

PromethION 2 Solo IT requirements

Overview

The PromethION 2 (P2) device maintains the flexibility associated with PromethION 24 and PromethION 48 devices but in a compact, accessible form factor. It is designed for small- to medium-sized academic and research labs, as well as core labs and commercial service providers — delivering the benefits of high-coverage nanopore sequencing to users with lower sample processing requirements (i.e. up to four full-length sequencing runs per week or 200 flow cells per year). Each flow cell has the capacity to deliver hundreds of gigabases of nanopore sequencing data and presents a convenient way to rapidly enable the coverage required for sequencing larger genomes. Additionally, P2 devices allow users to explore applications and projects enabled by high-output nanopore sequencing, before committing to a P24 or P48 device.

The P2 Solo is a sequencing device that requires external compute to run the MinKNOW software. Users can plug a P2 Solo into a GridION or user-provided computer; more details are provided below.



Specifications - GridION

GridION is a convenient compute resource for running the P2 Solo – offering plug-and-play compatibility taking advantage of the GPU contained within the GridION to basecall data generated from the P2 Solo. Furthermore, the operating system, SSD and memory are all compatible with the P2 solo out-of-the-box, and engineered to have the optimum amount of compute resources for sequencing.

Item	Specification
1x GridION (GXB02xxx or greater)	GridION is used by the P2 Solo as the compute resource. A GridION Mk1 with serial number GXB02xxx or greater is required to use all basecalling models with a PromethION flow cell.
1 Gb/s RJ45 port with DHCP service	Internet connection and long-term data transfer. Details on using networked storage can be found in the GridION data transfer section below.
Networked storage	See below for detailed instructions for additional storage.

Item	Specification
1x RJ45 Ethernet cable (recommended cat5e or better for 1 Gbps)	Internet connection is required to use MinKNOW. Customers require Ethernet connection for data transfer to remote storage (NFS/SMB).
IP region access for software updates	The following is required to access specific servers for software updates. Your IT infrastructure/firewall must allow access to the following locations: <ul style="list-style-type: none"> - HTTPS port 443, HTTP port 80 - AWS-global IP ranges for SW updates - DNS cdn.oxfordnanoportal.com
100-240 V 50/60 Hz AC power outlet	AC Mains power required for both P2 Solo and GridION – sequencer and basecaller cannot be powered by DC batteries.
1x USB keyboard	USB Type-A keyboard is needed to operate the GridION (I/O on rear of device).
1x USB mouse	USB Type-A mouse is needed to operate the GridION (I/O on rear of device).
1x HDMI or Display Port cable	A Display Port or HDMI cable is required to view the MinION user interface.
1x monitor	GridION currently supports output with a Display Port-compatible monitor. Additionally, a DP-to-HDMI adapter is provided with the original GridION.

Specifications - workstation/laptop

Consumer workstations can include desktop towers and high-performance laptops. Workstations can be used if you are comfortable installing the MinKNOW software and do not already have a GridION. The basecalling performance of your setup will depend heavily on the components.

Item	Specification
Operating System (OS)	Computer running one of the following: <ul style="list-style-type: none"> - Ubuntu 18.04 (Bionic Beaver)/ 20.04(Focal Fossa) - MAC OS 10.14 or newer - Windows 10
Storage	<p>Minimum</p> <p>2 TB of internal SSD storage + 6 TB external SSD storage (when sequencing data is basecalled using minimum spec GPU and moved to external SSD storage over high-speed USB/Thunderbolt/Ethernet).</p> <p>Recommended</p> <p>8 TB*+ of useable, internal fast SSD storage. *Note if using 2x 4 TB internal SSDs, they need to appear as a single drive to the OS (e.g. using RAID 0).</p> <p>For more information on data sizes, see the information above about storage required for data generated by the P2 Solo.</p>

Item	Specification
GPU	<p>Minimum</p> <p>NVIDIA RTX GPU with at least 12 GB of GPU memory (e.g., RTX3080 12 GB)</p> <p>Recommended</p> <p>NVIDIA A6000 GPU or better</p>
Memory	<p>Minimum</p> <p>16 GB of DDR4+ RAM</p> <p>Recommended</p> <p>64 GB of DDR4+ RAM</p>
CPU	<p>Minimum</p> <p>8-cores/16-threads Intel i7/i9 7th gen processor / AMD Ryzen processor</p> <p>Recommended</p> <p>12-core/24-thread Intel i7/i9 10th gen or newer processor / AMD Ryzen processor</p>
100-240 V 50/60 Hz AC power outlet	P2 Solo is powered by mains power. The maximum power draw is from the device is 60 W. The power draw of the device will depend on the sequencing run conditions (e.g. temperature, number of flow cells etc).
1x Type-A/C port	<p>For transferring sequencing data for up to two PromethION flow cells. Designed for a maximum data transfer rate of 5 Gb/s.</p> <p>Note: A USB Type-C to Type-A cable will be included in the box for the production unit.</p>
IP region access for software updates	<p>The following is required to access specific servers for software updates. Your IT infrastructure/firewall must allow access to the following locations:</p> <ul style="list-style-type: none"> - HTTPS port 443, HTTP port 80 - AWS-global IP ranges for SW updates - DNS cdn.oxfordnanoportal.com

Networking explanations

The two requirements for firewall permissions are:

1. Access to the AWS eu-west-1 region for telemetry feedback and use of the EPI2ME platform through port 443.
2. DNS rules ping.oxfordnanoportal.com, mirror.oxfordnanoportal.com, and *.nanoporetech.com

The updates are triggered as pull requests, therefore outbound-only access is required.

Telemetry

Telemetry information is collected by MinKNOW during sequencing runs as per the Terms and Conditions to allow monitoring of device performance and enable remote troubleshooting. Some of this information is obtained from free-form text entry fields, therefore no personally identifiable information should be included. No sequence data is collected. Telemetry information is required for Oxford Nanopore to confirm Flow Cell Check results and enable replacement of flow cells should they fall below warranty.

The EPI2ME platform is hosted within AWS and provides cloud-based analysis solutions for multiple applications. Use of this platform

entails users uploading sequence data in FASTQ format from the EPI2ME Agent, which is then processed through defined pipelines within the EPI2ME Portal. Downloads from EPI2ME are either in Data+Telemetry or Telemetry form. Telemetry information is used to populate reports within the EPI2ME portal.

Software updates

The IP address from which you receive software updates will depend upon your geographical location. Updates are performed through the software UI or through **apt** on the terminal, therefore outbound-only access is required. Users are notified of the availability of software updates through the Nanopore Community and full instructions for updating are provided on each release note.

Storage

File types

Nanopore sequencing data is stored in two file types, .fast5 and FASTQ. Basecalling summary information is stored in a *sequencing_summary.txt* file:

- .fast5 is a customized file format based upon the .hdf5 file type, which is designed to contain all information needed for analysing nanopore sequencing data, including raw signal data, and tracking it back to its source. As default, each .fast5 file will contain 4000 reads although this can be configured when starting a run. The size of a .fast5 file size will depend on the number of reads contained and the length of DNA/RNA sequenced. See the table below for estimates of storage space required for a typical sequencing run.
- FASTQ is a universal text-based sequence storage format, generated when the nanopore signal data is basecalled and containing both the sequence of DNA/RNA and its quality scores. By default, the device saves up to 4000 sequences in one FASTQ file. The size of a FASTQ file will depend on the number of reads contained and the length of DNA/RNA sequenced. See the table below for estimates of storage space required for a typical sequencing run.
- *sequencing_summary.txt* contains metadata about all basecalled reads from an individual run. Information includes read ID, sequence length, per-read q-score, duration etc. The size of a sequence summary file will depend on the number of reads sequenced.

Typically, 1 Gbase of sequence data takes up approximately 11 Gbytes of storage. This typically comprises of 90% .fast5 files, 9% FASTQ files and 1% sequence summary file. The P2 Solo supports up to two PromethION Flow Cells each with up to 2675 nanopores; this is approximately 5-6 times the number of pores (and output) of a MinION/GridION Flow Cell. Additionally, there are intermediate files generated during a sequencing run which are temporarily written to drive during a sequencing run.

Below is an example of storage utilisation for different outputs and devices:

P2 Solo output (Gbases)	Typical .fast5 + FASTQ storage (Tbytes)	Usable storage utilised (%) - 3.5 TB GridION	Usable storage utilised (%) - 8 TB custom workstation	Usable storage utilised (%) - 2 TB laptop
200	~2.2	~63%	~28	Full (data offload required)
400	~4.4	Full (data offload required)	~55	Full (data offload required)
580 (theoretical maximum device output)	~6.4	Full (data offload required)	~80	Full (data offload required)

Example file sizes above are based on different outputs from a flow cell. Theoretical maximum output (TMO) is from two PromethION R9.4.1 Flow Cells with Kit 10/11 chemistry.

Data transfer and long-term storage

It is essential that data is streamed from the device in real-time to prevent runs from terminating due to lack of storage space (this is common for high specification laptops). For this, a customer site must ensure that connectivity to the local infrastructure/external SSD

is of sufficient bandwidth to prevent data backing up. We recommend storage presented as NFS or CIFS. To stream data to storage in real-time, SSD is required due to its high write speed compared to HDD. After initial writing to networked SSD drives, data can be moved to storage with a slower write speed for long-term storage.

The form and volume of data to be stored will depend on customer requirements and whether you wish to basecall your sequencing data in the future when more advanced basecalling algorithms are available:

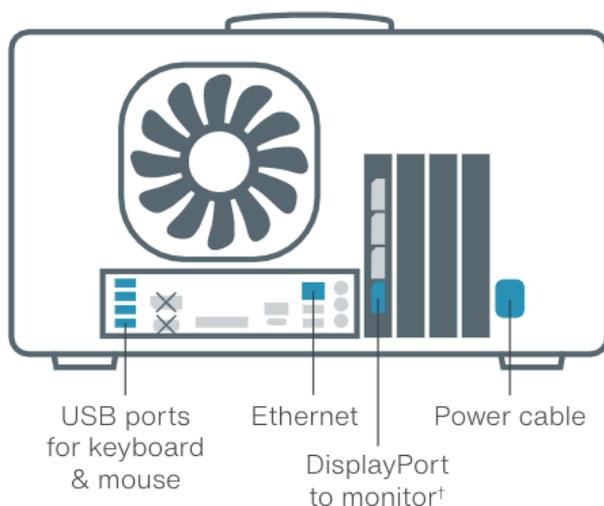
- Storing .fast5 files with raw read data in will permit re-basecalling of data when new algorithms are released by Oxford Nanopore Technologies. In such cases, new releases of basecallers have enabled significant improvements in basecalling accuracy of existing datasets through re-basecalling. Further, selected Oxford Nanopore and third-party tools use the raw signal information contained within the .fast5 to extract additional information from the raw signal, e.g. calling modified bases, reference-guided SNP calling, or polishing of data.
- Retaining only FASTQ files will allow use of standard downstream analysis tools using the DNA/RNA sequence, but no further sequence data can be generated when improvements in basecalling become available.

Oxford Nanopore is unable to provide exact recommendations for storage, as these will be site-specific. The requirements below should be taken into consideration.

GridION data transfer

If you are using a P2 Solo in combination with a GridION and require additional SSD storage, ensure you are using the correct USB port/Ethernet on the rear of the device. Do NOT use either the USB port with a white, rectangular centre or those at the front of the device (if your GridION has front-facing USB ports).

Instead, use the blue USB Type-A ports on the rear of the GridION (see the image below for reference). Alternatively if using Ethernet, ensure that you are using a cable capable of at least 1 Gbps (CAT5e) and the minimum length for reduced latency.



Only those ports and connections shaded in blue should be used.

× These ports cannot be used to connect to a monitor.

† If using HDMI-only monitor, use the included DisplayPort to HDMI adapter.

Included software

Oxford Nanopore Technologies build and provide numerous software types involved in acquisition, orchestration and analysis, available from the [Software Downloads](#) page on the Nanopore Community.

MinKNOW

MinKNOW is the core software provided by Oxford Nanopore and provides an intuitive graphical user interface (GUI) to run Nanopore sequencing devices. MinKNOW carries out several core tasks:

- Data acquisition
- Real-time analysis and feedback

- Data streaming
- Device control, including run parameter selection
- Sample identification and tracking
- Ensuring chemistry is performing correctly

Data from MinKNOW is packaged into read .fast5 files, which are a customised file format based upon the .hdf5 file type. These .fast5 files can be re-basecalled at a later date. FASTQ files are also produced, containing both the basecalled sequence of DNA/RNA and its quality scores. MinKNOW is updated regularly, and users should keep their software up to date to receive access to the latest chemistry, basecallers and features.

Guppy

Guppy is a data processing toolkit that contains the Oxford Nanopore Technologies' production basecalling algorithms and several bioinformatic post-processing features. It is run from the command line in Windows, Mac OS, and on multiple Linux platforms. Guppy is also integrated with our sequencing instrument software, MinKNOW, and a subset of Guppy features are available via the MinKNOW UI.

EPI2ME

The EPI2ME platform is a cloud-based data analysis service offering a number of workflows for end-to-end analysis of nanopore data. The basecalled read files from a sequencing run can be uploaded to the EPI2ME platform via another piece of software called the Desktop Agent. The user selects a workflow in the Agent such as 16S alignment or Barcoding, and the Agent transfers the read files one by one into the cloud for real-time analysis. The analysis results are generated and updated in real-time in the form of a report in the EPI2ME platform.

Frequently asked questions

Can I plug a P2 Solo into a PromethION 24/48 Data Acquisition Unit as the compute?

No, this is not recommended.

Can I plug two P2 Solo devices into a GridION at the same time?

No this is not recommended.

My workstation does not have a USB-C port, do I need to purchase a new workstation?

No, it is possible to use a USB Type-A port on your workstation with the use of a Type-A to Type-C adapter (one will be included in the released product box). You will need to ensure your compute device has a USB Type-A port capable of at least 5 Gb/s.

Should I use networked storage or local storage for post-run basecalling?

Basecalling from local .fast5 files on your compute will reduce the time spent copying files. If using networked storage, we recommend basecalling using local .fast5 files and transferring the basecalled FASTQ data off the compute.

Is my compute GPU-compatible?

If using the Guppy basecaller, basecalling acceleration is compatible with Nvidia GPUs using CUDA compute compatibility 6.1 or greater. Future iterations of the stand-alone basecalling software will have support for Apple Silicon compute.

Can I run PromethION and MinION/GridION flow cells on my GridION at the same time?

We are working on implementing software to enable running a mix of flow cells at the same time. For initial Early Access devices, we recommend only running PromethION flow cells or MinION/GridION flow cells at the same time.

My GridION has a GXB01nnn serial number, can I still run the P2 Solo?

It is not recommended to run the P2 Solo on this GridION. To receive an upgrade to the latest GridION, the device must have a valid Software license and Device warranty and an open flow cell order on the account. Please contact support@nanoporetech.com for more details.

Change log

Date	Version	Changes made
May 2022	V1	Initial publication