High Performance Computing - Fall 2025

MiniWeather

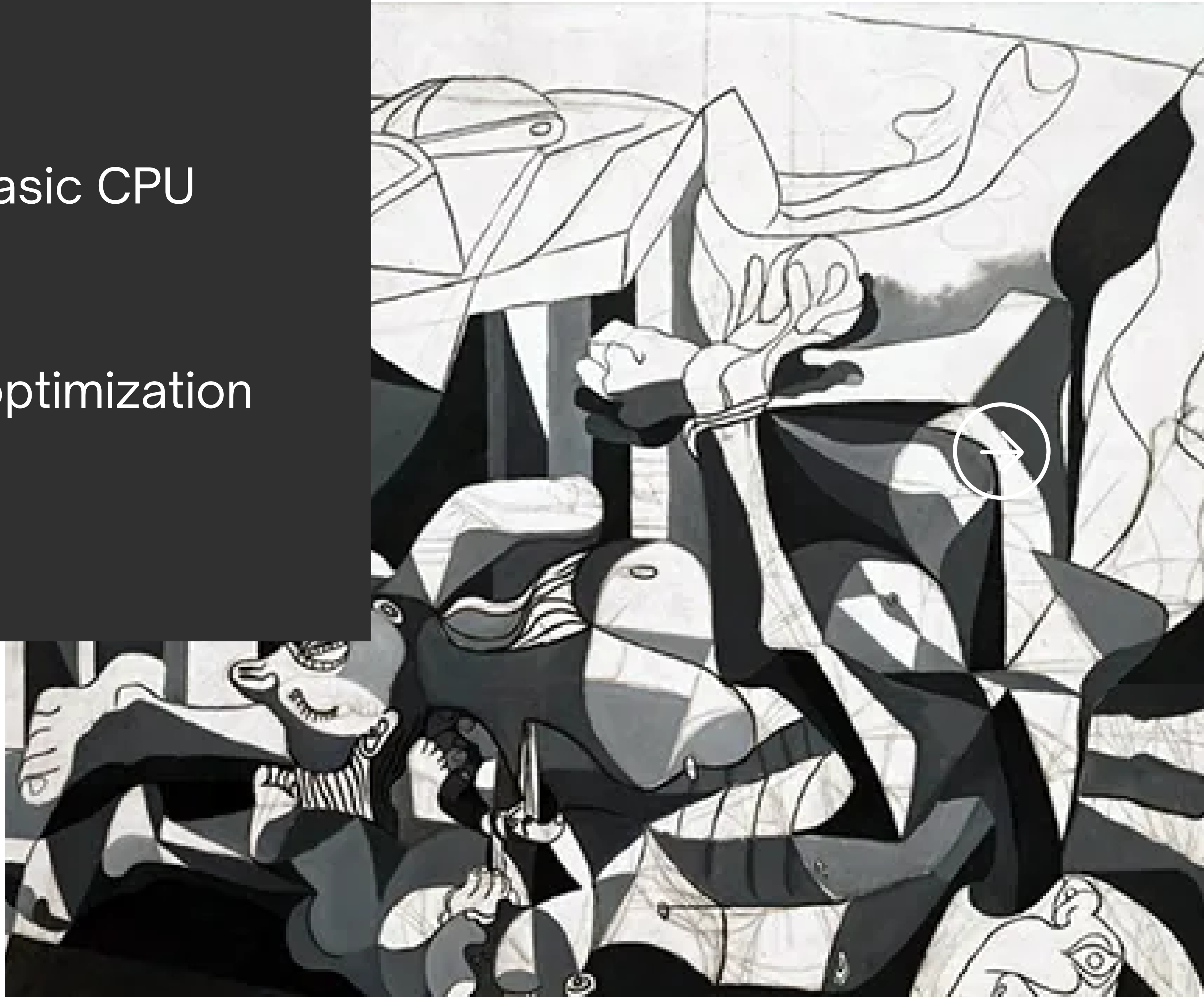
Parallel Performance Analysis & EuroHPC Proposal

Diana Cordovez, Gabriela Vega, Sabina Bacaoanu, Hibba Kamas, Diego Oliveros, Cayetana Hinostroza

MiniWeather models 3D stencil evolution representative of weather forecasting kernels

- **Challenge:** Achieve efficient parallel scaling for operational weather prediction
- **Current limitation**: Teaching cluster provides only basic CPU nodes and 2 GPUs
- **Impact**: Understanding scaling bottlenecks guides optimization of larger atmospheric models (ICON, IFS, FV3)

Problem & Impact



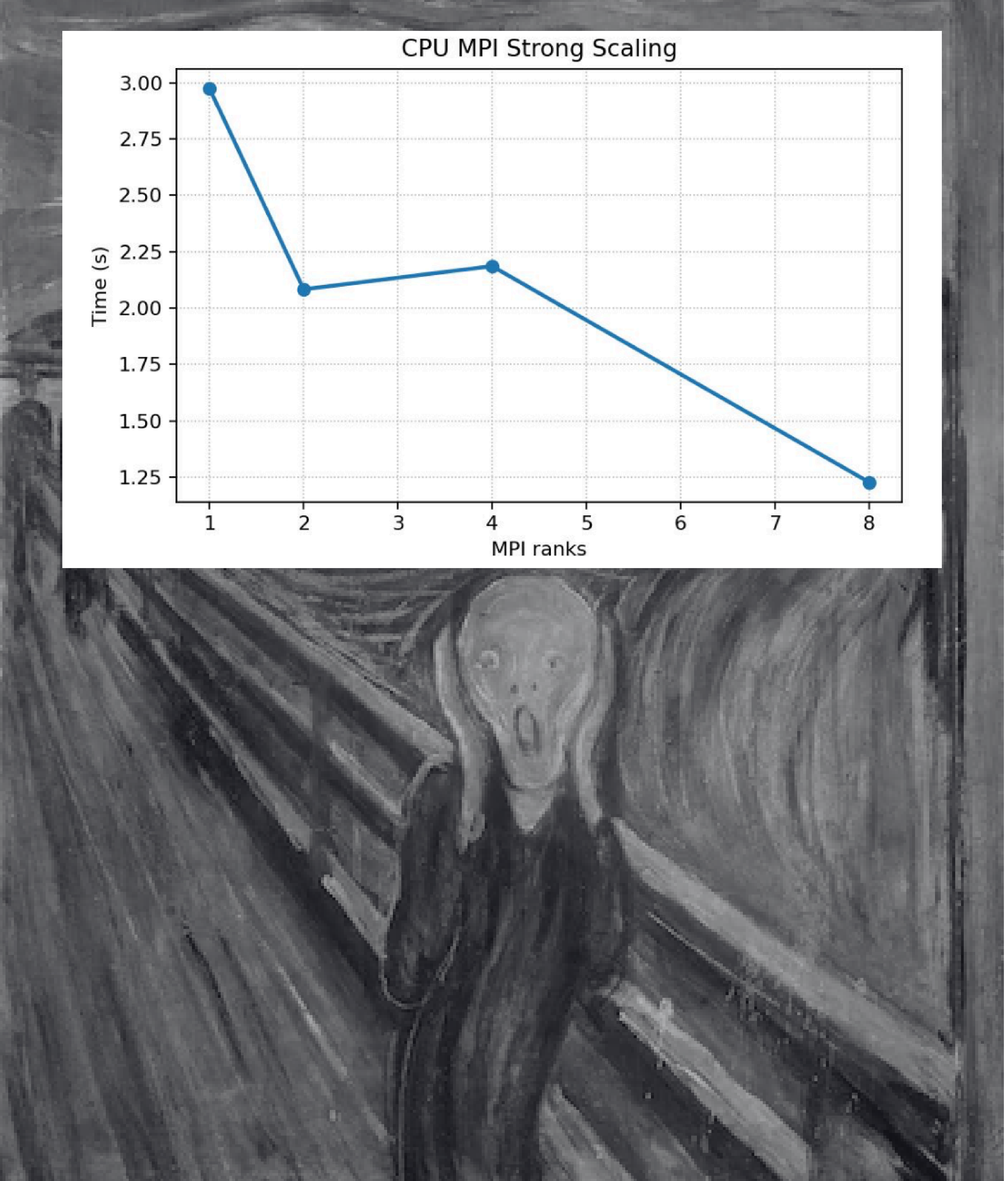
Approach & Prototypes

- **Implementations**: Serial, OpenMP, MPI, Hybrid MPI+OpenMP, MPI+OpenACC (attempted)

- **Test configuration:** Baseline grid 256×128×128 cells, 50 time steps

- **Experiments conducted:**
  - **Strong scaling:** 1-8 MPI ranks across 1-4 nodes
  - **Weak scaling:** Grid grows with rank count
  - **Hybrid:** MPI + OpenMP threading

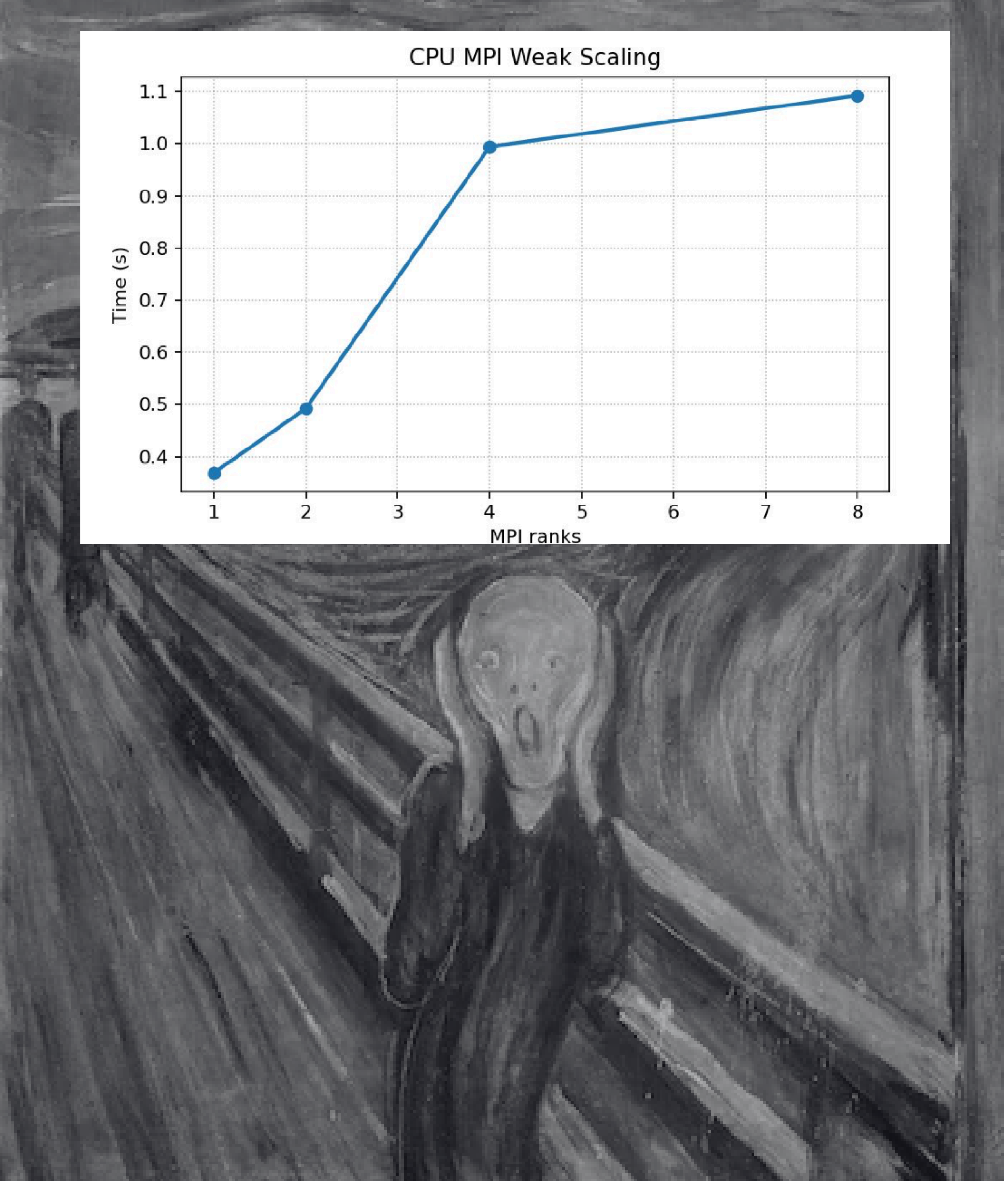
- **Infrastructure:** Slurm job scripts, automated CSV logging, Python plotting pipeline



Key Results: Scaling Limitations Identified

Strong scaling (MPI):

- **Good:** 1→2 ranks (2.98s → 2.08s)
- **Limited:** 2→4 ranks (2.08s → 2.19s, performance degradation)
- **Best**: 8 ranks/2 nodes (1.12s), but 8 ranks/4 nodes slower (1.33s)



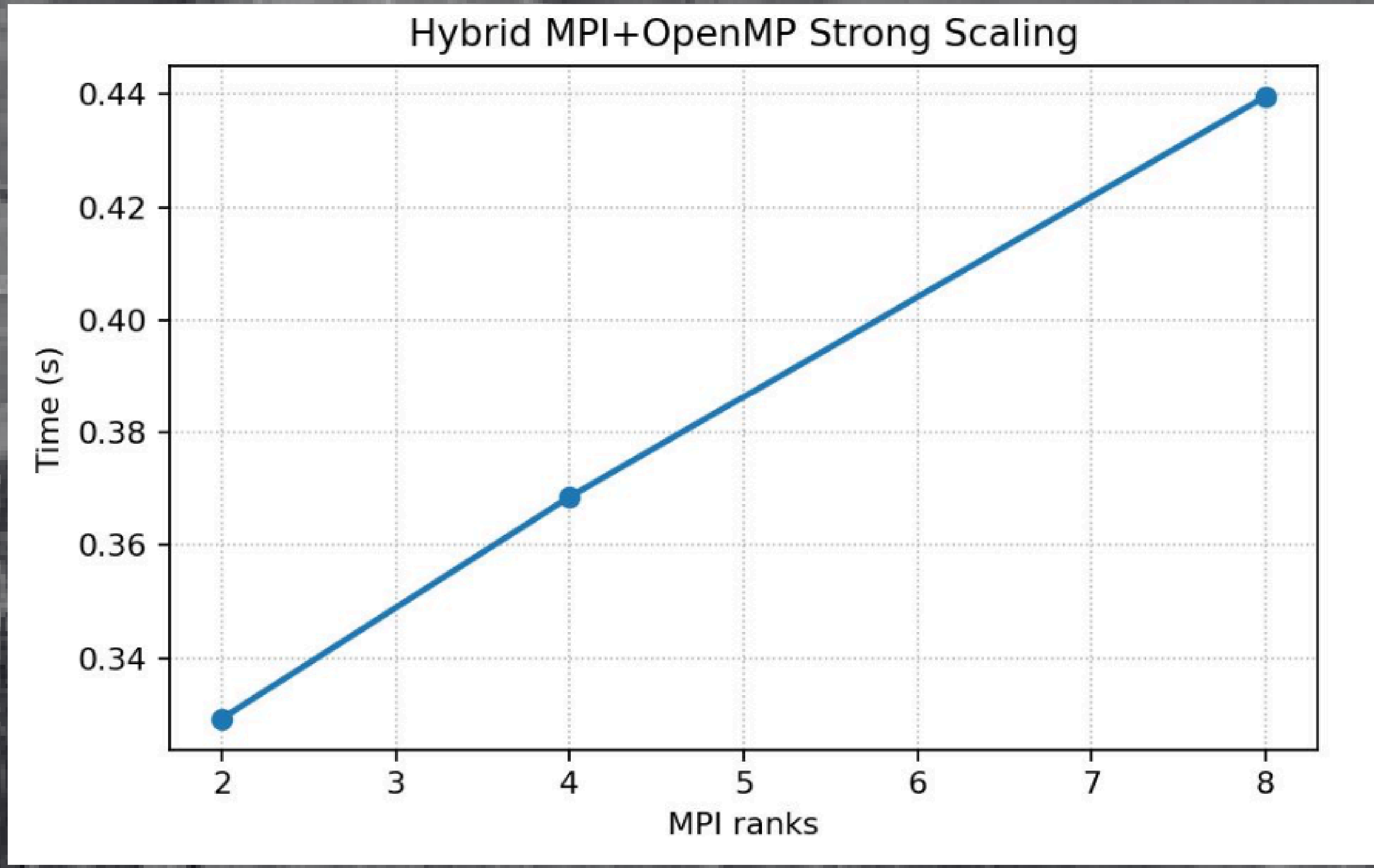
Key Results: Scaling Limitations Identified

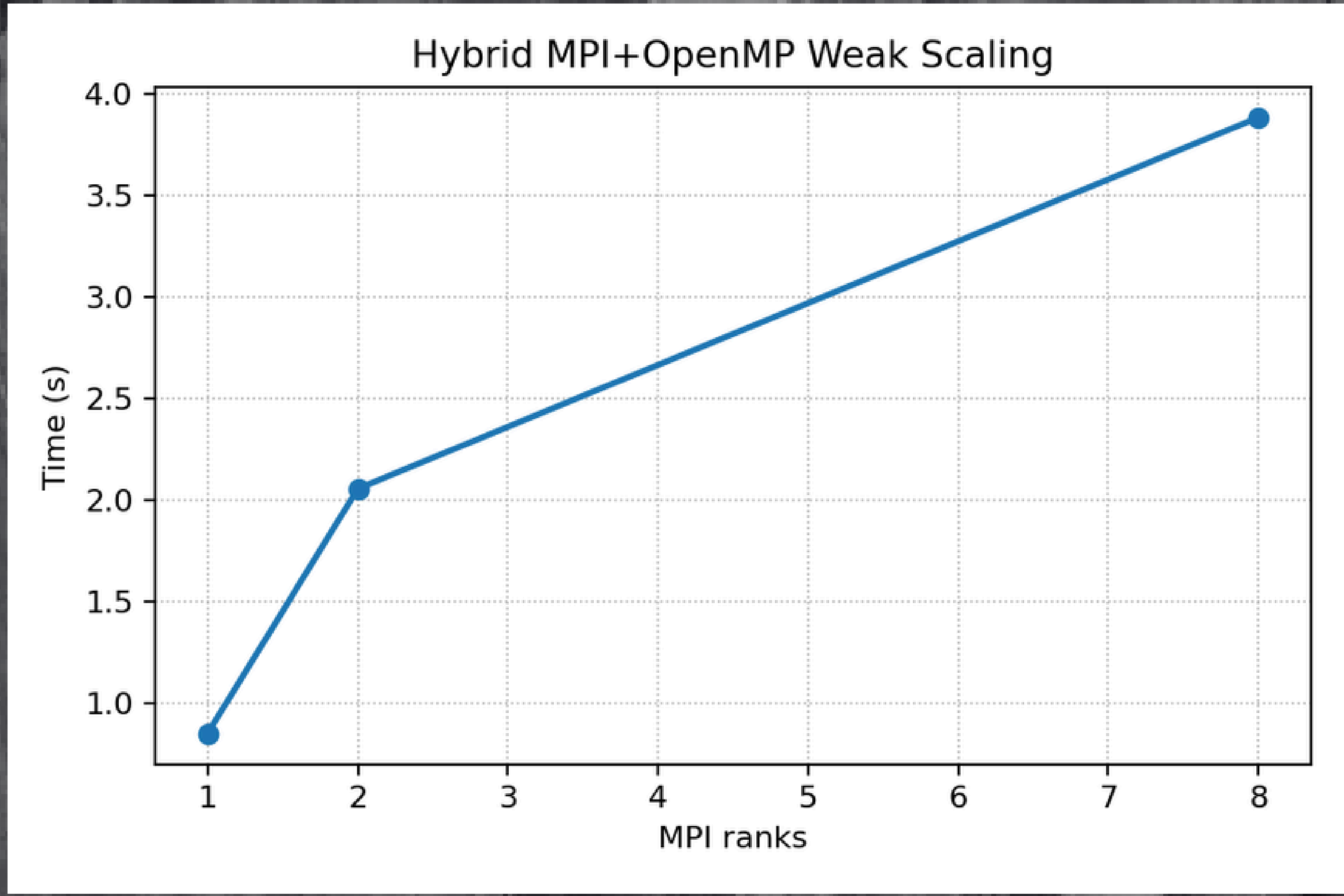
Strong scaling (MPI):

- **Good:** 1→2 ranks (2.98s → 2.08s)
- **Limited:** 2→4 ranks (2.08s → 2.19s, performance degradation)
- **Best**: 8 ranks/2 nodes (1.12s), but 8 ranks/4 nodes slower (1.33s)

Weak scaling:

- Runtime grows with ranks instead of staying constant (0.37s → 1.08s for 1→4 ranks)





Key Results: Scaling Limitations Identified

Strong scaling (MPI):

- **Good:** 1→2 ranks (2.98s → 2.08s)
- **Limited:** 2→4 ranks (2.08s → 2.19s, performance degradation)
- **Best**: 8 ranks/2 nodes (1.12s), but 8 ranks/4 nodes slower (1.33s)

Weak scaling:

- Runtime grows with ranks instead of staying constant (0.37s → 1.08s for 1→4 ranks)

Hybrid MPI+OpenMP:

- Slight improvement at 2 ranks, degradation at 8 ranks

Root causes: Communication overhead, memory bandwidth saturation, inter-node latency

- **Target systems:** LUMI-G (AMD MI250X) or Leonardo Booster (NVIDIA A100)
- **Scale objective:** 1-64 GPU nodes with strong/weak scaling analysis
- **Resource request:**
  - 2000 CPU core-hours (baseline + hybrid runs)
  - 1500 GPU-hours (scaling sweeps + profiling)
  - <50 GB storage for logs and traces
- **Deliverables:** Optimized code, profiling reports, scaling plots, public repository with reproducible workflows

Next Steps: EuroHPC Development Access

Work Plan: 4-Month Timeline

- **Month 1:** Port & validate code on target system, establish baseline correctness
- **Month 1-2:** Profiling with Nsight Systems/Compute, rocprof, LIKWID - identify hotspots
- **Month 2-3:**
  - Implement optimizations:
  - Communication-computation overlap
  - Cache blocking strategies
  - GPU stream concurrency
  - Target: >1.5× speedup on ⩾16 nodes
- **Month 3:** Large-scale validation: 64-node strong/weak scaling sweeps
- **Month 4:** Documentation, short paper, public dataset release (Zenodo)

Optimization Strategy & Milestones

Risks & Mitigation Strategy

Risk 1 - Compiler compatibility: Different toolchains (Cray/NVHPC/ROCm)

- **Mitigation:** Vendor support tickets, simplified kernel versions

Risk 2 - Queue wait times: Limited access windows

- **Mitigation:** Night runs, request reservations if needed

Risk 3 - Hardware failures at scale: Instability on 64 nodes

- **Mitigation:** Schedule contingency buffer, checkpoint/restart

Support needed:

- Site application expert guidance (profiling counters, submission best practices)
- Brief onboarding session with EuroHPC staff

Thank You