# Deep Learning Optimisé - Jean Zay

## Conclusion

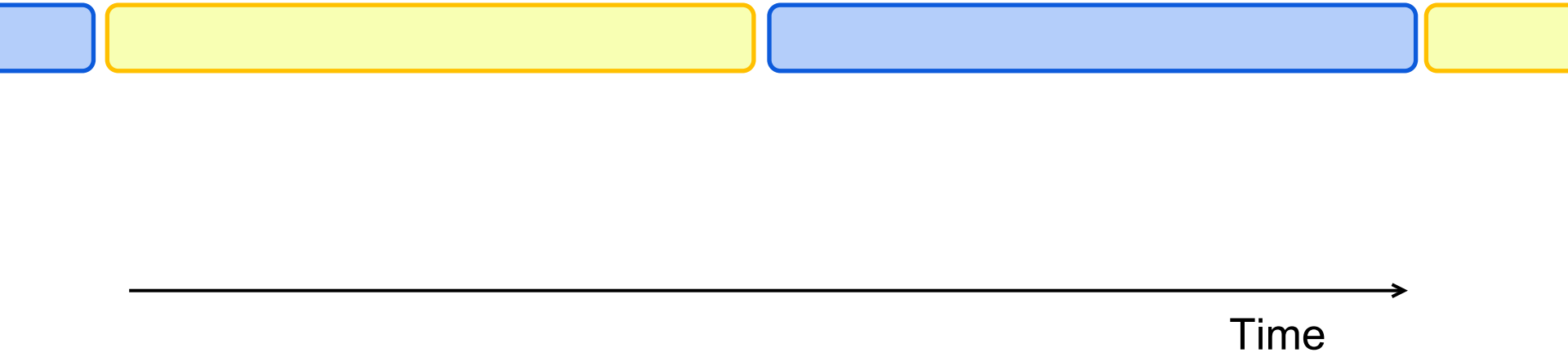

INSTITUT DU
DÉVELOPPEMENT ET DES
RESSOURCES EN
INFORMATIQUE
SCIENTIFIQUE

# Conclusion

**Synchronous : num_worker = 0**

DataLoader

Forward/Backward

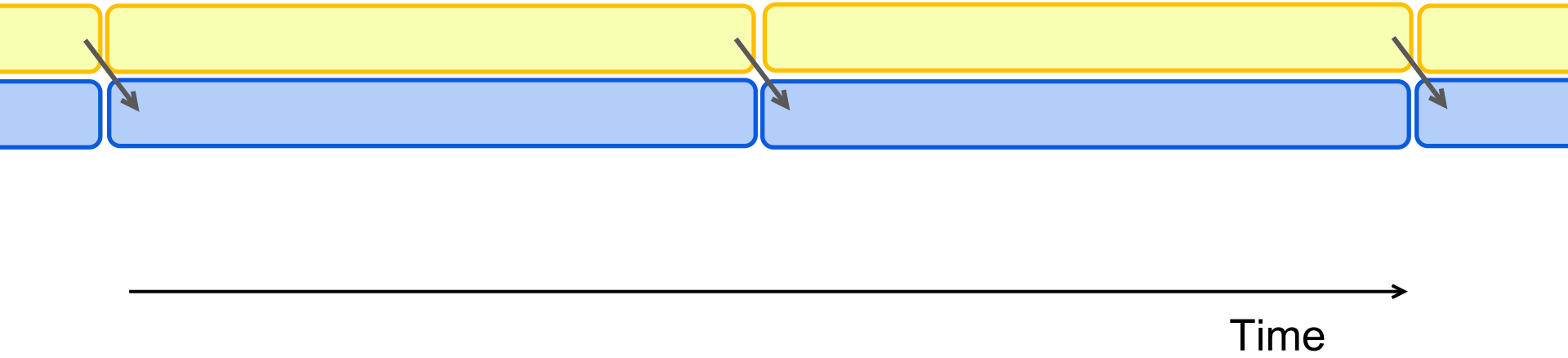


# Conclusion

**Asynchronous : num_worker > 0**

DataLoader

Forward/Backward

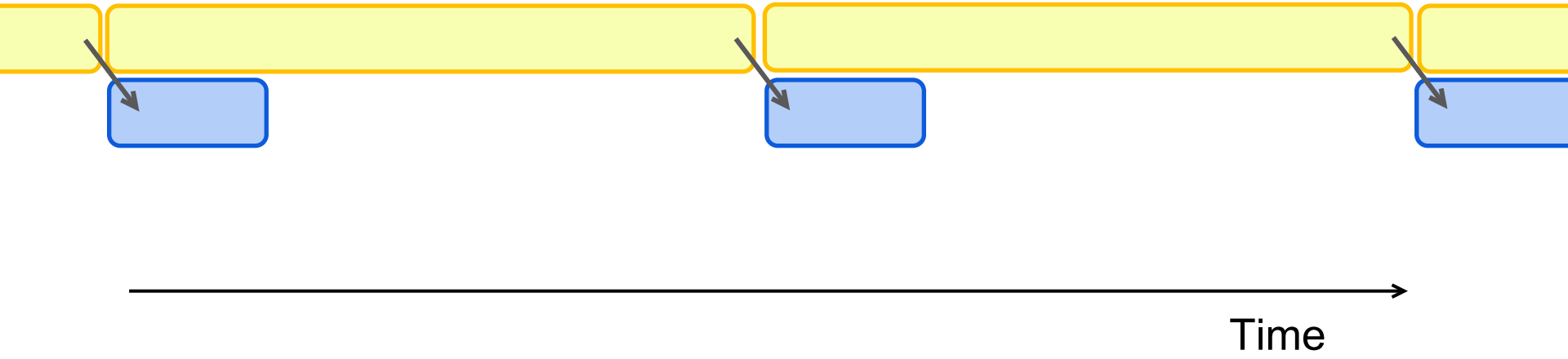


# Conclusion

**GPU Computing, Mixed Precision, torch.compile, ...**

DataLoader

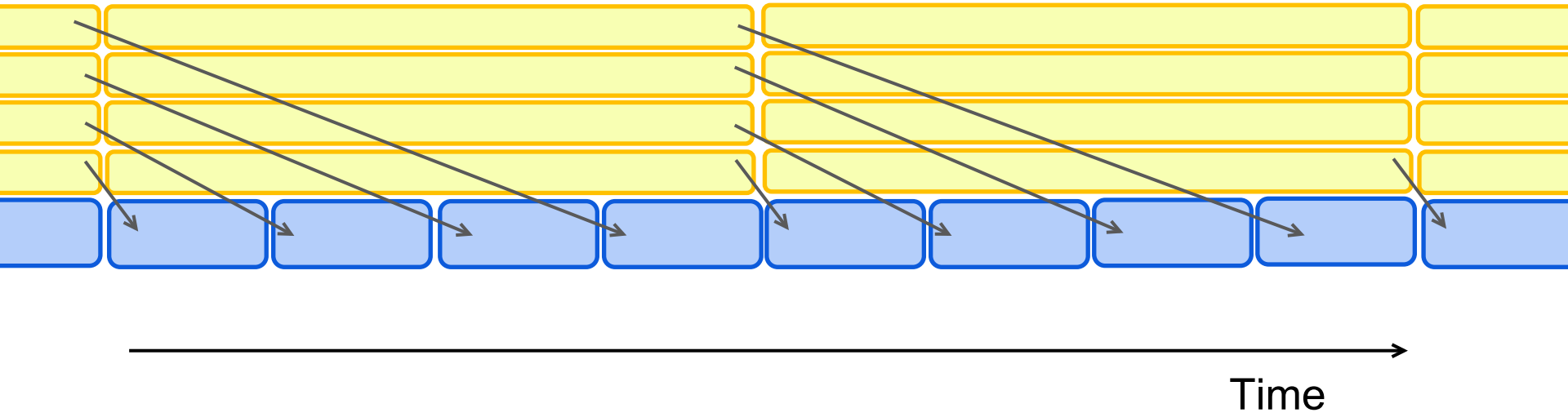
Forward/Backward



# Conclusion

**DataLoader Optim. : num_worker > 1, ...**

DataLoader Forward/Backward



# Conclusion

Training take too long !!!

Increase your batch size

# Conclusion

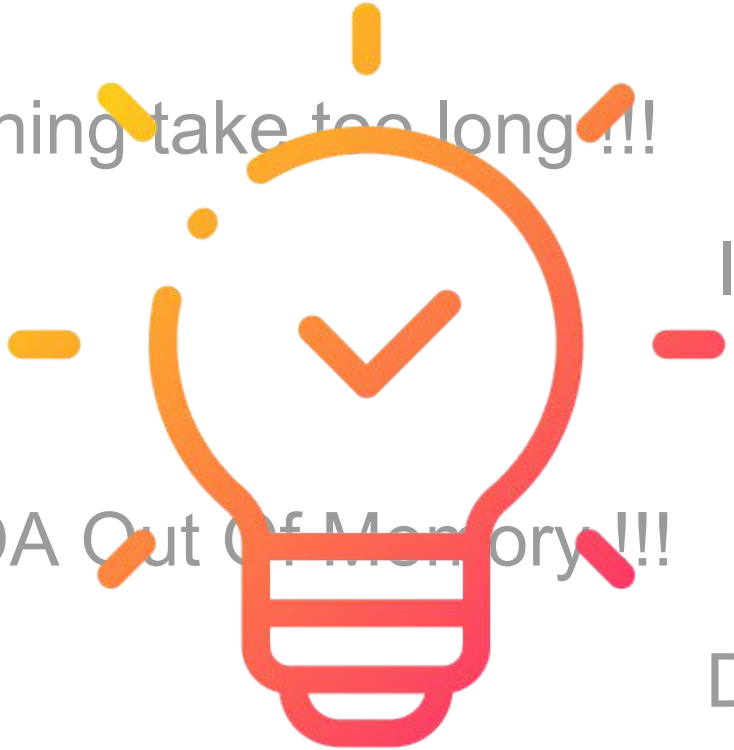
Training take too long !!!

Increase your batch size

CUDA Out Of Memory !!!

Decrease your batch size

# Conclusion



Training take too long !!!

Increase your batch size

CUDA Out Of Memory !!!

Decrease your batch size

# Conclusion

For Small Model !!!

10s or 100s M Params

**Distributed Data Parallelism**

# Conclusion

For Small Model !!!

10s or 100s M Params

**Distributed Data Parallelism**

⟹ Large Batch !!

# Conclusion

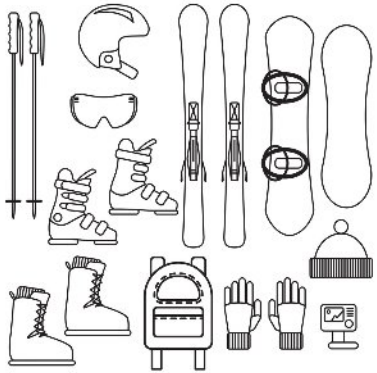
SGD AdamW

Weight Decay

LARS

LAMB

Batch Scheduler



LR Scheduler

LR scaling

Warmup LR Decay

# Conclusion

For Large Model !!!

> 1G Params

ZeRO

FSDP

**Model Parallelisms**

Pipeline Parallelism

Tensor Parallelism