

# BioExcel Webinar Series #37

## More bang for your buck: Improved use of GPU Nodes for GROMACS 2018

**Presenter:** Carsten Kutzner (Max Planck Institute)

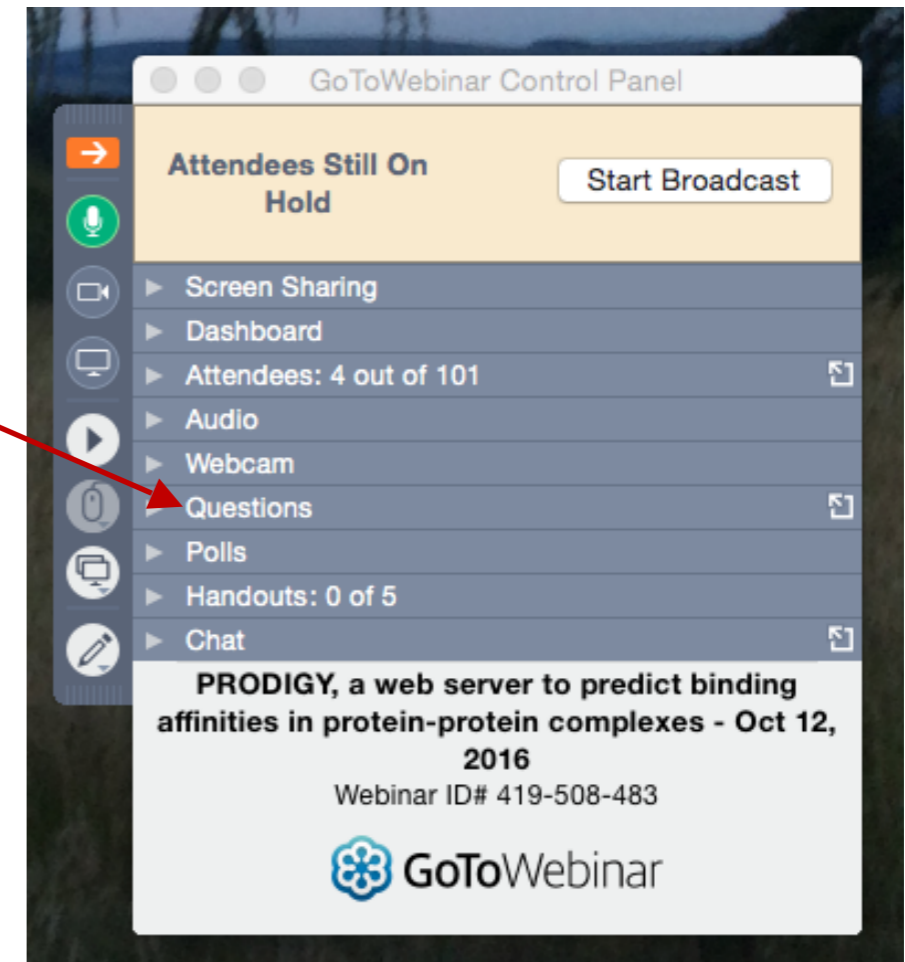
**Host:** Rossen Apostolov (KTH Royal Institute of Technology)



This webinar is being recorded

# Audience Q&A session

- Please use the Questions function in GoToWebinar application
- Any other questions or points to discuss after the live webinar? Join the discussions at <http://ask.bioexcel.eu>.



# Today's Presenter



**Carsten Kutzner**

*Max Planck Institute for Biophysical Chemistry*

Carsten studied physics at the University of Göttingen. For his PhD he focused on numerical simulations of Earth's magnetic field, which brought him in contact with high performance and parallel computing. After a stay at the MPI for Solar System Research he moved to computational biophysics. Since 2004 he has been working at the Max Planck Institute for Biophysical Chemistry in the lab of Helmut Grubmüller. He is interested in method development, high performance computing, and atomistic biomolecular simulations.

**Twitter:**

@kutznercarsten <https://twitter.com/kutznercarsten>

@CompBioPhys <https://twitter.com/CompBioPhys>

**Homepage:**

<https://www.mpibpc.mpg.de/grubmueller/kutzner>

Carsten Kutzner<sup>1</sup>, Szilárd Páll<sup>2</sup>, Martin Fechner<sup>1</sup>, Ansgar Esztermann<sup>1</sup>,  
Bert de Groot<sup>1</sup>, Helmut Grubmüller<sup>1</sup>

**MORE**



**FOR YOUR BUCK!**

---

Improved use of GPU Nodes for GROMACS 2018

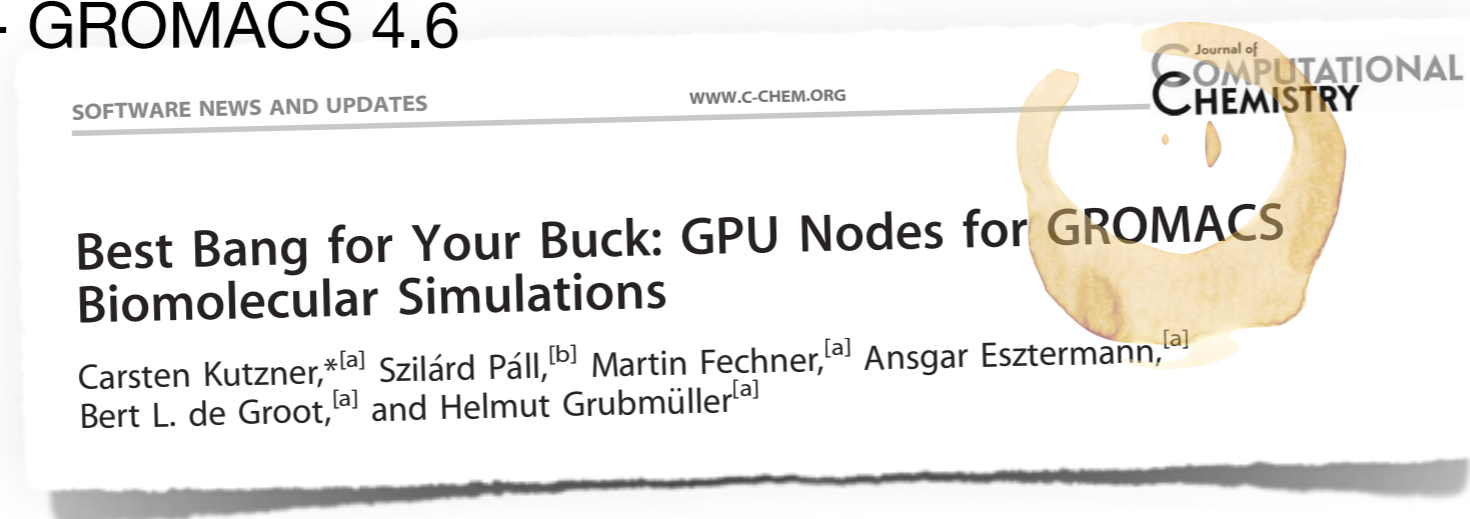
# Motivation

- Many MD groups buy small compute clusters from a **fixed budget**
- How to optimally make use of that?
  - We run mostly **GROMACS** MD,  
→ tailor nodes for GROMACS,  
maximise cost-efficiency by specialisation
  - queue is always full → optimise for  
**throughput / single-node performance**
  - (scaling → HPC centres)
- **Given a fixed budget,**  
**how can we produce as much MD trajectory as possible?**



# Outline

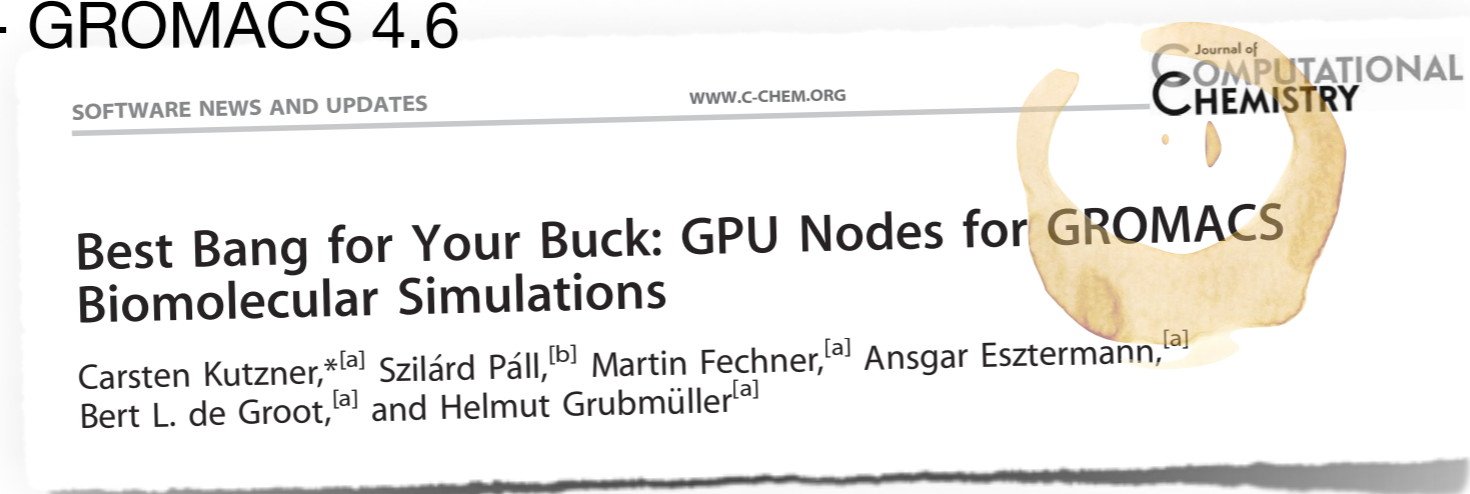
- 2014 - GROMACS 4.6



- recap: what were our conclusions in 2014/15?
- hardware & software developments and their impact
- update

# Outline

- 2014 - GROMACS 4.6



- recap: what were our conclusions in 2014/15?
- hardware & software developments and their impact
- update





# Approach

- from ~10 CPU types + ~10 GPU models we **assemble** and **benchmark** various compute nodes
  - CPU nodes
  - GPU nodes with 1, 2, 3, and 4 GPUs
  - **consumer** and **professional** GPUs
- determine **performance-to-price (P/P) ratio**

**GTX 980**  
**GTX 1070**  
**GTX 1070Ti**  
**GTX 1080**  
**GTX 1080Ti**  
**RTX 2070**  
**RTX 2080**  
**RTX 2080Ti**

**consumer GPUs (GeForce)**

**Quadro P6000**  
**Tesla V100**

**professional GPUs (Tesla)**

**Ryzen (16 core)**  
**Epyc (24 core)**  
**Core i7 (4 core)**  
**Xeon (4, 6, 8, 10, and 20 core)**

**CPUs**

# Approach

- from ~10 CPU types + ~10 GPU models we **assemble** and **benchmark** various compute nodes
  - CPU nodes
  - GPU nodes with 1, 2, 3, and 4 GPUs
  - **consumer** and **professional** GPUs
- determine **performance-to-price (P/P) ratio**
- no comprehensive evaluation of currently available hardware!
  - but aim to uncover HW with good P/P ratio
- no strong scaling!

**GTX 980**  
**GTX 1070**  
**GTX 1070Ti**  
**GTX 1080**  
**GTX 1080Ti**  
**RTX 2070**  
**RTX 2080**  
**RTX 2080Ti**

**consumer GPUs (GeForce)**

**Quadro P6000**  
**Tesla V100**

**professional GPUs (Tesla)**

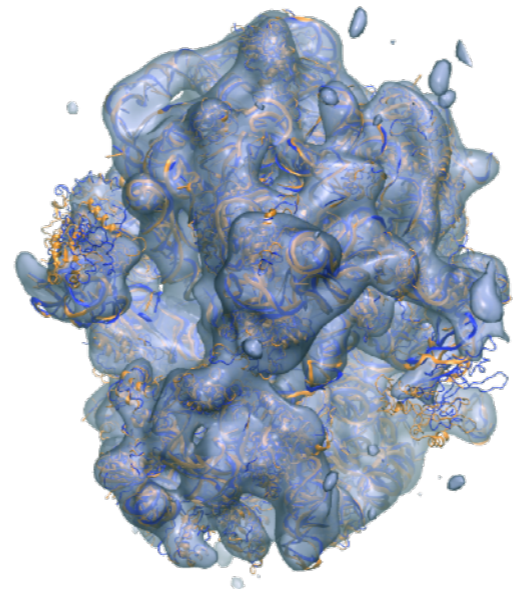
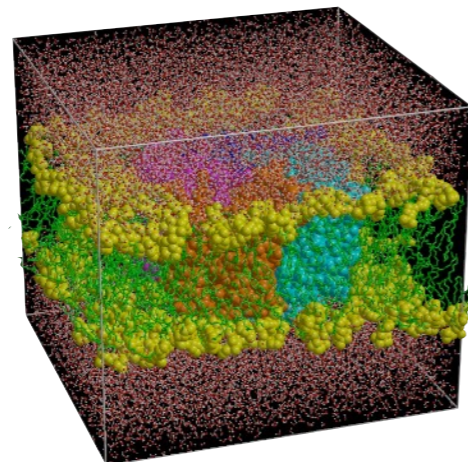
**Ryzen (16 core)**  
**Epyc (24 core)**  
**Core i7 (4 core)**  
**Xeon (4, 6, 8, 10, and 20 core)**

**CPUs**

# Approach

- from ~10 CPU types + ~10 GPU models we **assemble** and **benchmark** various compute nodes
  - CPU nodes
  - GPU nodes with 1, 2, 3, and 4 GPUs
  - **consumer** and **professional** GPUs
- determine **performance-to-price (P/P) ratio**
- no comprehensive evaluation of currently available hardware!
  - but aim to uncover HW with good P/P ratio
- no strong scaling!
- benchmark MD systems:

**80k atom MEM benchmark**  
channel in membrane +  
water + ions, PME, 2 fs  
time step



**GTX 980**  
**GTX 1070**  
**GTX 1070Ti**  
**GTX 1080**  
**GTX 1080Ti**  
**RTX 2070**  
**RTX 2080**  
**RTX 2080Ti**

**consumer GPUs (GeForce)**

**Quadro P6000**  
**Tesla V100**

**professional GPUs (Tesla)**

**Ryzen (16 core)**  
**Epyc (24 core)**  
**Core i7 (4 core)**  
**Xeon (4, 6, 8, 10, and 20 core)**

**CPUs**

**2M atoms RIB benchmark**  
ribosome in solution,  
PME, 4 fs time step

# What do we really want?

## Hardware requirements:



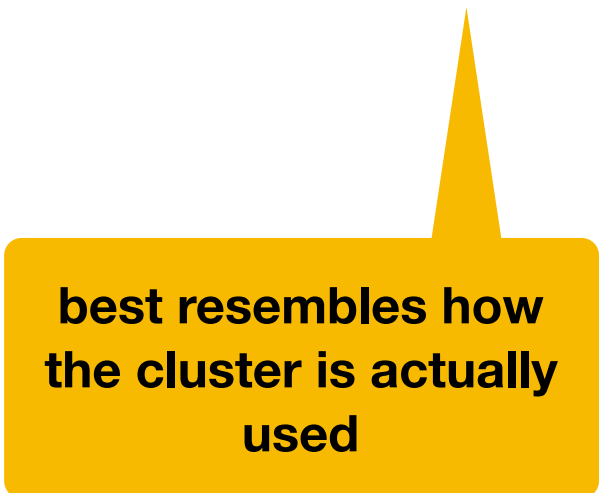
1. high performance-to-price (P/P) ratio
2. low energy consumption
3. low rack space requirements  
packing density at least 1 GPU per U
4. reasonably high performance of a single simulation  
→ one simulation per GPU on GPU nodes,  
one simulation per node on CPU nodes

# What do we really want?

## Hardware requirements:



1. high performance-to-price (P/P) ratio
2. low energy consumption
3. low rack space requirements  
packing density at least 1 GPU per U
4. reasonably high performance of a single simulation  
→ one simulation per GPU on GPU nodes,  
one simulation per node on CPU nodes



**best resembles how  
the cluster is actually  
used**

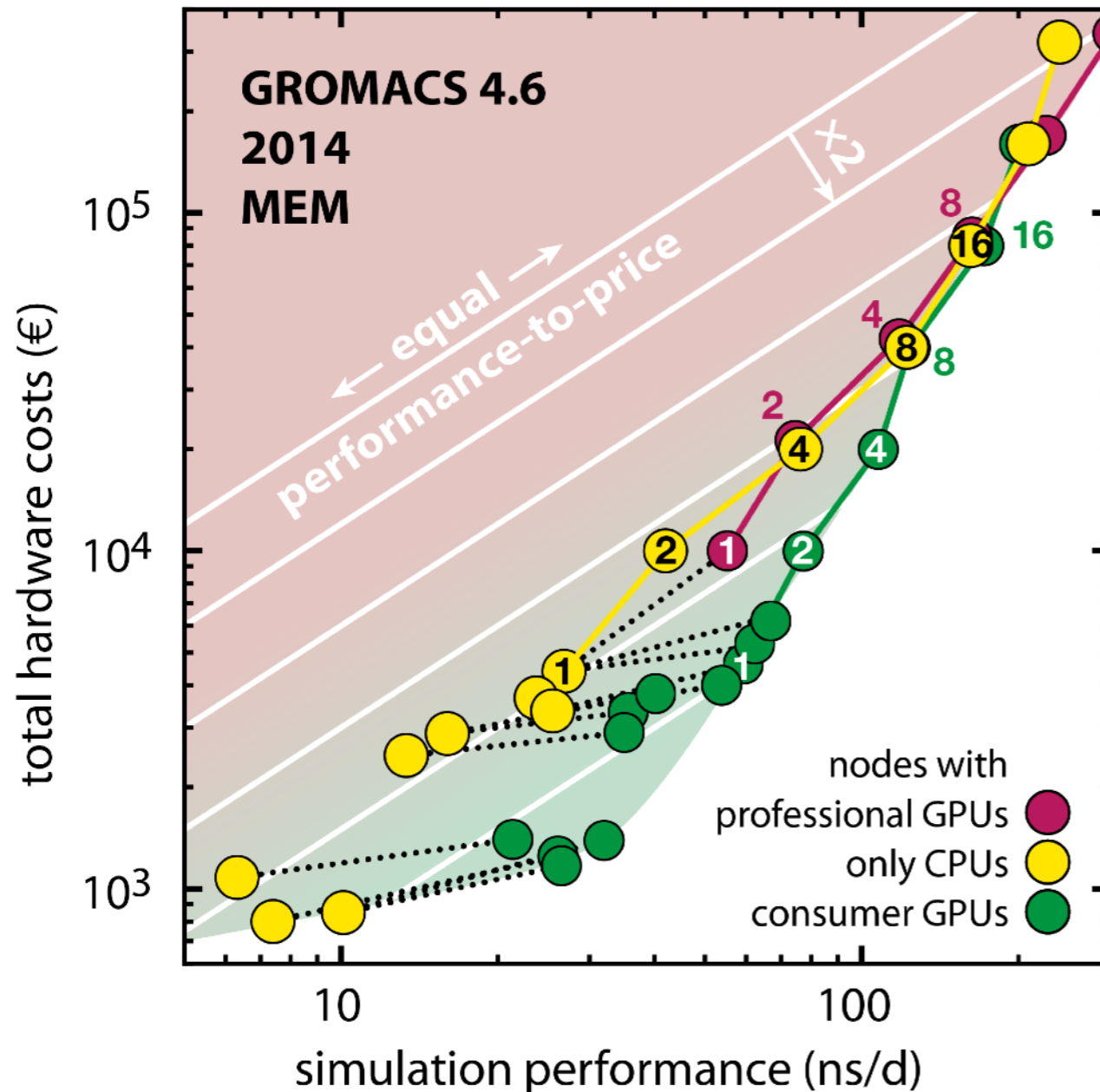
# Details for the hardware comparison benchmarks

- **GROMACS 2018**
- GCC 6.4 + CUDA 9.1
- GCC 5.4 + CUDA 8.0  
(~2.5% slower, taken into account)
- AVX2\_128 SIMD for AMD CPUs
- AVX2\_256 SIMD for recent Intel CPUs
  - (AVX\_256 SIMD for old Intel CPUs)
- OpenMP enabled
- Nodes with 2, 3, or 4 GPUs:
  - using Intel MPI 2017
- **Nodes booted from a common software image** (Scientific Linux 7.4)

# Details for the hardware comparison benchmarks

- **GROMACS 2018**
- GCC 6.4 + CUDA 9.1
- GCC 5.4 + CUDA 8.0  
(~2.5% slower, taken into account)
- AVX2\_128 SIMD for AMD CPUs
- AVX2\_256 SIMD for recent Intel CPUs
  - (AVX\_256 SIMD for old Intel CPUs)
- OpenMP enabled
- Nodes with 2, 3, or 4 GPUs:
  - using Intel MPI 2017
- **Nodes booted from a common software image** (Scientific Linux 7.4)
- benchmarks
  - average of two runs
  - MEM: 20,000 steps, average over last 5,000
  - RIB: run for 10,000 steps, average over last 2,000
- on multi-GPU nodes, benchmarks use 1 simulation per GPU (via -multidir),
  - reported node performance (ns/d) is sum of the performances of the individual simulations  
**(“aggregate” performance)**

# 2014: First Comprehensive Hardware Evaluation



- Main 2014 result:

- nodes with GeForce consumer GPUs

produce **2–3x** as much MD trajectory per invested € as

- CPU nodes

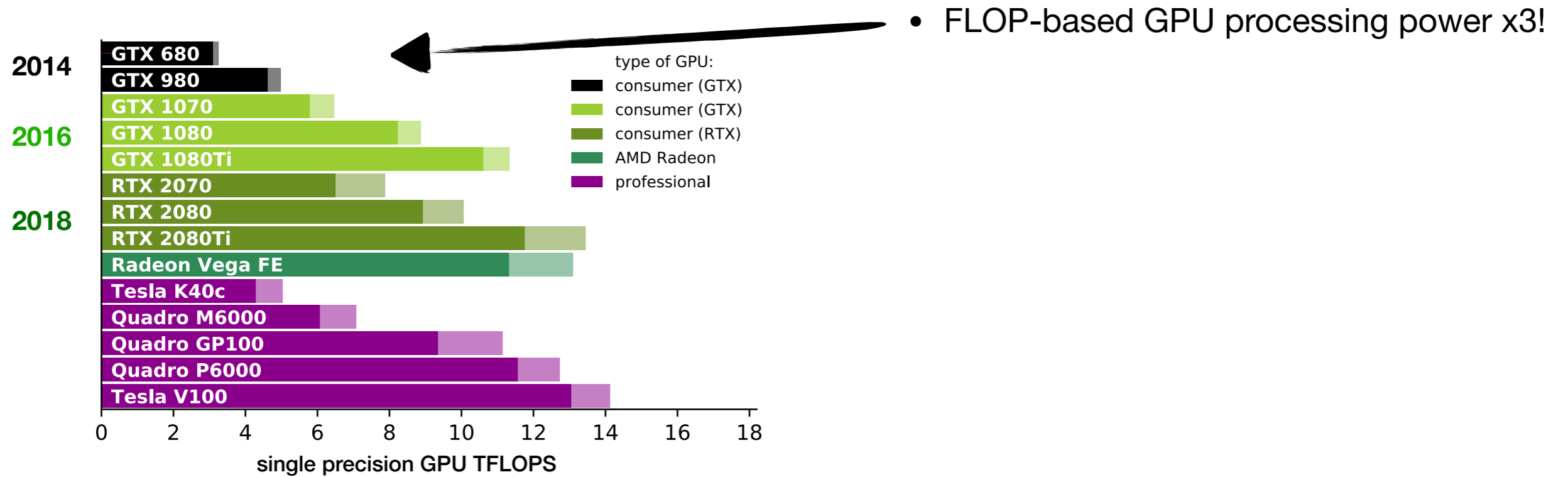
C Kutzner, S Páll, M Fechner, A Esztermann, BL de Groot, H Grubmüller.

**Best bang for your buck: GPU nodes for GROMACS biomolecular simulations.**

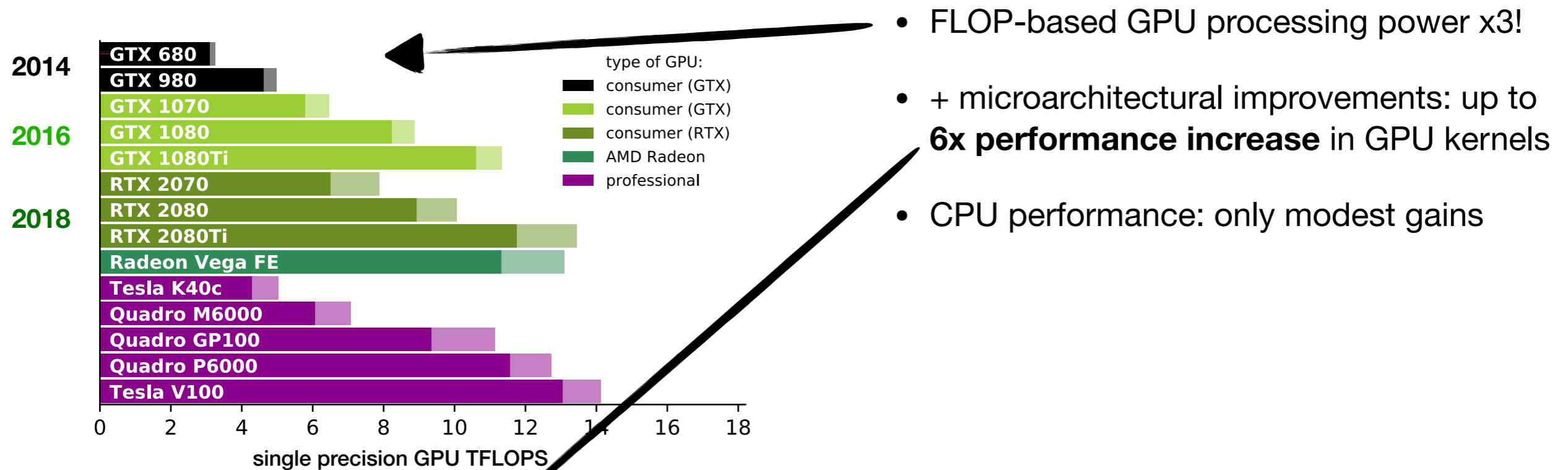
JCC 36 (26), pp. 1990 - 2008 (2015)



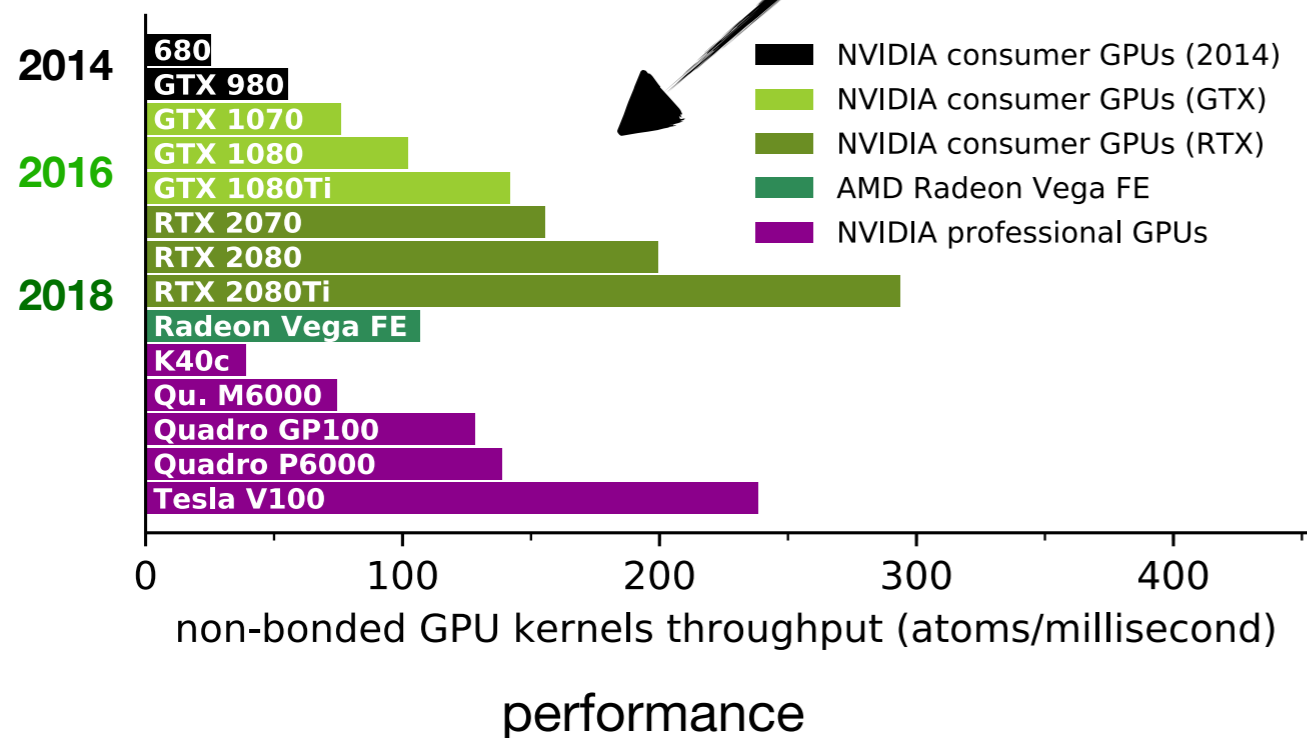
# Hardware Developments Since 2014



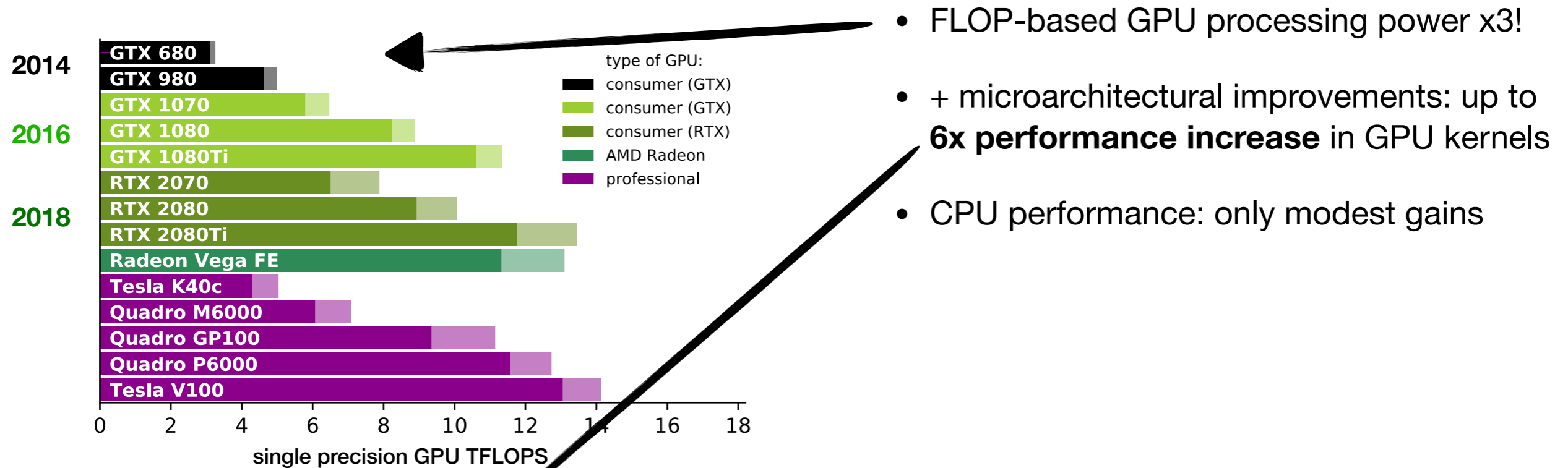
# Hardware Developments Since 2014



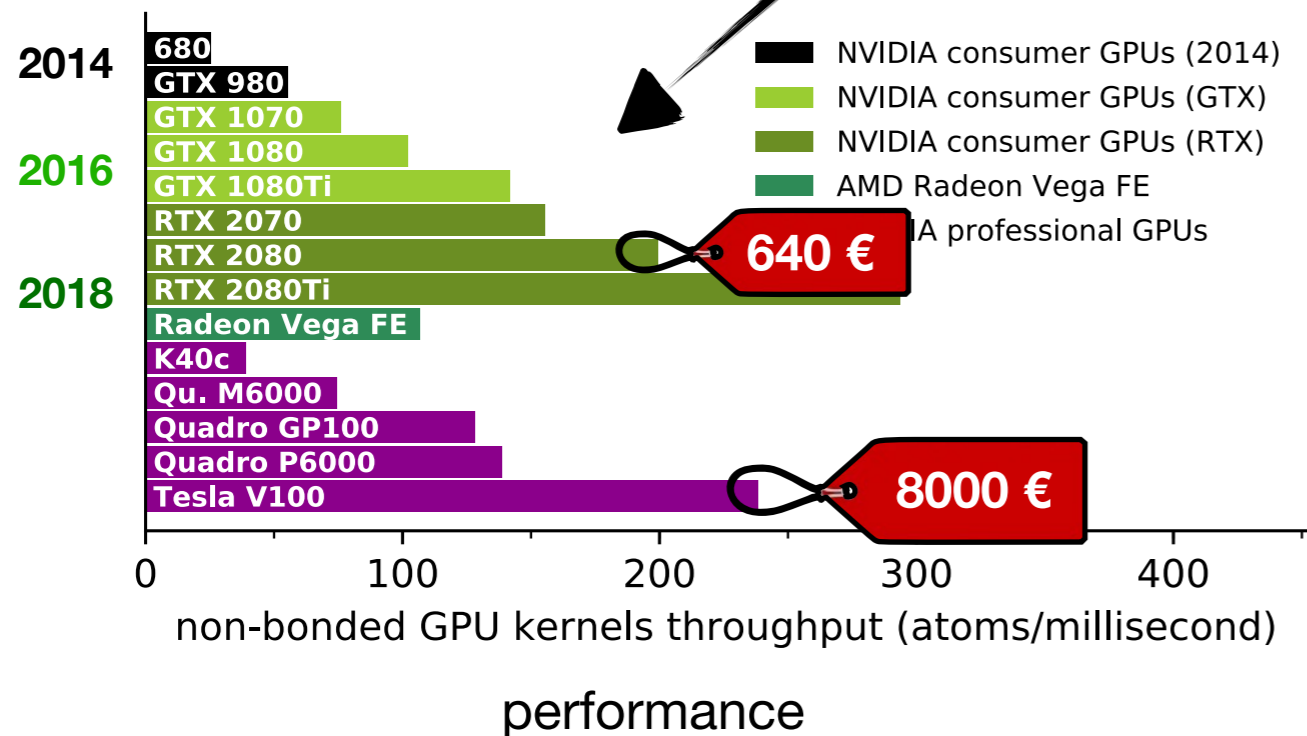
- FLOP-based GPU processing power x3!
- + microarchitectural improvements: up to **6x performance increase** in GPU kernels
- CPU performance: only modest gains



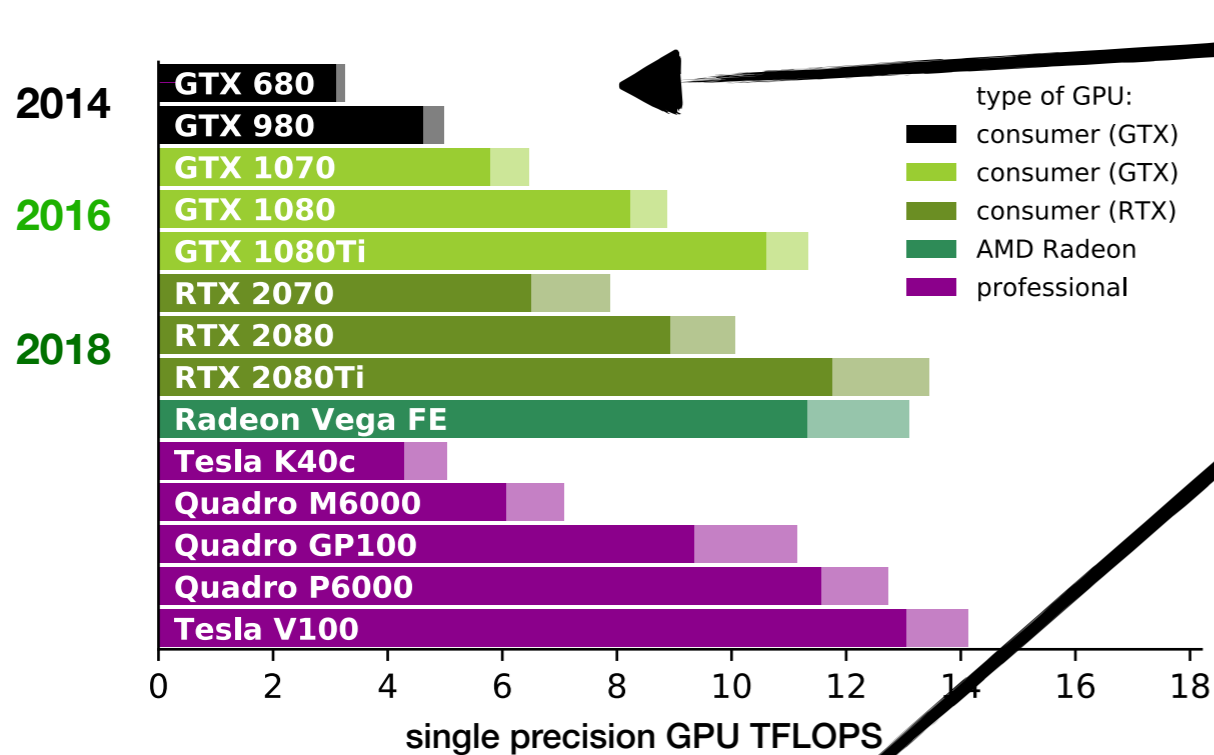
# Hardware Developments Since 2014



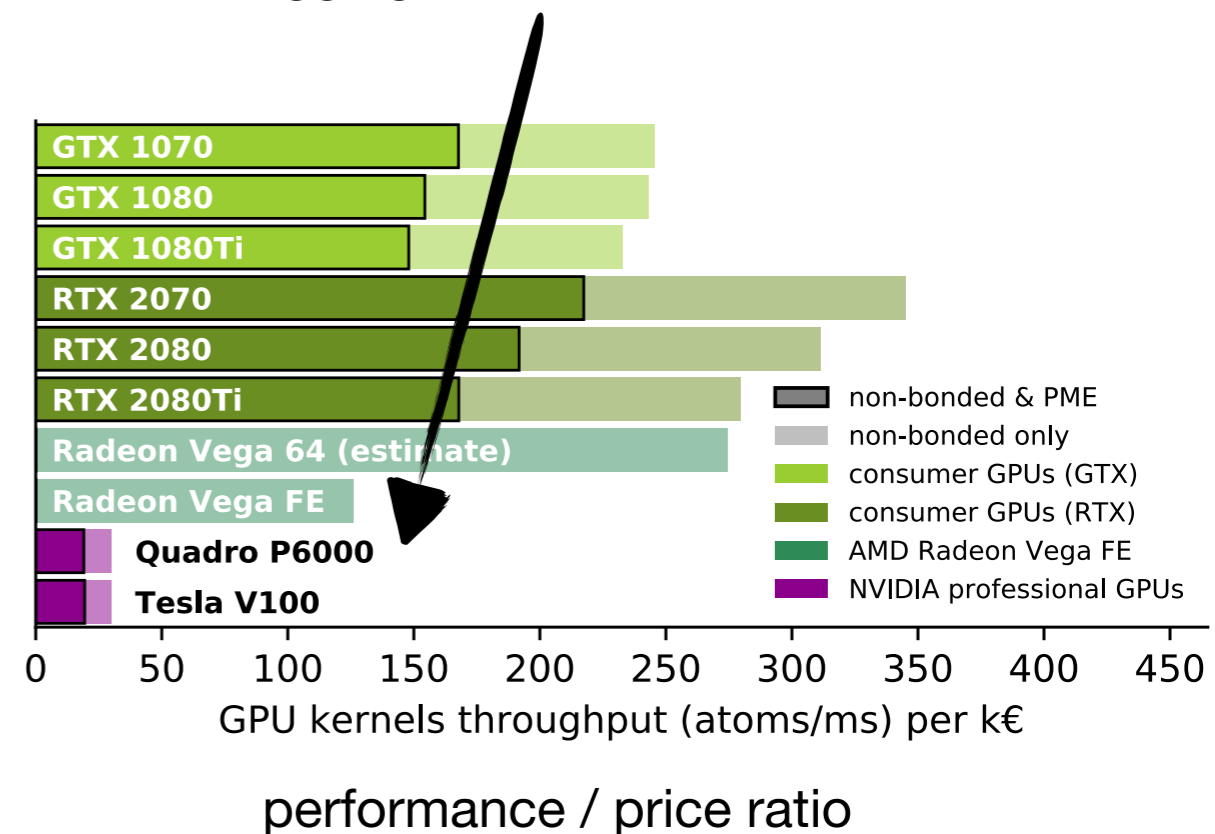
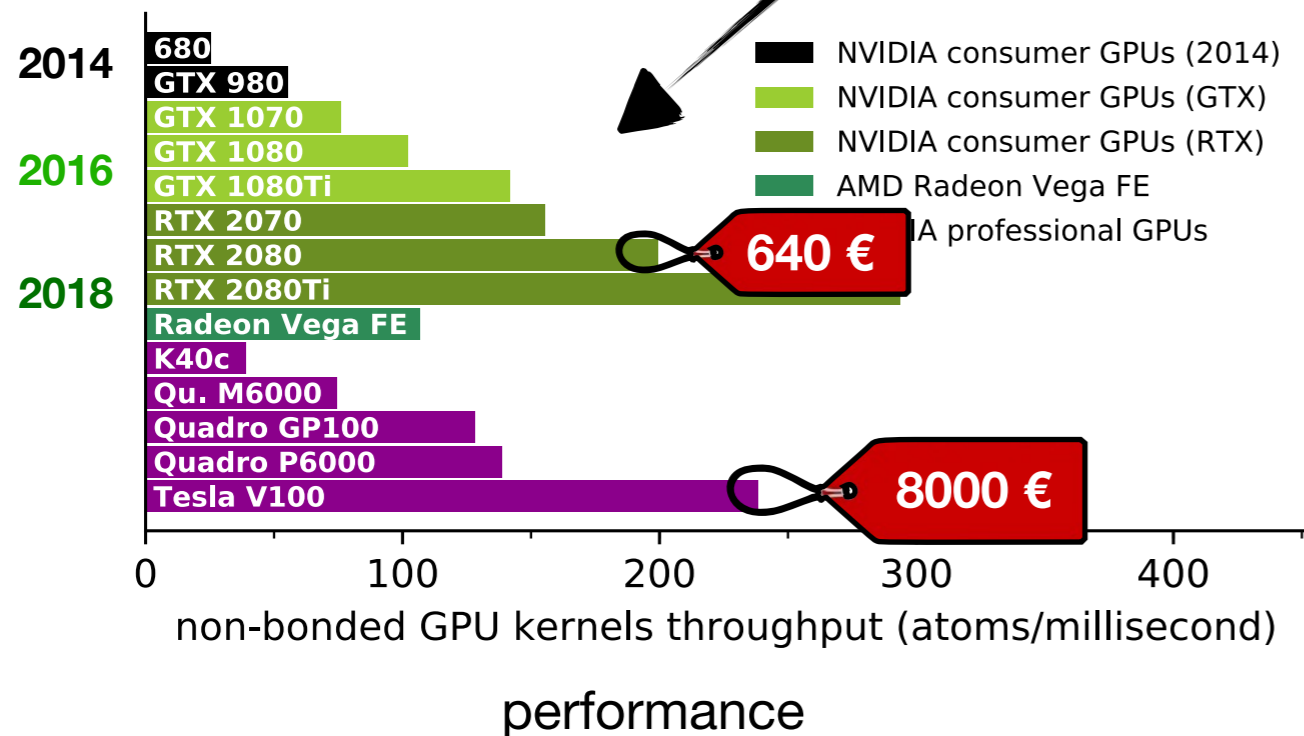
- FLOP-based GPU processing power x3!
- + microarchitectural improvements: up to **6x performance increase** in GPU kernels
- CPU performance: only modest gains



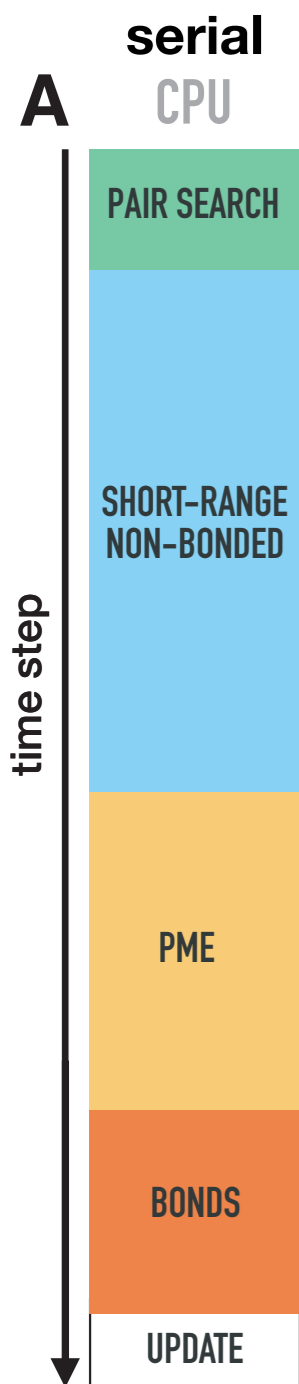
# Hardware Developments Since 2014



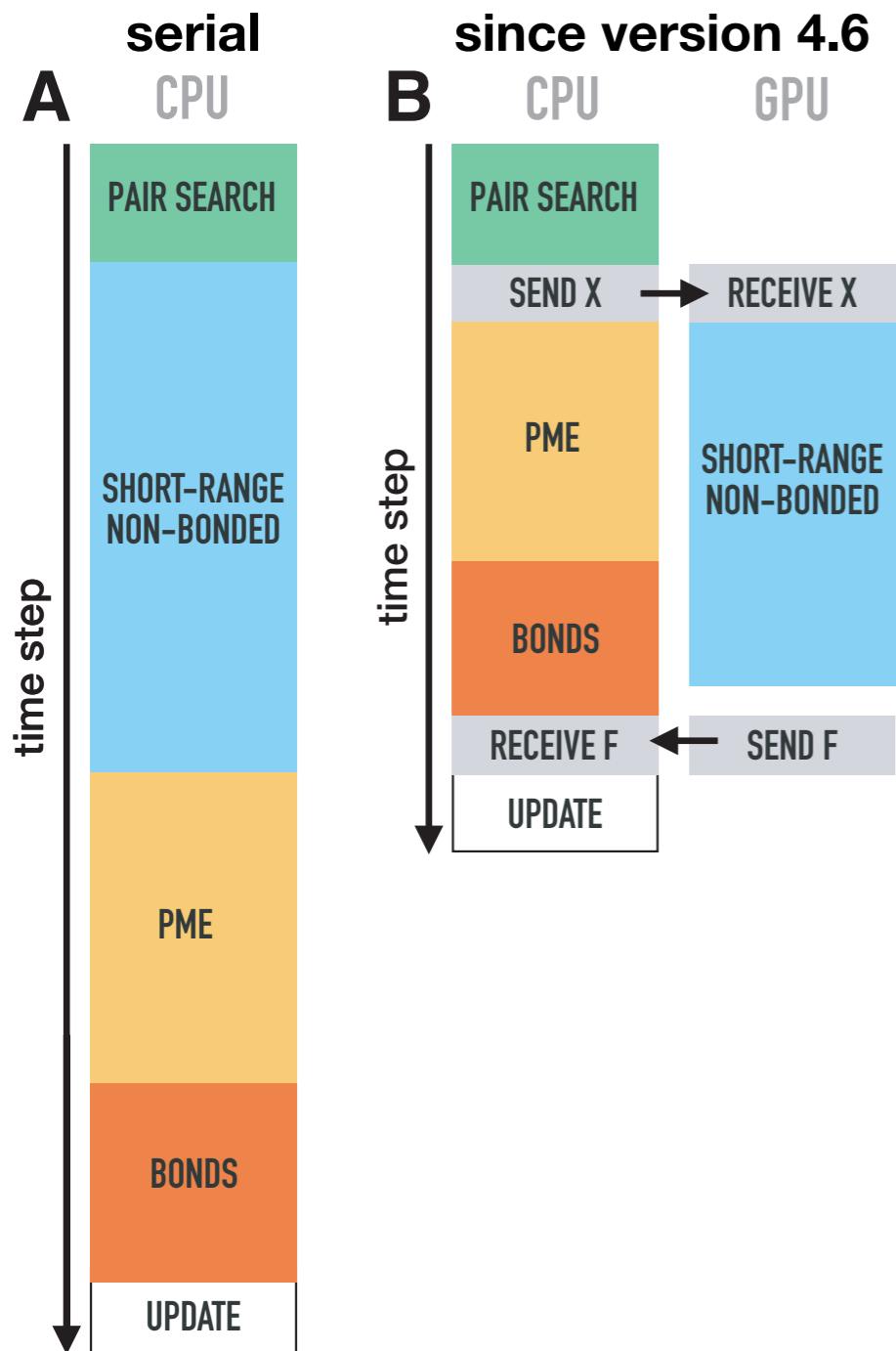
- FLOP-based GPU processing power x3!
- + microarchitectural improvements: up to **6x performance increase** in GPU kernels
- CPU performance: only modest gains
- **professional Tesla GPUs** compete with **consumer GPUs** in terms of performance, but are lagging far behind in terms of P/P



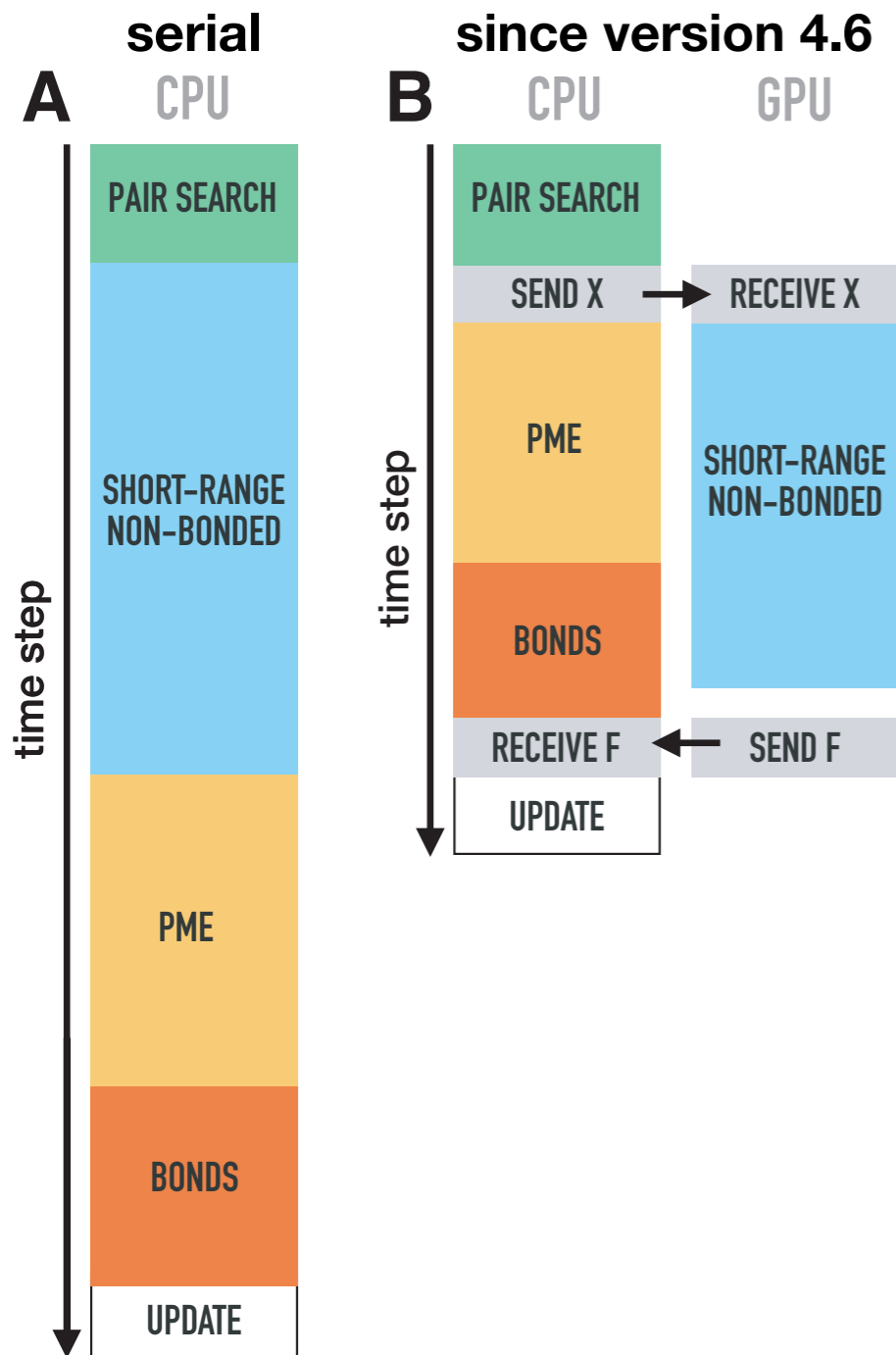
# Software Developments



# Software Developments



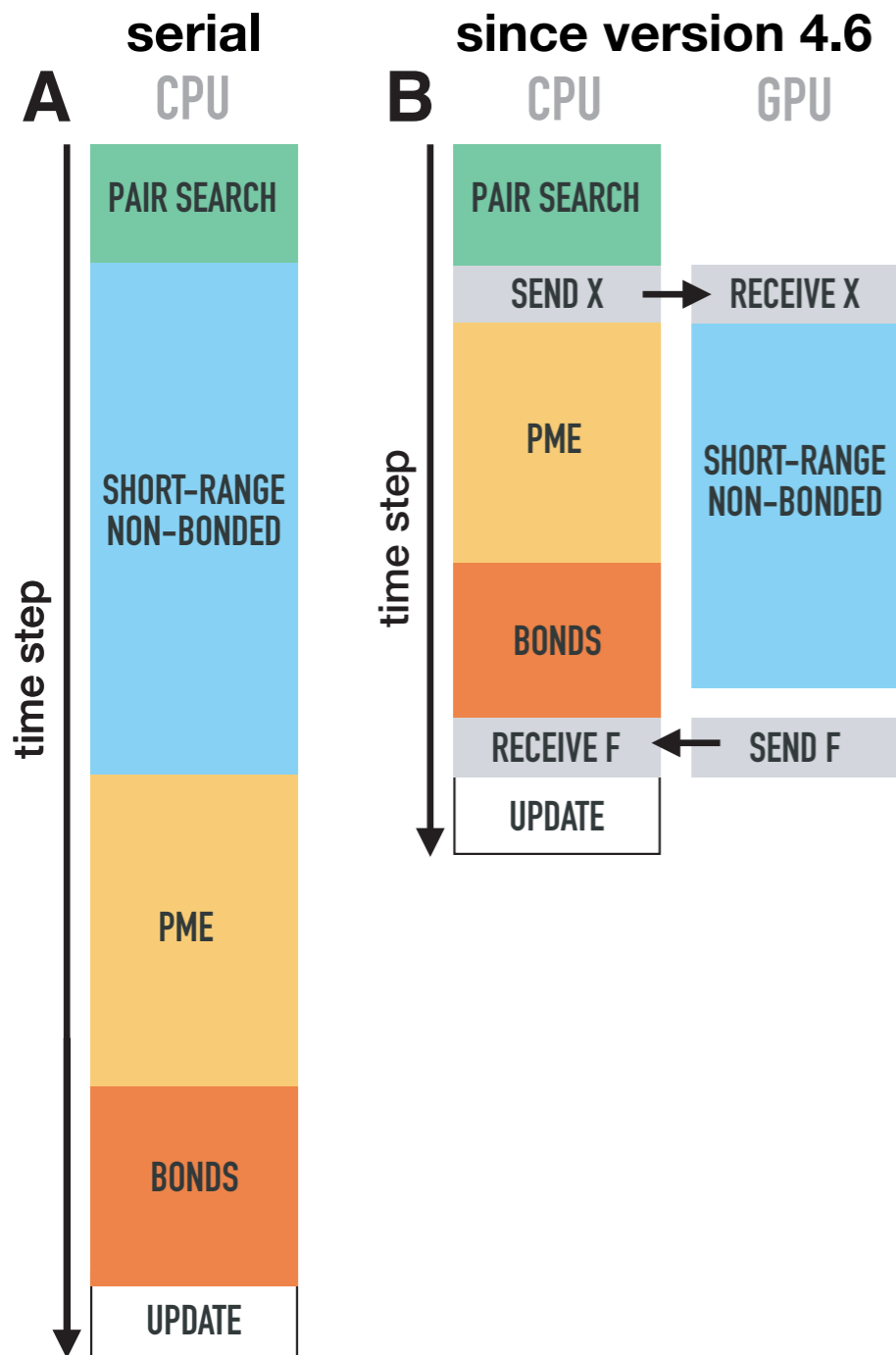
# Software Developments



since version 2018

1. dual pair lists with dynamic pruning
2. PME offloading

# Software Developments

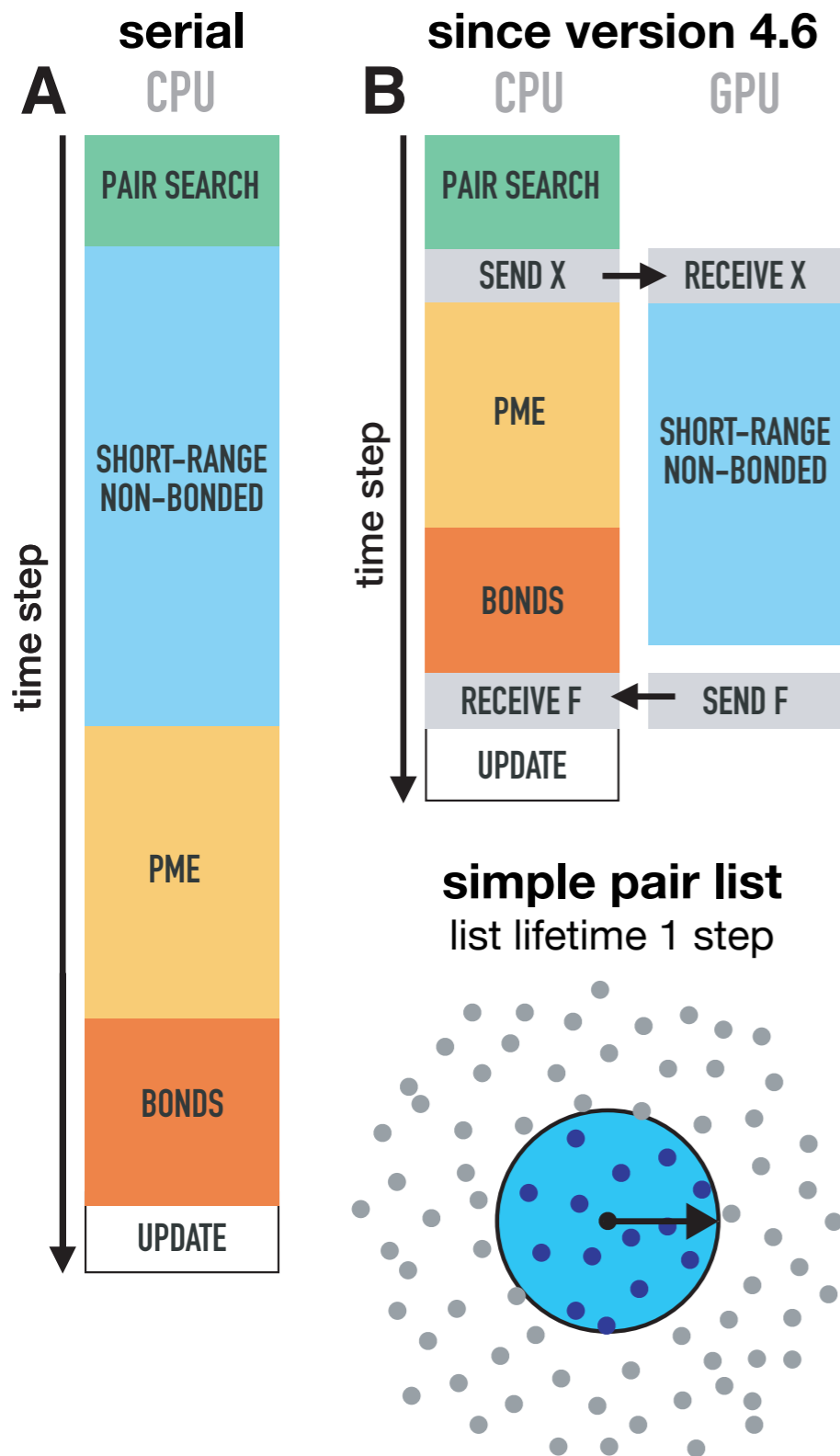


since version 2018

1. dual pair lists with dynamic pruning
2. PME offloading



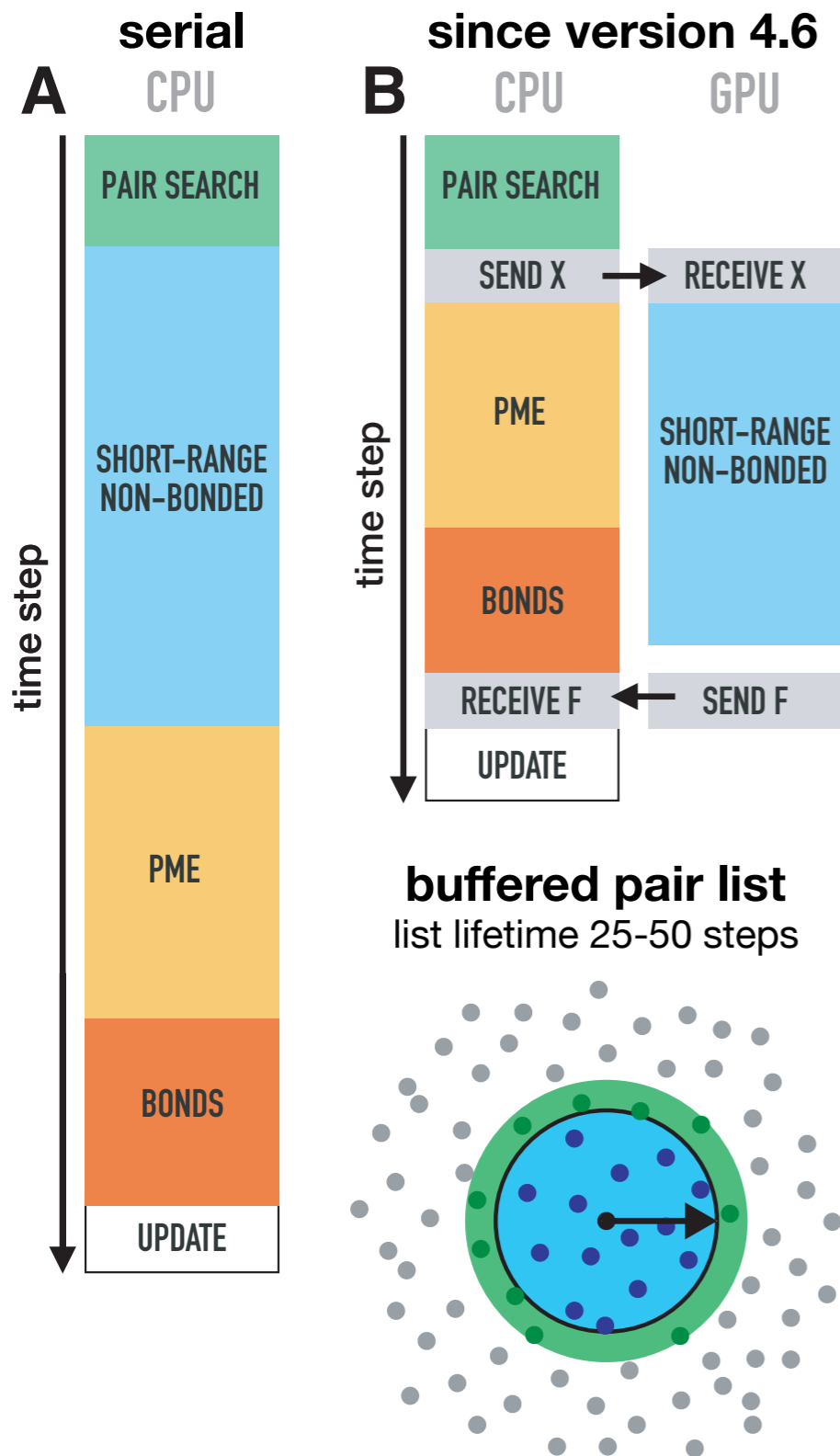
# Software Developments



since version 2018

1. dual pair lists with dynamic pruning
2. PME offloading

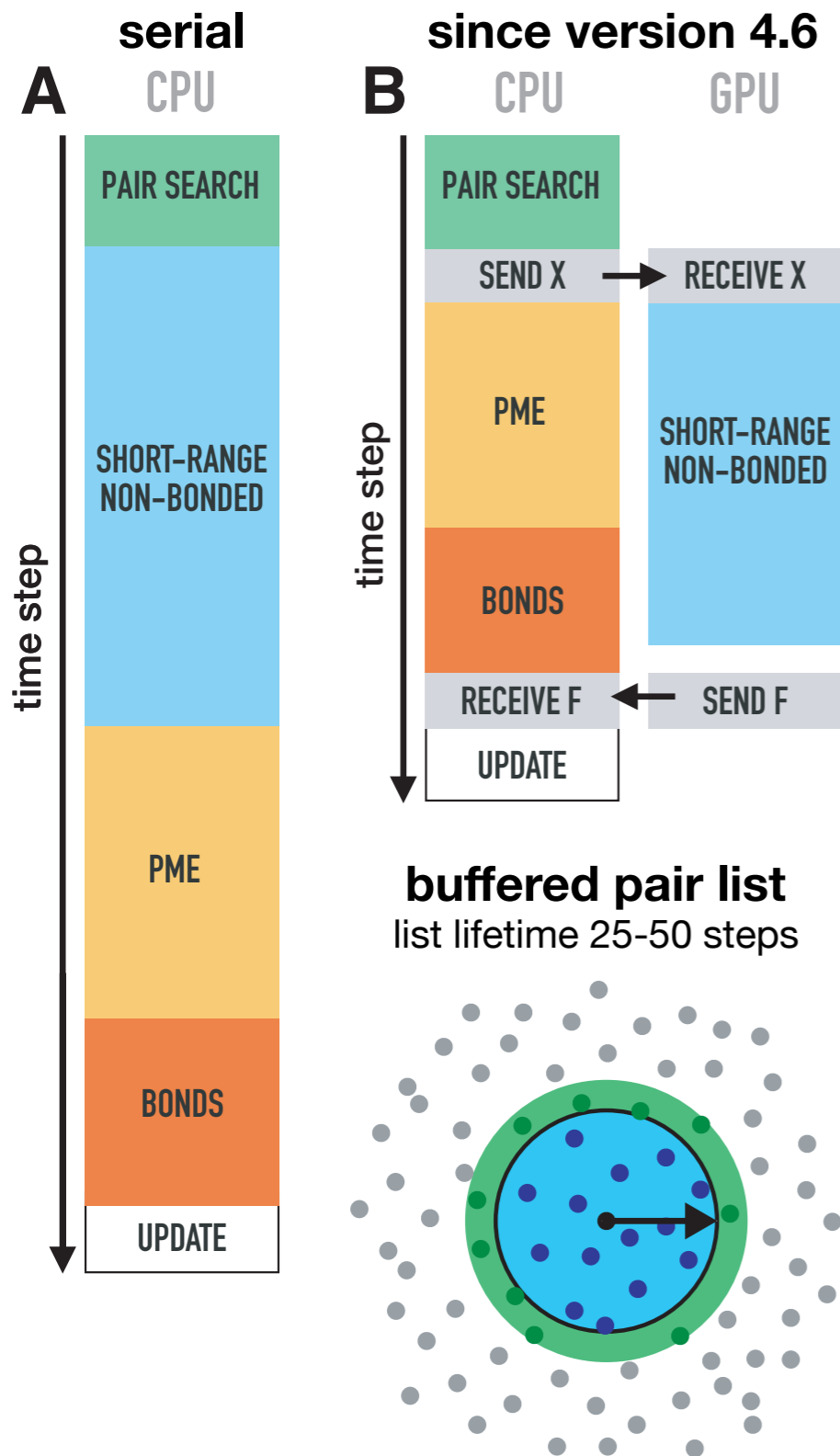
# Software Developments



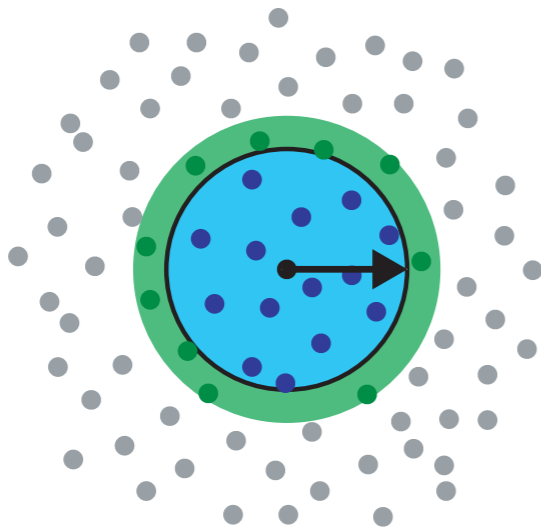
since version 2018

1. dual pair lists with dynamic pruning
2. PME offloading

# Software Developments



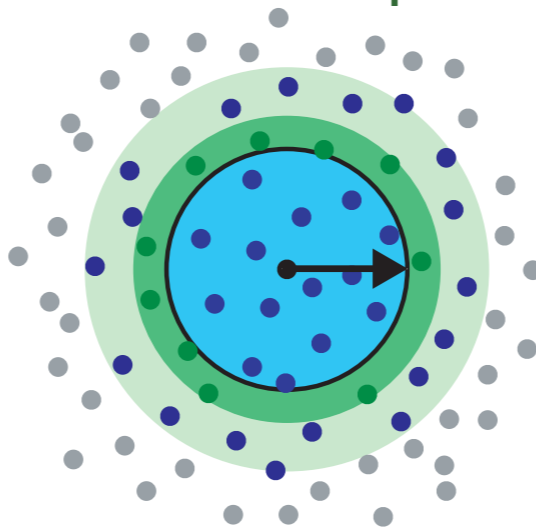
**buffered pair list**  
list lifetime 25-50 steps



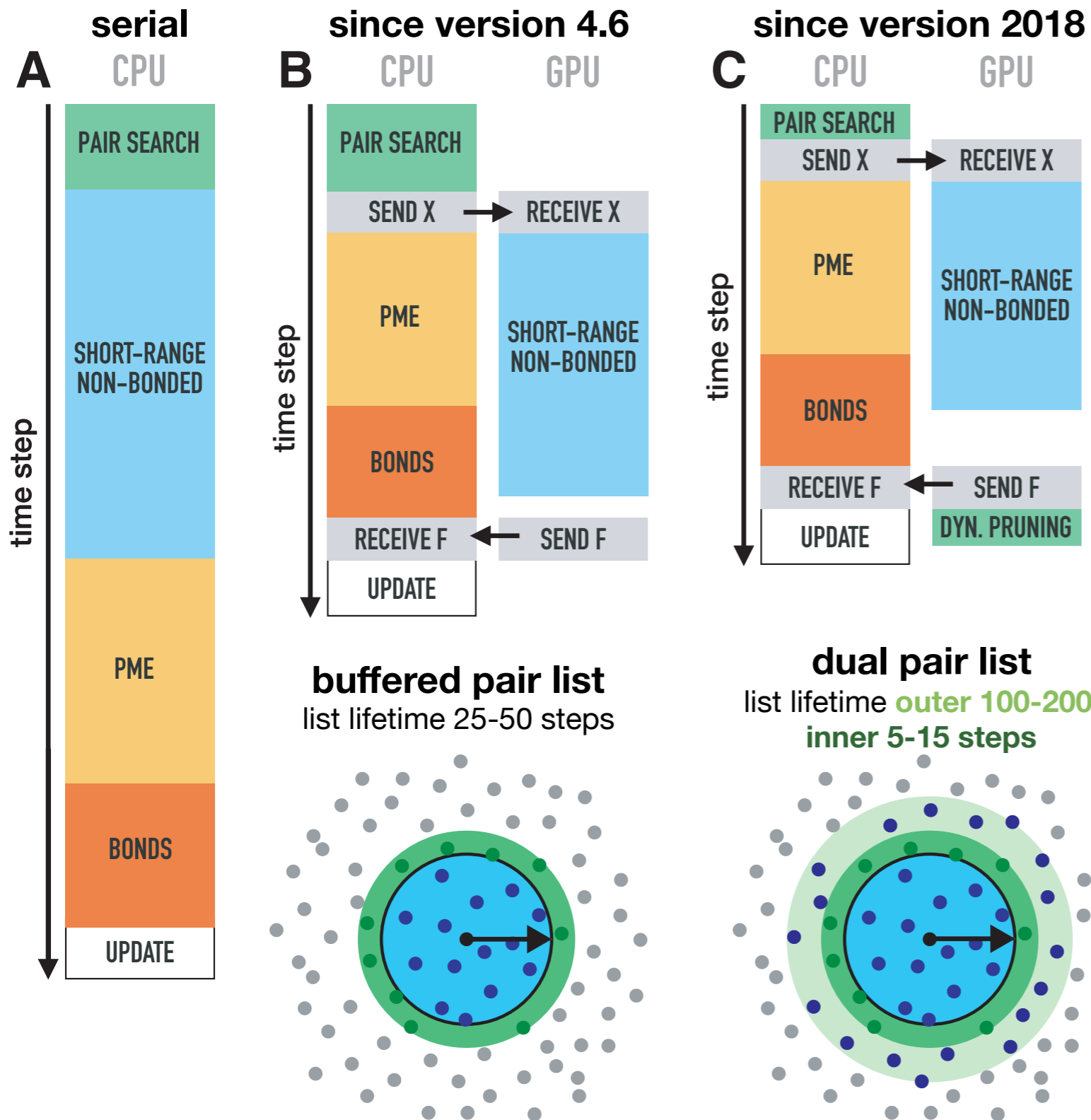
since version 2018

1. dual pair lists with dynamic pruning
2. PME offloading

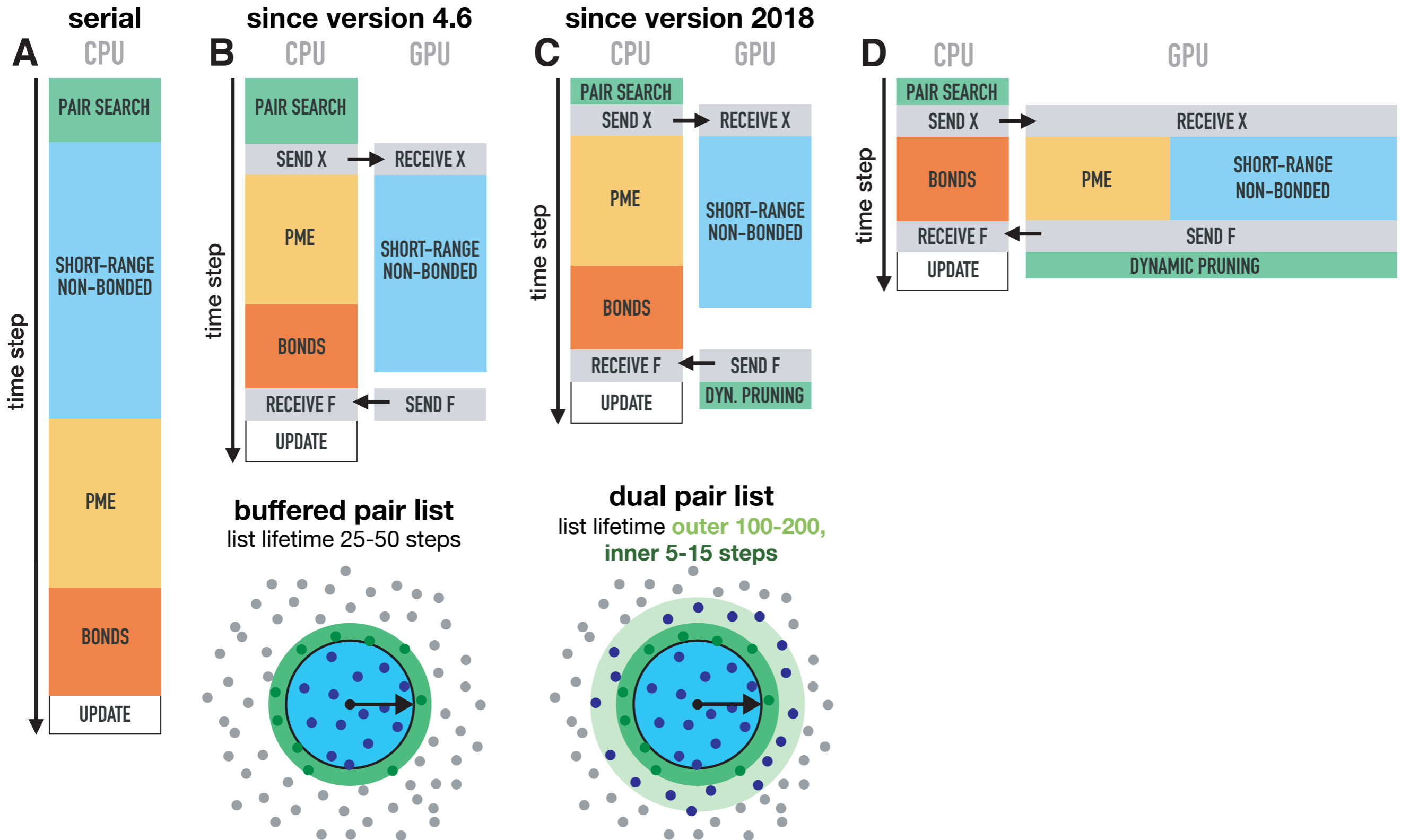
**dual pair list**  
list lifetime **outer 100-200,**  
**inner 5-15 steps**



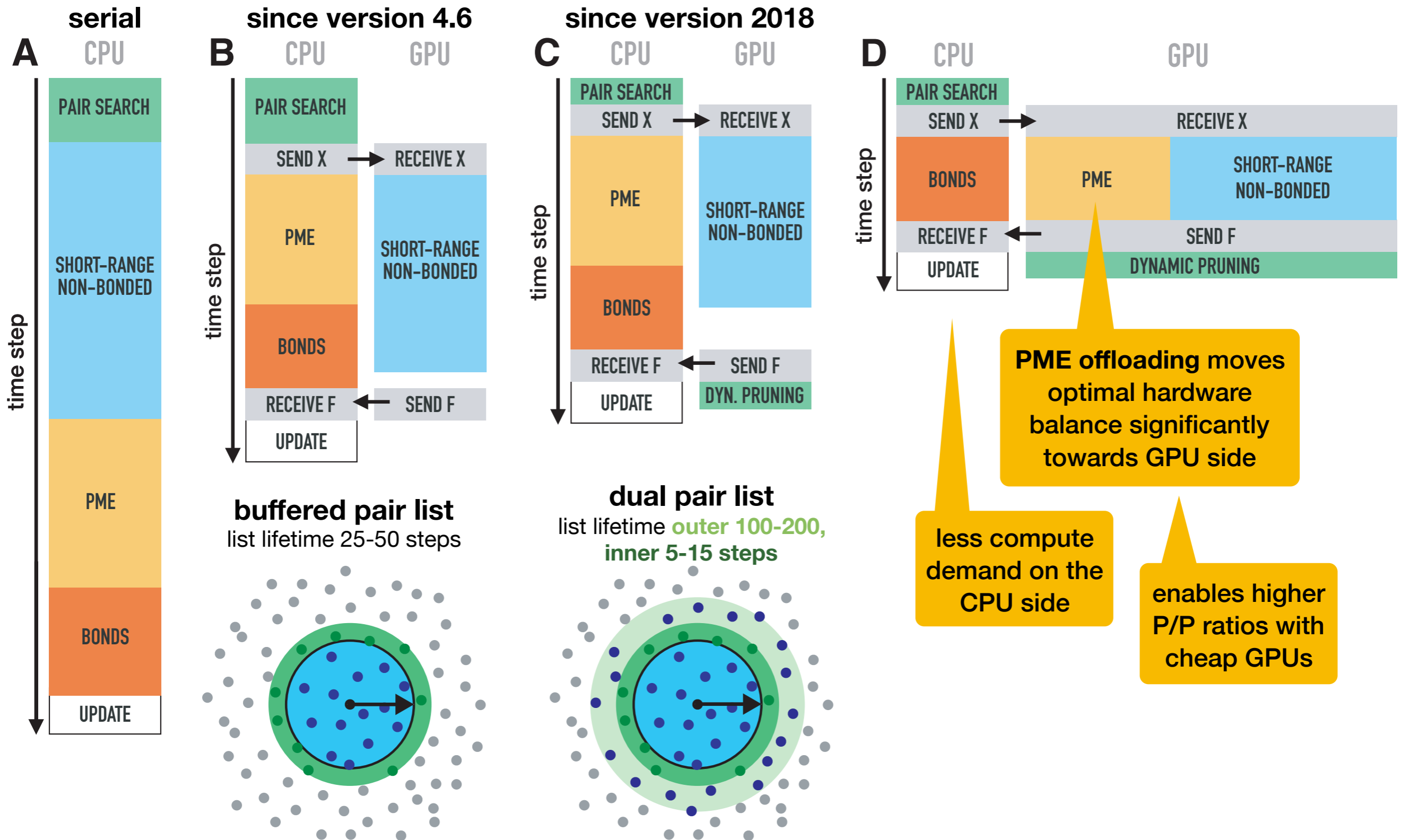
# Software Developments



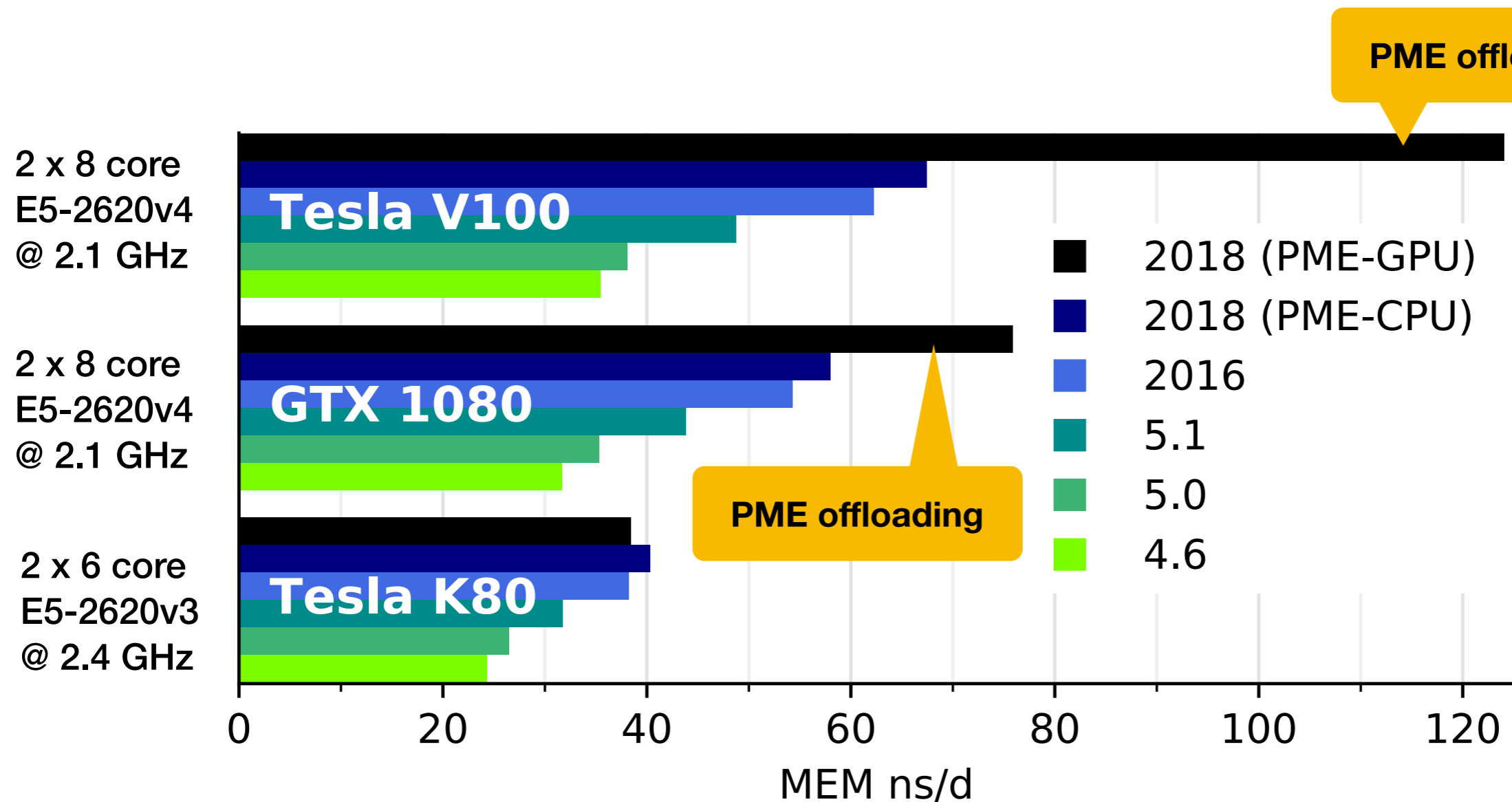
# Software Developments



# Software Developments

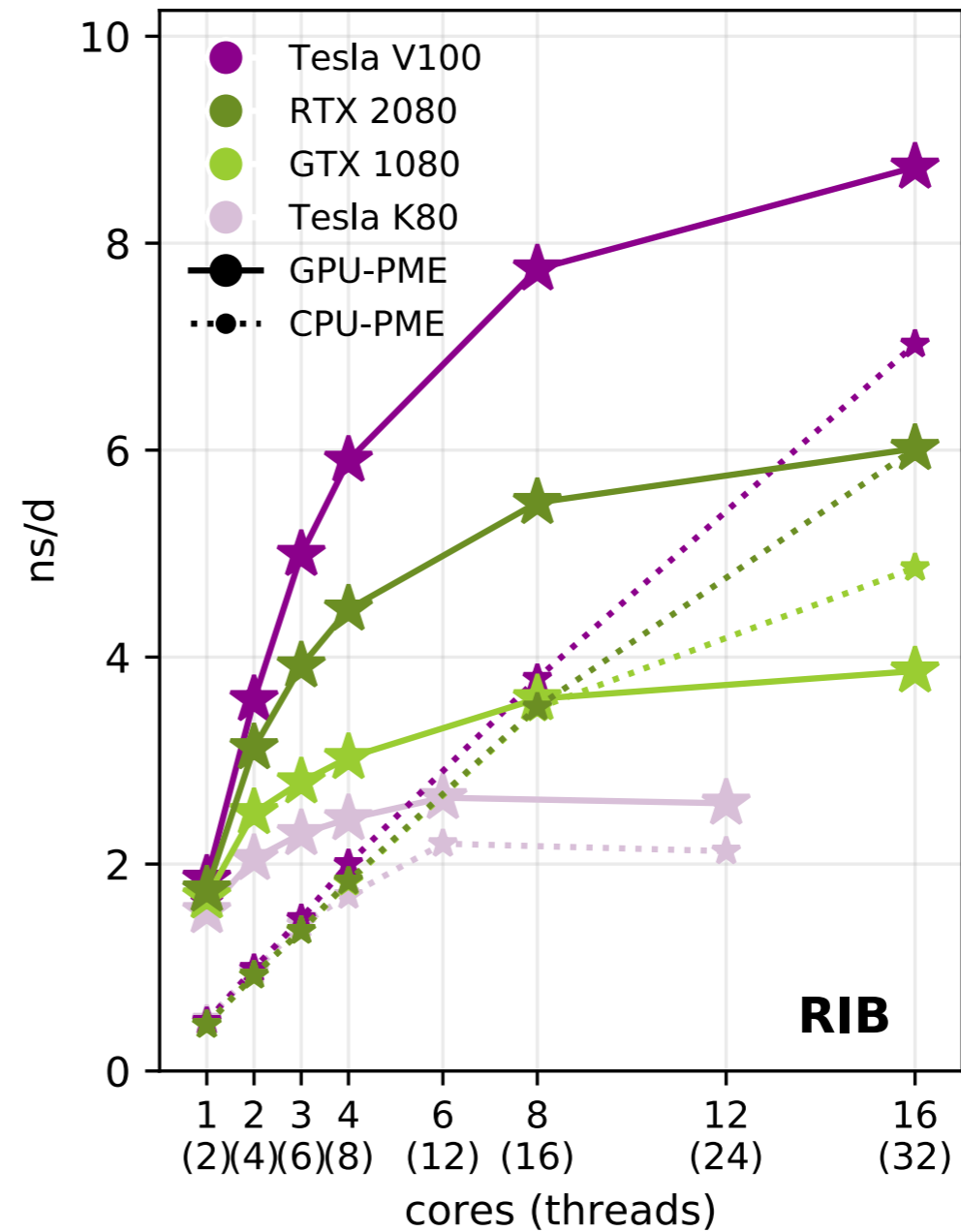
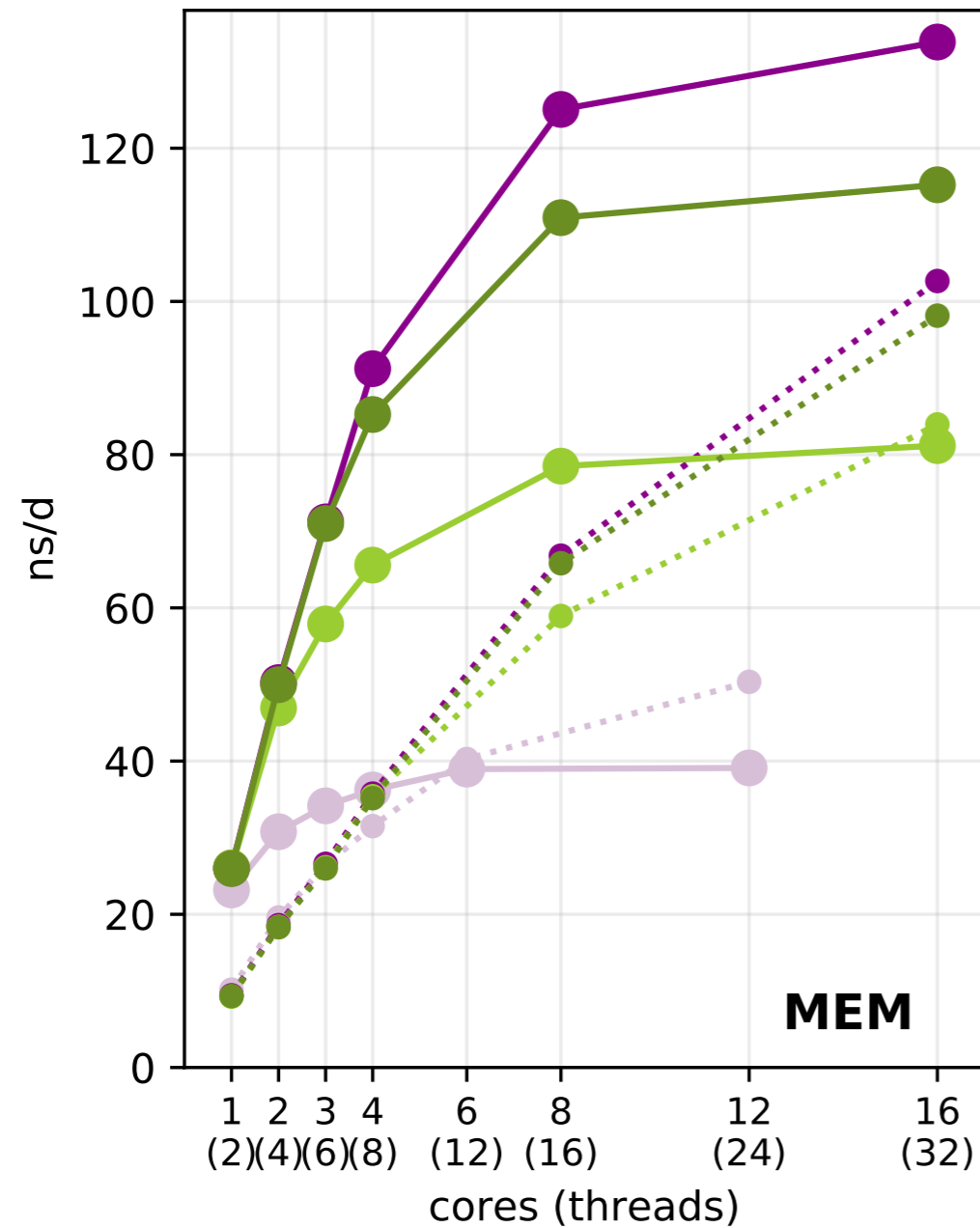


# GROMACS performance evolution on GPU nodes



- most pronounced increase in performance with PME offloading (given a strong enough GPU)

# Performance as a function of CPU cores per GPU

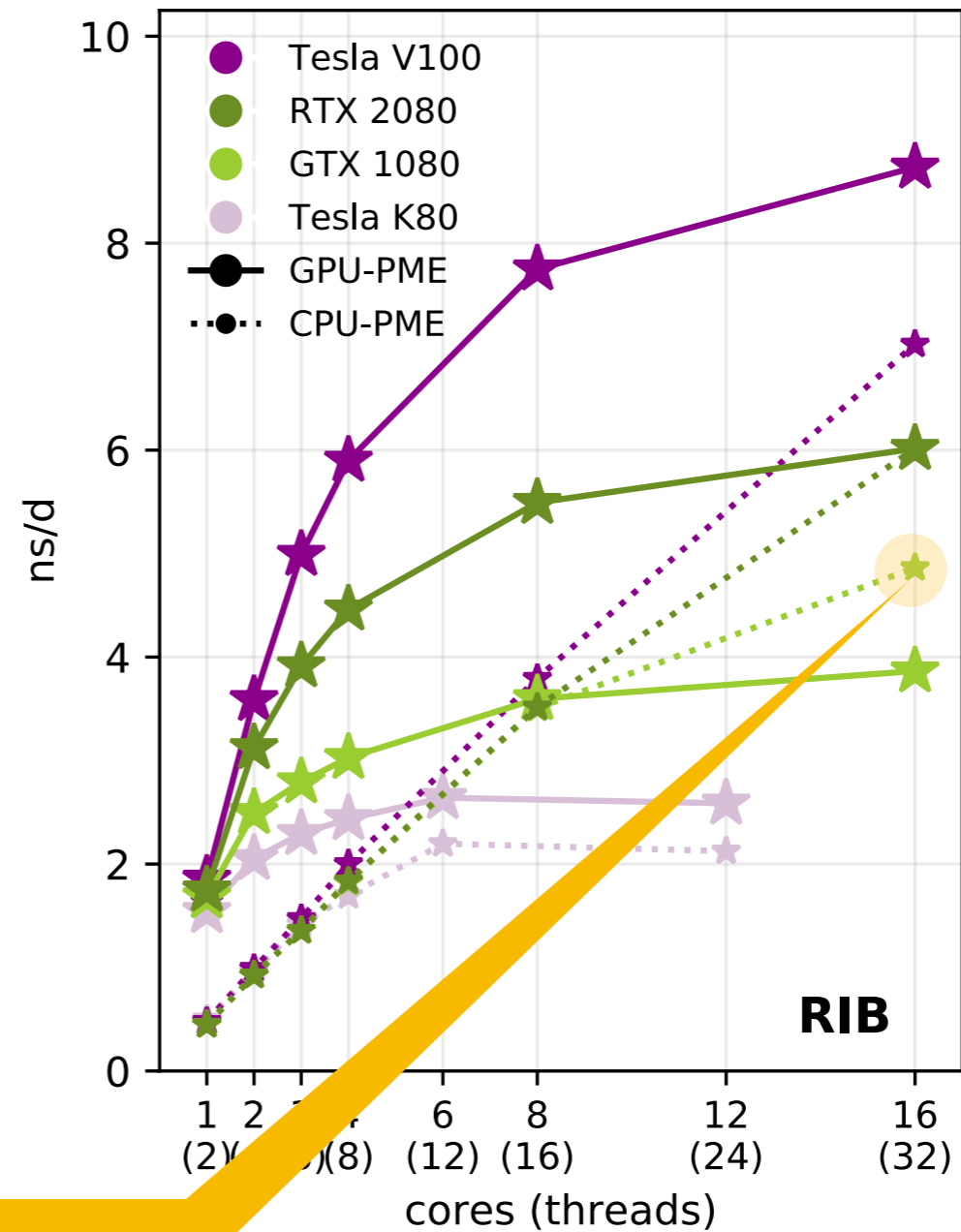
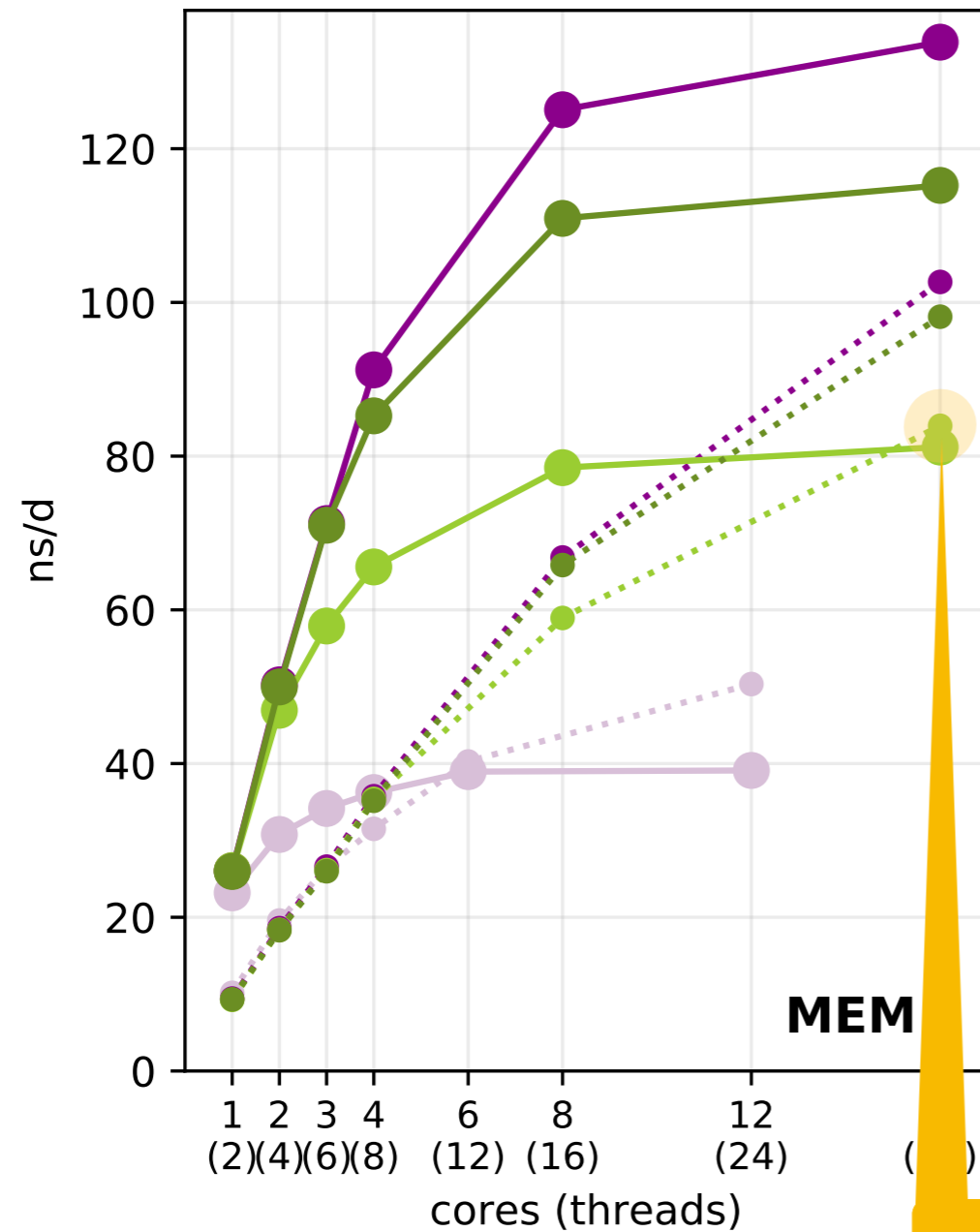


2 x 8 core  
E5-2620v4  
@ 2.1 GHz

2 x 6 core  
E5-2620v3  
@ 2.4 GHz



# Performance as a function of CPU cores per GPU

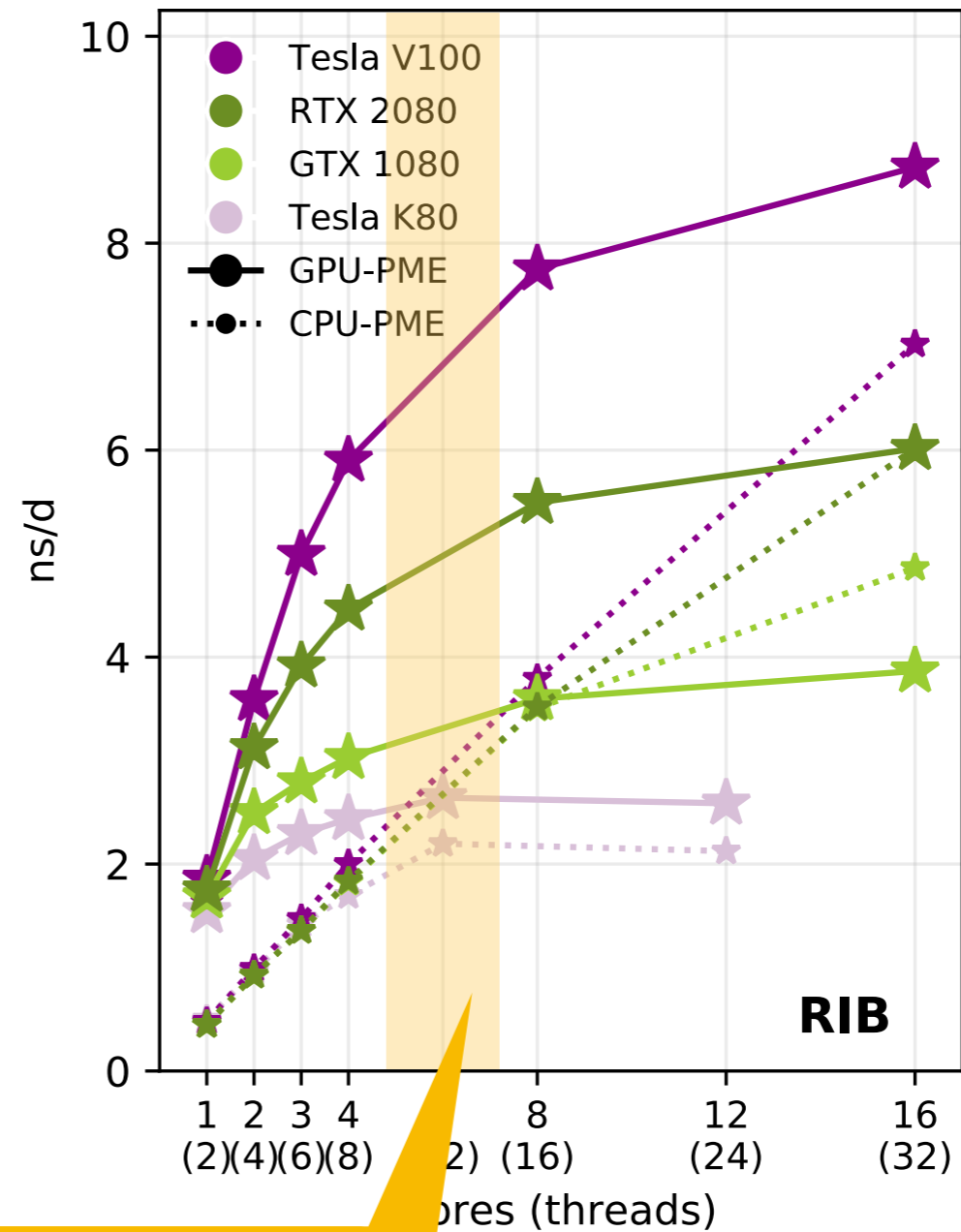
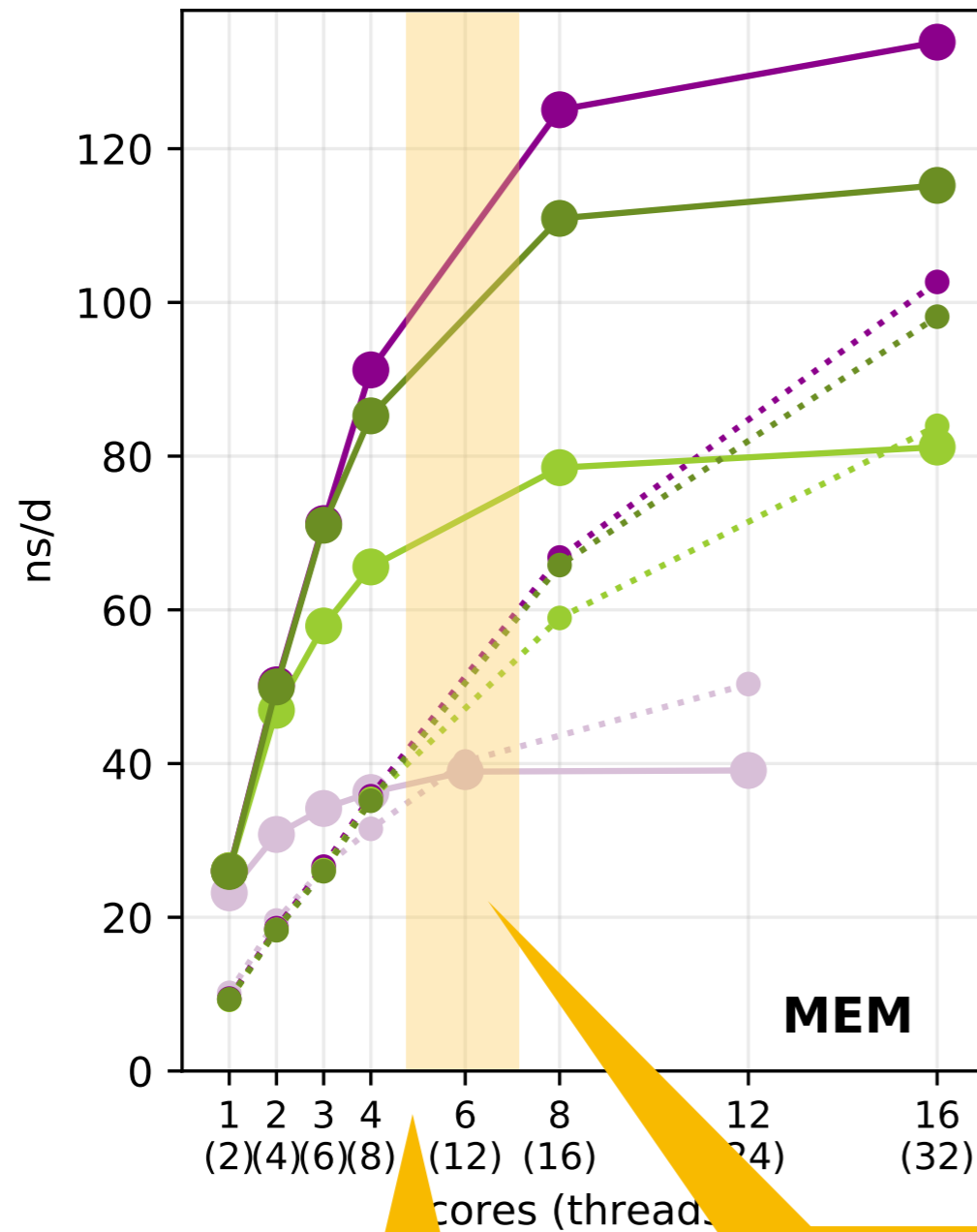


2 x 8 core  
E5-2620v4  
@ 2.1 GHz

2 x 6 core  
E5-2620v3  
@ 2.4 GHz

PME on CPU  
faster only for  
strong CPU

# Performance as a function of CPU cores per GPU



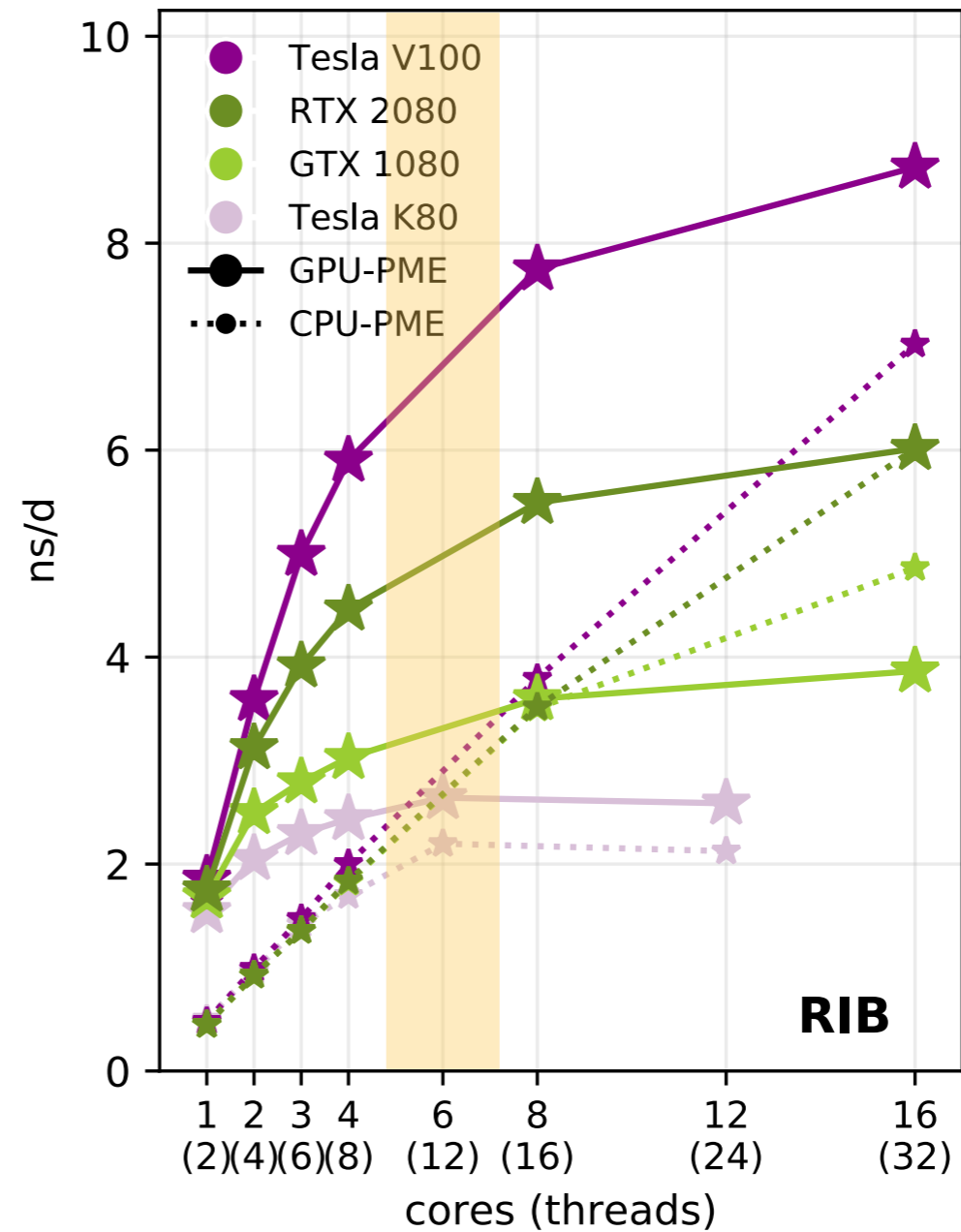
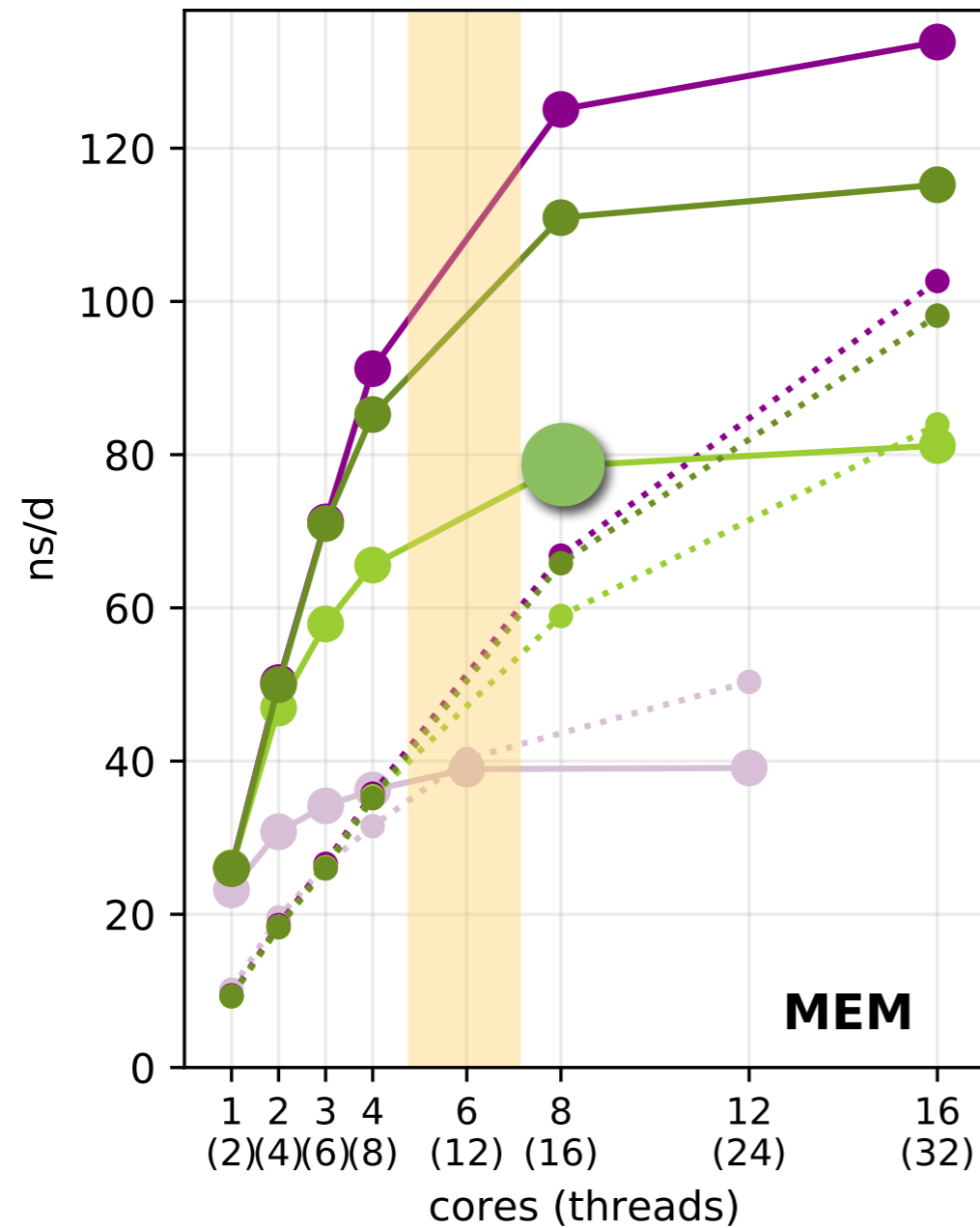
2 x 8 core  
E5-2620v4  
@ 2.1 GHz

2 x 6 core  
E5-2620v3  
@ 2.4 GHz

with PME offloading, far less (4-6) cores are needed to reach >80% peak simulation performance

10-15 „core-GHz“ suffice with a mid- to high-end GPU

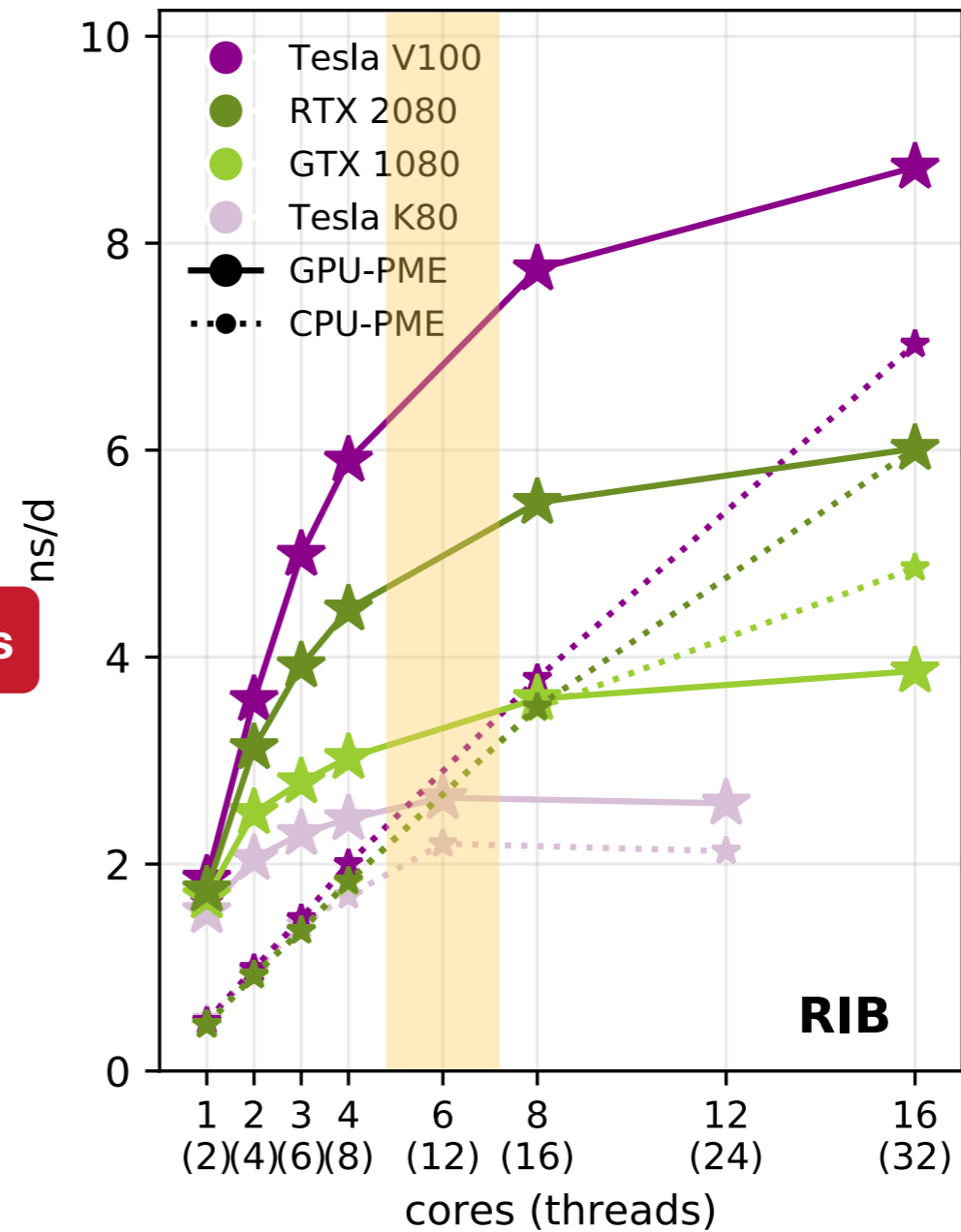
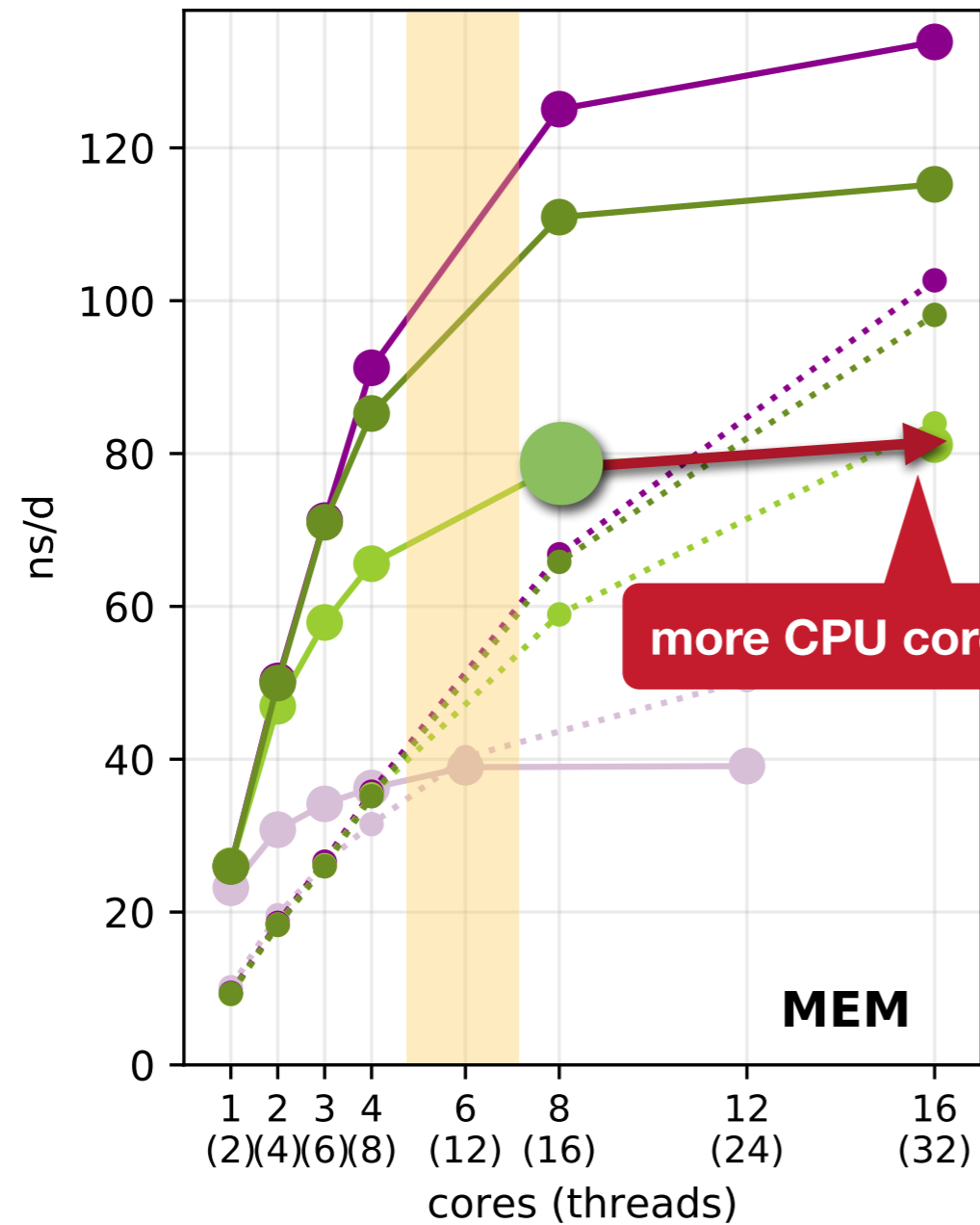
# Performance as a function of CPU cores per GPU



2 x 8 core  
E5-2620v4  
@ 2.1 GHz

2 x 6 core  
E5-2620v3  
@ 2.4 GHz

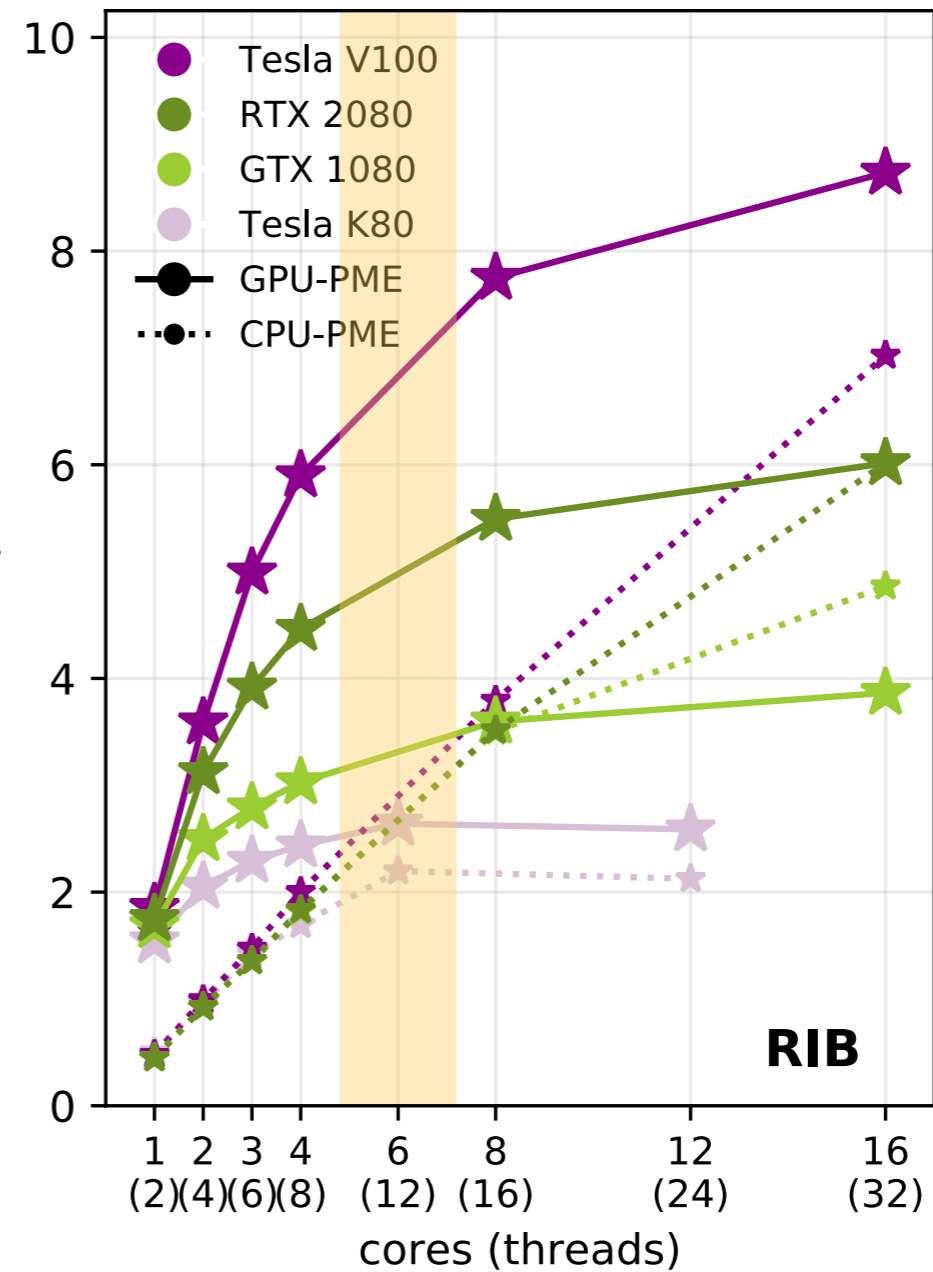
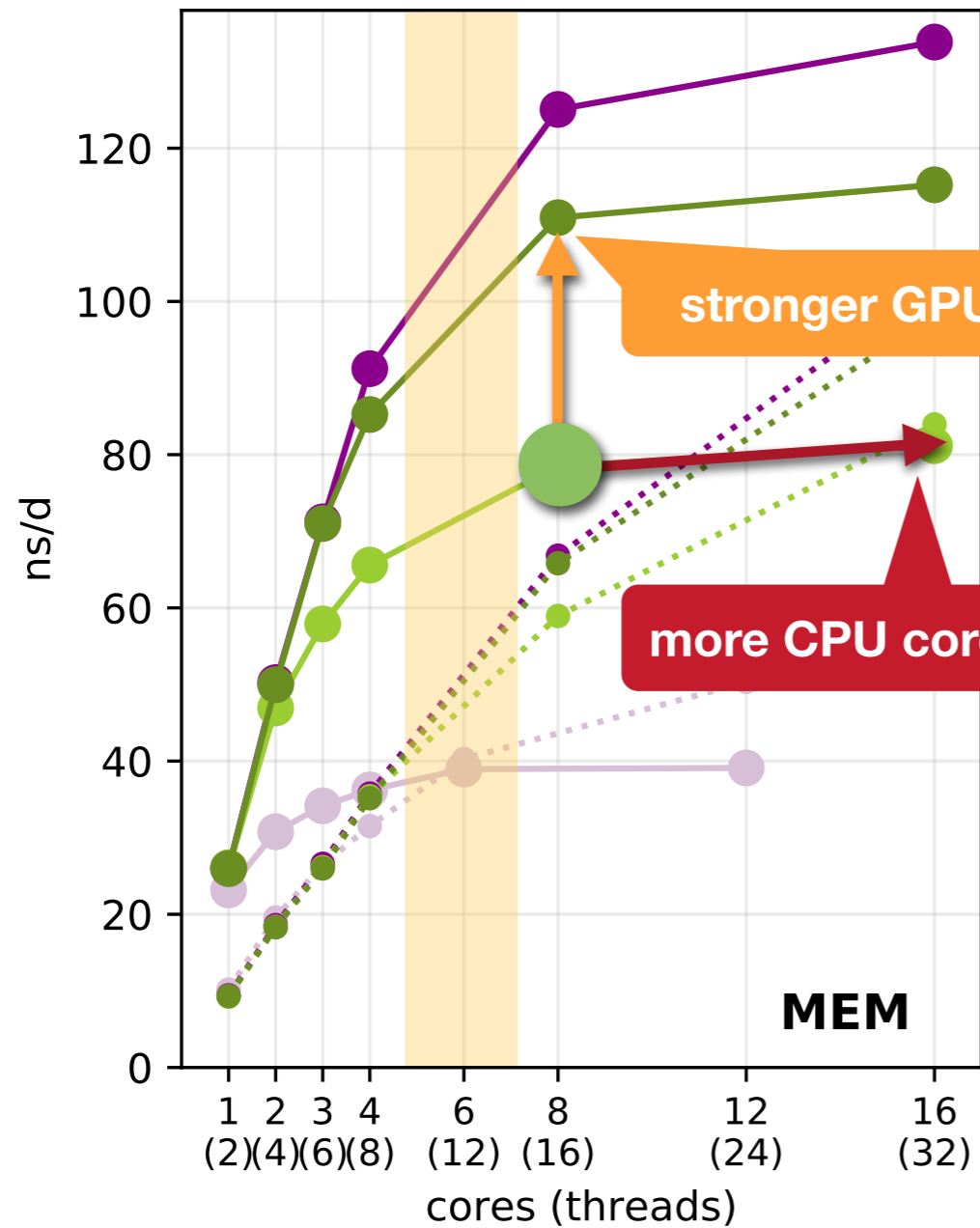
# Performance as a function of CPU cores per GPU



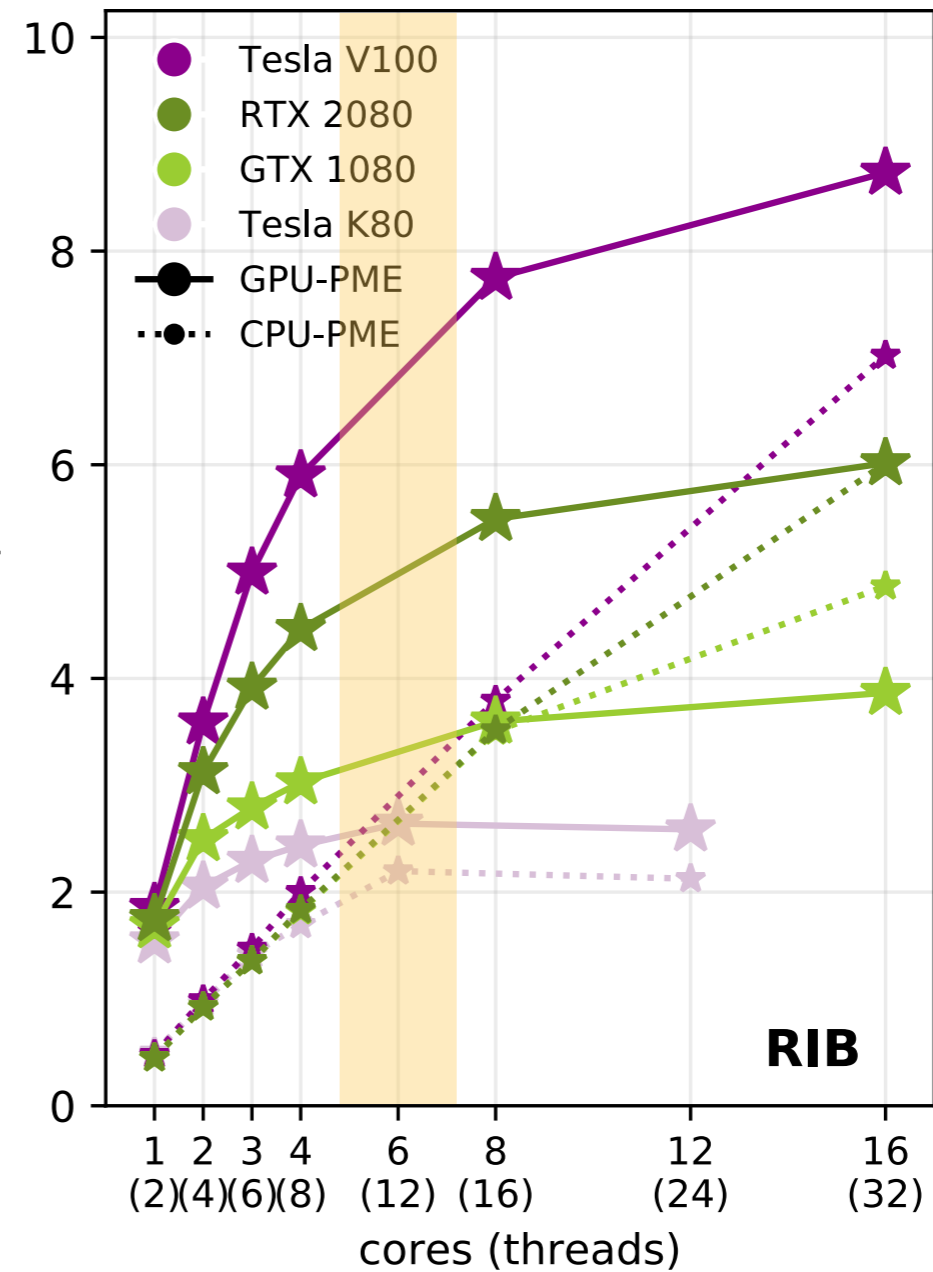
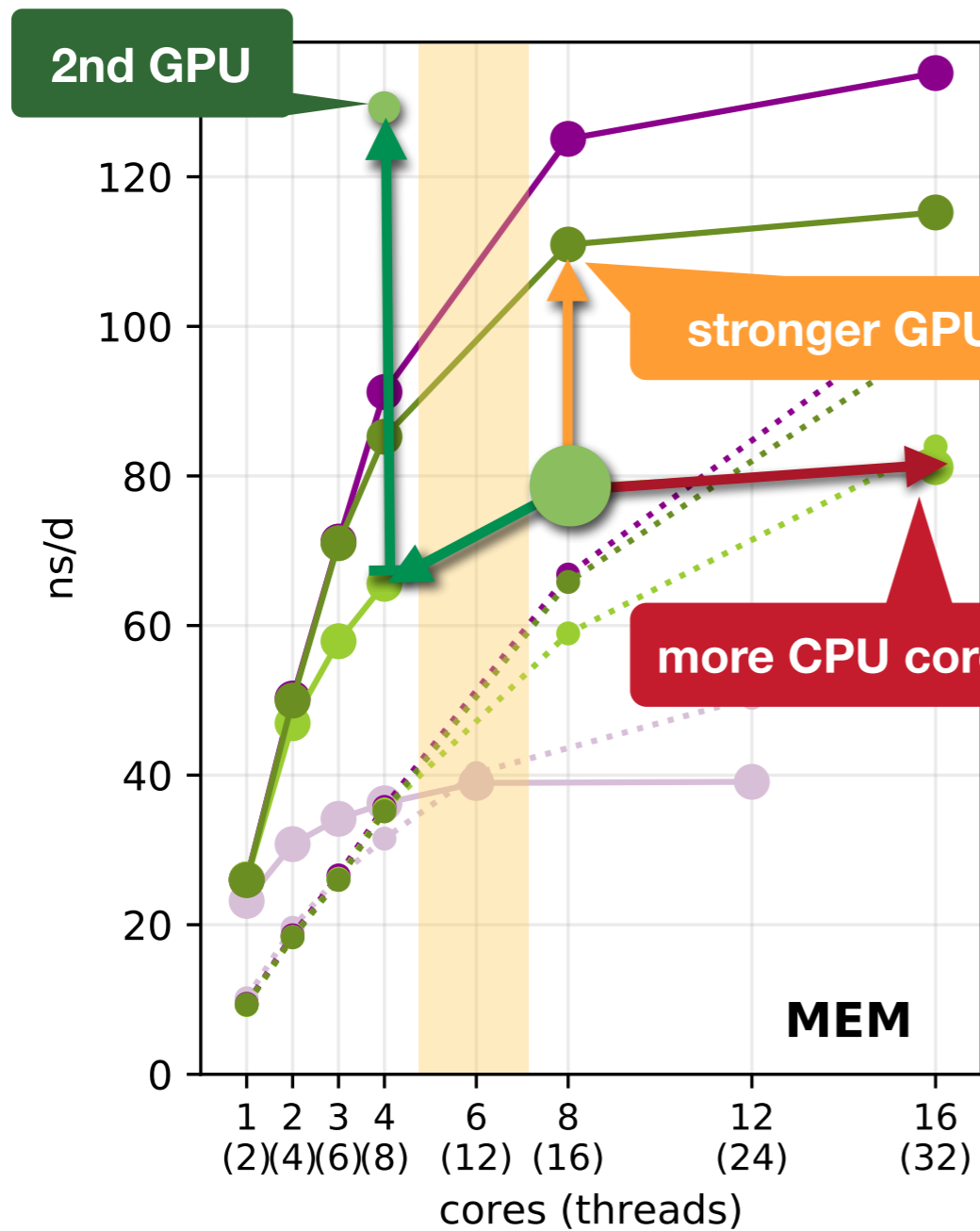
2 x 8 core  
E5-2620v4  
@ 2.1 GHz

2 x 6 core  
E5-2620v3  
@ 2.4 GHz

# Performance as a function of CPU cores per GPU



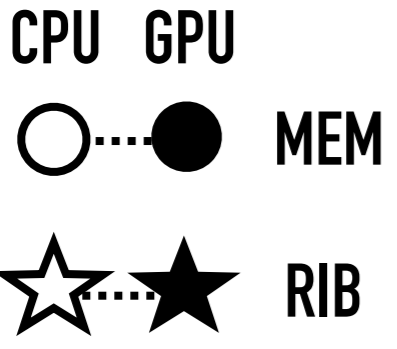
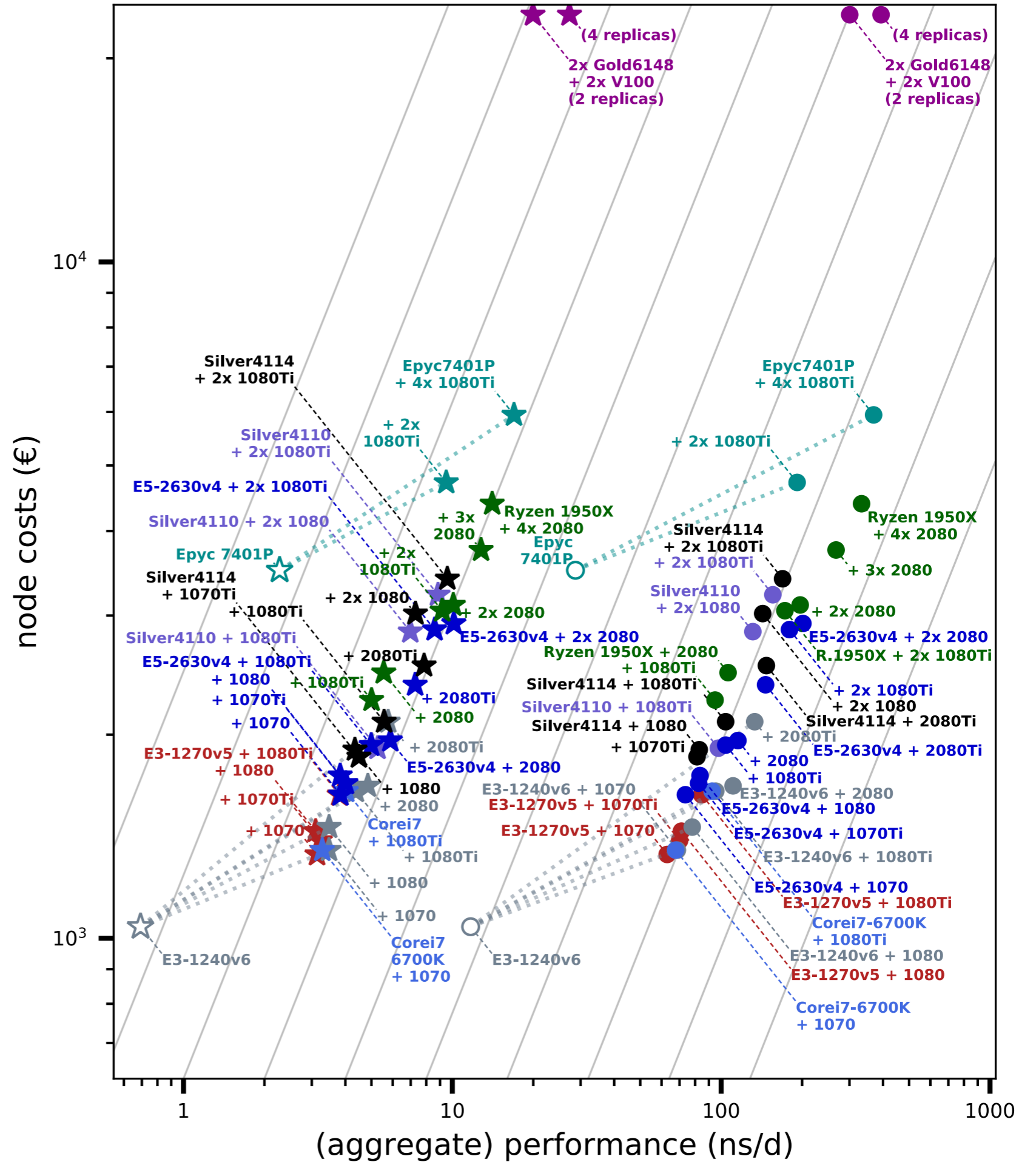
# Performance as a function of CPU cores per GPU



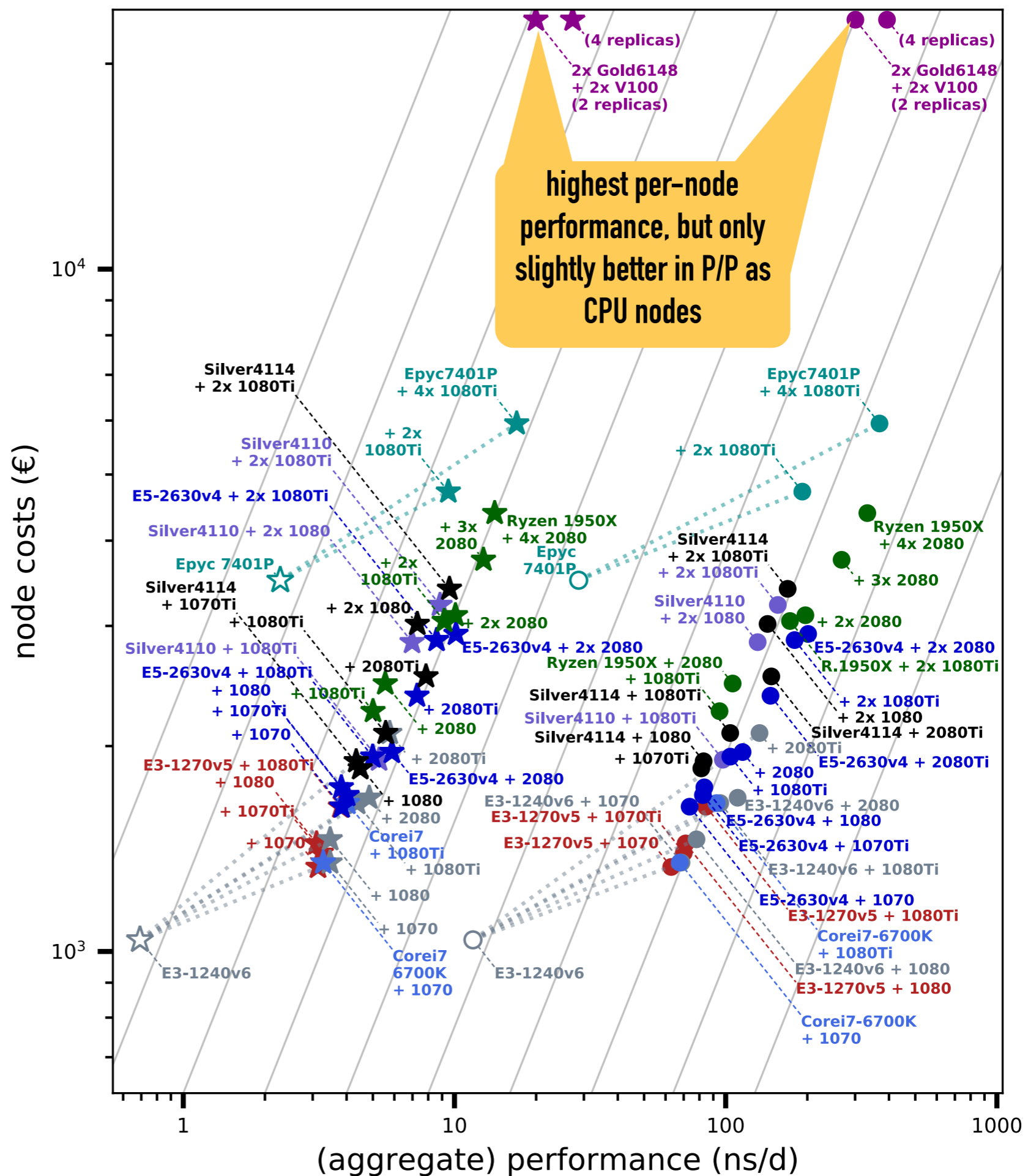
2 x 8 core  
E5-2620v4  
@ 2.1 GHz

2 x 6 core  
E5-2620v3  
@ 2.4 GHz

# Performance in relation to node costs



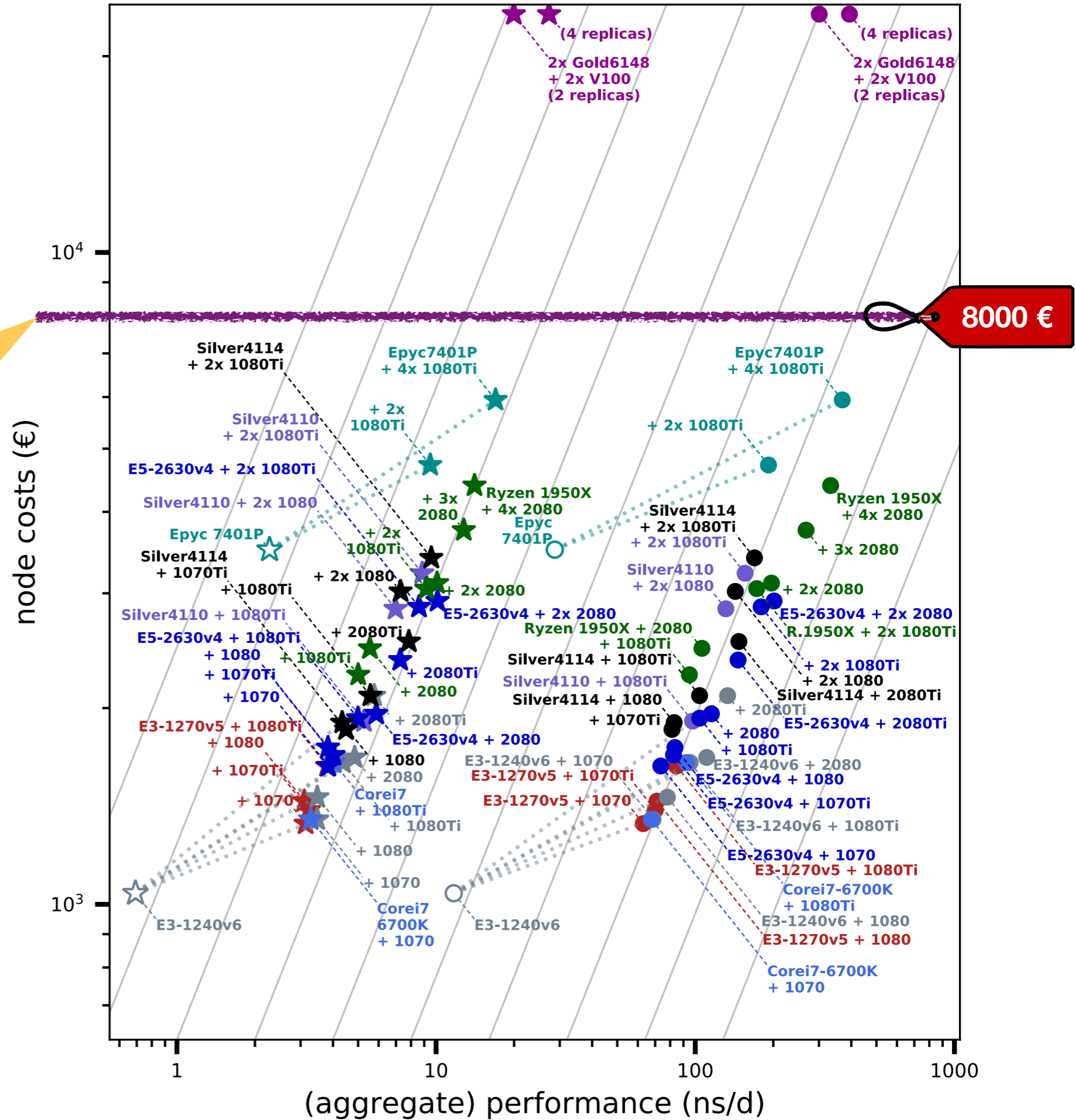
# Performance in relation to node costs





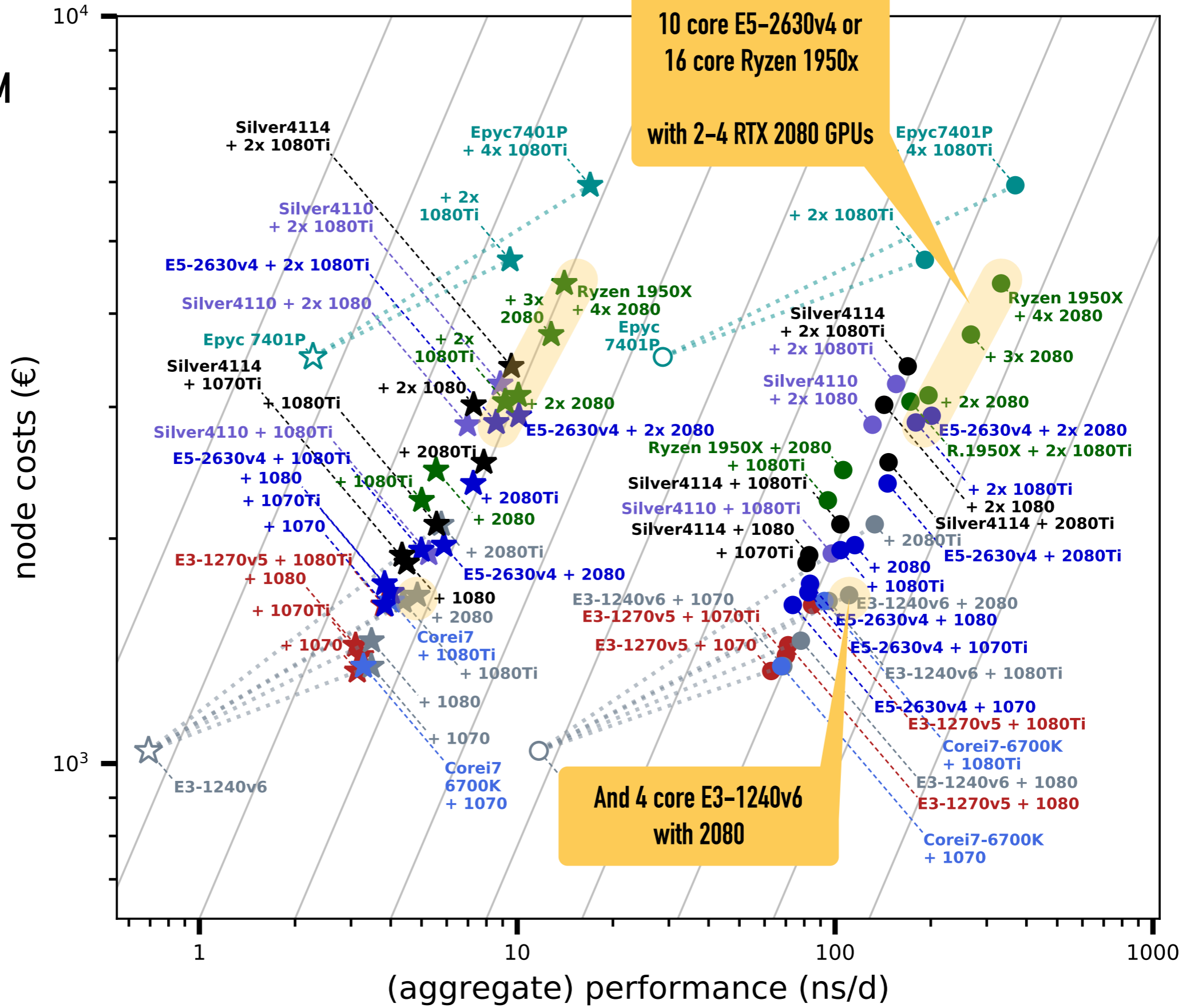
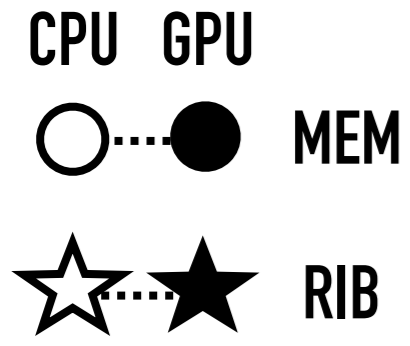
# Performance in relation to node costs

cost of 1 Tesla V100 GPU

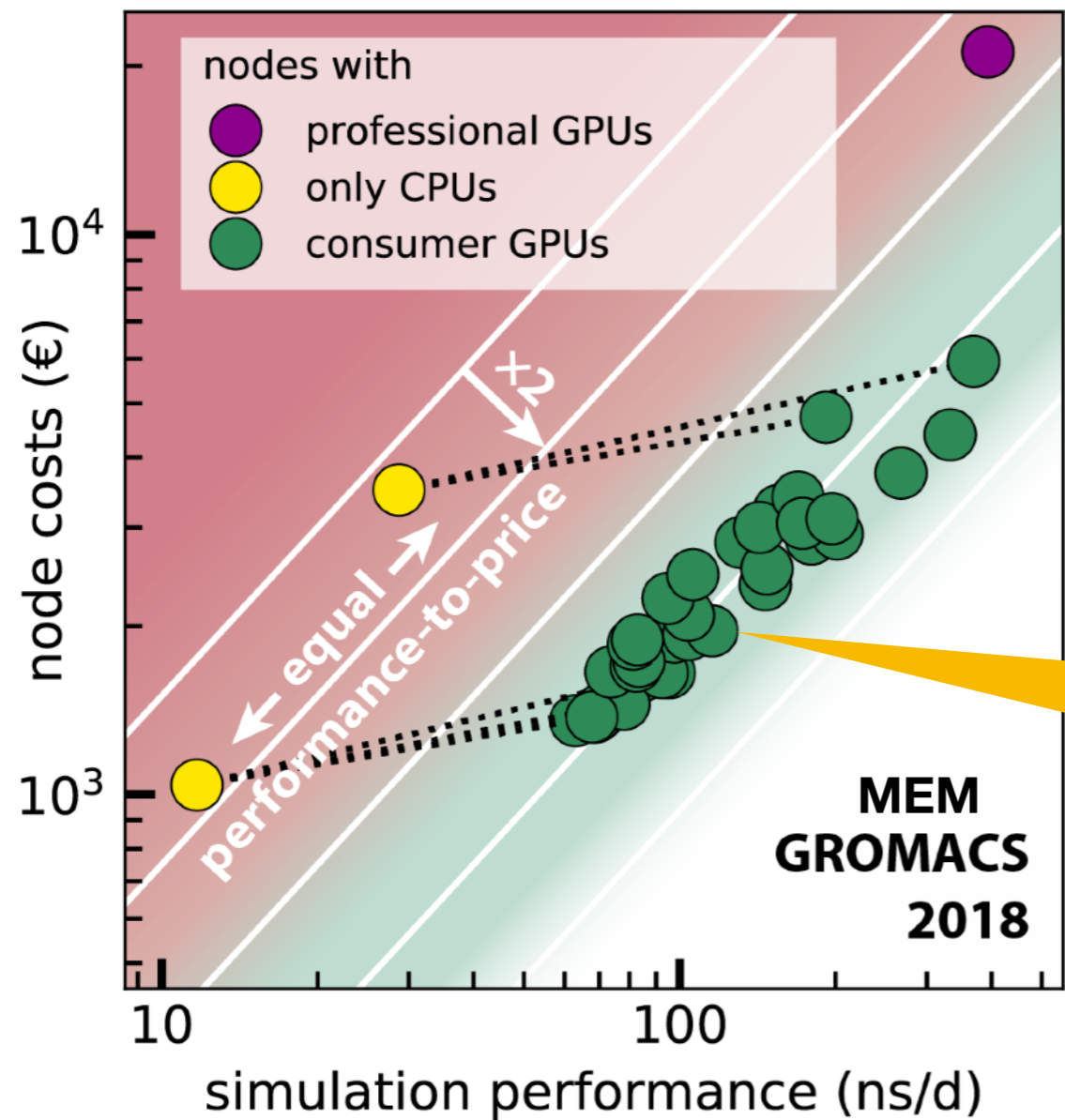








# The Gap Widens With GROMACS 2018



- Main 2014 result:

● nodes with GeForce consumer GPUs

produce **2–3x** as much MD trajectory per invested € as

● CPU nodes

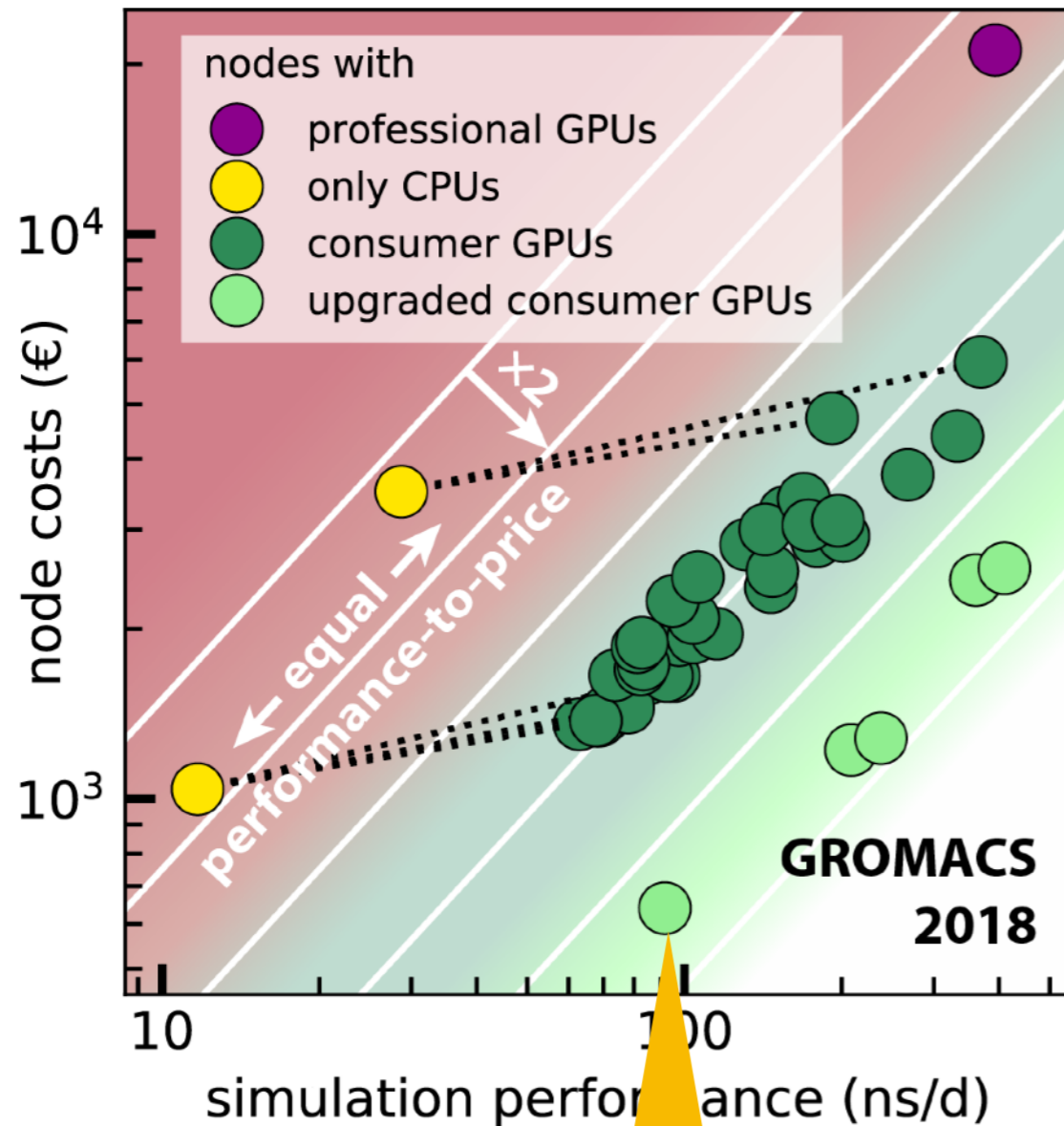
**3–7 x with  
GROMACS 2018**

C Kutzner, S Páll, M Fechner, A Esztermann, BL de Groot, H Grubmüller.

**More bang for your buck: Improved use of GPU nodes for GROMACS 2018.**

JCC 40 (27), pp. 2418-2431 (2019)

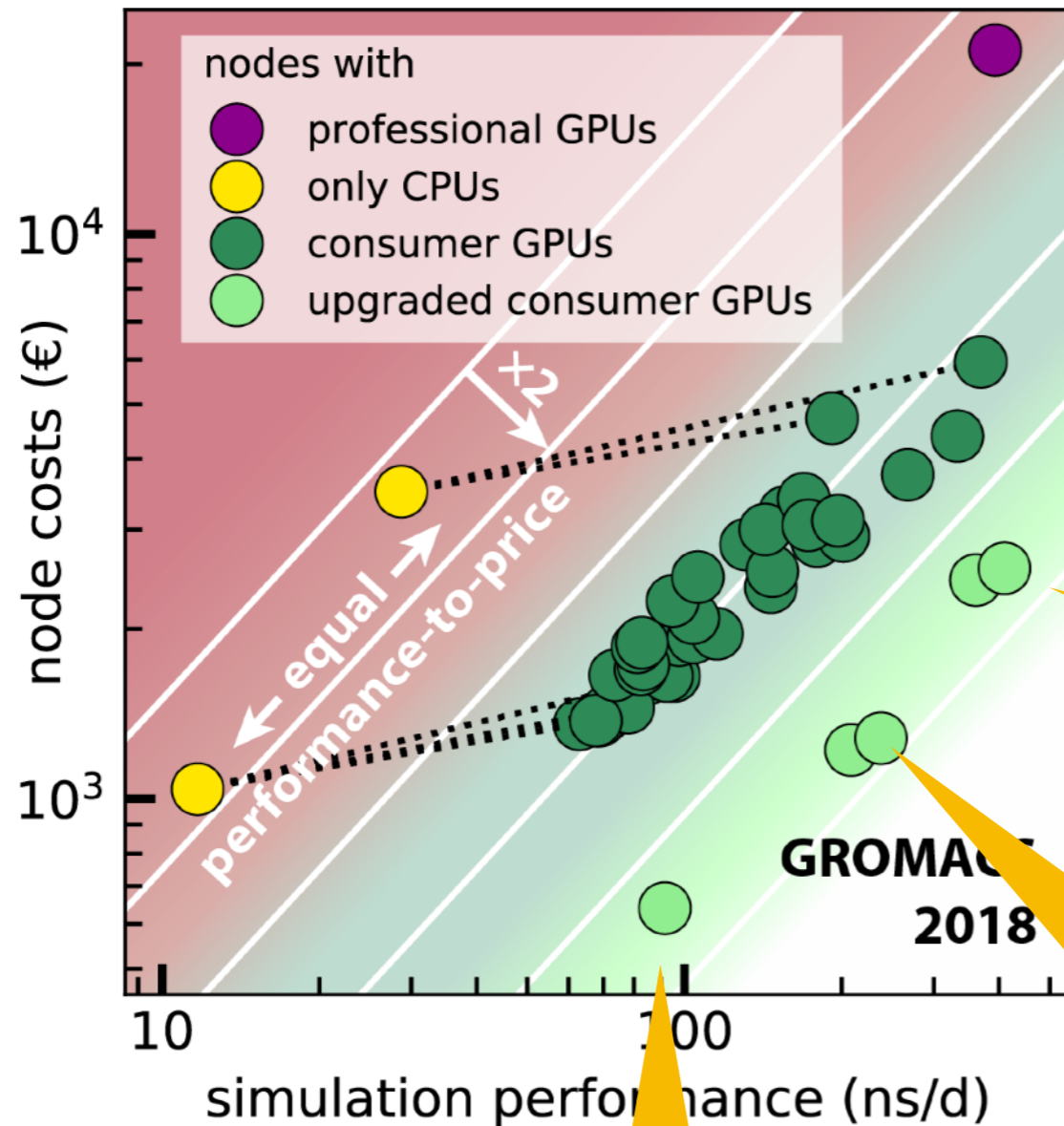
# Free Lunch! GPU Upgrades



4 core E3-1270v2  
RTX 2080

- shift CPU → GPU allows to upgrade old nodes with recent GPUs!
- e.g. E3-1270v2 CPU (4 cores @3.5 GHz)  
+ GTX 680 (27 ns/d)  
+ (●) RTX 2080 (92 ns/d) → 3.4x perf!

# Free Lunch! GPU Upgrades



- shift CPU → GPU allows to upgrade old nodes with recent GPUs!

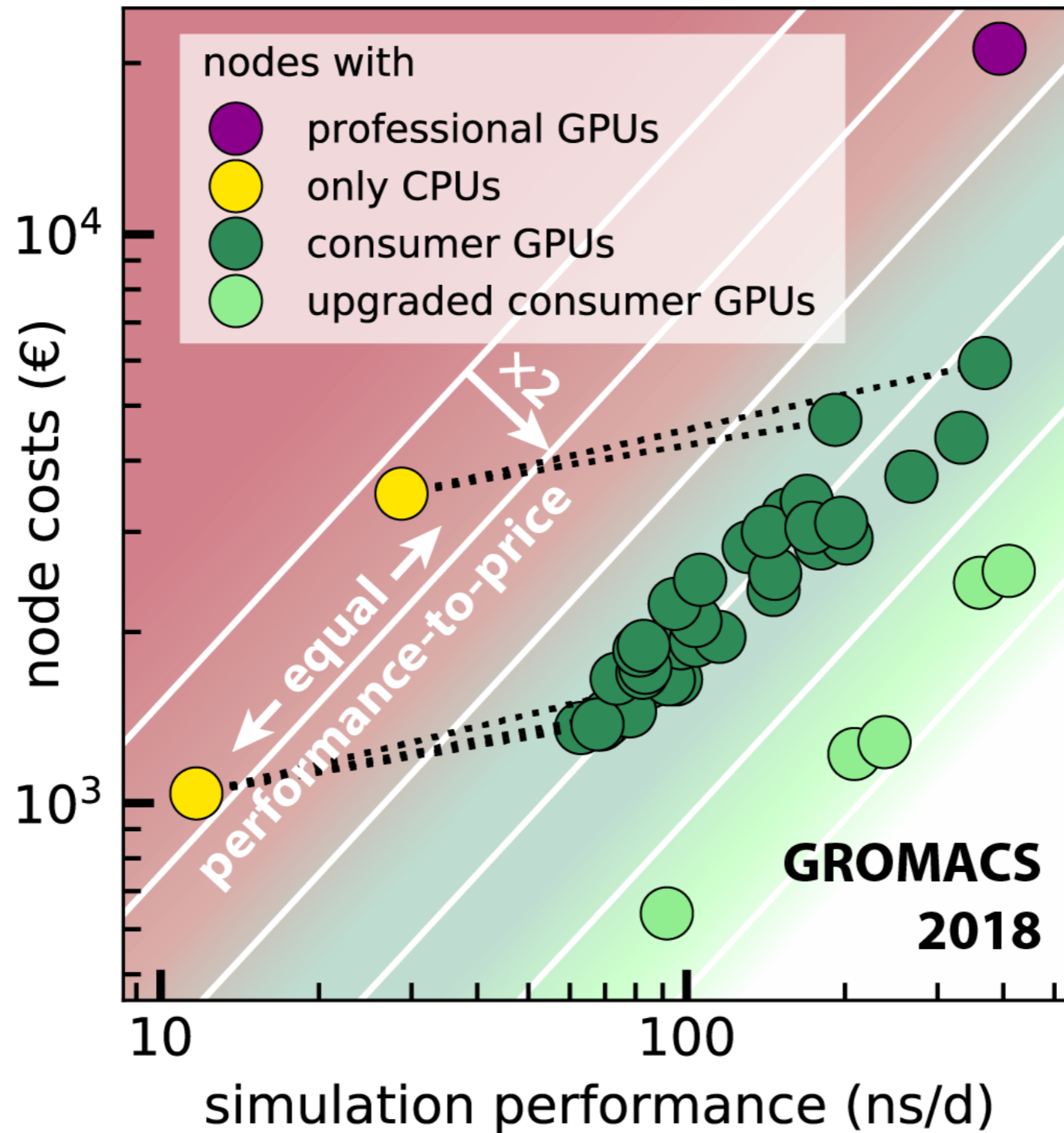
- e.g. E3-1270v2 CPU (4 cores @3.5 GHz)  
+ GTX 680 (27 ns/d)  
+ (●) RTX 2080 (92 ns/d) → 3.4x perf.

4 core E3-1270v2  
2080

2x10 core E5-2670v2  
2x 1080Ti or  
2x 2080

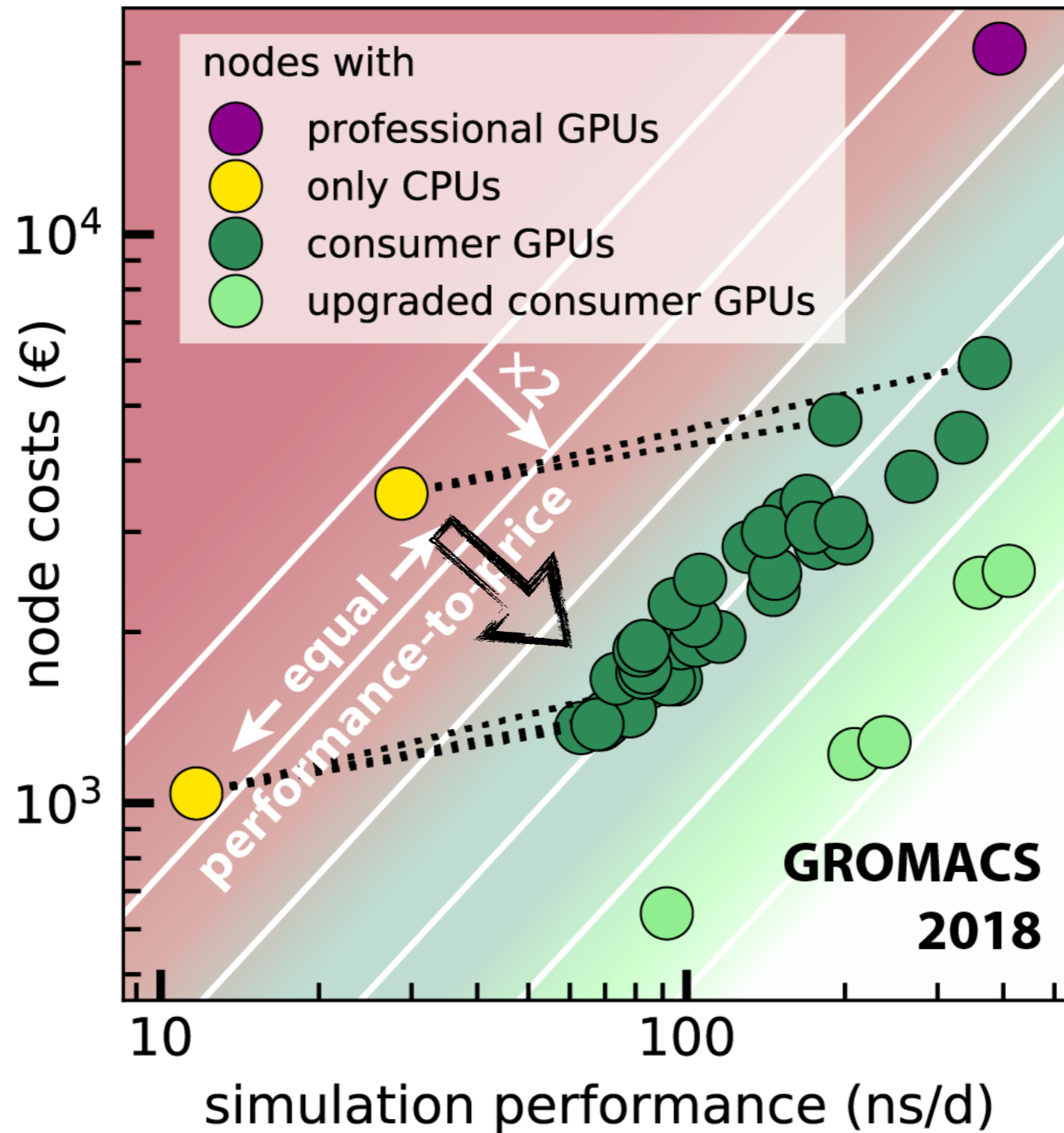
2x10 core E5-2680v2  
4x 1080Ti or  
4x 2080

# Free Lunch! GPU Upgrades

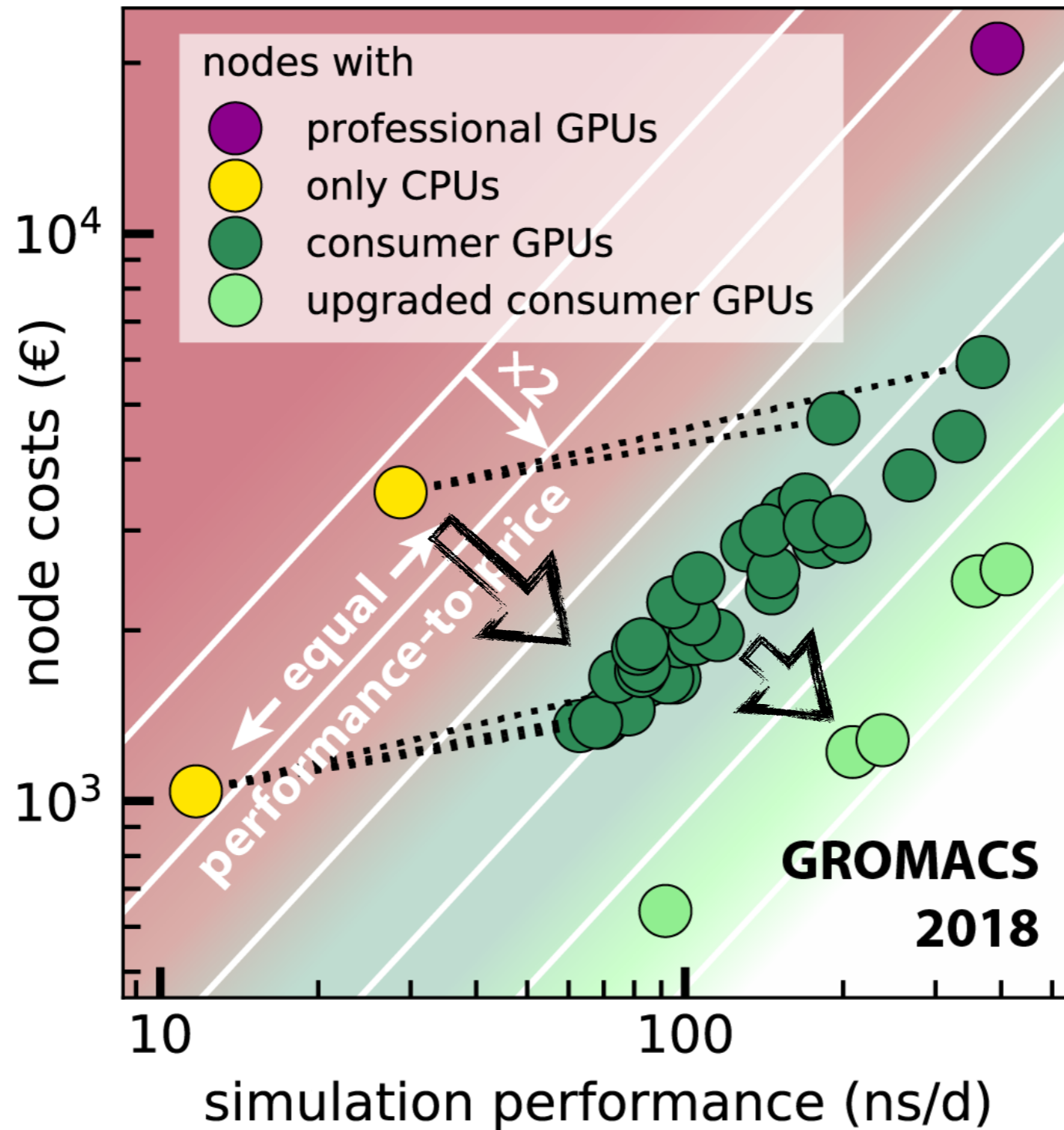




# Free Lunch! GPU Upgrades



# Free Lunch! GPU Upgrades



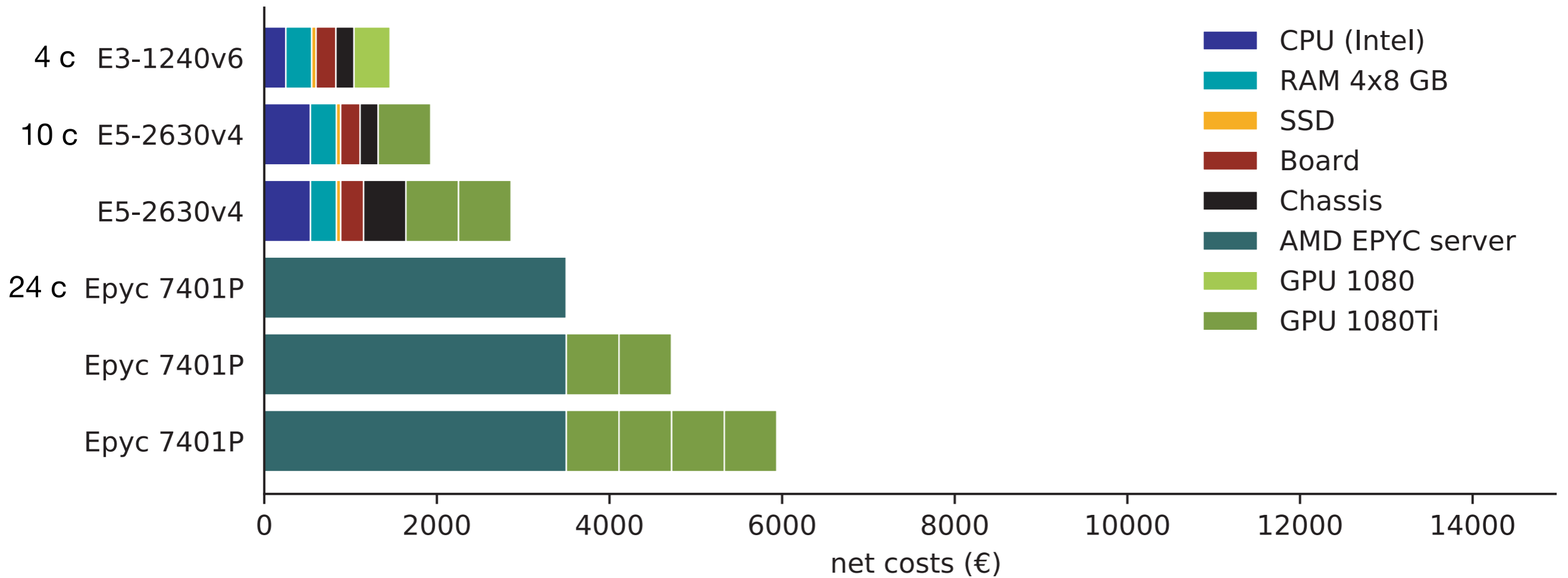
# Energy Efficiency

Add energy costs to the bill



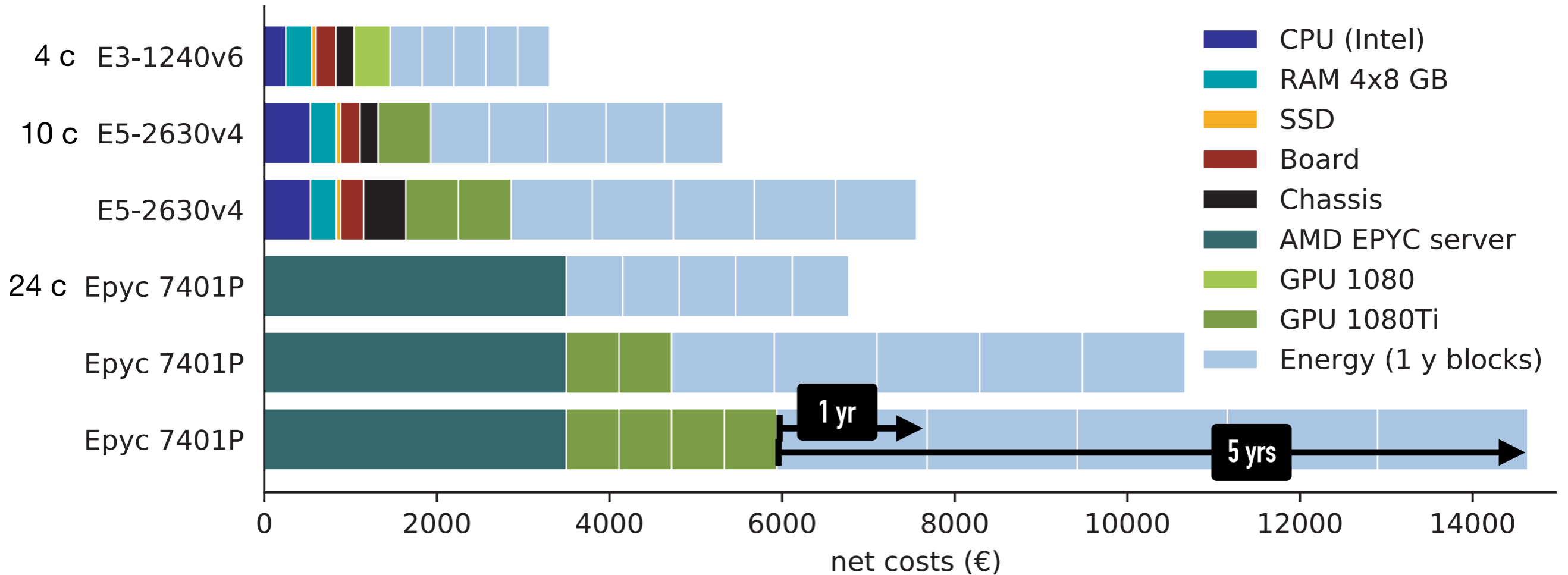
# Add energy costs to the bill

## Node costs



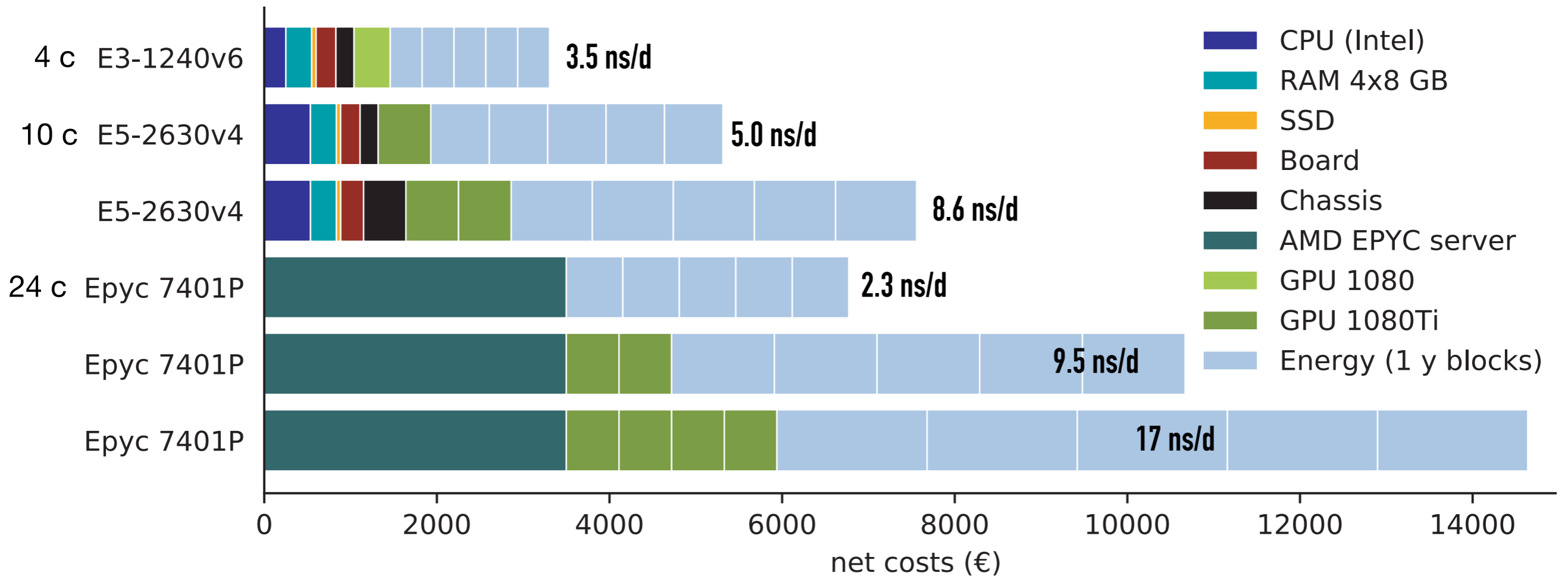
# Add energy costs to the bill

Node costs taking into account energy + cooling (0.2 EUR / kWh) RIB



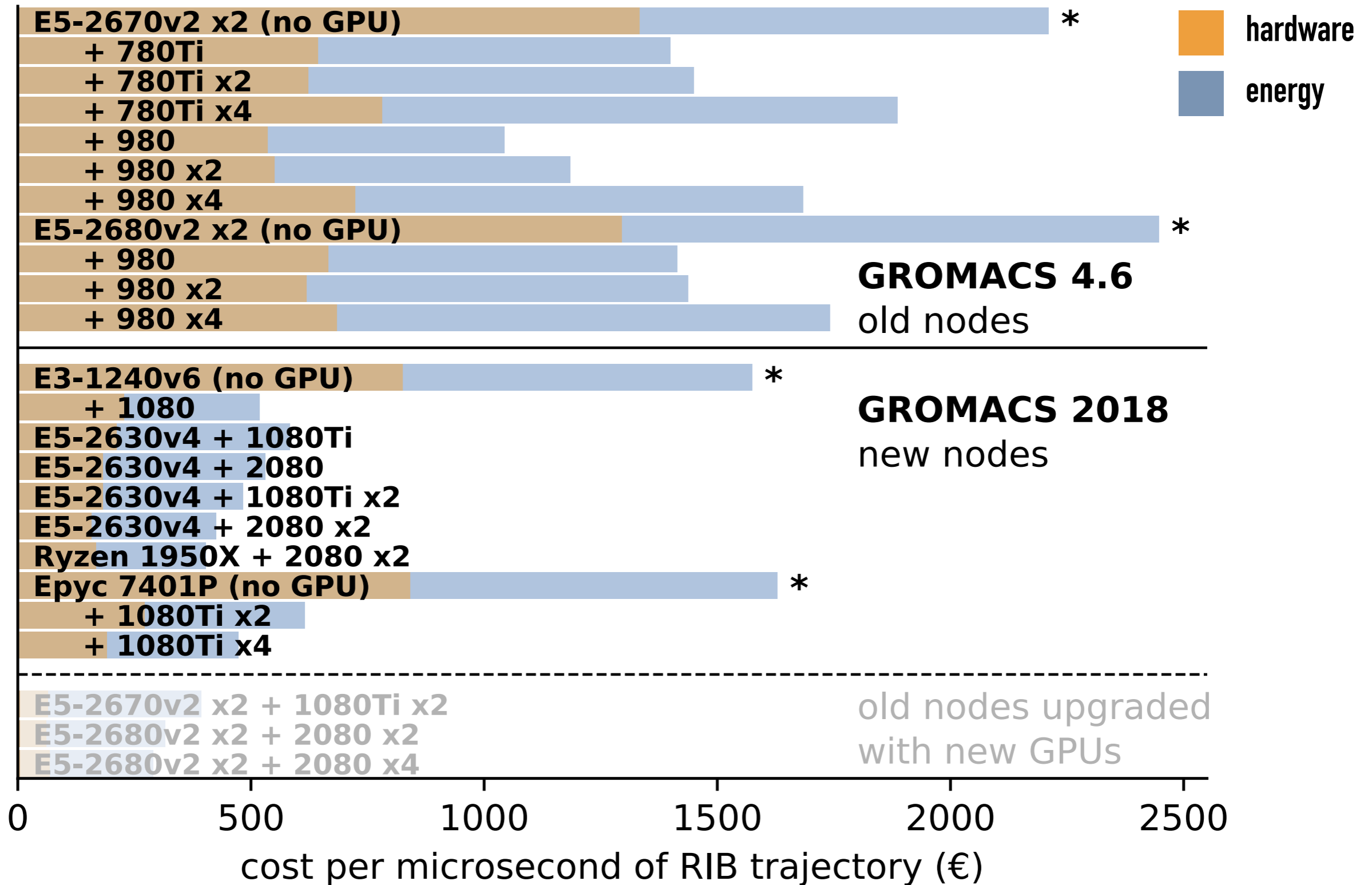
# Add energy costs to the bill

Node costs taking into account energy + cooling (0.2 EUR / kWh) RIB



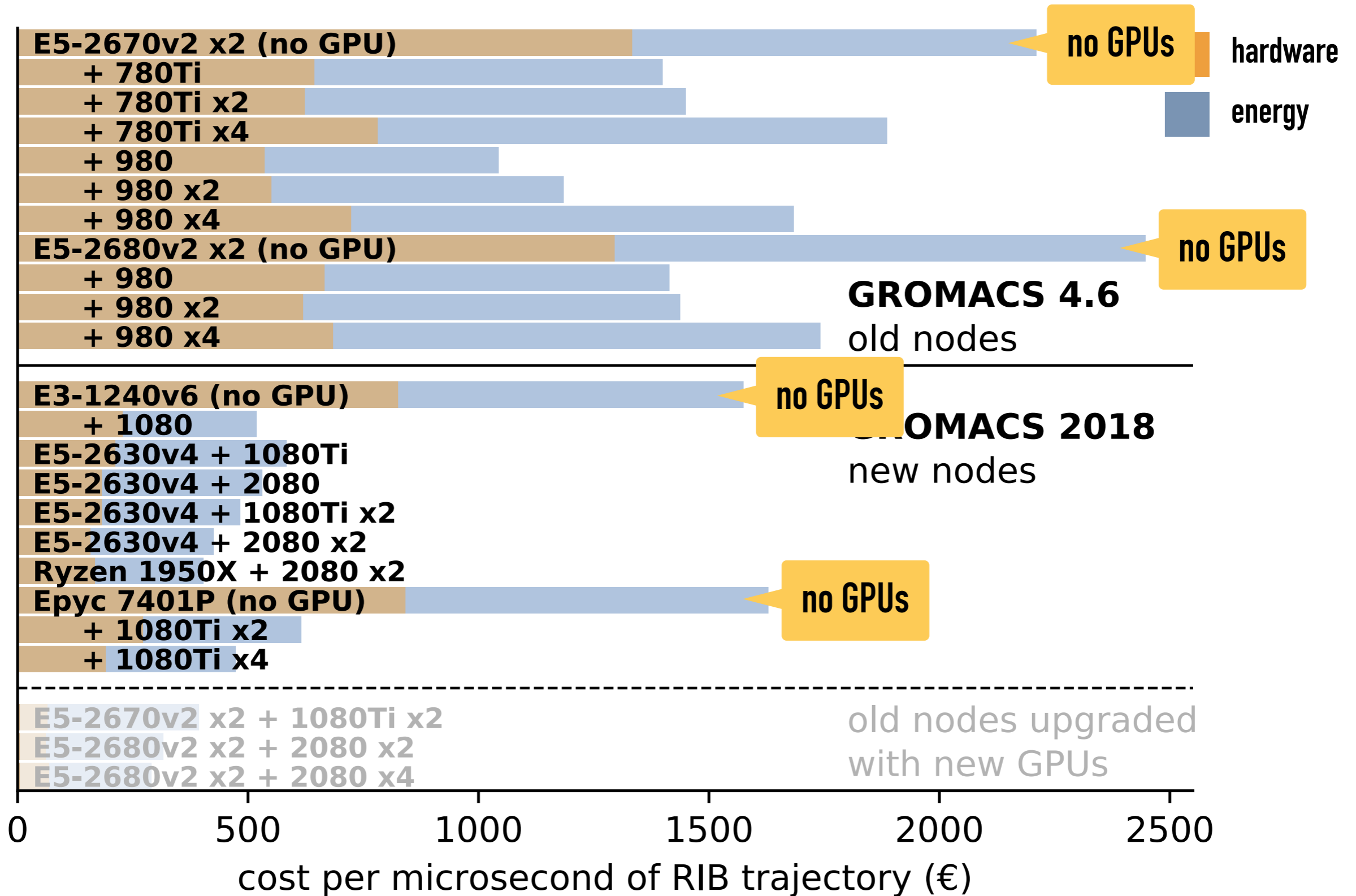
# Total trajectory costs hardware+energy

for 5 years of operation



# Total trajectory costs hardware+energy

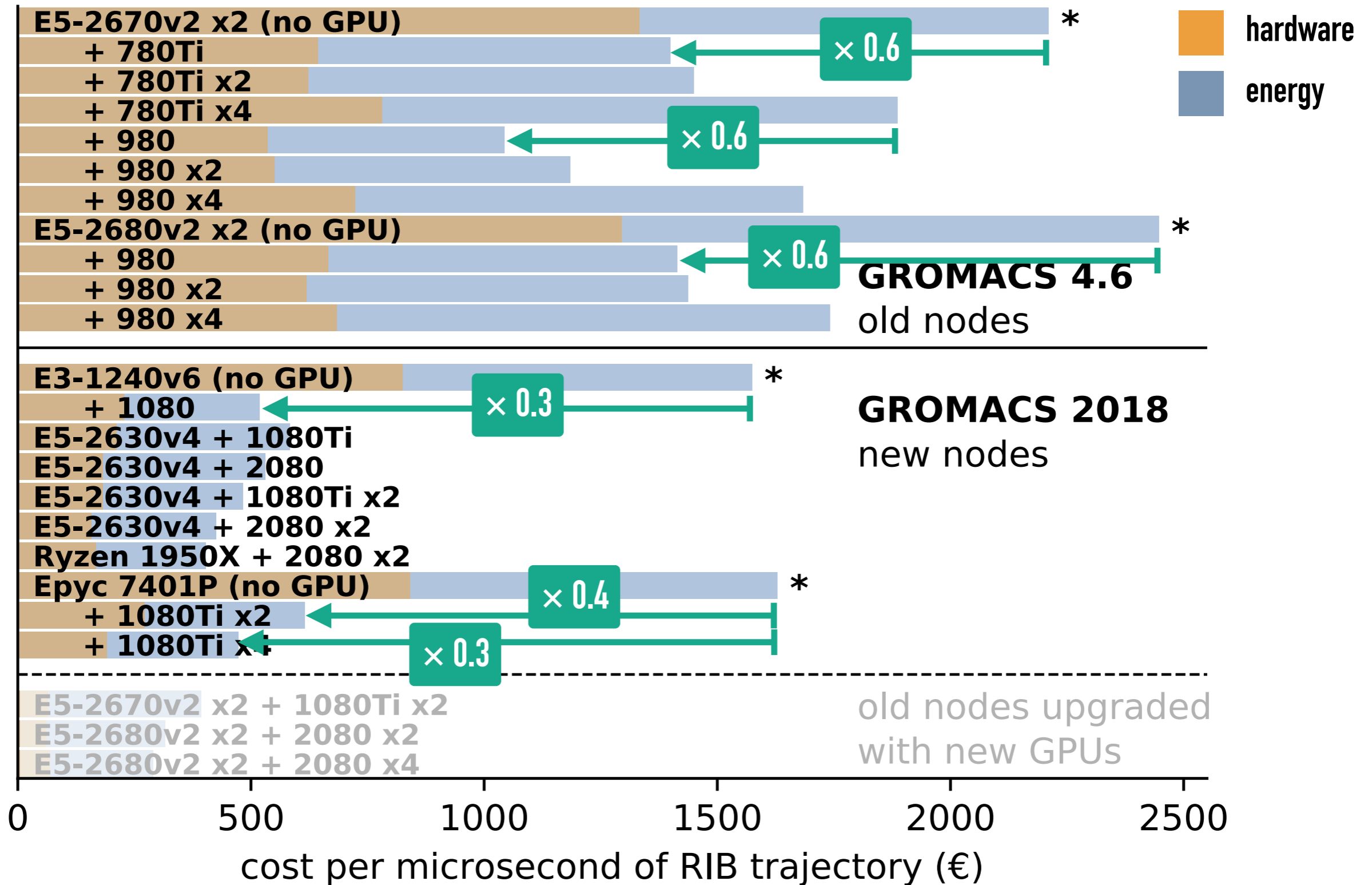
for 5 years of operation





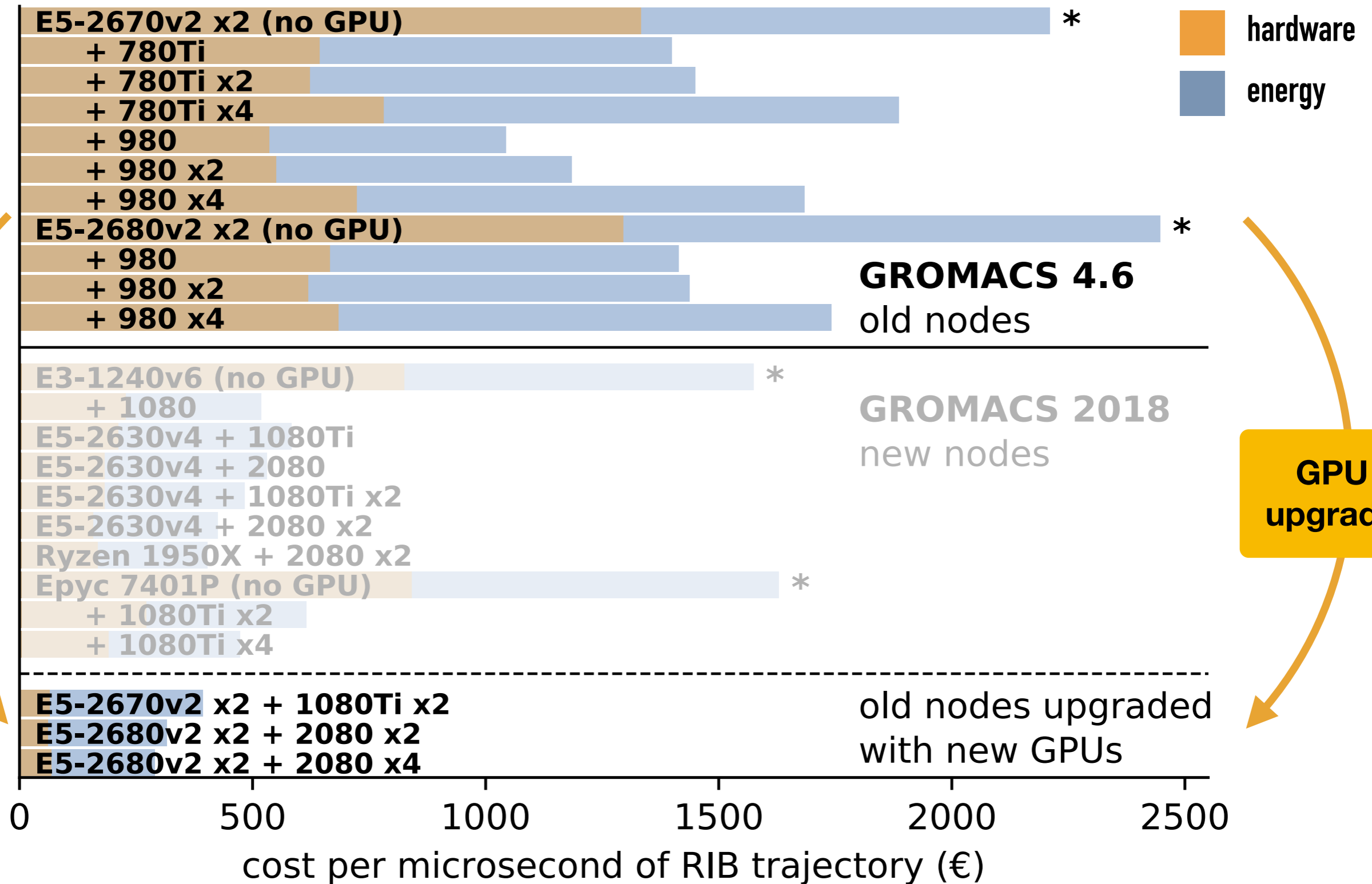
# Total trajectory costs hardware+energy

for 5 years of operation



# Total trajectory costs hardware+energy

for 5 years of operation



**GPU upgrade**

# Conclusions

## Buying new nodes:

- Consumer GPU nodes have a **much higher performance-to-price ratio** than CPU nodes
  - raw node price: 2–3 x for GROMACS 4.6, and **3–7 x for GROMACS 2018**
  - + energy costs: 2 x for GROMACS 4.6, and **3 x for GROMACS 2018**

# Conclusions

## Buying new nodes:

- Consumer GPU nodes have a **much higher performance-to-price ratio** than CPU nodes
  - raw node price: 2–3 x for GROMACS 4.6, and 3–7 x for GROMACS 2018
  - + energy costs: 2 x for GROMACS 4.6, and 3 x for GROMACS 2018

## Recycle old nodes if you can! As a result of CPU → GPU work shifting (PME on GPU)

- **upgrading the GPU** yields large performance increase, whereas
- exchanging the rest of a node (CPU, ..) can be a **waste of money**

# Conclusions

## Buying new nodes:

- Consumer GPU nodes have a **much higher performance-to-price ratio** than CPU nodes
  - raw node price: 2–3 x for GROMACS 4.6, and 3–7 x for GROMACS 2018
  - + energy costs: 2 x for GROMACS 4.6, and 3 x for GROMACS 2018

## Recycle old nodes if you can! As a result of CPU → GPU work shifting (PME on GPU)

- **upgrading the GPU** yields large performance increase, whereas
- exchanging the rest of a node (CPU, ..) can be a **waste of money**
- optimal hardware balance: ~15 core-GHz per 2080 GPU

# Conclusions

## Buying new nodes:

- Consumer GPU nodes have a **much higher performance-to-price ratio** than CPU nodes
  - raw node price: 2–3 x for GROMACS 4.6, and 3–7 x for GROMACS 2018
  - + energy costs: 2 x for GROMACS 4.6, and 3 x for GROMACS 2018

## Recycle old nodes if you can! As a result of CPU → GPU work shifting (PME on GPU)

- **upgrading the GPU** yields large performance increase, whereas
- exchanging the rest of a node (CPU, ..) can be a **waste of money**
- optimal hardware balance: ~15 core-GHz per 2080 GPU
- results transfer to GROMACS 2019 as well
  - bonded interactions → CUDA GPU
  - PME offload with OpenCL → AMD GPUs

# Additional Material

- want to compare your own hardware and contribute to benchmarking?  
<https://www.mpibpc.mpg.de/grubmueller/bench> has various **benchmark .tprs for download** (CC licensed)
- Related publications:
  - GROMACS 2018/2019:  
More Bang for Your Buck: Improved use of GPU Nodes for GROMACS 2018
    - JCC <https://onlinelibrary.wiley.com/doi/10.1002/jcc.26011>
    - arXiv <https://arxiv.org/abs/1903.05918>
  - Summary **poster**: <https://www.mpibpc.mpg.de/grubmueller/kutzner/posters>
  - GROMACS 4.6/5.0:  
Best bang for your buck: GPU nodes for GROMACS biomolecular simulations
    - JCC <https://onlinelibrary.wiley.com/doi/full/10.1002/jcc.24030>
    - arXiv <https://arxiv.org/abs/1507.00898>

# Acknowledgments



The Department of Theoretical & Computational Biophysics  
@ MPI for Biophysical Chemistry Göttingen

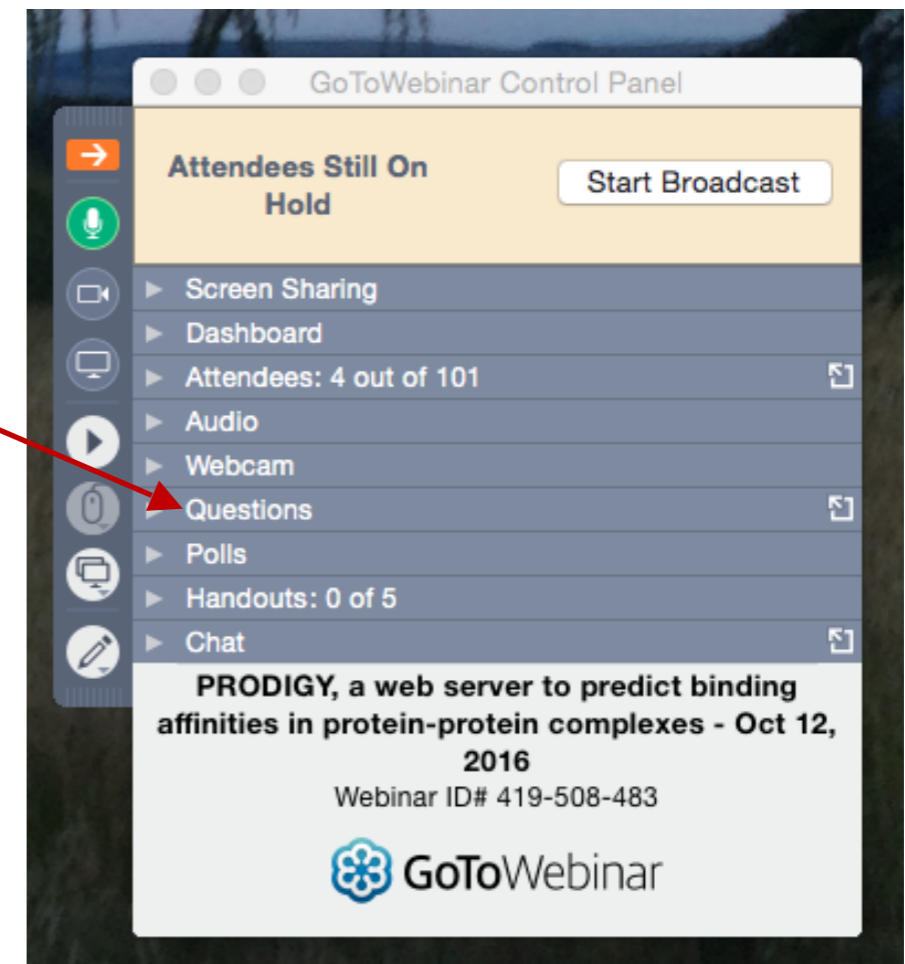
Markus Rampp, Hermann Lederer (Max Planck Computing & Data Facility)





# Audience Q&A session

- Please use the Questions function in GoToWebinar application
- Any other questions or points to discuss after the live webinar? Join the discussions at <http://ask.bioexcel.eu>.



# Coming up next!



12 Sept 2019

Enhanced molecular simulations with PLUMED