

PromethION P24/P48 IT requirements

Checklist

This checklist represents the minimal requirements for installing the PromethION in your institution. For full explanation of requirements, please continue to the document. Please note either Fibre **OR** Copper connections can be used.

Item/setup required	Reason	Provided?
2x SFP+ modules supporting 10GBASE-SR or 10GBASE-LR (not required if copper used)	Connection to IT infrastructure and internet	
Appropriate 10 Gbps cabling for chosen SFP+ modules, OR Appropriate cabling for 10 Gbps Ethernet	Connection to IT infrastructure and internet	
2x 10 Gbps fibre ports with DHCP service running, OR 2x 10 Gbps Ethernet ports with DHCP service running	Connection to IT infrastructure and internet	
HTTPS/port 443 TCP access to AWS eu-west-1 IP ranges listed at http://docs.aws.amazon.com/general/latest/gr/aws-ip-ranges.html	Telemetry feedback, EPI2ME analysis	
HTTPS/port 443 and HTTP/port 80 TCP access to Linode-hosted IP addresses 178.79.175.200 and 96.126.99.215 for software distribution	Software updates	
Storage: *Sufficient infrastructure for required storage option	Streaming and long-term data storage	
USB mouse	Device control	
USB keyboard	Device control	
VGA compatible monitor	Device control	
3x Power supply capable of delivering 2x 2 kW and 1x 1.2 kW, respectively	Device power	

PromethION P24/P48

Overview

The PromethION is a benchtop device for nanopore sequencing designed to run and analyse up to 24 or 48 flow cells. It is ideal for labs with multiple projects that need the advantages of nanopore sequencing:

- Simple library preparation
- Real-time analysis
- Biological insights from long reads

In addition, the PromethION P24/P48 also enables users to offer nanopore sequencing as a service when certified.

The PromethION benefits from the inclusion of on-board compute which permits device control, data acquisition, basecalling and data streaming, all without placing any additional burden on existing IT infrastructure.

All device control, data acquisition and basecalling on the PromethION is carried out by pre-installed custom software created by Oxford Nanopore Technologies. The default data analysis workflow when using the PromethION is as below:

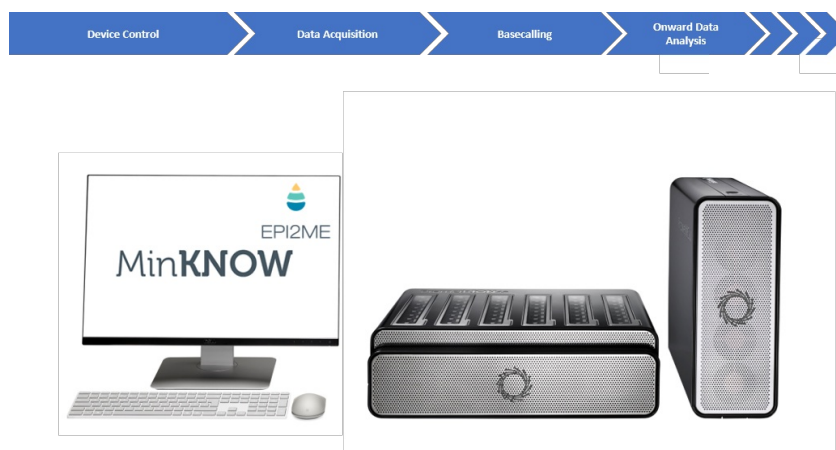


Figure 1: Default data analysis workflow of the PromethION P24/P48 device

Specifications

The PromethION is designed around a simple user interface on top of cutting-edge custom electronics providing real-time analysis solutions. The specification of your device will depend on the P24 or P48 model:

Component	Specification
Operating system	Ubuntu 16.04 running off Intel CPU Customers are advised to keep up to date with all software and security patches
Storage	P24: 32 TB internal SSD P48: 64 TB internal SSD
GPU cards	P24: 2x NVIDIA GV100 P48: 4x NVIDIA GV100
Memory	384 GB RAM
Size and weight	Sequencing Unit: - H 190 x W 590 x D 430 mm - Weight 28 kg Data Acquisition Unit: - H 440 x W 178 x D 470 mm - Weight 25 kg
Environmental ranges	Designed to sequence at +18° C to +22° C*
Maximum heat output	Sequencing Unit: 4,094 BTU/hr Data Acquisition Unit: 6,824 BTU/hr

*Functional range of electronics -5° C to +40° C

Site requirements

Installing the PromethION in your institution is similar to installing any new server. Requirements for the device are as follows:

Component	Requirements
2x 10 Gbp/s fibre port (SFP+ modules must be provided) OR 2x 10 Gbp/s RJ45 Port	<p>IP address via either DHCP service OR statically assigned</p> <p>TCP running on port 80 HTTP and port 443 HTTPS</p> <p>Requires perimeter firewall permissions:</p> <ul style="list-style-type: none"> - To access AWS eu-west-1 IP ranges listed at https://ipranges.amazonaws.com/ip-ranges.json for telemetry feedback and EPI2ME analysis - To 178.79.175.200 and 96.126.99.215 for software updates <p>Connected to local storage infrastructure</p> <p>Optional: For higher bandwidth connection into local infrastructure can bond either actively/passively, or using LACP</p>
3x power supply	<p>2x for Data Acquisition Unit, C13 cables included with device</p> <ul style="list-style-type: none"> - Maximum power draw from Data Acquisition Unit of 2 kW <p>1x for Sequencing Unit, C13 cable included with device</p> <ul style="list-style-type: none"> - Maximum power draw from Sequencing Unit of 1.2 kW
1x monitor	Compatible with VGA for device interaction. After initial setup the device may be configured for use with DisplayPort if desired
1x keyboard	Compatible with USB connection for device interaction
1x mouse	Compatible with USB connection for device interaction

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. Access to the Linode-hosted IP addresses 178.79.175.200 and 96.126.99.215 for software distribution.

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.

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 **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.

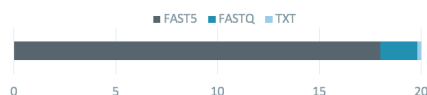
Storage

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 20 Gbytes of storage. This typically comprises of 90% .fast5 files, 9% FASTQ files and 1% sequence summary file.



Example file sizes below are based on different throughputs from an individual flow cell, with a run saving both .fast5 and FASTQ files with a read N50 of 25 kb. The P24 and P48 can run 24 or 48 flow cells simultaneously.

Output (Gbases)	.fast5 storage (Gbytes)	FASTQ storage (Gbytes)	.fast5 + FASTQ storage (Gbytes)
50	800	100	900
100	1600	200	1800
200	3200	400	3600

As an experiment progresses, .fast5 files are produced for all reads. If basecalling is chosen, these reads are utilised by the on-board software (more information below) to generate sequence data which is then stored in FASTQ files.

Data transfer and long-term storage

The storage capacity of the P24 or P48 Data Acquisition Unit is approximately 24 hours of 24 or 48 flow cells respectively, with .fast5 and FASTQ output. A standard sequencing run lasts 64 hours, therefore it is essential that this data is streamed from the device in real-time to prevent runs from terminating due to lack of storage space. For this, a customer site must ensure connectivity to the local infrastructure is of sufficient bandwidth to prevent data backing up. The PromethION offers two 10 Gbit per second ports for this purpose, with the customer able to choose between Ethernet or Fibre solutions. USB ports do not provide sufficient bandwidth for real-time data transfer so should not be used. Below is a worked example showing idealised data transfer speeds; real transfer speeds could be slower depending on network configuration.

	1 Gbit/s	10 Gbit/s
1x 200 Gbase flow cell	~11 hours	~1 hour
48x 200 Gbase flow cell	~520 hours	~52 hours

The PromethION runs on Ubuntu 16.04 and is able to mount multiple filesystem types. We recommend storage presented as NFS or CIFS. This storage streamed to in real-time should be SSD for 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:

- Storing .fast5 files with raw read data in will permit re-basecalling of data when new algorithms are released by Oxford Nanopore. 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 modified bases calling, reference-guided SNP calling or polishing of data.
- Retaining just 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 above guidelines and requirements should be taken into consideration.

Included Software

Oxford Nanopore Technologies build and provide numerous software types involved in acquisition, orchestration and analysis:

MinKNOW

MinKNOW carries out several core tasks:

- Device control, including run parameter selection
- Data acquisition
- Real-time analysis and feedback
- Data streaming
- Basecalling (through integrated Guppy)

The MinKNOW software carries out several core tasks: data acquisition, real-time analysis and feedback, basecalling, data streaming, controlling the device, and ensuring that the platform chemistry is performing correctly to run the samples. MinKNOW takes the raw data and converts it into reads by recognition of the distinctive change in current that occurs when a DNA strand enters and leaves the pore. MinKNOW then basecalls the reads, and writes out the data into .fast5 or FASTQ files.

Guppy

Guppy is a data processing toolkit that contains the Oxford Nanopore Technologies' basecalling algorithms, and several bioinformatic post-processing features. It is run from the command line, but has a low entry level for use. A selection of configuration files allow basecalling of DNA and RNA libraries, made with Oxford Nanopore Technologies current sequencing kits, in a varied range of flow cells.

If rebasecalling off the GridION, it is possible to use an external GPU to basecall with Guppy, and this will dramatically increase basecalling speed over using CPU. Guppy currently works with only NVIDIA GPUs, with the following specific models supported:

- NVIDIA Volta GV100
- NVIDIA Tesla V100
- NVIDIA GTX1080Ti
- NVIDIA Jetson TX2

It is possible to use other NVIDIA GPUs for basecalling, but support for those is limited.

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 use an uninterruptible power supply (UPS) with the PromethION P24/P48?

Yes you can use a UPS, however we are unable to recommend a specific product to use.

Which SFP+ modules are compatible with the fibre ports on the PromethION P24/P48?

Any SFP+ modules will be compatible, please select the SFP+ module that fits with the infrastructure of your institution.

Do you have any recommendations for how to move data off the PromethION 24/48 in real-time during a sequencing run?

Currently we recommend moving the data off the box using **rsync** run hourly through **crontab**. For further details, please email support@nanoporetech.com

Do you have recommendations for storage I can connect to the PromethION P24/P48?

The PromethION can communicate with any NAS device; we suggest you discuss your storage requirements with your local IT team based on the information provided above to get the right solution for your needs.

Change log

Date	Version	Changes made
July 2019	V1	Initial version