A birds eye view of a decade of

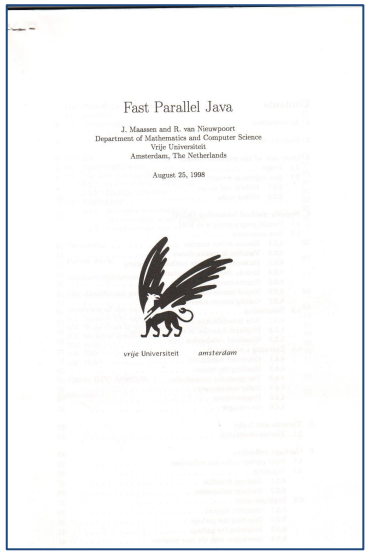
efficient computing

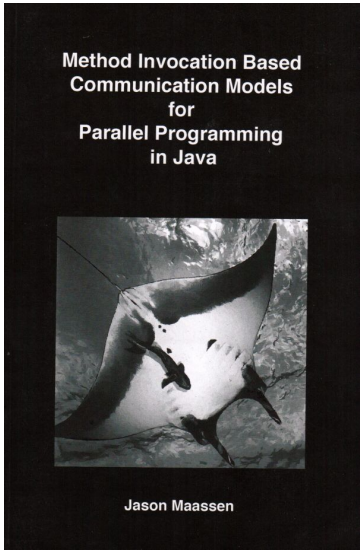
at the

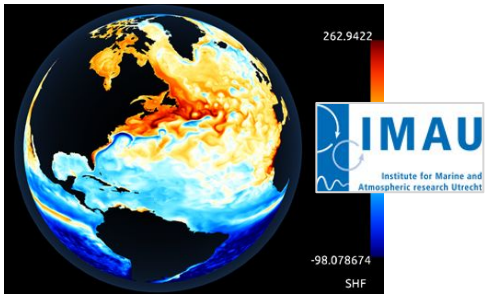
netherlands eScience center

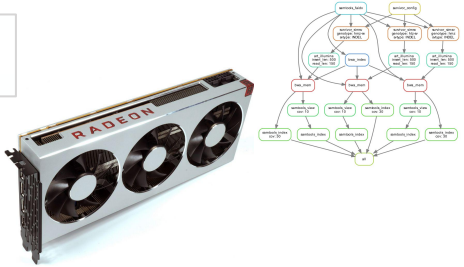
Dr. Jason Maassen

My background









1993-1998
MSc Computer Science
HPC

1998-2012
PhD / Postdoc Computer Science
HPC / Grid computing

2012+
Research Software Engineer
Ocean Modelling / Digital Forensics

2016+
Technology Lead
Efficient Computing

VU VRIJE UNIVERSITEIT AMSTERDAM

VU VRIJE UNIVERSITEIT AMSTERDAM

UNIVERSITEIT VAN AMSTERDAM

eScience center

eScience center

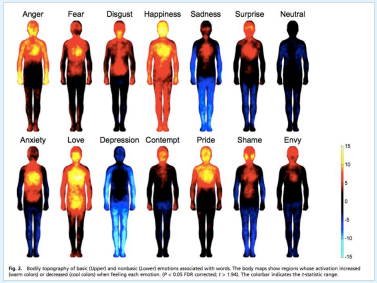
netherlands eScience center ?

netherlands eScience center

National center for the development and application of research software

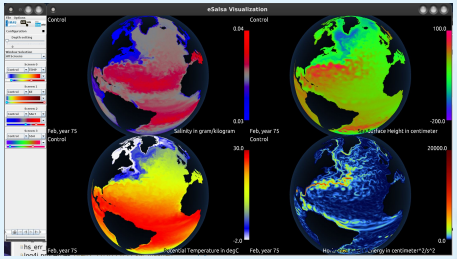
Founded in 2012 by NWO and SURF

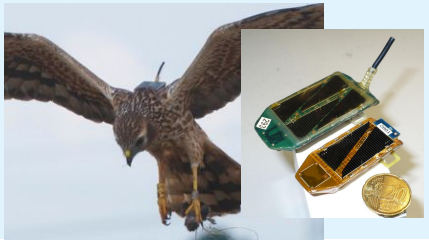
Science Park Amsterdam



325 projects

(on many different topics)





Humanities & Social Sciences

incl. SMART cities, text analysis, creative technologies

Physics & Beyond

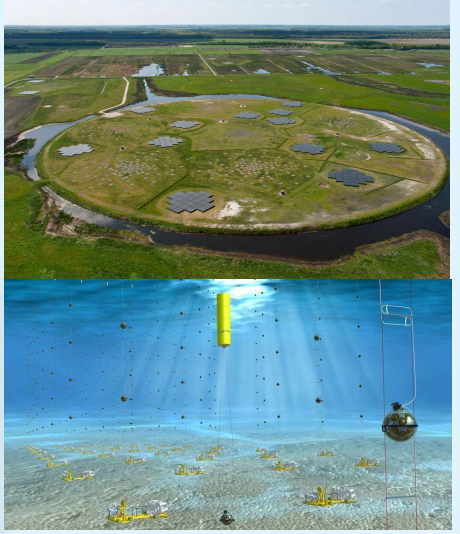
incl. astronomy, high-energy physics, advanced materials

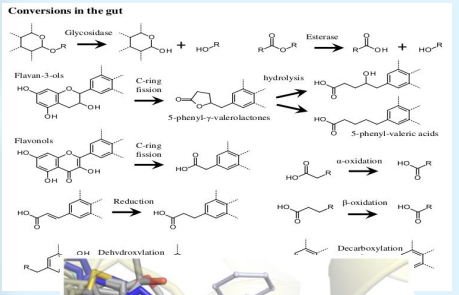
Sustainability & Environment

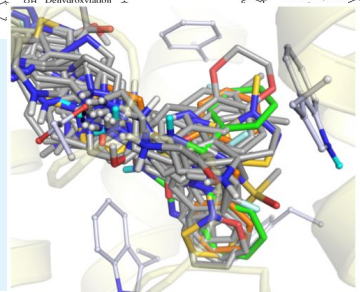
incl. climate, ecology, energy, logistics, water management

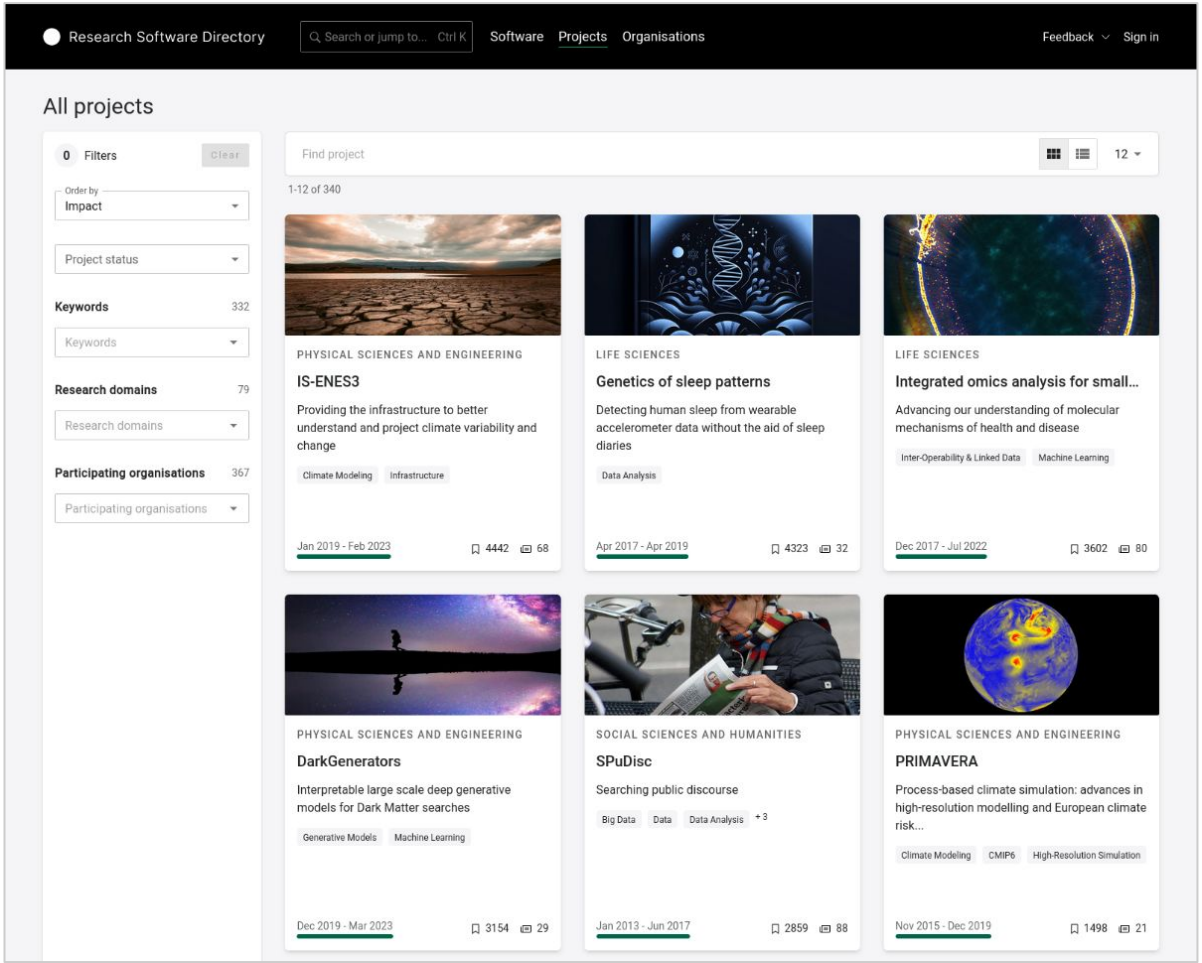
Life Sciences & eHealth

incl. bio-imaging, next generation sequencing, molecules

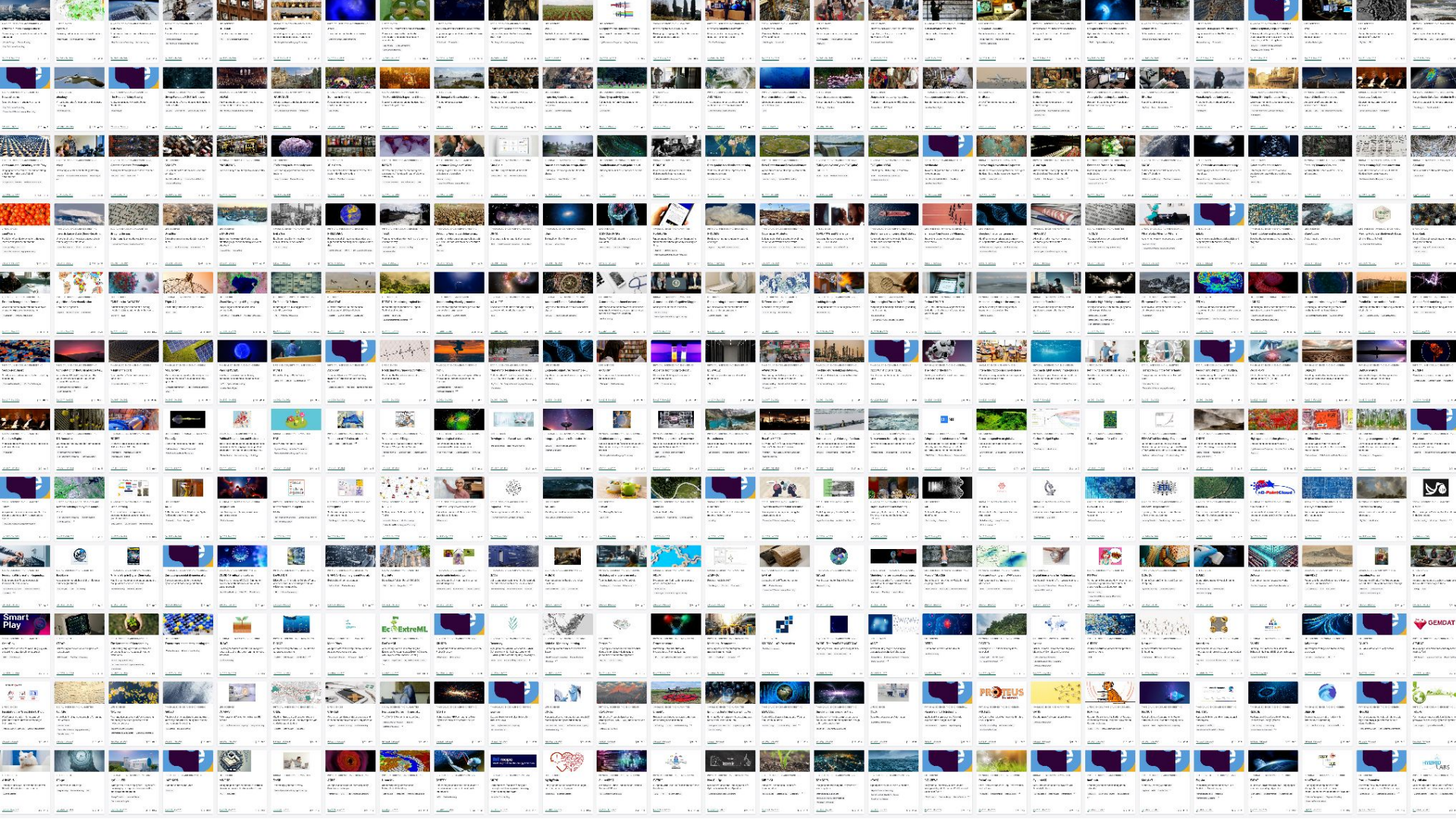








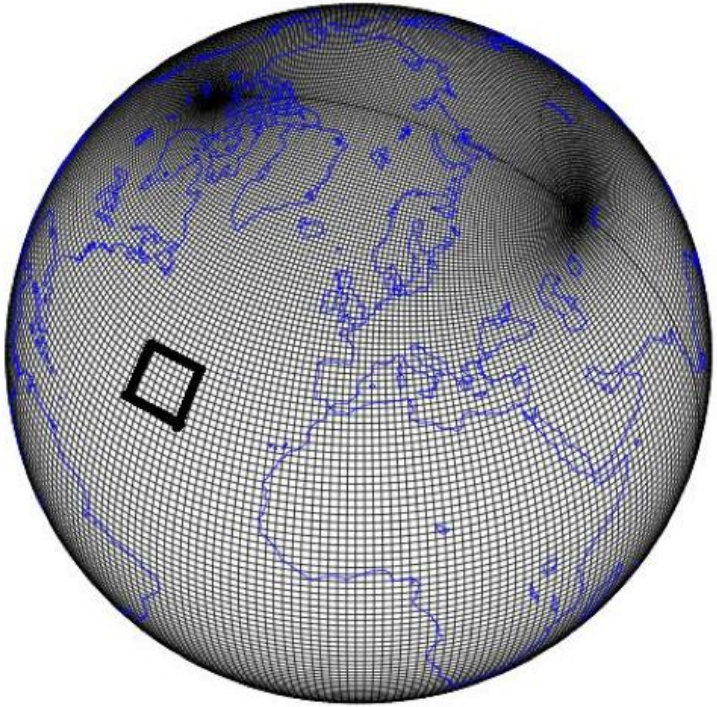
https://research-software-directory.org/projects




trends
observations
lessons learned
interesting crossovers

FORTRAN is here to stay

eSalsa (2012-2016)

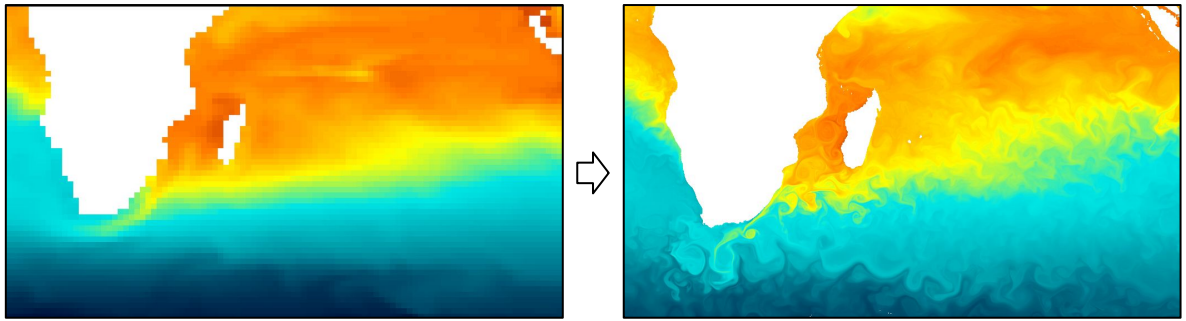


Source: Los Alamos National Laboratory

Ocean modelling using Parallel Ocean Program (POP)

Very traditional Fortran/MPI application (1992)

Goal: go from 1° to 0.1° resolution (100x100km to 10x10km)



IMAU Institute for Marine and Atmospheric research Utrecht

Universiteit Utrecht

VU UNIVERSITY AMSTERDAM

PHYSICAL SCIENCES AND ENGINEERING

eSALSA

Predicting local sea level changes

Climate Modeling · GPU · Ocean-Circulation-Models

Jan 2012 - Feb 2016 · 79 · 27

PHYSICAL SCIENCES AND ENGINEERING

Extreme climate changes due to...

Showcasing an extreme high resolution climate simulation

Climate Change · Climate Modeling · High Performance Computing

Nov 2012 - Dec 2013 · 0 · 1

LIFE SCIENCES, PHYSICAL SCIENCES AND...

Summer in the City

Forecasting and mapping human thermal comfort in urban areas

Data Analysis

Jan 2013 - Sep 2016 · 266 · 31

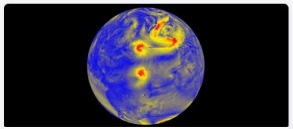
PHYSICAL SCIENCES AND ENGINEERING

Towards Large-Scale Cloud-Resolvin...

Understanding the interaction between clouds and the large-scale circulation

Atmospheric Physics · Clouds · Convection · +2

Dec 2015 - Oct 2019 · 200 · 38



PHYSICAL SCIENCES AND ENGINEERING

PRIMAVERA

Process-based climate simulation: advances in high-resolution modelling and European climate risk...

Climate Modeling · CMIP6 · High-Resolution Simulation

Nov 2015 - Dec 2019 · 1497 · 21

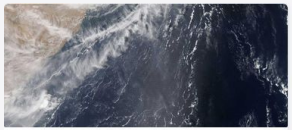
LIFE SCIENCES

Blue-Action

Arctic impact on weather and climate

Climate Change

Dec 2016 - Feb 2021 · 97 · 11



PHYSICAL SCIENCES AND ENGINEERING

MAGIC

Metrics and Access to Global Indices for Climate Projections

Optimized Data Handling · Workflow Technologies

Oct 2016 - Mar 2019 · 0 · 4

PHYSICAL SCIENCES AND ENGINEERING

EUropean Climate Prediction system...

An innovative European regional ensemble climate prediction system

Visualization

Dec 2017 - May 2022 · 2 · 15

LIFE SCIENCES

MOSAIC

Modelling sea level and inundation for cyclones

Multi-Scale & Multi Model Simulations

Dec 2018 - May 2024 · 748 · 24

PHYSICAL SCIENCES AND ENGINEERING

ESiWACE2

For future exascale climate and weather predictions

GPU

Jan 2019 - Dec 2022 · 120 · 41

PHYSICAL SCIENCES AND ENGINEERING

ESiWACE3

Centre of Excellence in Simulation of Weather and Climate in Europe

Climate Change · Climate Modeling · GPU · +1

Jan 2023 - Dec 2025 · 0 · 5

PHYSICAL SCIENCES AND ENGINEERING

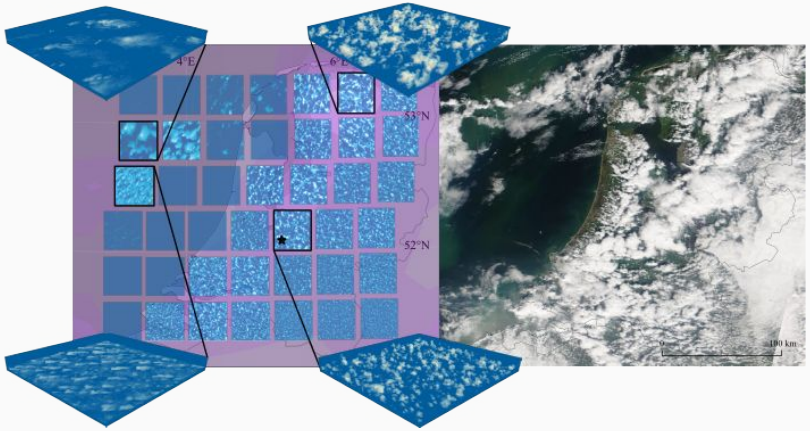
eTAOC

Tipping of the Atlantic Ocean Circulation

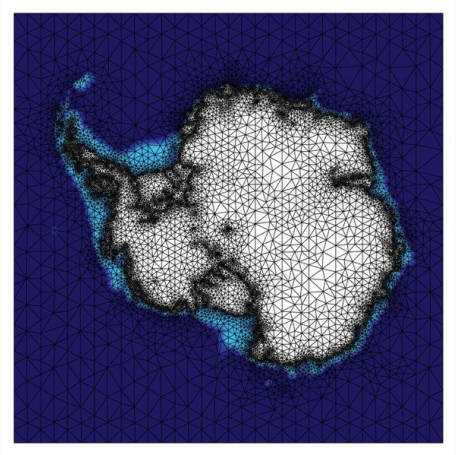
High Performance Computing · Multi-Scale & Multi Model Simulations · Rare Event Simulations

Jan 2023 - Sep 2027 · 0 · 2

Since then … a long string of FORTRAN based projects!

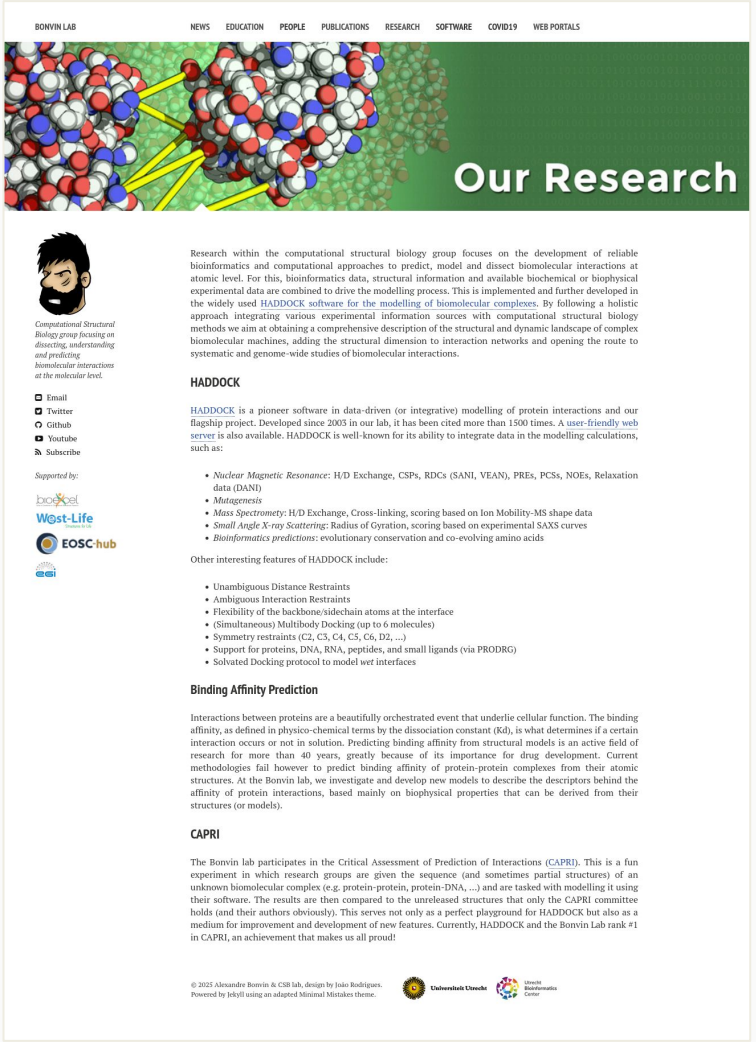


DALES: Atmospheric Large Eddy Simulation



UFEMISM 2.0 Ice sheet modelling

HADDOCK3:
High Ambiguity Driven
protein-protein DOCKing



HPC = GPU

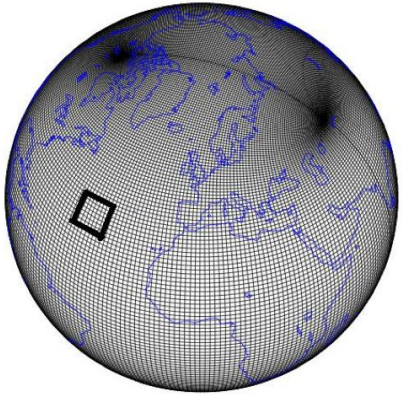
Rank	System	Cores	Rmax (PFlop/s)	Rpeak (PFlop/s)	Power (kW)	
1	**El Capitan** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	11,039,616	1,742.00	2,746.38	29,581	AMD GPU
2	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE Cray OS, HPE DOE/SC/Oak Ridge National Laboratory United States	9,066,176	1,353.00	2,055.72	24,607	AMD GPU
3	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel DOE/SC/Argonne National Laboratory United States	9,264,128	1,012.00	1,980.01	38,698	Intel GPU
4	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Microsoft Azure Microsoft Azure United States	2,073,600	561.20	846.84		NVIDIA GPU
5	**HPC6** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9, HPE Eni S.p.A. Italy	3,143,520	477.90	606.97	8,461	AMD GPU
6	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu RIKEN Center for Computational Science Japan	7,630,848	442.01	537.21	29,899	ARM CPU
7	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE Cray OS, HPE Swiss National Supercomputing Centre (CSCS) Switzerland	2,121,600	434.90	574.84	7,124	NVIDIA GPU
8	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE EuroHPC/CSC Finland	2,752,704	379.70	531.51	7,107	AMD GPU
9	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64 GB, Quad-rail NVIDIA HDR100 Infiniband, EVIDEN EuroHPC/CINECA Italy	1,824,768	241.20	306.31	7,494	NVIDIA GPU
10	**Tuolumne** - HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS, HPE DOE/NNSA/LLNL United States	1,161,216	208.10	288.88	3,387	AMD GPU

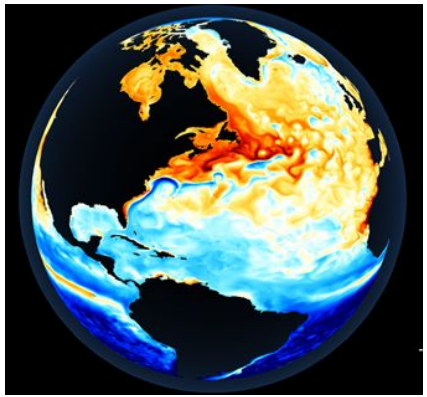
source: https://www.top500.org/lists/top500/2024/11/

The best illustration of this …

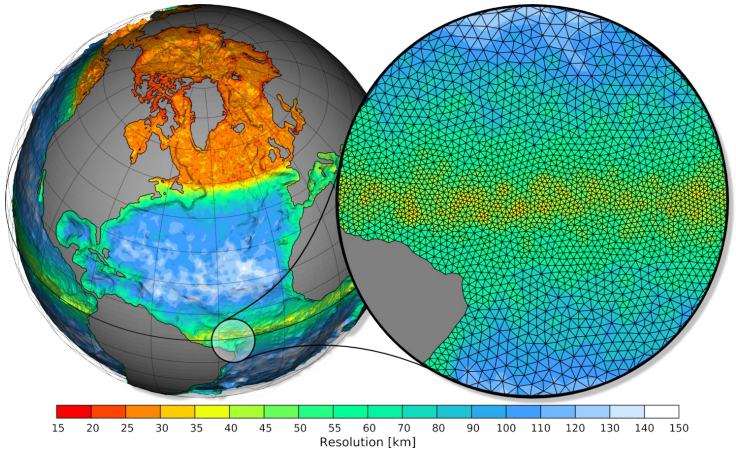
1997
ASCI Red
1.3 teraflops
850 KW / $46M
150 m2

2025
AMD Radeon RX 9070 XT
1.5 teraflops
317 W / €799
0.04 m2





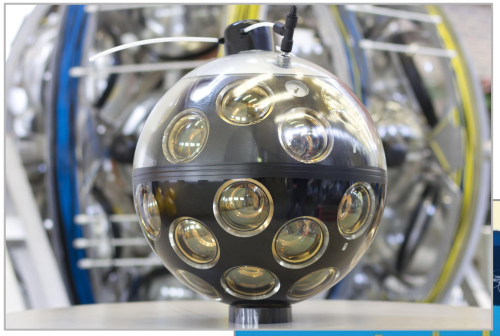
eSalsa (2012-2016)
Porting an Ocean Model to GPU

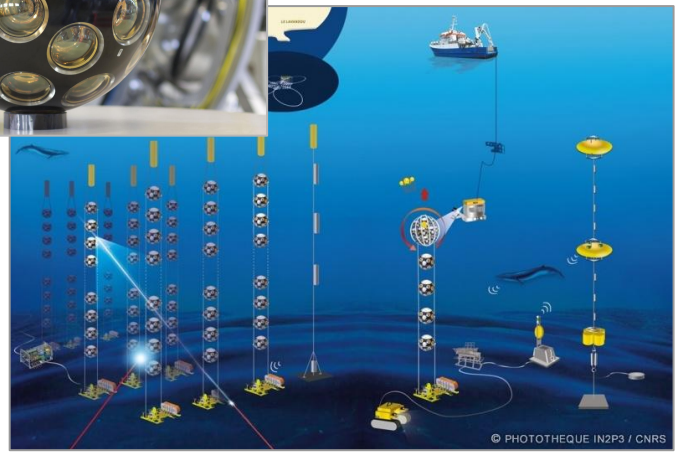


source: https://fesom.de/applications/paleoclimate/

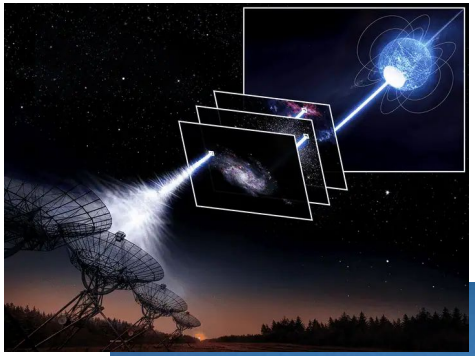
ESiWACE2 (2019-2022)
Porting an Ocean Model to GPU

(yes, we've seen a lot of repeated efforts over the years!)



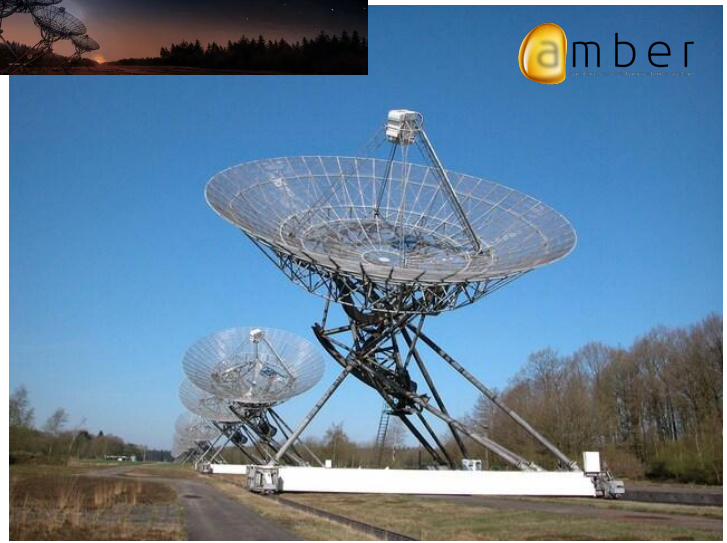


KM3NeT neutrino detector
12.000 spheres with 31 photo detectors
on 600 strings in the Mediterranean.
Filter 55 GB/s in real time to detect neutrinos



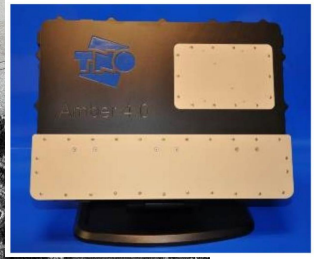
ASTRON

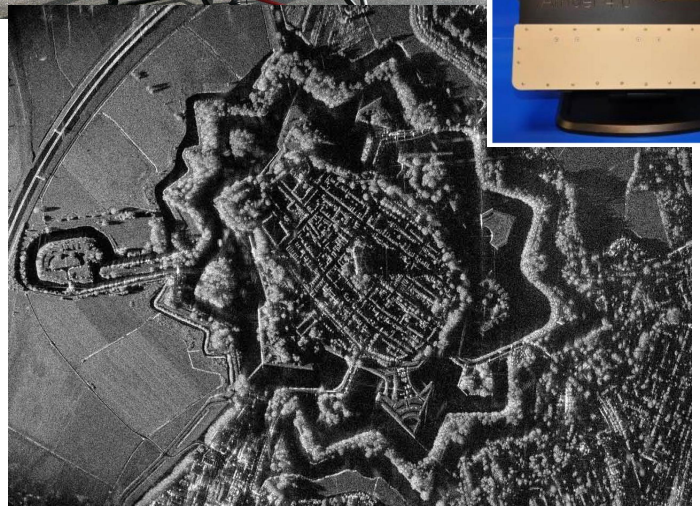
University of Amsterdam



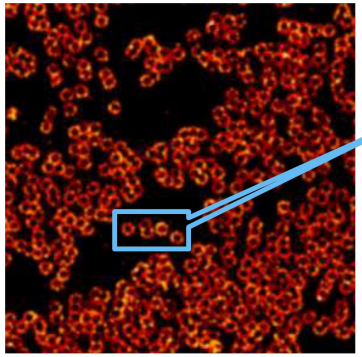
APERTIF Westerbork Synthesis Radio Telescope
12 telescopes produce 500 GB/s
Processed in real time to detect fast radio bursts

TNO innovation for life

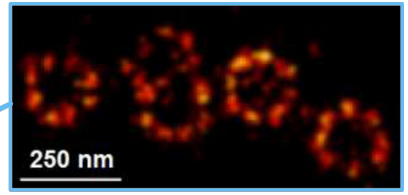


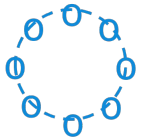


Real-time processing of Synthetic Aperture Radar (SAR) data
Energy budget is limited to 50W

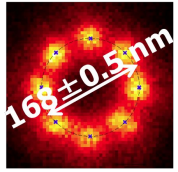


dSTORM of the Nuclear Pore complex



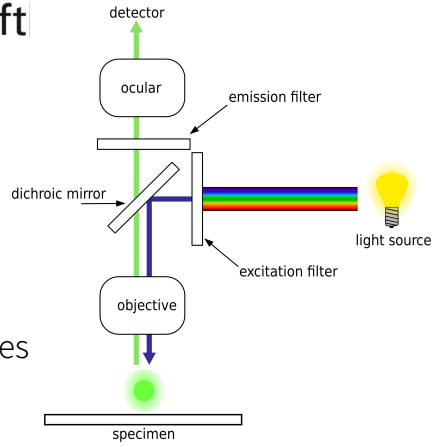


Template



426 fused NPCs

TUDelft



Super-resolution microscopy
Combine a large number of images from fluorescence microscope into a single super-resolution reconstruction

(Portable) performance is hard!

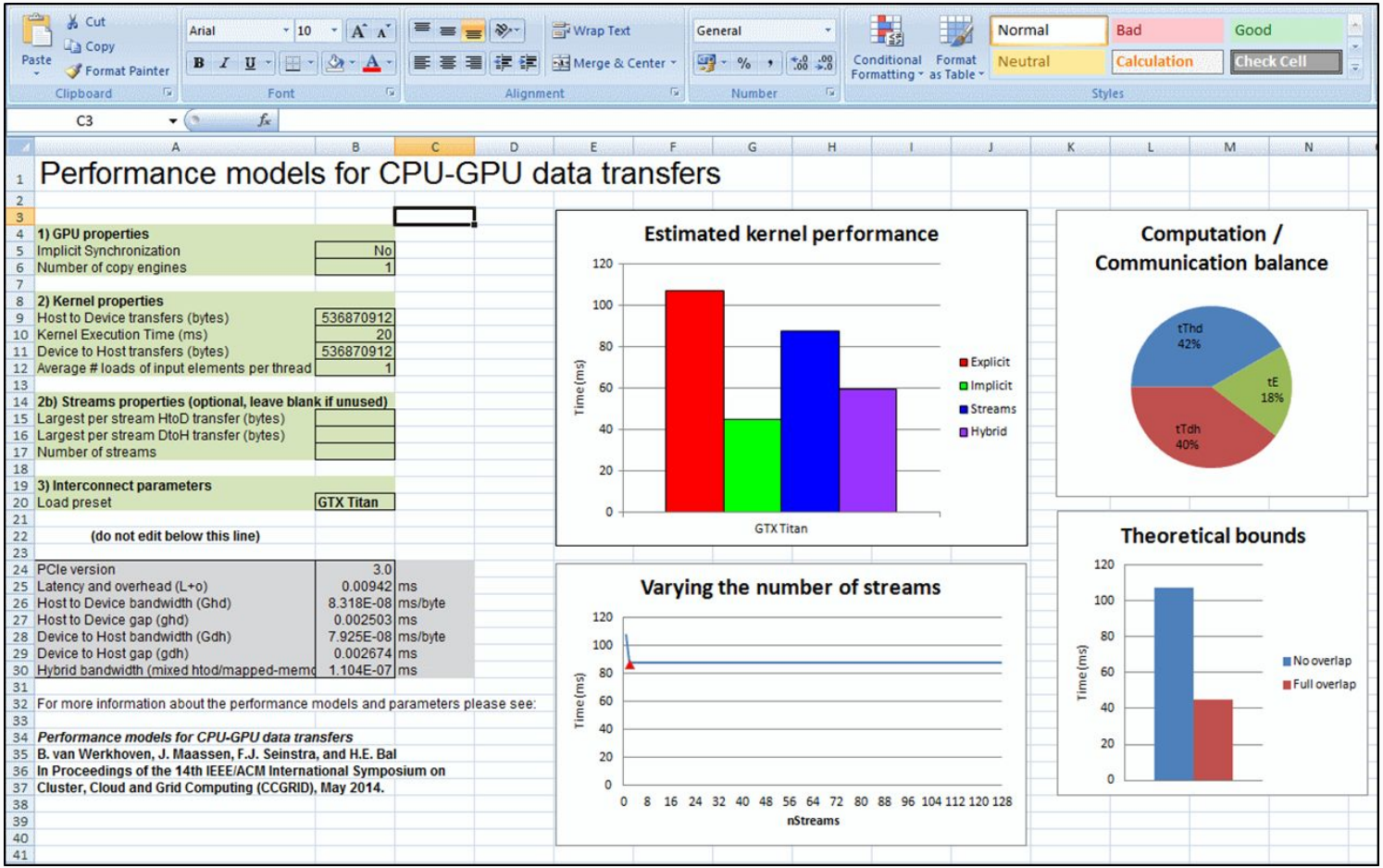
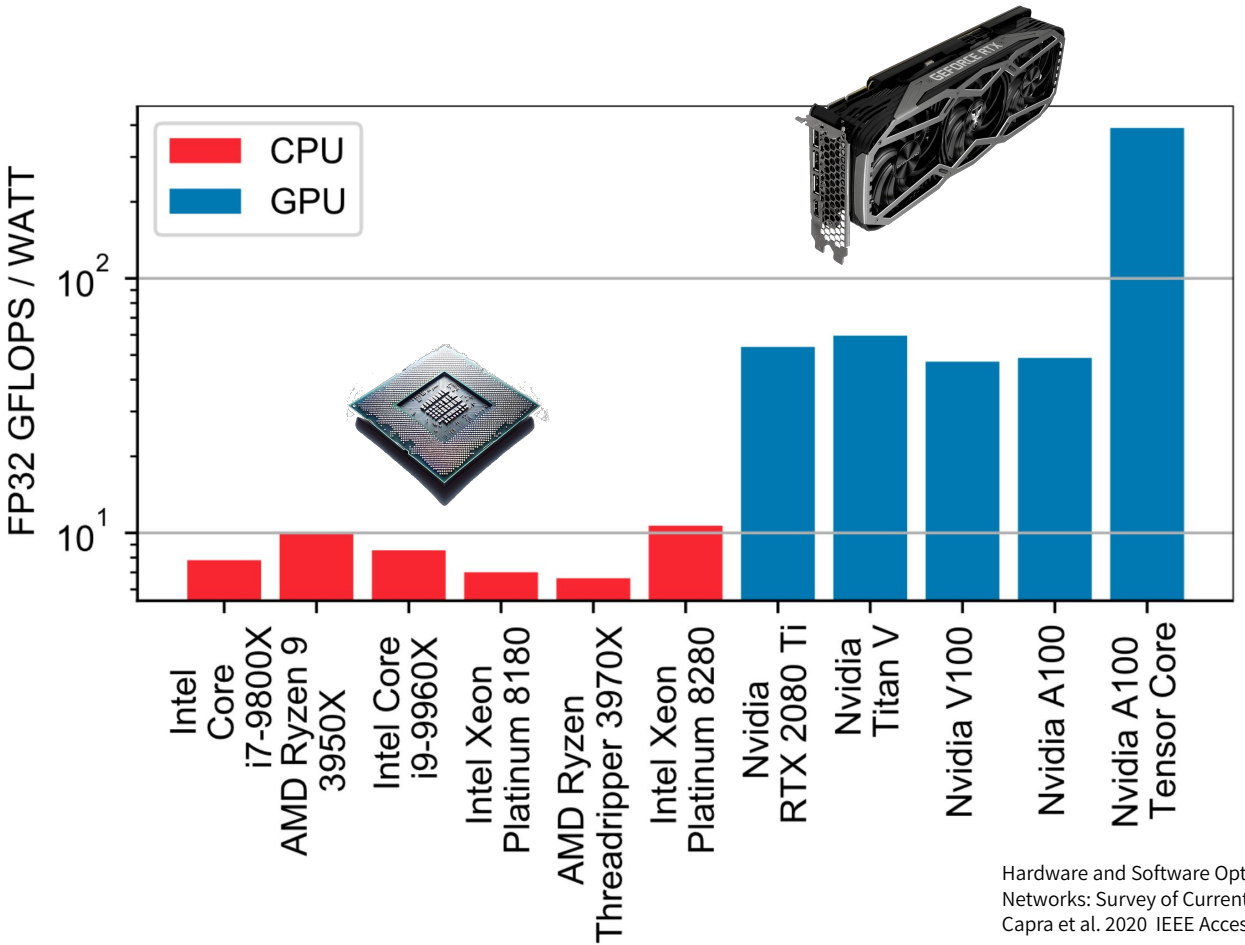


Image: Ben van Werkhoven

GPU performance used to be predictable…

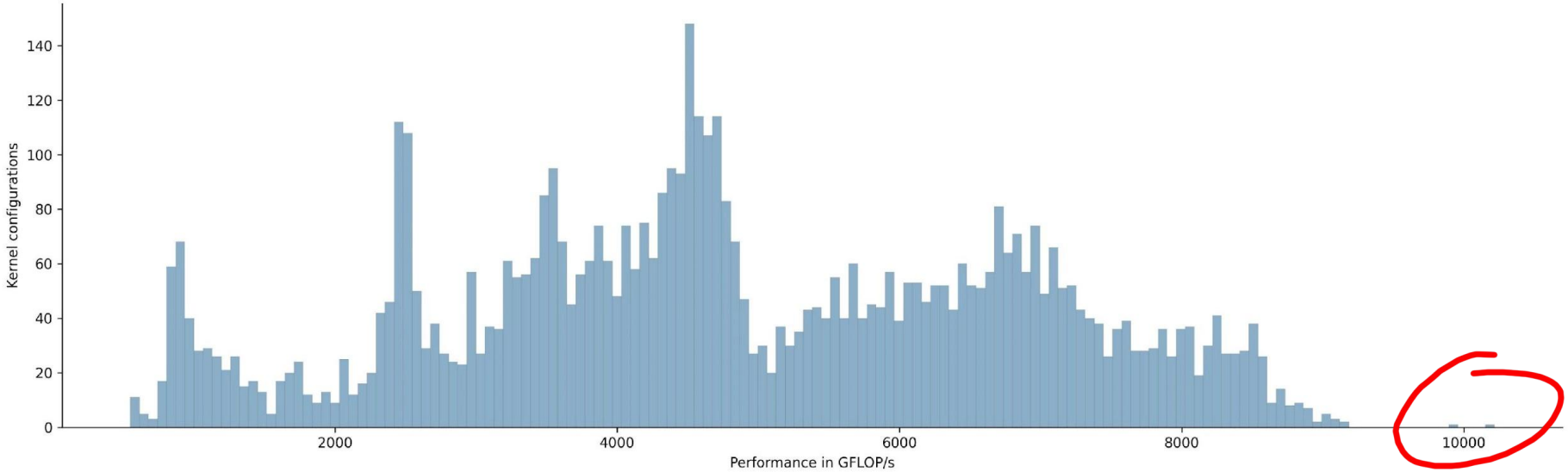
Today, performance depends on hardware…



Hardware and Software Optimizations for Accelerating Deep Neural Networks: Survey of Current Trends, Challenges, and the Road Ahead
Capra et al. 2020 IEEE Access

… and on the configuration of the software!

Different configurations of a convolution program on an Nvidia A100 GPU

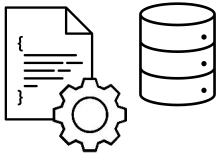


Solution: Autotuning of GPU kernels!

Tuning script (python)

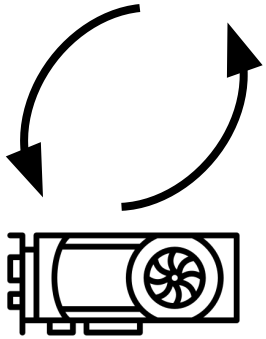
1

Kernel Data Parameters

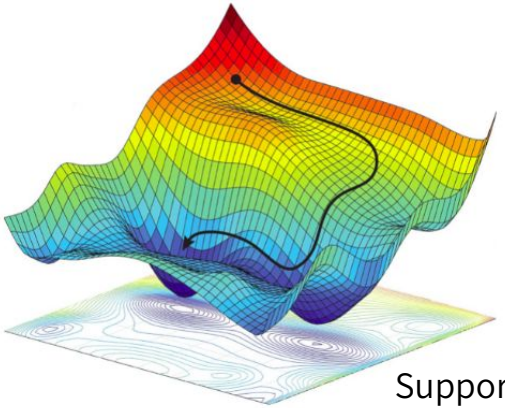


Kernel Tuner

Compile & benchmark



2



Supports many optimization strategies to find the best solution!

3

Optimal parameter values

Kernel Tuner

A tool for automatic performance tuning of GPU kernels

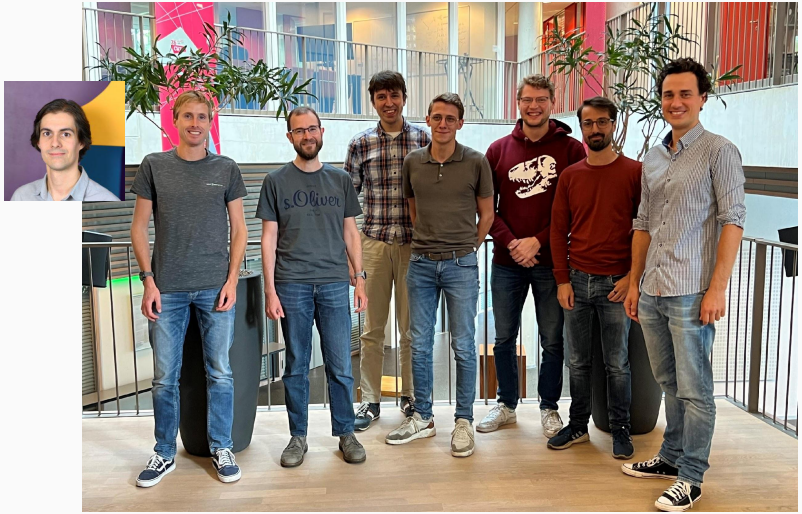
Developed open source since 2016

Used by 10+ universities & organizations

Funded by several NL and EU projects

Supports:

- CUDA, HIP, OpenCL, C++, Fortran, OpenACC
- 20+ search optimization algorithms
- Energy measurement of GPU kernels
- Many different use cases



netherlands eScience center · CWI · ASTRON · Universiteit Leiden The Netherlands

https://github.com/KernelTuner/kernel_tuner

Kernel Dashboard

https://github.com/KernelTuner/dashboard

Kernel Launcher

Extracts & integrates
auto-tuned kernels
into C++ applications

cudawrappers

Portable API for
resource management
on NVIDIA & AMD GPUs

Think about energy use!

Energy cost of supercomputers

Frontier:

#2 in TOP500 list

20 Megawatt continuously

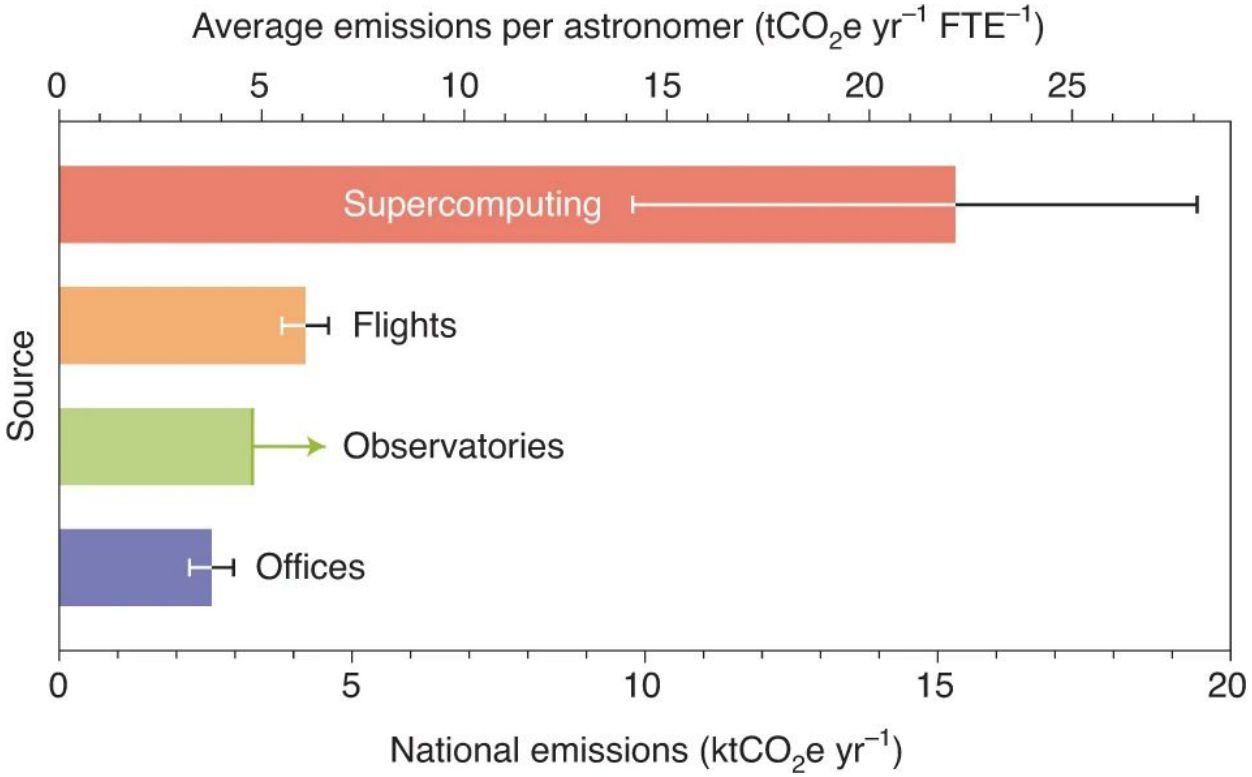
$40 million annual electricity bill

100,000 metric tons of CO2 annually

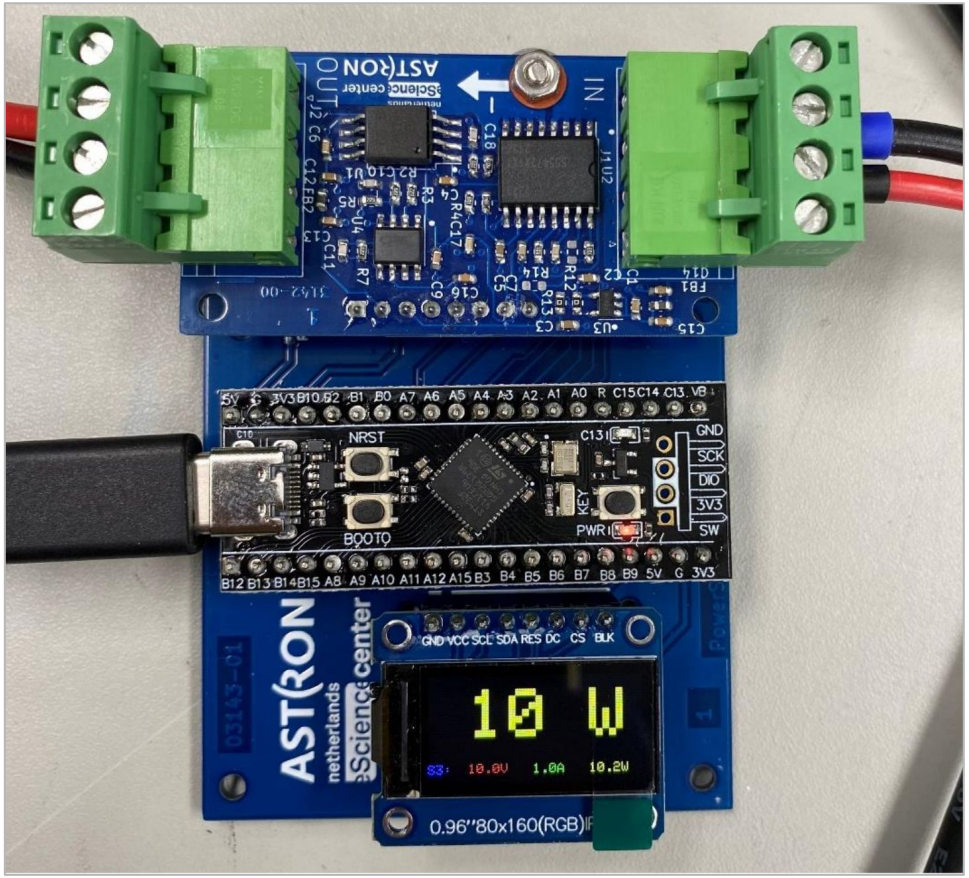
~20,000 cars on the road for a year in US

Efficient Computation through Tuned Approximation
David Keyes, SIAG/SC Supercomputing Spotlights 2022
M. Stachowski, A. Fiebig, and T. Rauber, Journal of Supercomputing, 2020.

The carbon footprint of an astronomer



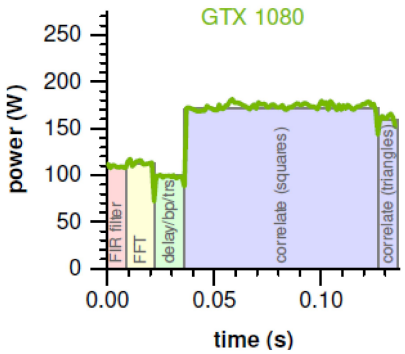
The imperative to reduce carbon emissions in astronomy
A.R.H Stevens et al. Nature Astronomy 2020



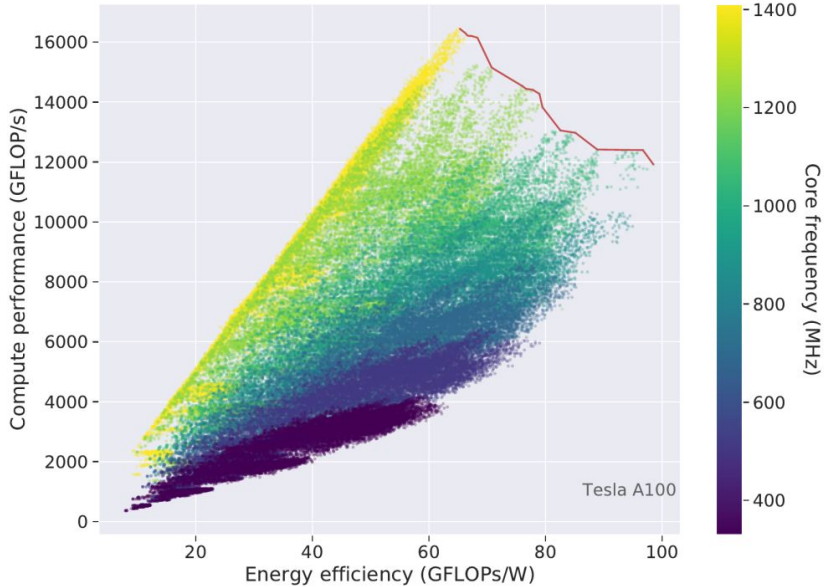
source: https://casper.astro.berkeley.edu/workshop2023/agenda/presentations/day3/11_SV.pdf

powersensor3

ASTRON



Kernel Tuner



Going green: optimizing GPUs for energy efficiency through model-steered auto-tuning
https://doi.org/10.1109/PMBS56514.2022.00010

Less precision for more performance?

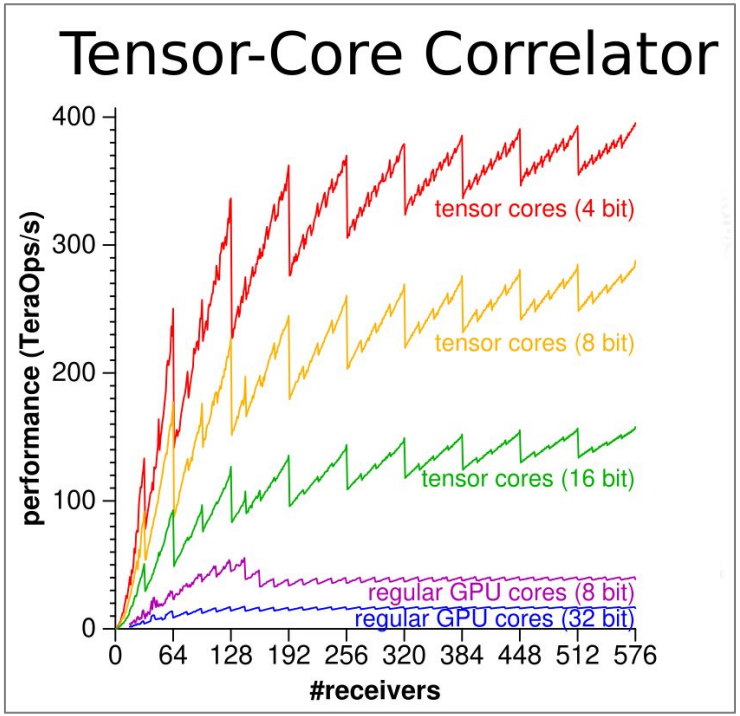


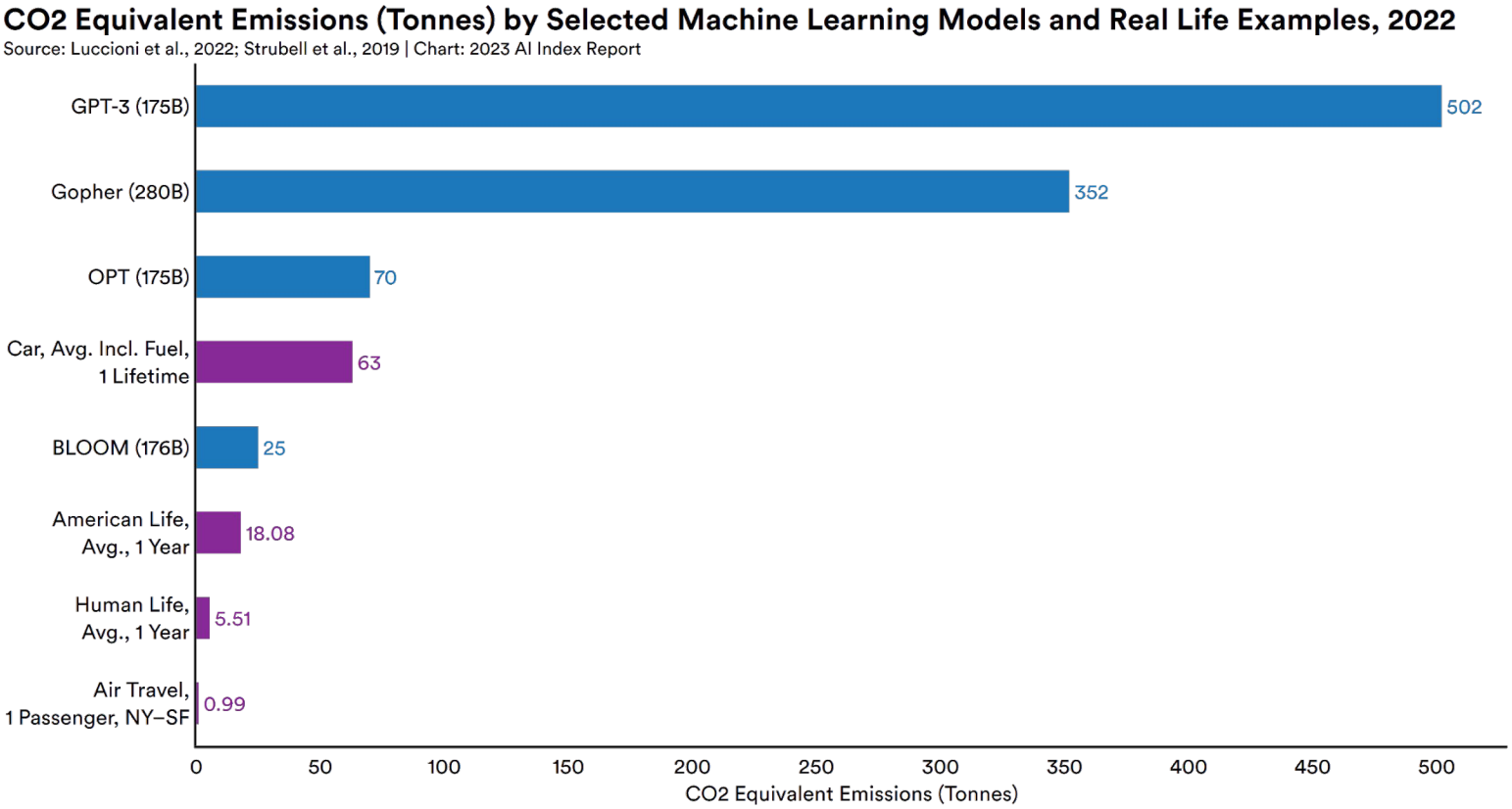
Image: John Romein / ASTRON

Kernel Float

C++ data types for mixed-precision GPU kernel programming

AI is the new HPC ?

LLM Training emissions



Source: Stanford AI Index report 2023

Medium Search Write Sign up Sign in

Netherlands eScience Cen...

Benchmarking LLM fine-tuning on different HPC systems

Flavio Hafner Follow 5 min read · Apr 1, 2025

5

Authors: Flavio Hafner (Netherlands eScience center), Mattie Niznik (Princeton Research Computing), Malte Lüken (Netherlands eScience center), Alessandra Maranca (Princeton University), Matthew Salganik (Princeton University).

We have developed a benchmark that compares the compute performance of fine-tuning LLMs on multiple high-performance computing (HPC) systems, including systems designed for working with sensitive data. In this blog post, we introduce the benchmark, describe the lessons learned developing it and make it open-source so that it can be used and improved by others.

Photo by Nana Dua on Unsplash

Our use case

Our team researches how Large Language Models (LLMs) can be leveraged to predict social outcomes with large-scale social and survey data. We use the ODISSEI Secure Supercomputer (OSSC), which SURF operates in the Netherlands. The OSSC is a virtual private cluster running on the hardware of Snellius, the Dutch national supercomputer. The OSSC makes it possible to use GPU and CPU nodes to analyze sensitive data from Statistics Netherlands. This makes the OSSC one among very few systems globally that bring the power of national computing clusters to large-scale social data.

https://blog.esciencecenter.nl/benchmarking-llm-fine-tuning-on-different-hpc-systems-0ed7efcac646

ASTRON

Self-learning machines hunt for explosions in the Universe

The National Science Agenda has awarded a 5 million euro grant to CORTEX – the Center for Optimal, Real-Time Machine Studies of the Explosive Universe. The CORTEX consortium of 13 partners from academia, industry and society will make self-learning machines faster: to figure out how massive cosmic explosions work, and to innovate systems that benefit our society.

Published by the editorial team, 11 June 2019

Machine learning has rapidly become an integral part of our lives. It is now commonly used for speech recognition and information retrieval. This is also true in science, for detecting patterns in nature and the Universe. But the need is growing rapidly for such machines to respond quickly, for example in self-driving cars and for responsive manufacturing. On a more fundamental level, self-learning machines help us unveil a dynamical Universe we did not know existed until recently. Bright explosions appear all over the radio and gravitational-wave sky. Many citizens and scientists are curious to understand where these come from.

"In CORTEX we aim to solve these open problems by bridging fundamental research to society", says dr. Joeri van Leeuwen (ASTRON), the project lead. "We can only reach these ambitious goals if academic, applied, public and industry partners work together."

The 5 million euro grant from the Nationale Wetenschapsagenda: Onderzoek op Routes door

https://www.astron.nl/self-learning-machines-hunt-for-explosions-in-the-universe/

AMS Journals JOURNALS BROWSE PUBLISH SUBSCRIBE ABOUT Sign in Sign up Search

Editorial Type: Article

Article Type: Research Article

Extended-Range Arctic Sea Ice Forecast with Convolutional Long Short-Term Memory Networks

Yang Liu, Laurens Bogaardt, Jisk Attema, and Wilco Hazeleger

Online Publication: 05 May 2021

Print Publication: 01 Jun 2021

DOI: https://doi.org/10.1175/MWR-D-20-0113.1

Page(s): 1673–1693

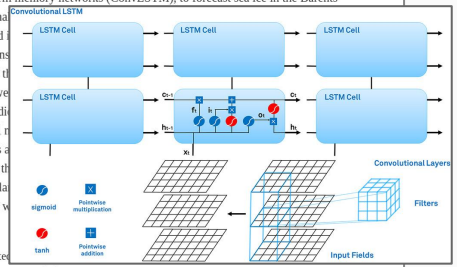
Article History | Download PDF | © Get Permissions

Open access

Abstract/Excerpt | Full Text | PDF | Supplementary Materials

Abstract

Operational Arctic sea ice forecasts are of crucial importance to science and to society in the Arctic region. Currently, statistical and numerical climate models are widely used to generate the Arctic sea ice forecasts at weather time scales. Numerical models require near-real-time input of relevant environmental conditions consistent with the model equations and they are computationally expensive. In this study, we propose a deep learning approach, namely convolutional long short-term memory networks (ConvLSTM), to forecast sea ice in the Barents Sea at weather to subseasona
use of historical records and i
spatial and temporal relations
ConvLSTM is able to learn t
concentration skillfully at we
between predictors and predi
persistence, and a statistical r
with different climate fields a
predictors on the quality of th
budget components have a la
This method is a promising w
future.



Supplemental information relate
https://doi.org/10.1175/MWR-D-20-0113.s1.

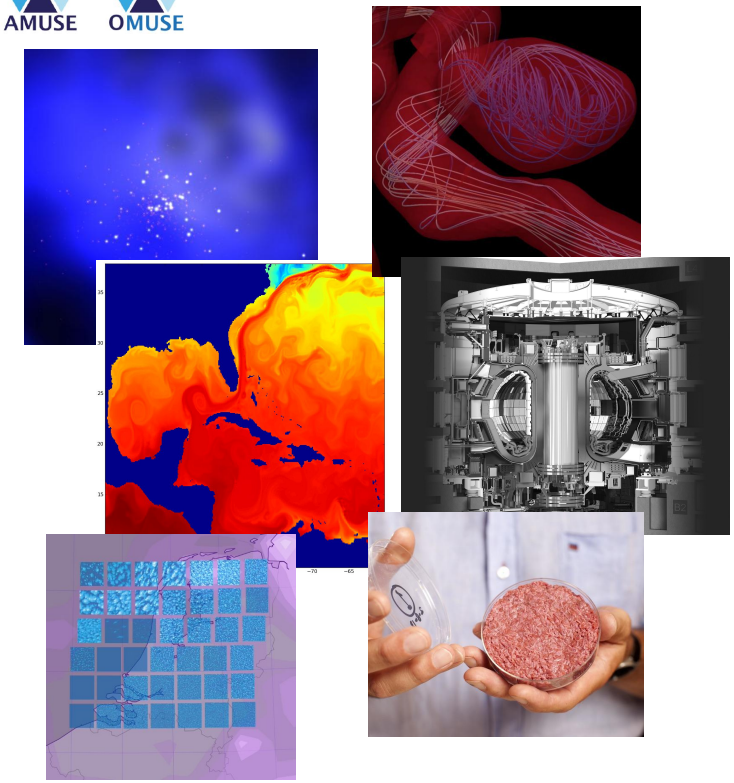
Denotes content that is immediately available upon publication as open access.

https://doi.org/10.1175/MWR-D-20-0113.1

Interesting Crossovers

AMUSE OMUSE

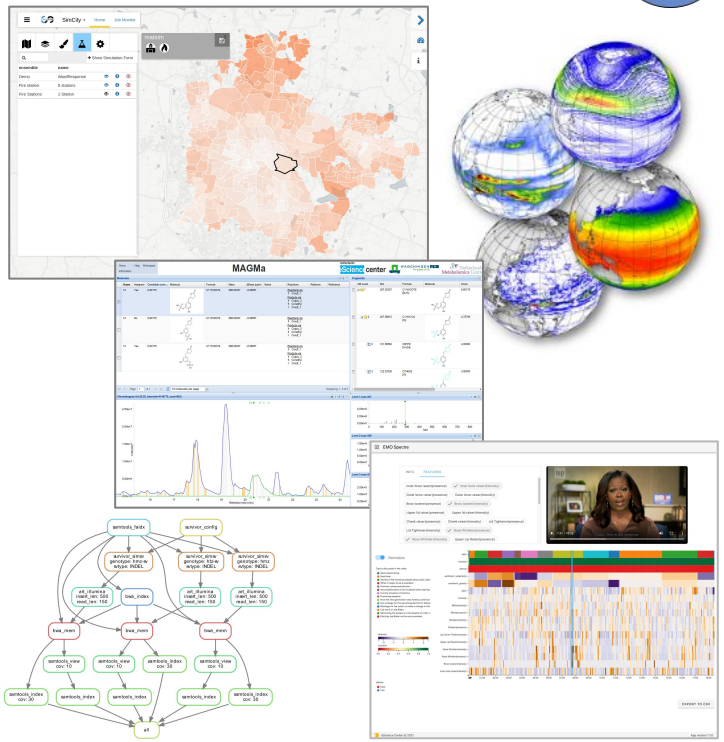
MUSCLE 3



multi-model multi-scale simulations

XENON

Grand Application Tool with JAVA



deployment & workflows

What’s next for the roaring 20’s?

HPC will get bigger … and smaller

The role of autotuning will only get bigger
(because hardware will diversify further)

AI will grow up

(scientific AI, physics informed AI, explainable AI, …)

Digital twins are the new hype

(multi-model + data assimilation + visualization)

FORTRAN will still be here!