MinION Mk1B IT requirements

Overview

The MinION™ is a small, self-contained device for nanopore sequencing. It plugs directly into a standard USB 3.0 port on a desktop or laptop computer.

The MinION device is controlled by MinKNOW™ software.

MinKNOW carries out several core tasks: data acquisition, real-time analysis and feedback, basecalling, data streaming, device control including selecting the run parameters, sample identification and tracking and ensuring that the platform chemistry is performing correctly to run the samples.

Despite its size, the MinION is capable of sequencing many gigabases (Gbases) in a single sequencing experiment. As such, it is recommended that users purchase a powerful PC to ensure all the technology features are accessible.

The default data analysis workflow when starting to use the MinION is below:

Host computer specification

We recommend using a GPU for optimal performance and for basecalling to keep up with the rate of data generation. Currently, GPU basecalling is supported on Linux, and there is a beta GPU support for Windows. The minimal CPU compute requirements for data acquisition and basecalling are also listed below.

Please note that to run MinKNOW basecalling on a GPU, The GPU version of Guppy must be installed on the same system to which the MinION is connected. See the instructions in the MinKNOW protocol for how to configure these settings.

Component	**Recommended specification: GPU high-accuracy basecalling**	**Minimal specification: data acquisition/CPU basecalling** **(note: CPU basecalling performance is limited - a GPU is recommended)**
	Have you considered using the MinION Mk1C (optimised for real-time basecalling using the Fast basecalling model) or GridION (optimised for real-time high-accuracy basecalling)?	
Operating system	Windows – 10 Linux – Ubuntu 18.04	Windows – 10 OSX – Mojave, Catalina Linux – Ubuntu 18.04
Memory/RAM	16 GB RAM	16 GB RAM

Component	Recommended specification: GPU high-accuracy basecalling	Minimal specification: data acquisition/CPU basecalling (note: CPU basecalling performance is limited - a GPU is recommended)
CPU	Intel i7†, i9, Xeon, or better Ryzen 5, 7, or better, with at least 4 cores/8 threads	Intel i7†, i9, Xeon, or better Ryzen 5, 7, or better, with at least 4 cores/8 threads
GPU	NVIDIA GPU RTX 2060 or better, with at least 8 GB of GPU memory. Theoretical performance is only a guide, but more is better. Technical information can be found on various websites, for example https://www.techpowerup.com/gpu-specs/. Widely-available examples include the RTX 2060, RTX 2070, RTX 3060, RTX 3070. If you are working with a different type of GPU than the models listed above, please ensure that it has a CUDA Compute Capability >6.1 (for more information about CUDA-enabled GPUs, see the NVIDIA website).	-
Storage	1 TB internal SSD	1 TB internal SSD
Ports	USB3.0‡	USB3.0‡

† *Users need to verify that their i7 is a four-core model or better*

‡ *The MinION device is CE marked using USB3.0. If a user wished to use USBC, they may, but this invalidates the CE marking*

We recommend that MinKNOW is installed to the internal SSD, rather than the main operating system drive. This avoids intermediate files being written to a non-SSD location during the experiment, potentially leading to a build-up of intermediate files that can cause the system to run out of space and stop.

The below requirements are to ensure correct operation of the MinION during a sequencing experiment:

Component	Minimum requirement
User account privilege level	Local Administrator for MinKNOW installation and updates. Local Administrator privileges are not required for running sequencing experiments.
Internet connection	An internet connection is required at all times for software updates and telemetry. Offline configurations can be made available for field use and expeditions. Please contact Support.
Antivirus settings	Antivirus software scanning the entire hard drive uses significant amounts of resources and may cause performance issues. For this reason, it is recommended that the scans are switched to manual setting, rather than to automatic, so that they can be performed when the MinION is not in use.
OS update settings	OS system updates should be switched to manual mode, as downloading them during a sequencing experiment may impact performance. Updates that require restart will fatally halt a run.
Telemetry feedback, EPI2ME analysis	Ethernet: 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
Software updates	Ethernet: 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 OR DNS rules ping.oxfordnanoportal.com, mirror.oxfordnanoportal.com, and *.nanoporetech.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. a. Access to the Linode-hosted IP addresses 178.79.175.200 and 96.126.99.215 for software distribution
b. OR DNS rules ping.oxfordnanoportal.com, mirror.oxfordnanoportal.com, and *.nanoporetech.com

Telemetry

Telemetry information is collected by MinKNOW sequencing runs as per the Terms and Conditions to allow monitoring of experiment performance and provide support in troubleshooting cases. 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 from Metrichor is hosted within AWS and provides cloud-based analysis solutions for multiple applications. Use of this platform entails users uploading sequence data in .fast5 or 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

Depending on your geographical region, only one of 178.79.175.200 or 96.126.99.215 will be used for provision of updates to device software. The Updates are triggered as pull requests, therefore outbound-only access is required.

File types

The MinION stores biological data in two file types, .fast5 and FASTQ. Basecalling summary information is stored in a sequencing_summary.txt file:

- .fast5 is a customised file format based upon the .hdf5 file type, which is designed to contain all information needed for analysing nanopore sequencing data and tracking it back to its source. A .fast5 file contains data from multiple reads (4000 reads as default), and is several hundred Mb in size. Storing .fast5 files with raw read data in will permit re-basecalling of data when new improvements are released by Oxford Nanopore.
- FASTQ is a universal text-based sequence storage format, containing both the sequence of DNA/RNA and its quality scores. By default, the MinION saves up to 4000 DNA sequences in one FASTQ file. File size can vary from < 1Mb to tens of Mb depending on the number and length of sequences. Retaining only FASTQ files will allow use of downstream analysis tools, but no further sequence data can be generated when improvements in basecalling become available.
- `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 90% .fast5 files, 9% FASTQ files and 1% sequence summary file.

Example file sizes below are based on different outputs from a flow cell, with a run saving both .fast5 and FASTQ files with a read N50 of 25 kb.

Output (Gbases)	.fast5 storage (Gbytes)	FASTQ storage (Gbytes)	.fast5 + FASTQ storage (Gbytes)
10	100	10	110
15	150	15	165
30	300	30	330

As an experiment progresses, .fast5 files are produced for all reads. If basecalling is chosen, these reads are utilised by the onboard software (more information below) to generate sequence data which is then stored in FASTQ files and as a FASTQ record within the .fast5 file.

Included software

Oxford Nanopore Technologies builds and provides numerous software types involved in acquisition, orchestration and analysis:

- MinKNOW
- Guppy
- EPI2ME

MinKNOW

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

MinKNOW provides an intuitive graphical user interface (GUI) and receives updates on a regular basis. This is the core software provided by Oxford Nanopore, without which the sequencing devices cannot be run. 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.

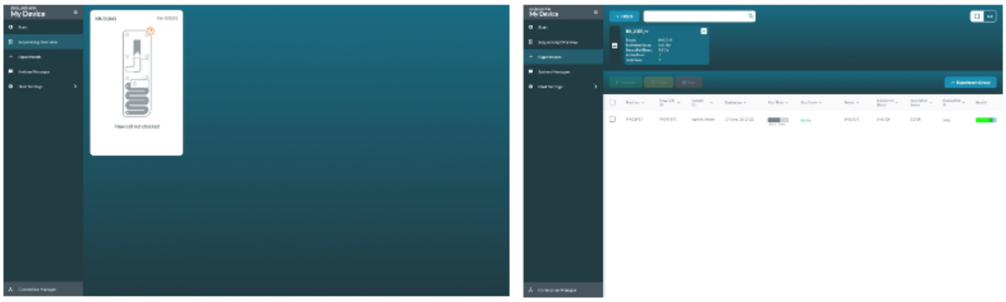


Figure 1: Example screen shots of the MinKNOW GUI. This permits selection and initiation of experiments, as well as providing real-time feedback on experiment progression.

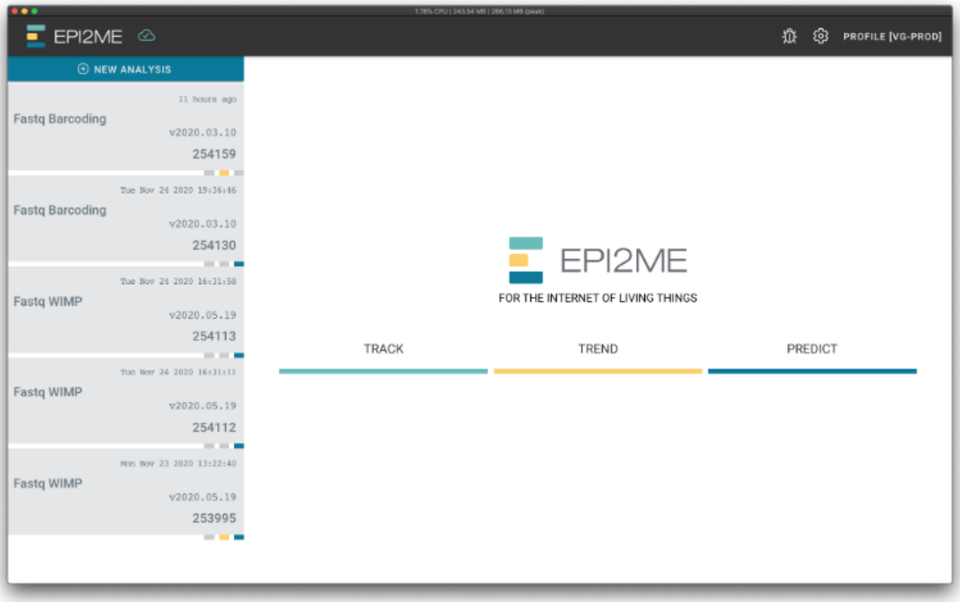
Guppy

Guppy is a production basecaller provided by Oxford Nanopore, and uses a command-line interface. It utilises the latest in Recurrent Neural Network algorithms to interpret the signal data from the nanopore, and basecall the DNA or RNA passing through the pore. It is optimised for running with basecall accelerators e.g. GPUs. Guppy implements stable features into Oxford Nanopore Technologies' software products, and is fully supported. It receives .fast5 files as an input, and is capable of producing:

- .fast5 files appended with basecalled information
- .fast5 files that have been processed, but basecall information present in a separate FASTQ file

EPI2ME

EPI2ME™ is an onwards data-analysis platform created by Oxford Nanopore's subsidiary company, Metrichor. It provides users with real-time analysis such as species identification, alignment workflows and other bioinformatics solutions. It is currently provided as a cloud-based analysis platform, which is initiated through the local EPI2ME Agent.



11 hours ago
Fastq Barcoding
v2020.03.10
254159

Tue Nov 24 2020 19:36:46
Fastq Barcoding
v2020.03.10
254130

Tue Nov 24 2020 16:31:58
Fastq WIMP
v2020.05.19
254113

Tue Nov 24 2020 16:31:11
Fastq WIMP
v2020.05.19
254112

Mon Nov 23 2020 13:22:40
Fastq WIMP
v2020.05.19
253995

Figure 2: The EPI2ME Agent. Installed on the device, this allows selection and initiation of workflows within the EPI2ME portal.

Users will be able to review basic experimental statistics as well as use prepared workflows such as real-time taxonomic identification, and reference alignment.

Taxon	Cumulative Reads
Staphylococcus	20,407
Escherichia	19,169
Bacillus	18,541
Salmonella	18,085
Listeria	16,886
Lactobacillus	16,793
Enterococcus	15,888
Pseudomonas	15,124
Saccharomyces	3,889
Cryptococcus	**3,723**
Metarhizium	482
Colletotrichum	190

Figure 3: Example reports and plots from the EPI2ME portal.

Change log

Date	Version	Changes made
10th September 2021	V11	Removed erroneous compatibility with Ubuntu 20.04
8th September 2021	V10	Added Windows 10 to the list of supported operating systems for GPU-enabled basecalling

Date	Version	Changes made
22nd July 2021	V9	In the table of minimal requirements, two extra fields have been added: "Telemetry feedback, EPI2ME analysis", and "Software updates"
30th June 2021	V8	- Checklist at the top of the document has been removed - Split IT requirements into Recommended (GPU) and Minimum (CPU) - Updated operating systems recommendations for Windows (10), Mac OS (Mojave, Catalina) and Linux (18.04 and 20.04) - Updated CPU recommendations to i7, i9, Xeon or better with 4+ cores - Updated description and use of .fast5 files in the MinKNOW section - Updated screenshot of the EPI2ME Agent - Added table of .fast5 and FASTQ file sizes with different flow cell outputs
23rd Dec 2020	V7	- Removed requirement for accessing IP address 106.187.40.102 - Updated the supported external NVIDIA GPUs to Tesla V100, Quadro GV100, Jetson TX2 and Jetson Xavier
23rd Oct 2020	V6	Updated OSX compatibility
3rd July 2020	V5	Updated Fig 1 to 4.0.8 UI
18th May 2020	V4	Removed Bluetooth compatibility for MinIT
Feb 2020	V3	More information about basecalling on GPUs using Guppy