-64k-hwe](#)

If you want to install a Nvidia kernel do it like this:

```
dpkg --list|grep linux-image
```

```
sudo apt purge linux-image-xxx (xxx = version number)
```

```
dpkg --list|grep linux-headers
```

```
sudo apt purge linux-headers-xxx (xxx = version number)
```

```
dpkg --list|grep linux-modules
```

```
sudo apt-get purge linux-modules-xxx (xxx = version number)
```

```
sudo apt update
```

```
sudo apt install linux-nvidia-64k-hwe-22.04-edge
```

```
sudo reboot
```

Drivers

There are two main ways to install the Nvidia drivers

1.) Via Nvidia driver installer

Download Nvidia GH200 drivers: <https://www.nvidia.com/Download/index.aspx?lang=en-us>

Select product type "data center", product series "HGX-Series" and operating system "Linux aarch64"

Copy the download url (may change with version and over time)

example:

```
curl -O https://us.download.nvidia.com/tesla/xxx/NVIDIA-Linux-aarch64-xxx.run (xxx = version number)
```

```
sudo apt install build-essential
```

```
sudo sh NVIDIA-Linux-aarch64-xxx.run -m=kernel-open (xxx = version number)
```

2.) Via package installer (example Ubuntu)

Check version if already installed:

```
cat /proc/driver/nvidia/version
```

List possible driver for install

For desktop:

```
sudo ubuntu-drivers list
```

or, for servers:

```
sudo ubuntu-drivers list --gpgpu
```

Automatic install (desktop):

```
sudo ubuntu-drivers install
```

Installing the drivers on servers

```
sudo ubuntu-drivers install --gpgpu
```

You will also want to install the following additional components:

```
sudo apt install nvidia-utils-xxx-server (xxx = version number)
```

Alternatively, manual driver installation (using APT)

```
sudo apt install nvidia-driver-xxx (xxx = version number)
```

Software

Nvidia CUDA

There are two ways to install the Nvidia CUDA

1.) Nvidia installation guide:

<https://docs.nvidia.com/cuda/cuda-installation-guide-linux/contents.html>

Downlad from:

https://developer.nvidia.com/cuda-downloads?target_os=Linux&target_arch=arm64-sbsa

2.) Via package intaller (example Ubuntu)

```
sudo apt install nvidia-cuda-toolkit
```

```
sudo apt install nvidia-cuda-dev (optional)
```

Nvidia resources

Official Nvidia GH200 Manual: <https://docs.nvidia.com/grace/#grace-hopper>

Official Nvidia Grace Manual: <https://docs.nvidia.com/grace/#grace-cpu>

Official Nvidia Grace getting started: <https://docs.nvidia.com/grace/#getting-started-with-nvidia-grace>

