FORTISSIMO

CESGA

Easy use of Distributed TensorFlow Training on supercomputing facilities

Gonzalo Ferro, CESGA
gferro@cesga.es

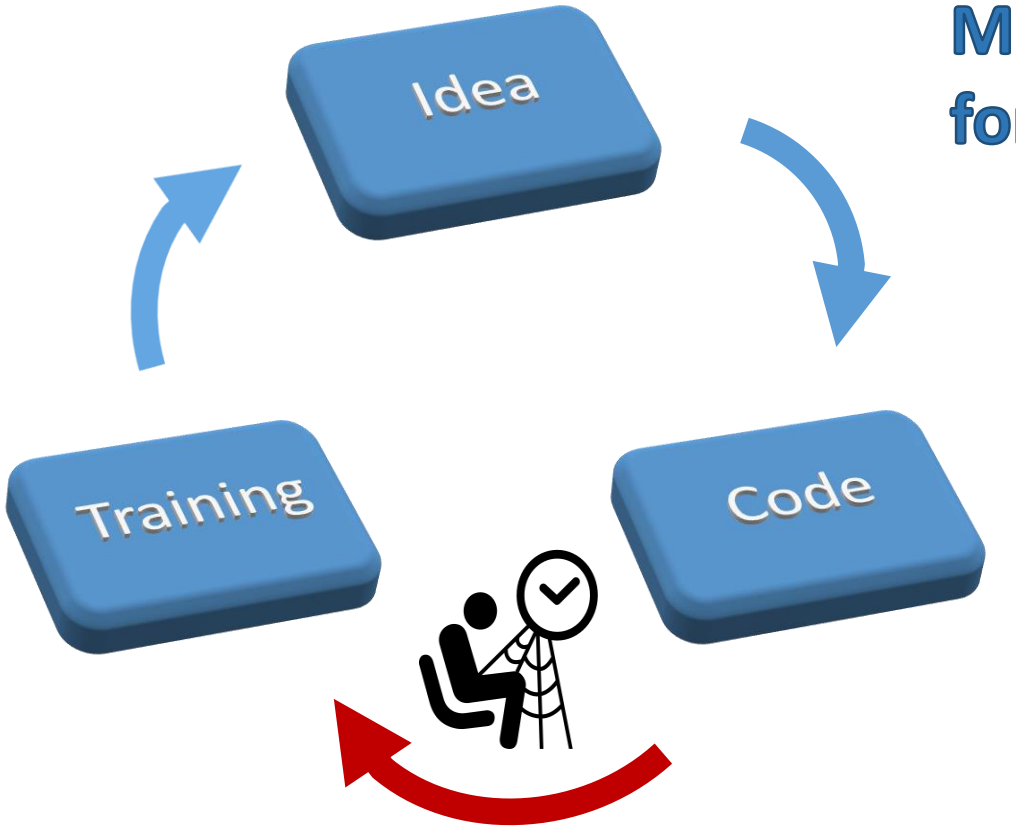
IBERGRID

IBERGRID 2018: Towards the European Open Science Cloud – EOSC.
11th - 12th October. Lisbon, ISCTE - University Institute of Lisbon (ISCTE-IUL)

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Machine Learning design cycle



Machine Learning (ML) is a powerful tool for science, industry and other sectors.

Performance of ML algorithms is improved by training them using large datasets.

HPC can help engineers to boost their algorithms.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

How exploit HPC for ML training?

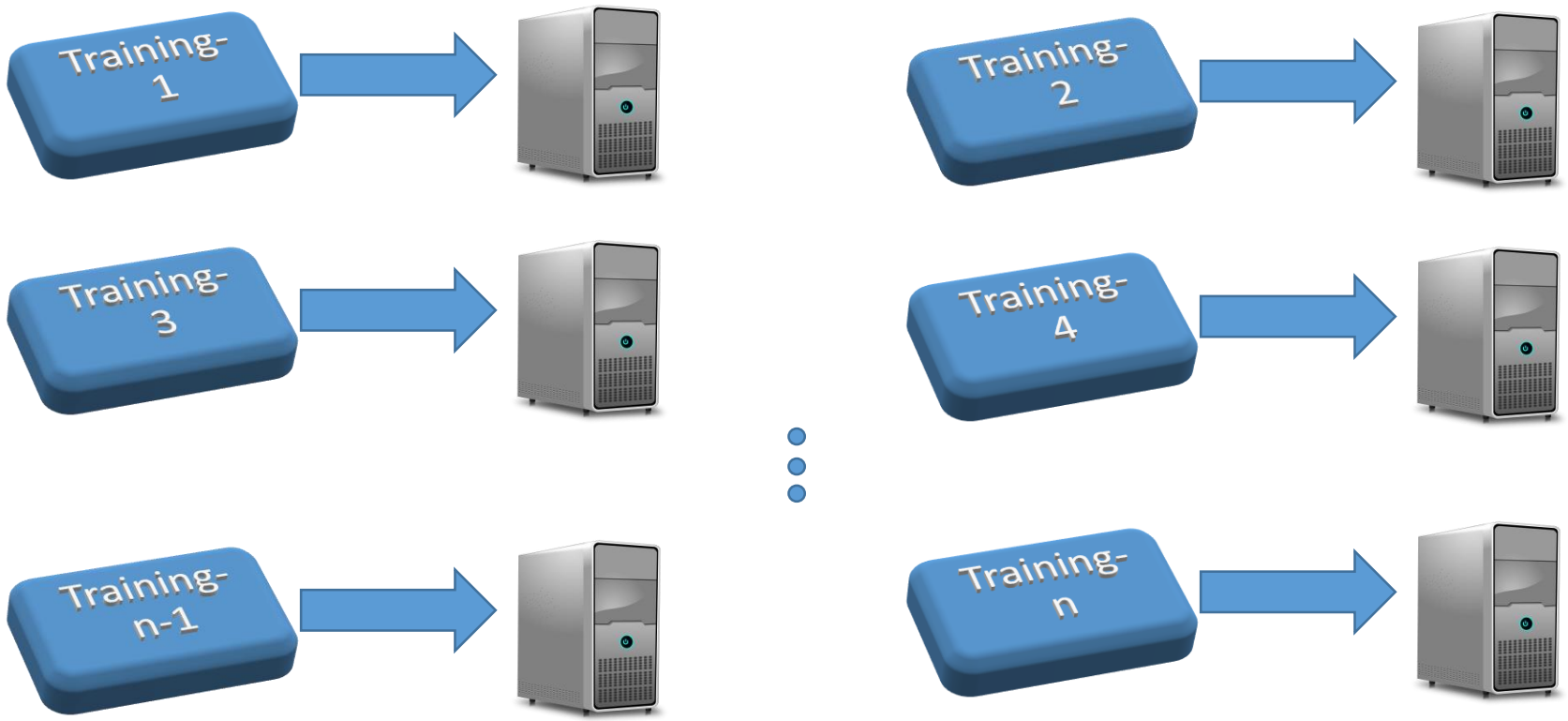
- Simultaneous Training.
- Parallel Distributed Training.
- Simultaneous + Parallel Distributed.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

How exploit HPC for ML training?

Simultaneous Training



Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

How exploit HPC for ML training?

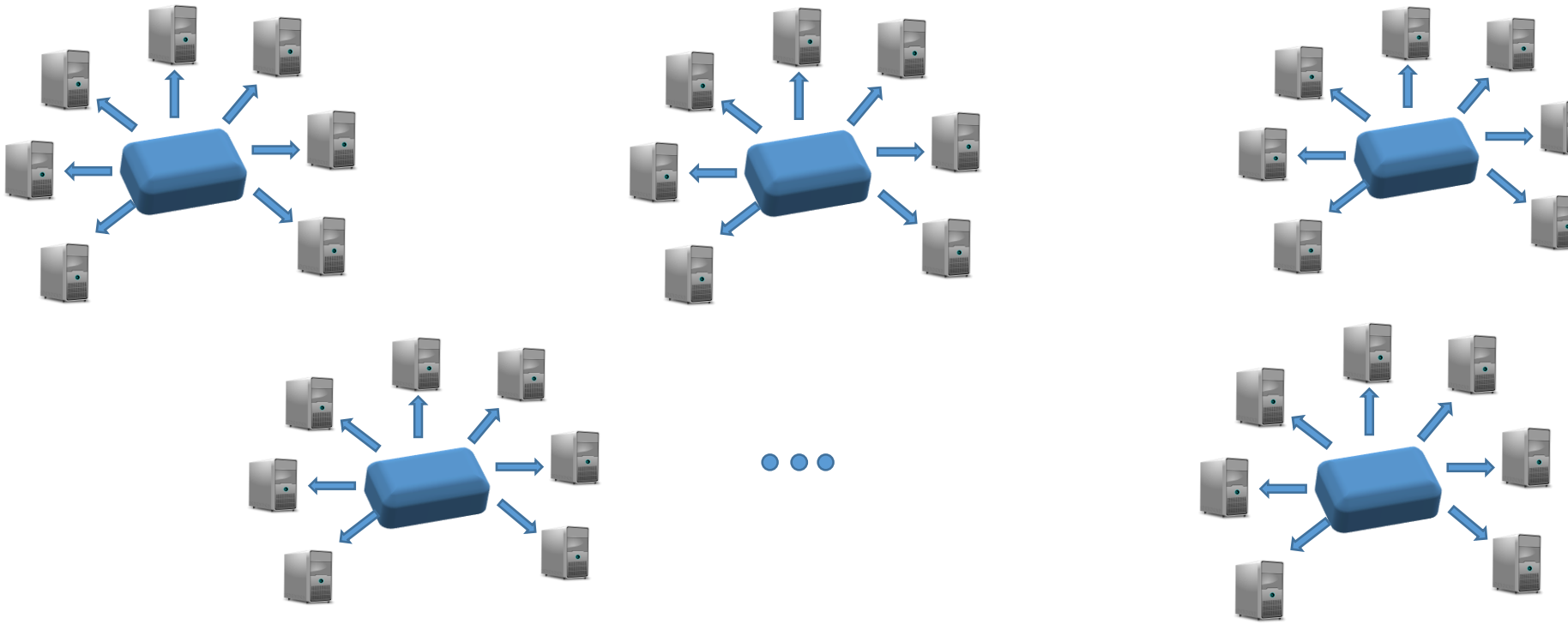
Parallel Distributed Training

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

How exploit HPC for ML training?

Simultaneous + Parallel Distributed



Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

How exploit HPC for ML training?

CESGA

FORTISSIMO

- Simultaneous Training
- **Parallel Distributed Training**
- Simultaneous + Parallel Distributed

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed Training API

TensorFlow (TF)

- A Machine Learning API developed by Google.
- One of the most widely of the tools used for developing and training of deep learning models.
- TF allows users to implement distributed computing capabilities in their training in an easy way.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed Training API

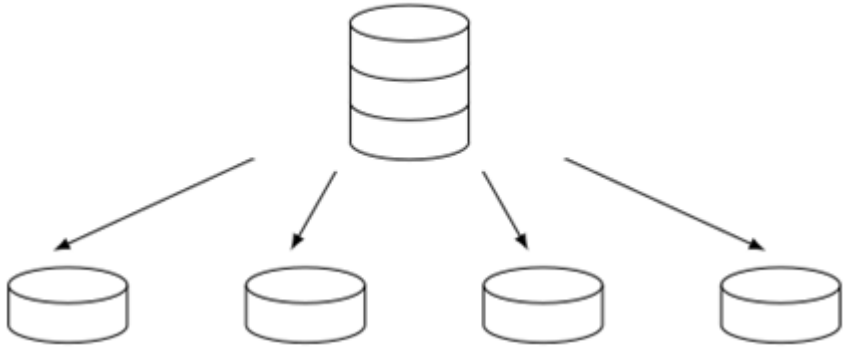
TensorFlow (TF)

- A Machine Learning API developed by Google.
- One of the most widely of the tools used for developing and training of deep learning models.
- TF allows users to implement **distributed computing capabilities** in their training in an easy way.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Data Parallelism.

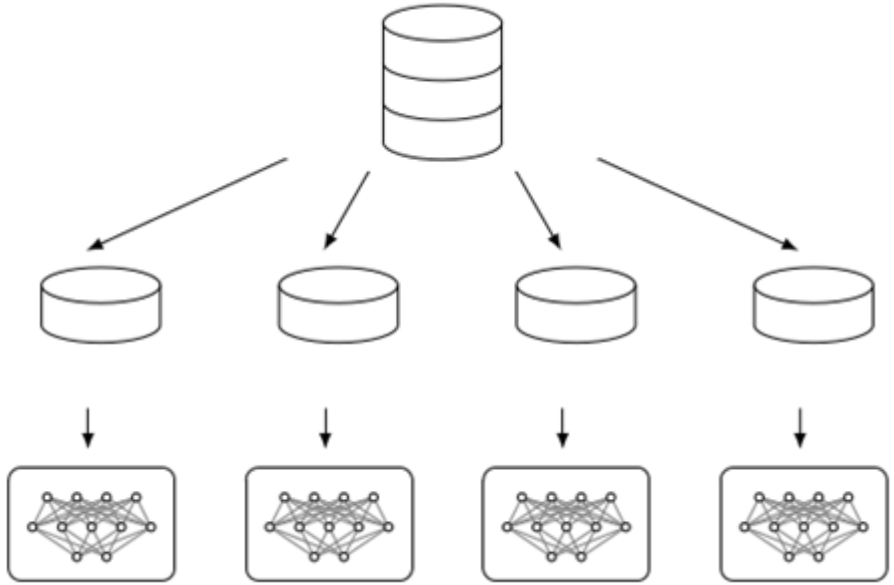


Splitting of Training Dataset.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Data Parallelism.



Splitting of Training Dataset.

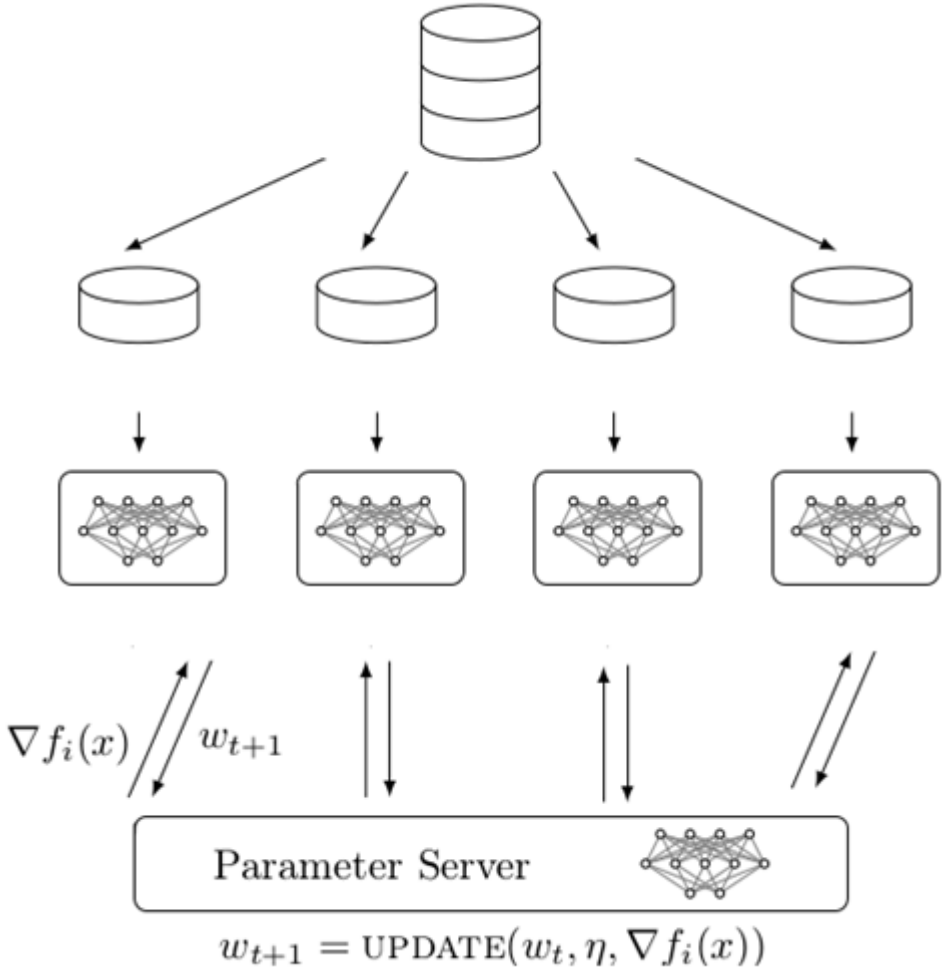
Workers:

- Copy of the computational graph.
- Calculate gradients over their correspondent part of dataset.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Data Parallelism.



Splitting of Training Dataset.

Workers:

- Copy of the computational graph.
- Calculate gradients over their correspondent part of dataset.

Parameter Servers:

- Store weights and bias of the model.
- Responsible for the aggregation of gradients calculated by Workers.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called gRPC.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called gRPC.

CESGA Finis Terrae II (FT2) uses Slurm for Resource Management.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called gRPC.

CESGA Finis Terrae II (FT2) uses Slurm for Resource Management.

- Detected issues when deploying Distributed TF on FT2:

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called **gRPC**.

CESGA Finis Terrae II (FT2) uses **Slurm** for Resource Management.

- Detected issues when deploying Distributed TF on FT2:
  1. Distributed TF needs **Network Addresses (IP:Port)** in advance.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called **gRPC.**

CESGA Finis Terrae II (FT2) uses **Slurm** for Resource Management.

- Detected issues when deploying Distributed TF on FT2:
  1. Distributed TF needs **Network Addresses (IP:Port)** in advance.

```
ServerDictionary={
        “ps”:[“hostname01:2222",“hostname02:2222"],
        “worker”:[“hostname03:2222",“hostname04:2222",“hostname05:2222"]
}
```

Co-funded by the European Union

Fortissimo has received funding from the European Union’s H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called **gRPC.**

CESGA Finis Terrae II (FT2) uses **Slurm** for Resource Management.

- Detected issues when deploying Distributed TF on FT2:
  1. Distributed TF needs **Network Addresses (IP:Port)** in advance.

```
ServerDictionary={
        “ps”:[“hostname01:2222",“hostname02:2222"],
        “worker”:[“hostname03:2222",“hostname04:2222",“hostname05:2222"]
}
```

Queue system does not provide it.

Co-funded by the European Union

Fortissimo has received funding from the European Union’s H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called **gRPC**.

CESGA Finis Terrae II (FT2) uses **Slurm** for Resource Management.

- Detected issues when deploying Distributed TF on FT2:
  1. Distributed TF needs **Network Addresses (IP:Port)** in advance.
     Queue system does not provide it.

  2. Distributed TF **Parameter Servers run forever**.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Distributed TF: Queue system issues.

Distributed TF is based on communication protocol called **gRPC**.

CESGA Finis Terrae II (FT2) uses **Slurm** for Resource Management.

- Detected issues when deploying Distributed TF on FT2:
  1. Distributed TF needs **Network Addresses (IP:Port)** in advance.
     Queue system does not provide it.

  2. Distributed TF **Parameter Servers run forever**.
     HPC Resources are wasted.
     User or Queue system have to stop them.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Solution: tf4slurm

CESGA has developed **tf4slurm** Python Package to solve the issues.

Python Module	Solved Issue
ServerDictionary	**IP:Port information**
DistributedTFQueueHook	**Close Distributed TF server gracefully***

* Adapted from https://gist.github.com/**yaroslavvb**/82a5b5302449530ca5ff59df520c369e

Technical Report:

https://www.cesga.es/es/biblioteca/downloadAsset/id/803

GitHub repository:

https://github.com/gonfeco/tf4slurm

Co-funded by the European Union

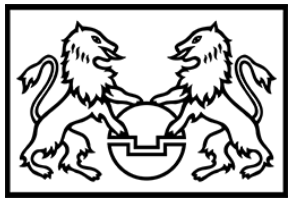
Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

tf4slurm: test with an Industrial Case

CESGA FORTISSIMO

Fortissimo H2020 Project

Experiment 707: Cyber-Physical Laser Metal Deposition (CyPLAM)



CESGA

Co-funded by the European Union

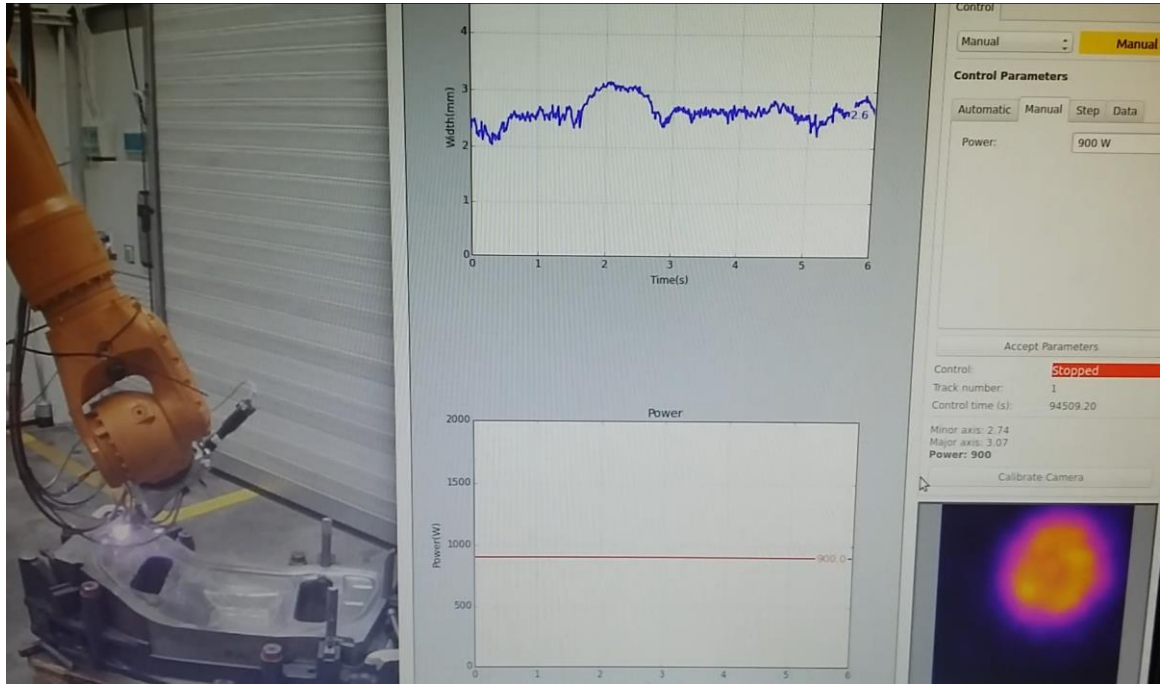
Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

CyPLAM description

Using Laser Metal Deposition (LMD) for building and repairing large metal parts.

LMD process recorder by Medium Wavelength Infrared (MWIR) sensors attached to laser header.

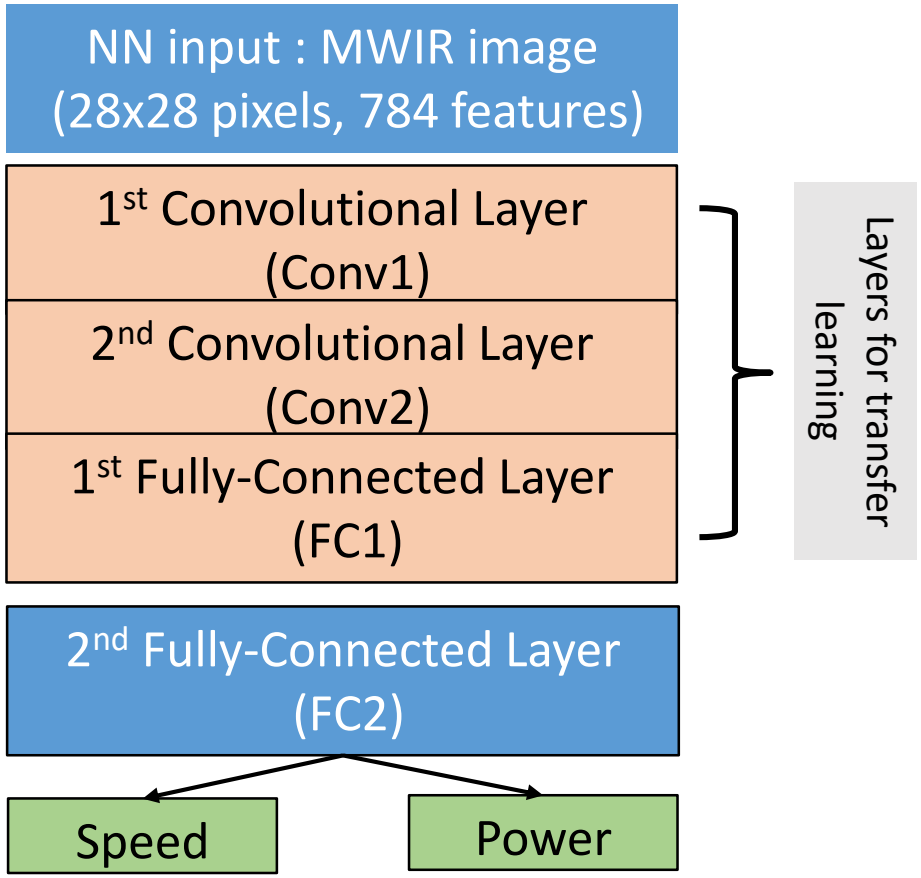
Use ML algorithms for monitoring the LMD process based on the MWIR images.



Co-funded by the European Union

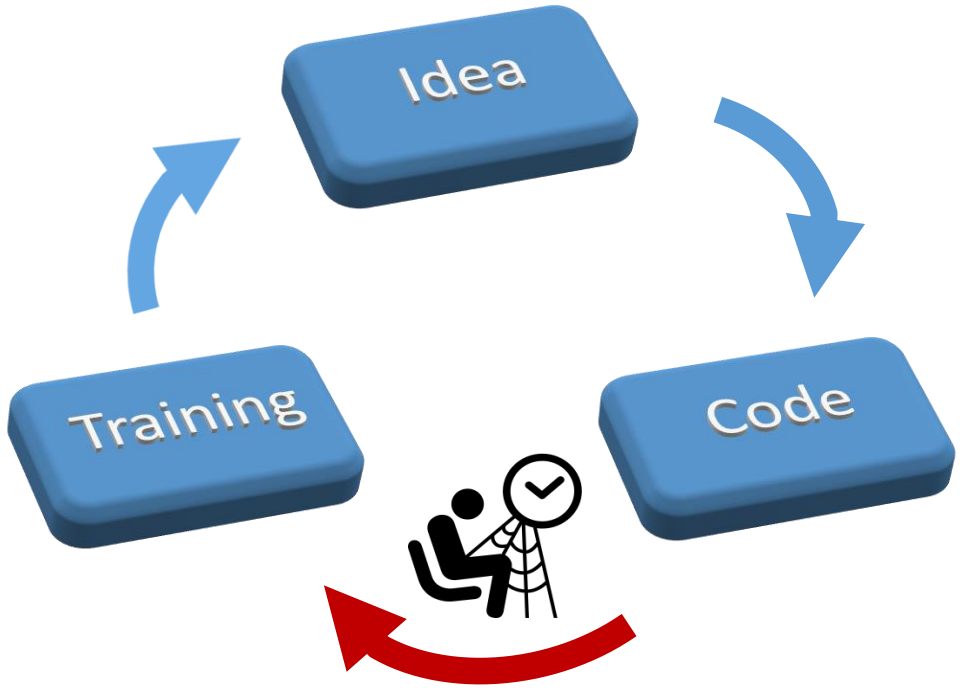
Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

CyPLAM Algorithm



NN Graph model (extracted from Tensorboard)

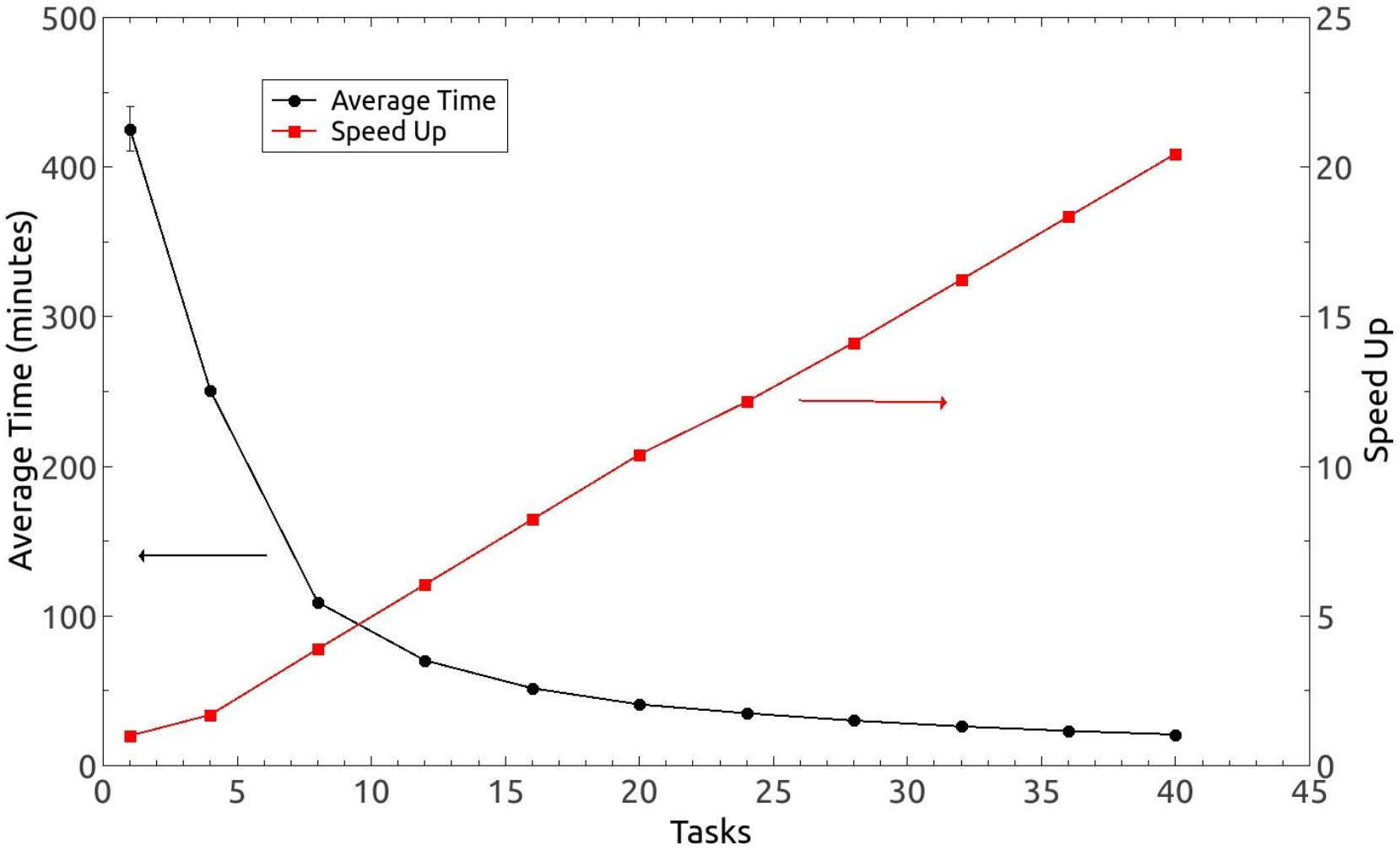
For Biggest tested model:
Training Time > 7 h.



Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

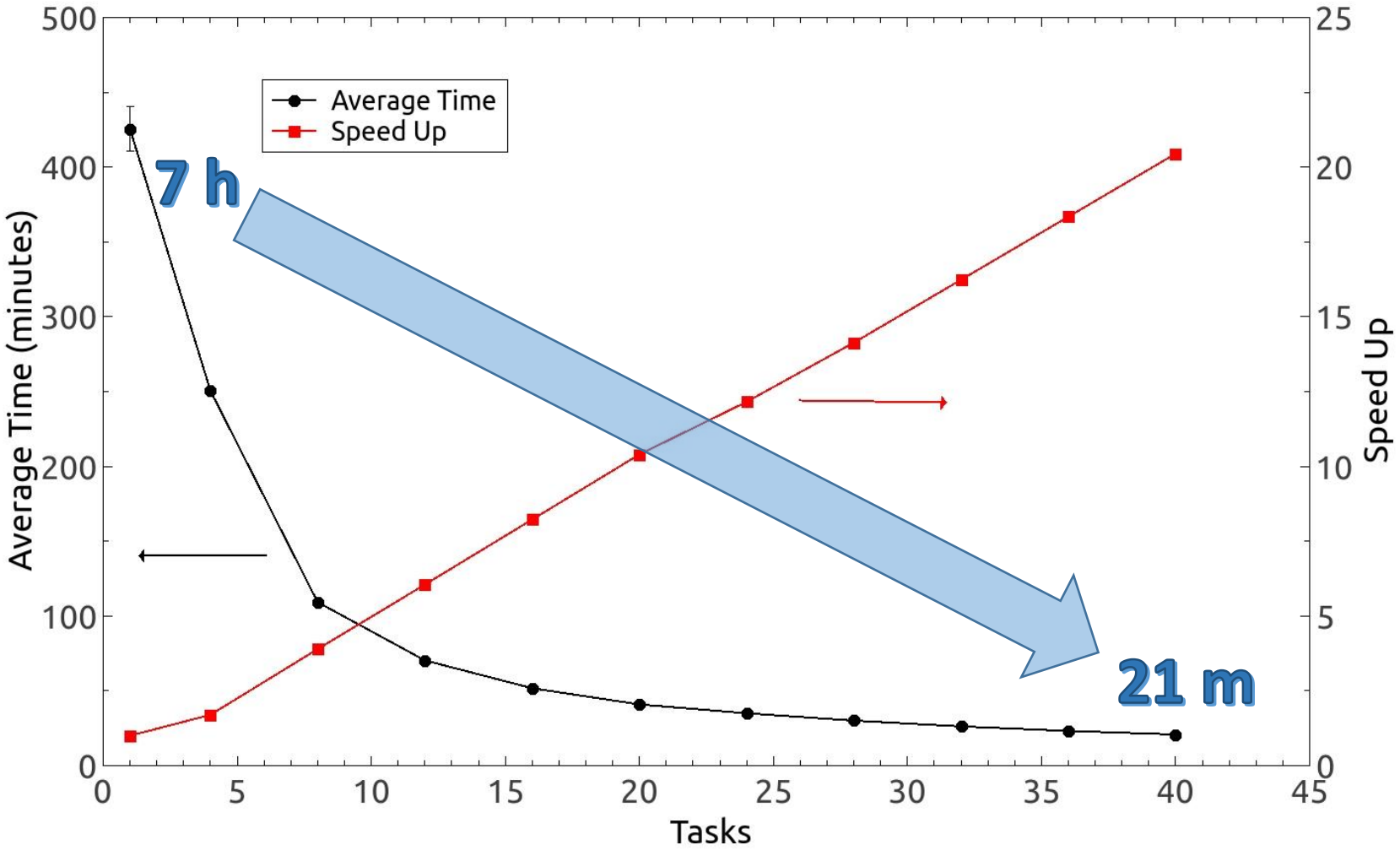
CyPLAM training using tf4slurm



Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

CyPLAM training using tf4slurm



7 hours

~20 minutes

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

Summary and Conclusions.

- Several issues deploying Distributed TensorFlow on CESGA Finis Terrae II were detected:
  - Queue system does not provide **mandatory IP:Port information** in advance.
  - HPC **Resources** are **wasted** due to Parameter **Servers running forever**.
- **tf4slurm** Python Package was developed to **solve** these detected **issues**.
- **tf4slurm** was **tested** using an Industrial Case:
  - Largest training **reduced from 7 hours to near 20 minutes**.
- **HPC can greatly decrease the design time** of ML algorithms boosting productivity.

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481

FORTISSIMO

CESGA

THANKS FOR YOUR ATTENTION !!!

Technical Report:

- https://www.cesga.es/es/biblioteca/downloadAsset/id/803

GitHub repository:

- https://github.com/gonfeco/tf4slurm

Co-funded by the European Union

Fortissimo has received funding from the European Union's H2020 research and innovation programme under grant agreement No 680481