

New directions in computing



Advanced Graduate Lectures on Practical Tools, Applications and Techniques in HEP Part-I

Jun 15 – 19, 2026
RAL, Visitors Centre

Brij Kishor Jashal
Rutherford Appleton Laboratory, Oxford

About me



Brij Kishor Jashal

Staff Scientist

Rutherford Appleton Laboratory, Oxford, UK

My research interests:

- ATLAS experiment at CERN (Past CMS and LHCb experiments)
- Everything that I don't know
- Intersection of experimental particle physics, software and computing
- HEP software, Distributed computing, Heterogeneous architectures, High level triggers, Dark sector physics,
- what AI can do for physics



Overview:

Part-I

- LHC Computing and scale
- Landscape of research software projects .
- Advancements in platforms and architectures.
- Languages and software engineering.

Break: Questions and discussion

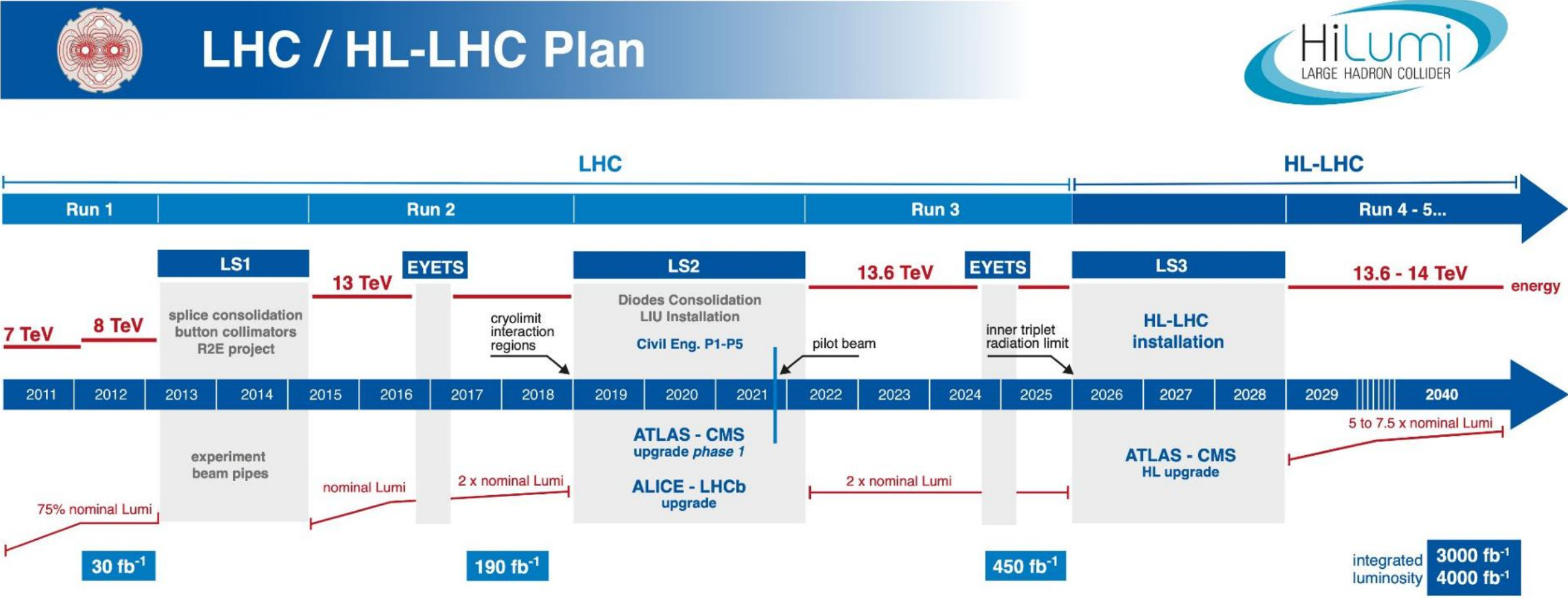
Part-II

- Triggers and Real-Time analysis (online)
- GPU programming.
- AI and future

End: Questions and discussion

LHC Computing and scale

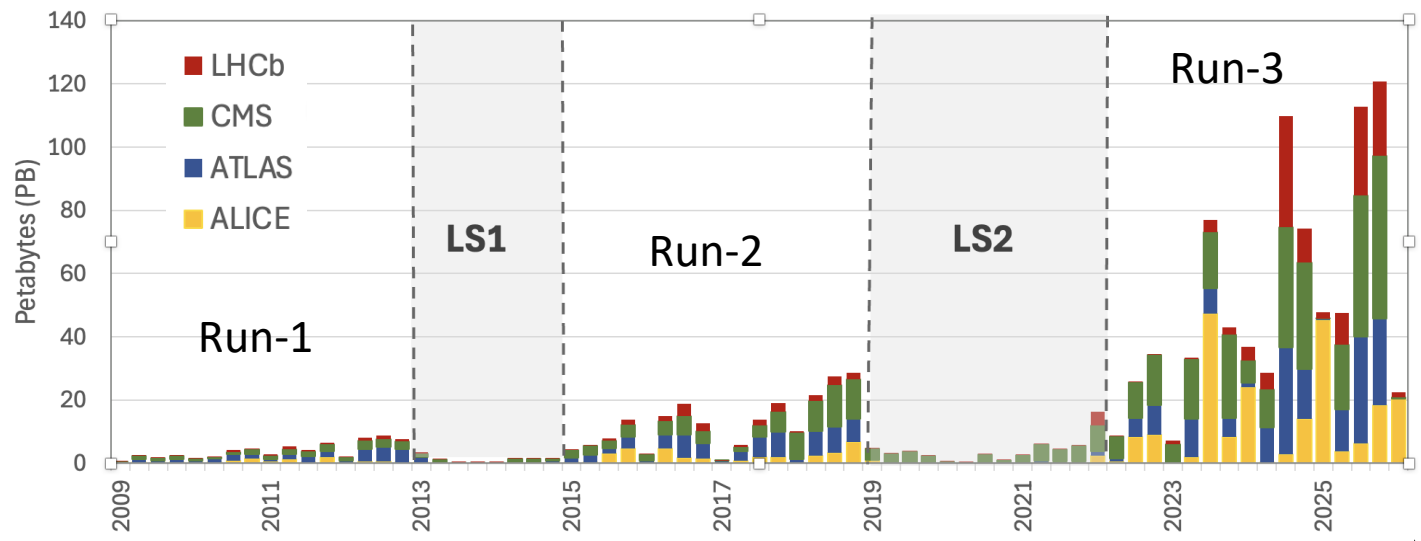
LHC Computing and scale



Experiment designs start far before data taking. CMS was formed in 1992 (more than 30 years ago!), expects to run through 2040 with continued data analysis beyond that date.

(HL-)LHC computing: scale

LHC data written to CERN archive (PB/quarter)



For every LHC run, the volume of collected data increases very steeply

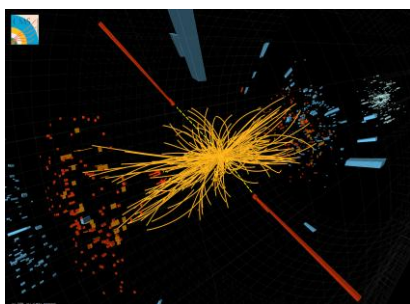
We collected ~ 20% of the expected data

HL-LHC: enormous challenge for storage and compute services, databases and networks

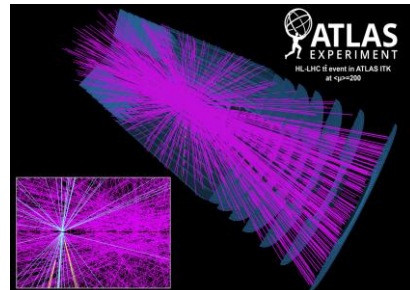


(HL-)LHC computing: scale

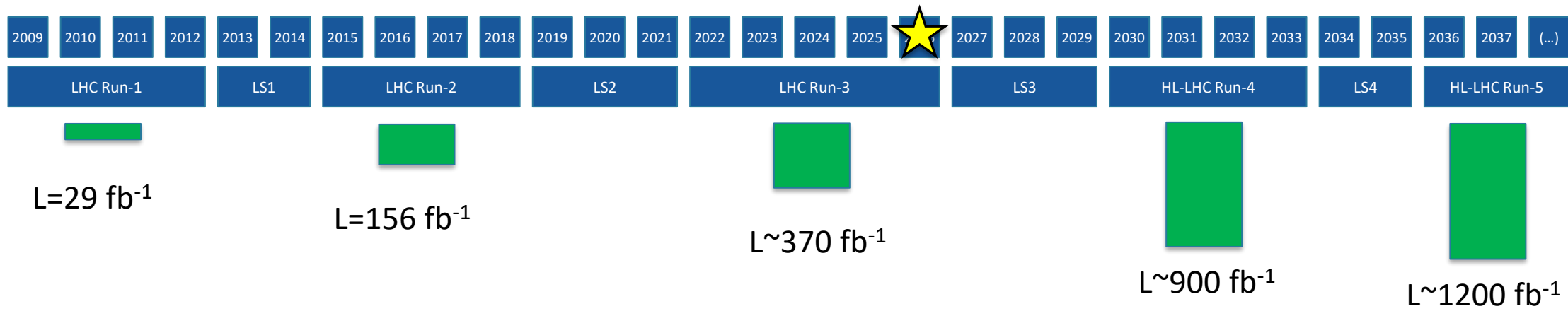
Run-1 Higgs candidate



HL-LHC simulated event



Increase in data volume, but also in **event complexity**
Major innovation needed in the experiments' software



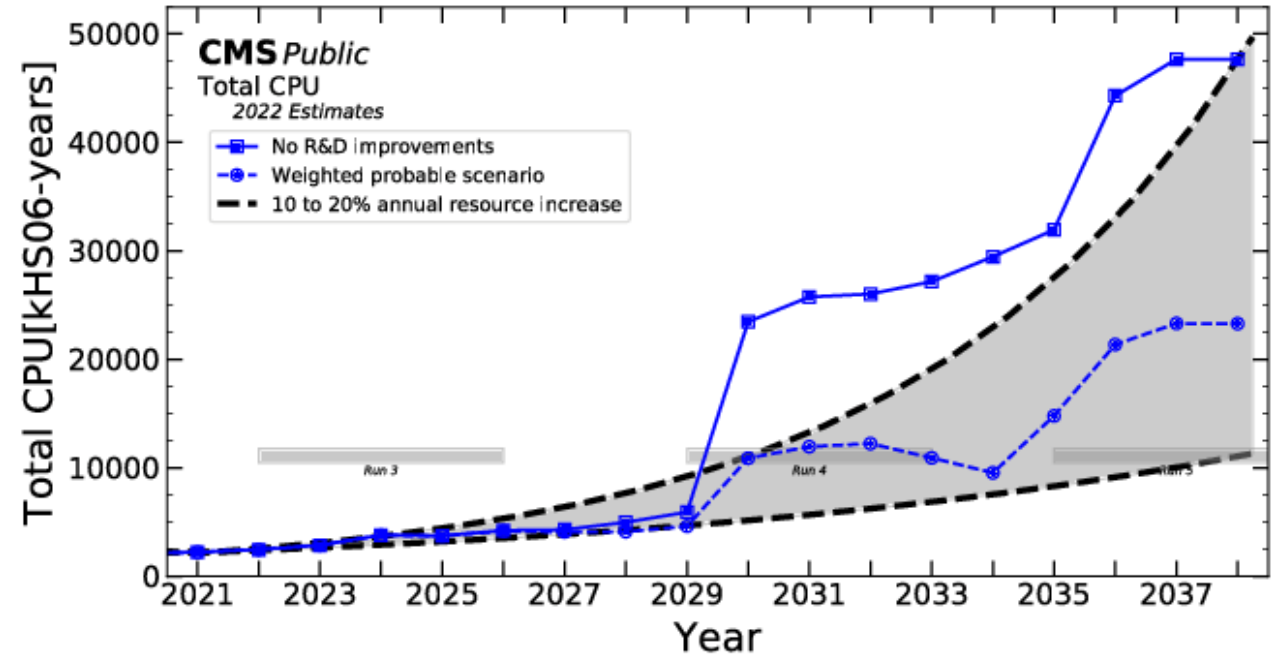
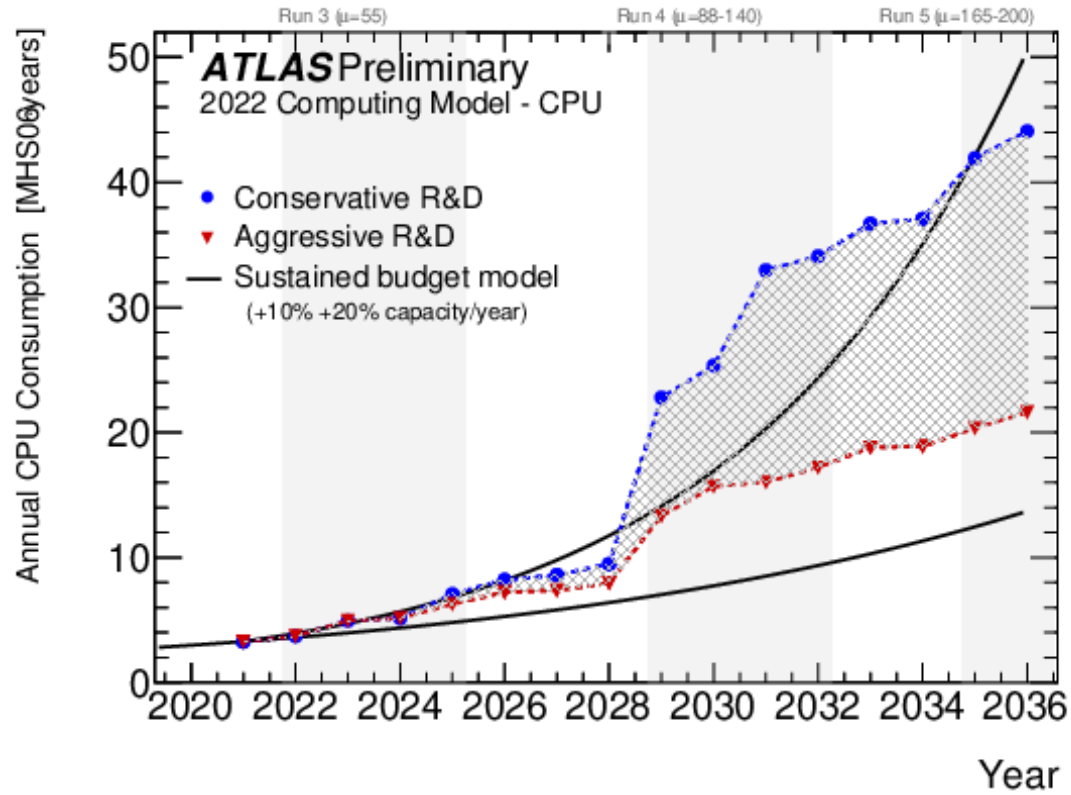
What Does fb^{-1} Mean, and Why Is a Barn Called a Barn?

- A barn is a particle-physics unit of cross-sectional area: $1 \text{ barn} = 10^{-28} \text{ m}^2$, roughly the scale of a heavy atomic nucleus.
- The name came from Manhattan Project slang in the 1940s: a uranium nucleus looked "as big as a barn door" to slow neutrons, so physicists jokingly called that large nuclear target area a barn. Once the barn stuck, the whole family followed standard prefixes: mb, microbarn, nb, pb, fb, and ab for increasingly smaller cross sections.
- **fb^{-1} means inverse femtobarn. It is a unit of integrated luminosity: how much collision data an experiment has accumulated, not the size of one event.**
- A femtobarn is extremely small:
$$1 \text{ fb} = 10^{-15} \text{ barn} = 10^{-43} \text{ m}^2$$
- The inverse, fb^{-1} , means "per femtobarn" and acts like the size of the dataset.
- Event yield is estimated by: $N = L_{int} \times \sigma$
 - Here: L_{int} = integrated luminosity in fb^{-1}
 σ = process cross section in fb
 - Example:
if a process has $\sigma = 10 \text{ fb}$ and the experiment collected 150 fb^{-1} , then $N \approx 150 \times 10 = 1500$ events

Why it matters: rare processes have tiny cross sections, so experiments need large integrated luminosity

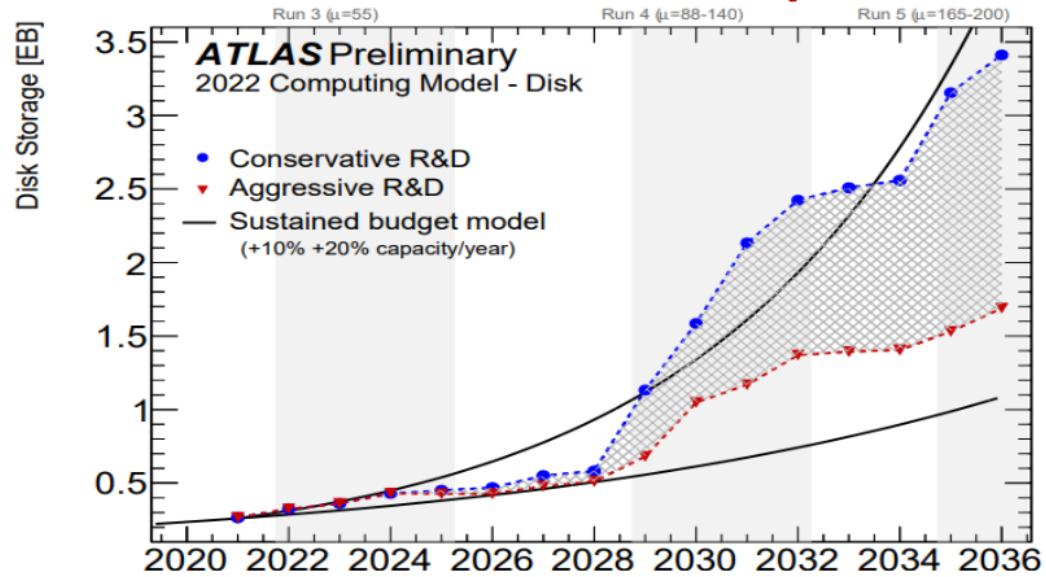
Takeaway: barns measure interaction probability; inverse barns measure dataset size.

(HL-)LHC computing

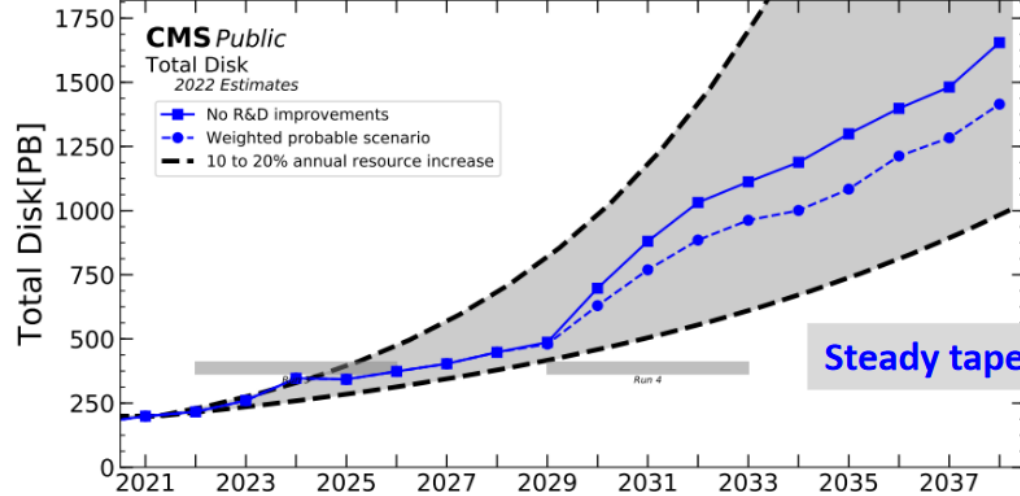
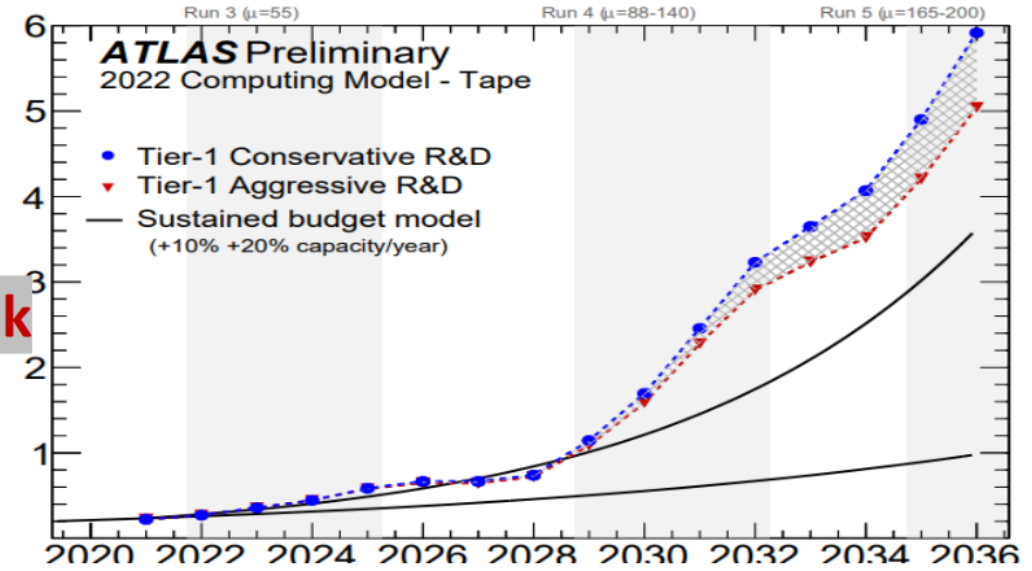


A naive extrapolation from today's computing model and techniques, even after assuming Moore's Law increases in capabilities, is insufficient to meet the expected resource needs for HL-LHC

(HL-)LHC computing:

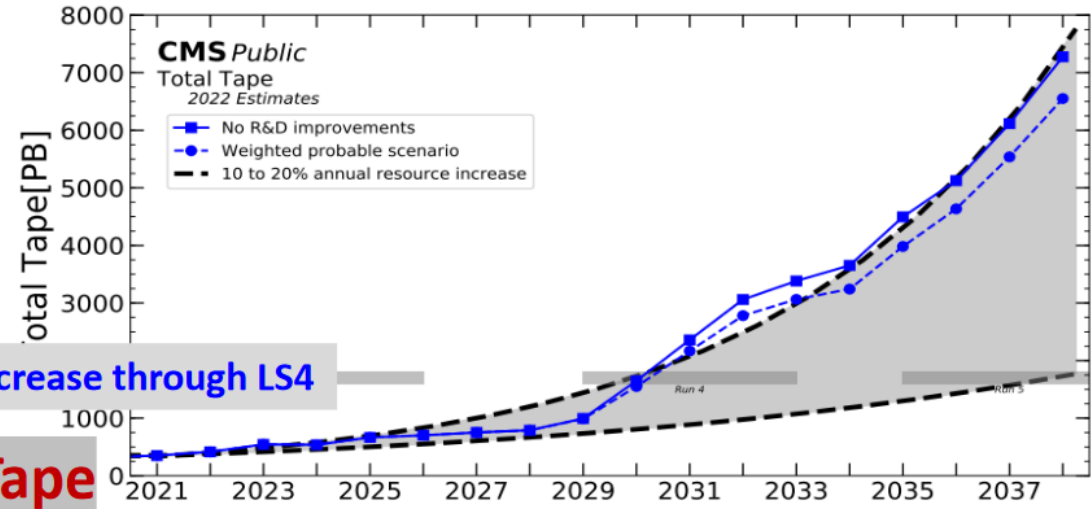


Disk

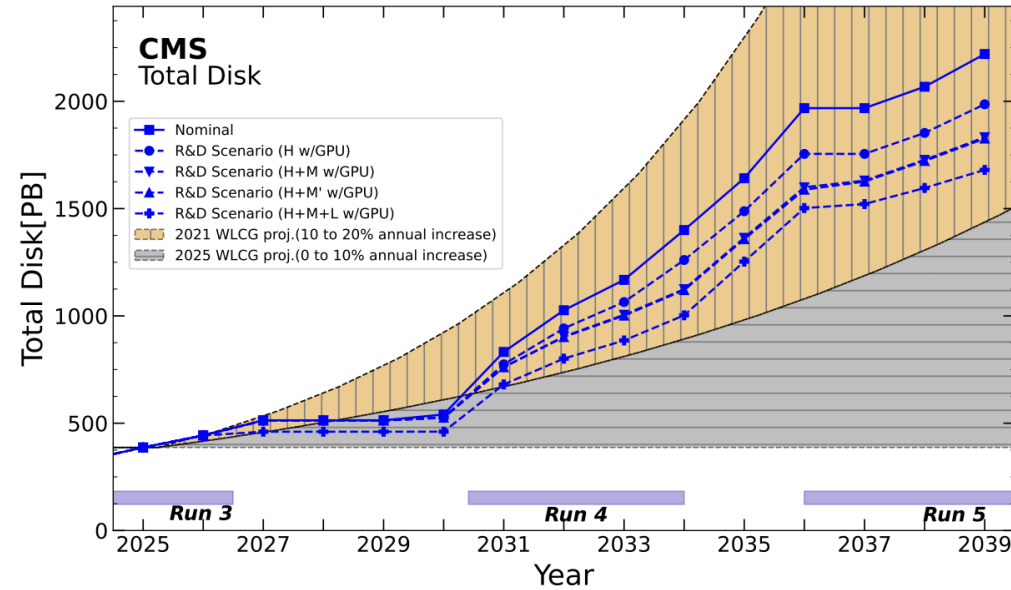
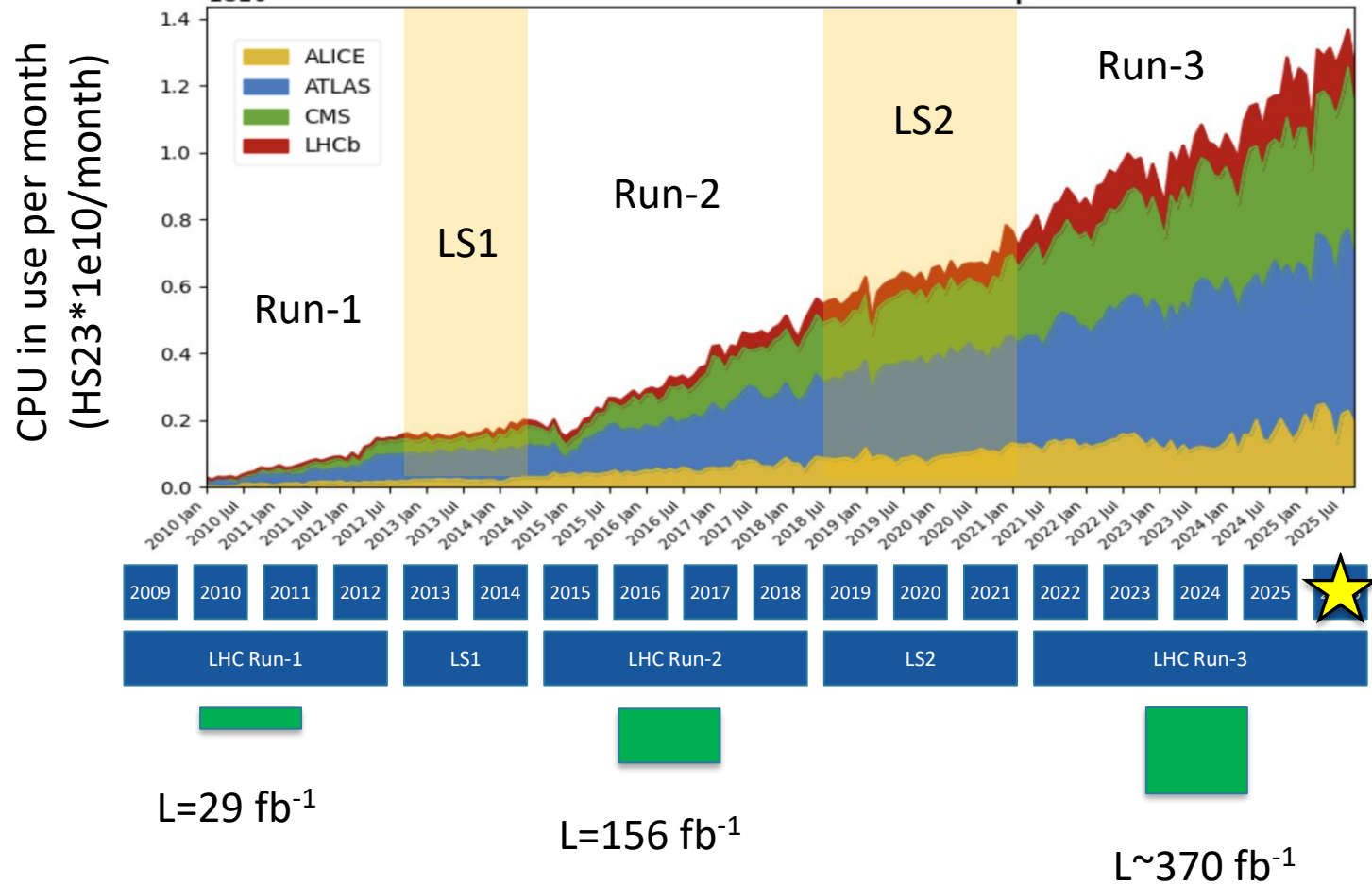


Steady tape increase through LS4

Tape



Evolution and Stability



From the CMS HL-LHC Computing Conceptual Design Report

Landscape of research software projects

Landscape of research software projects

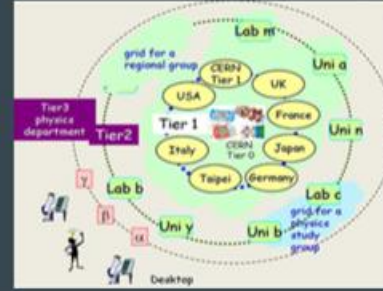
Detector



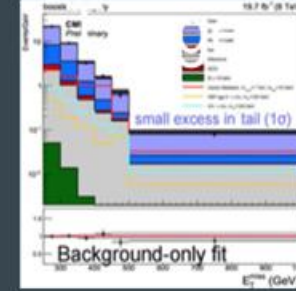
Trigger



Reconstruction





Analysis



Paper



  **Joint WLCG & HSF Workshop 2018**

26–29 Mar 2018
Napoli, Italy
Europe/Zurich timezone



A Roadmap for HEP Software and Computing R&D for the 2020s

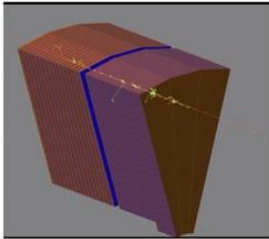
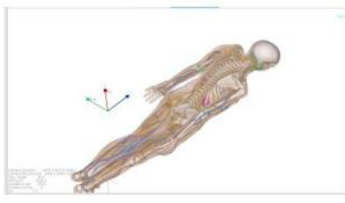
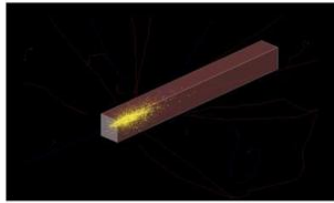
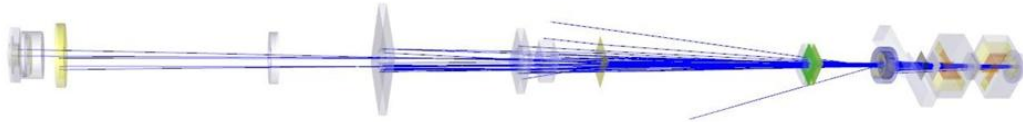
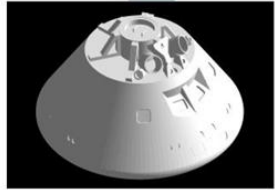
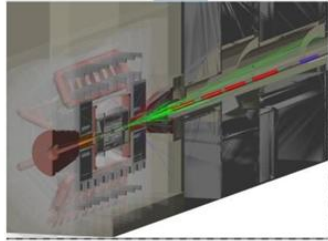
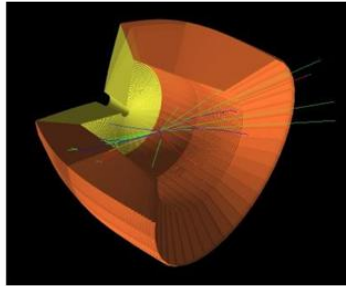
HEP Software Foundation¹

ABSTRACT: Particle physics has an ambitious and broad experimental programme for the coming decades. This programme requires large investments in detector hardware, either to build new facilities and experiments, or to upgrade existing ones. Similarly, it requires commensurate investment in the R&D of software to acquire, manage, process, and analyse the shear amounts of data to be recorded. In planning for the HL-LHC in particular, it is critical that all of the collaborating stakeholders agree on the software goals and priorities, and that the efforts complement each other. In this spirit, this white paper describes the R&D activities required to prepare for this software upgrade.

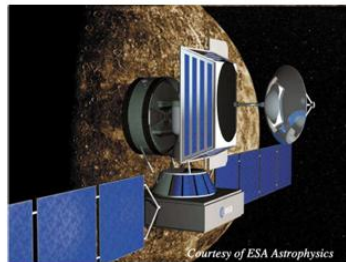
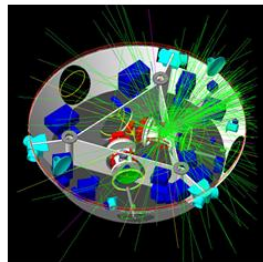
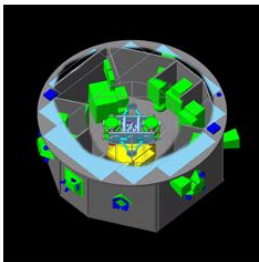
Contents

1	Introduction	2
2	Software and Computing Challenges	5
3	Programme of Work	11
3.1	Physics Generators	11
3.2	Detector Simulation	15
3.3	Software Trigger and Event Reconstruction	23
3.4	Data Analysis and Interpretation	27
3.5	Machine Learning	32
3.6	Data Organisation, Management and Access	36
3.7	Facilities and Distributed Computing	41
3.8	Data-Flow Processing Framework	45
3.9	Conditions Data	48
3.10	Visualisation	51
3.11	Software Development, Deployment, Validation and Verification	53
3.12	Data and Software Preservation	58
3.13	Security	61
4	Training and Careers	65
4.1	Training Challenges	66
4.2	Possible Directions for Training	67
4.3	Career Support and Recognition	68
5	Conclusions	69

Landscape of research software projects



Many applications beyond particle physics



LISA Pathfinder

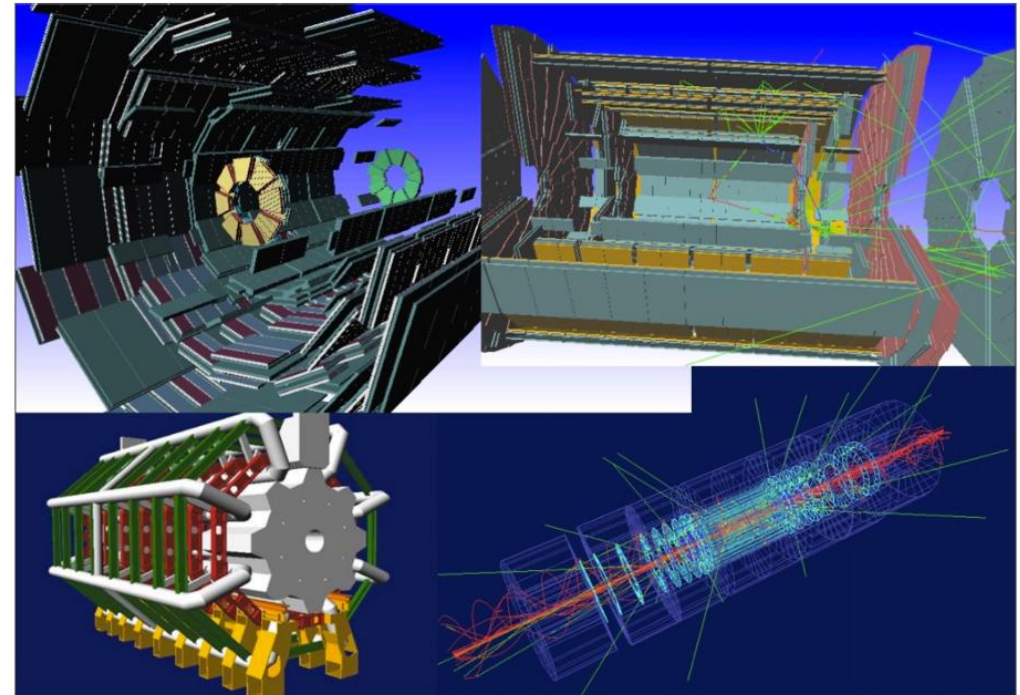
Bepi Colombo

MC Simulation

C++ 1995-



GEANT4
A SIMULATION TOOLKIT



Monte Carlo generators

Aim to generate events as in Nature
 → get average and fluctuations right
 → make random choices, as in nature

Integrals as averages:

$$I = \int_{x_1}^{x_2} f(x) dx = (x_2 - x_1) \langle f(x) \rangle$$

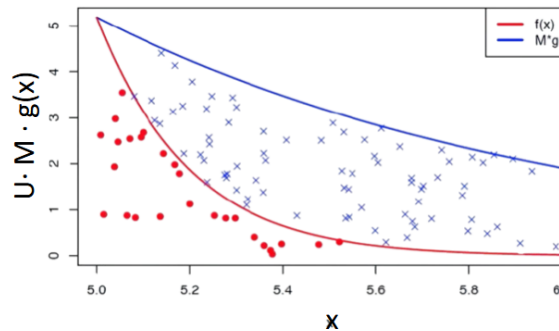
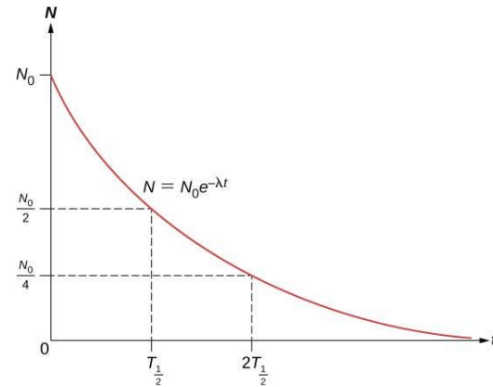
$$I \approx I_N \equiv (x_2 - x_1) \frac{1}{N} \sum_{i=1}^N f(x_i)$$

$$V = (x_2 - x_1) \int_{x_1}^{x_2} [f(x)]^2 dx - \left[\int_{x_1}^{x_2} f(x) dx \right]^2$$

$$I \approx I_N \pm \sqrt{V_N/N} \quad (\text{central limit theorem})$$

Accept-reject method:

- Create an envelope around $f(x)$ → $M \cdot g(x)$
- Accept event x' with probability $f(x')/[M \cdot g(x')]$



Monte Carlo generators

<https://montecarlonet.org/>



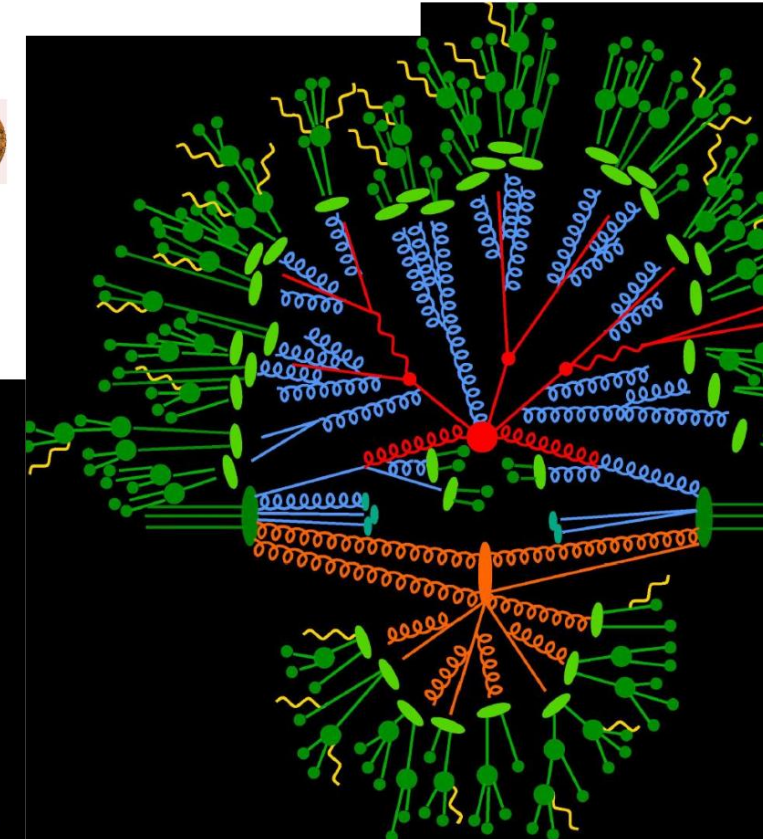
Herwig



Pythia

MadGraph/aMC@NLO

Sherpa



hard scatter: matrix elements from first principles - incoming partons from parton-distribution functions(PDFs)
radiative corrections: resumming logarithms to all orders
multiple parton interactions: additional interactions between proton remnants
hadronisation: going colourless
hadron decays: from excited states to final-state particles
photon radiation: QED corrections

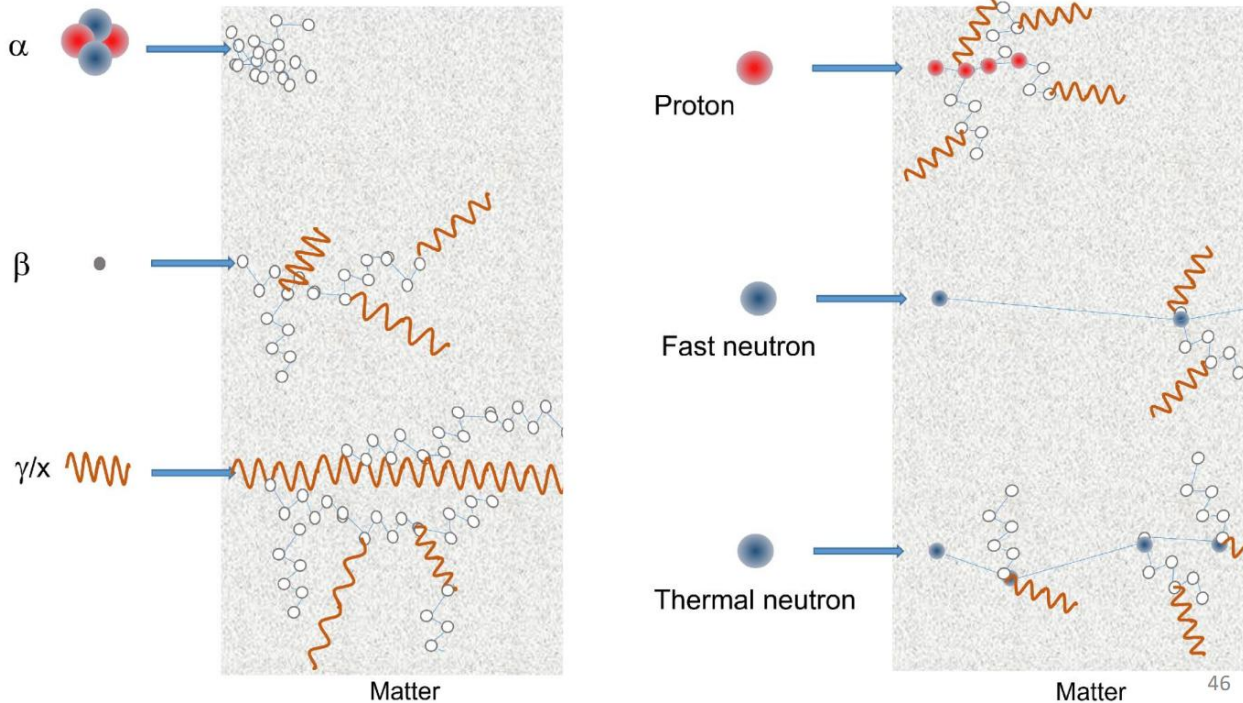
MadGraph: <https://launchpad.net/mg5amcnlo>

Pythia8: <https://pythia.org/>

MonteCarlo simulations

Interaction of charged particles with matter:

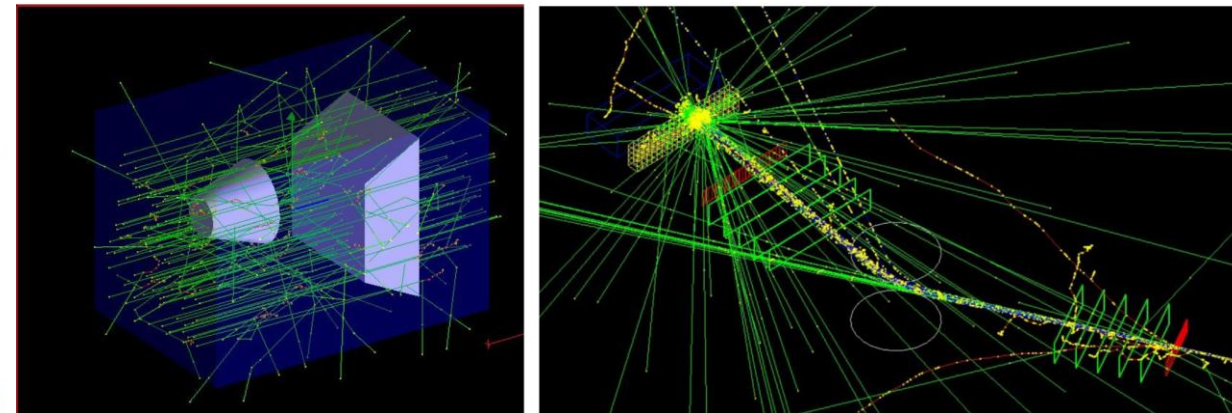
$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi}{m_e c^2} \cdot \frac{nz^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \cdot \left[\ln \left(\frac{2m_e c^2 \beta^2}{I \cdot (1 - \beta^2)} \right) - \beta^2 \right] \quad (\text{Bethe-Bloch})$$



<https://geant4.web.cern.ch/>

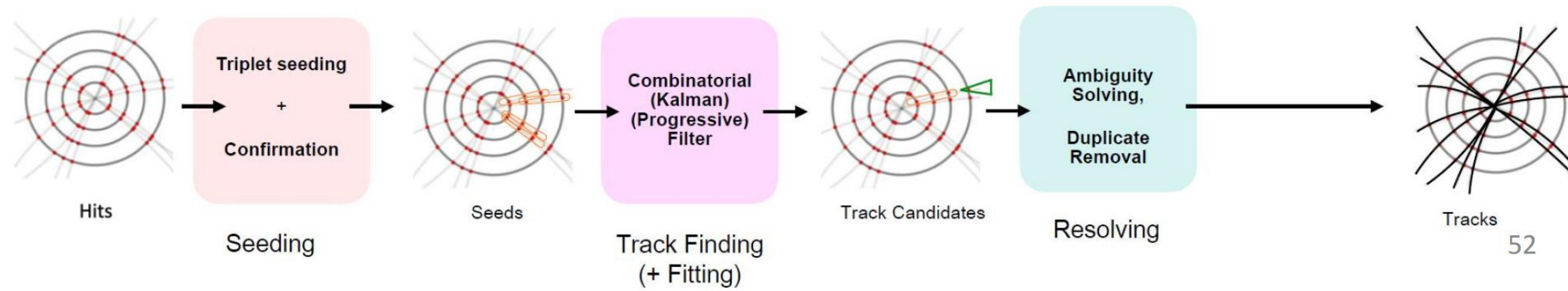
(GEometry ANd Tracking)

C++ simulation toolkit of the passage of particles through matter, using Monte Carlo methods

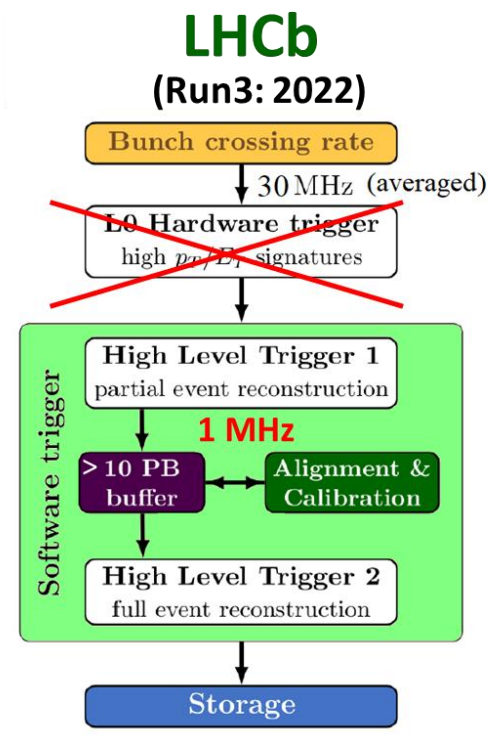
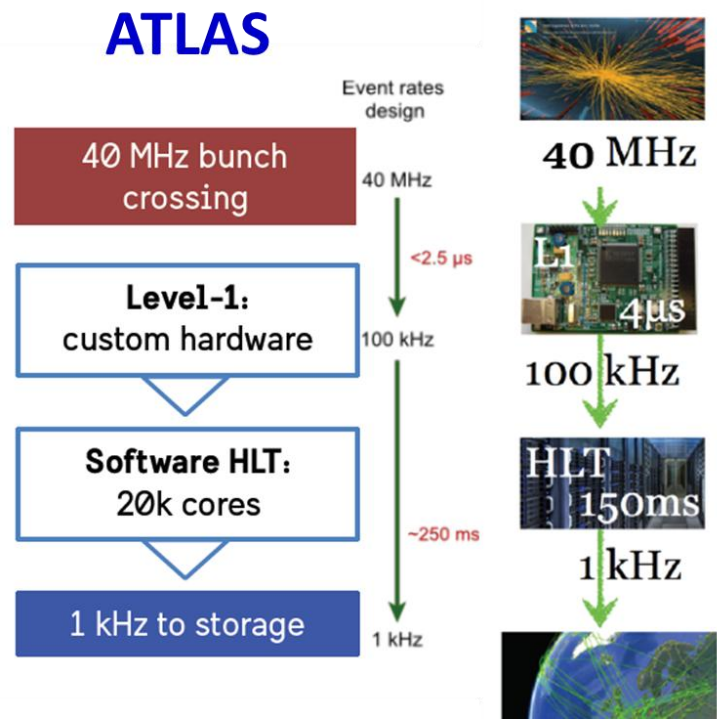


<https://gitlab.cern.ch/geant4/geant4/-/tree/master/examples>

Landscape of research software projects

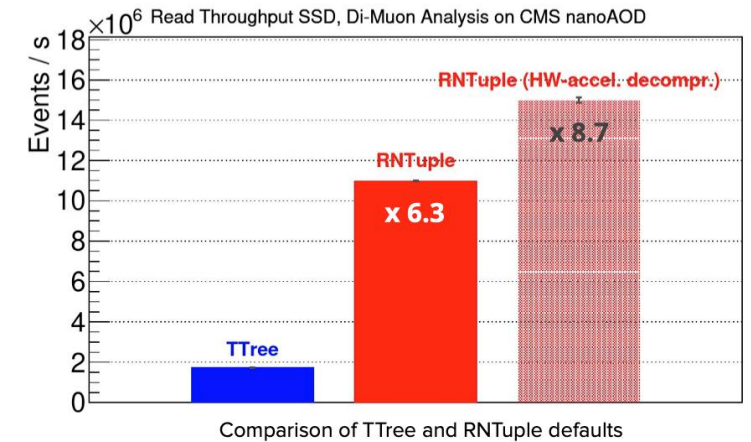
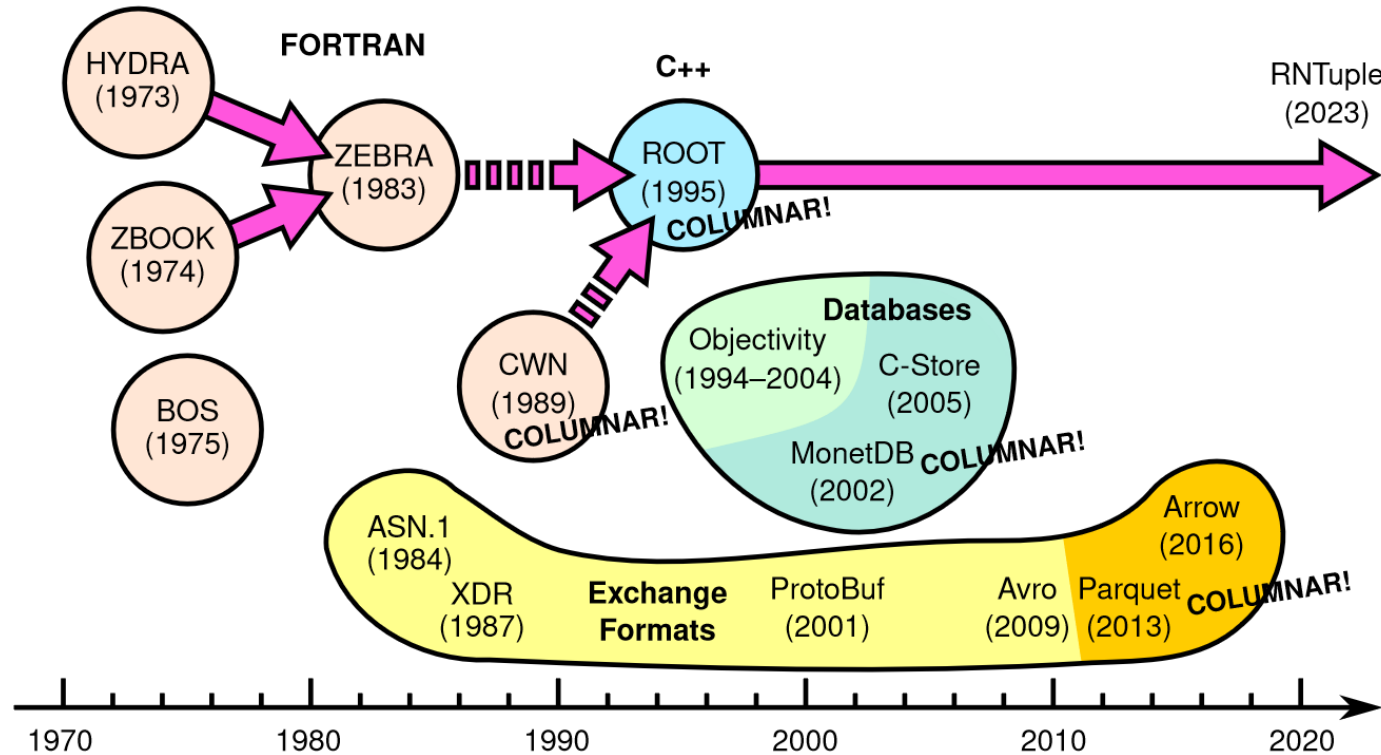


The trigger systems:



Landscape of research software projects

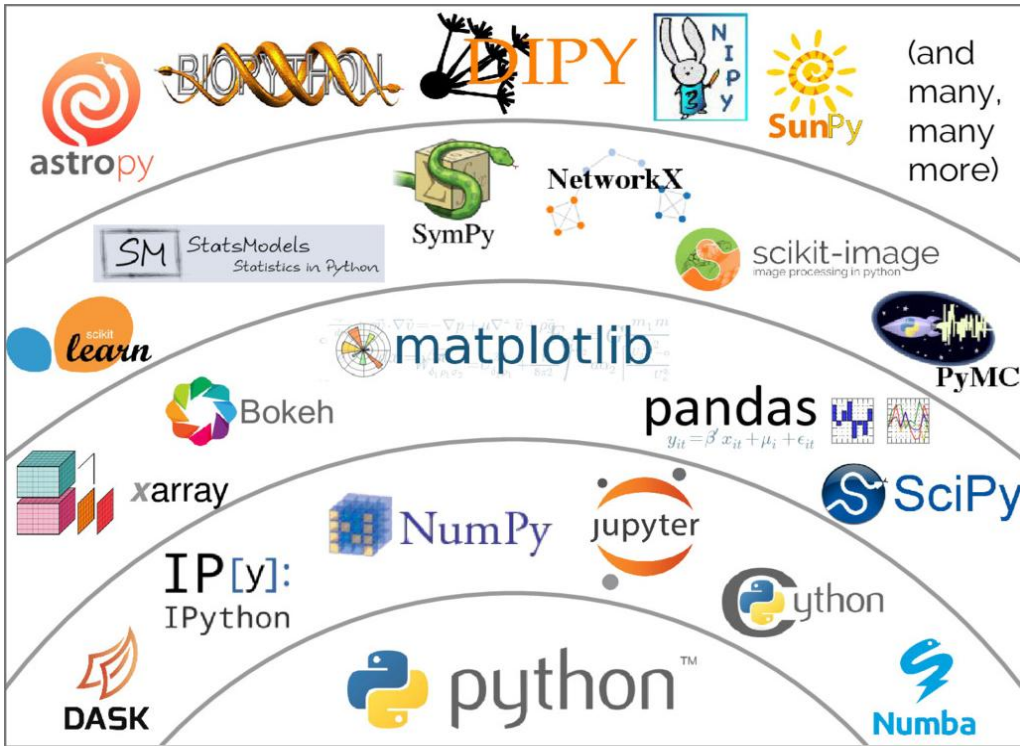
Internal ROOT data format: from TTree to RNTuple



- RNTuple is 10-20% smaller than TTree, resulting in storage saving
- Read throughput improves by x3-x5 with RNTuple

Landscape of research software projects

PyHEP ecosystem

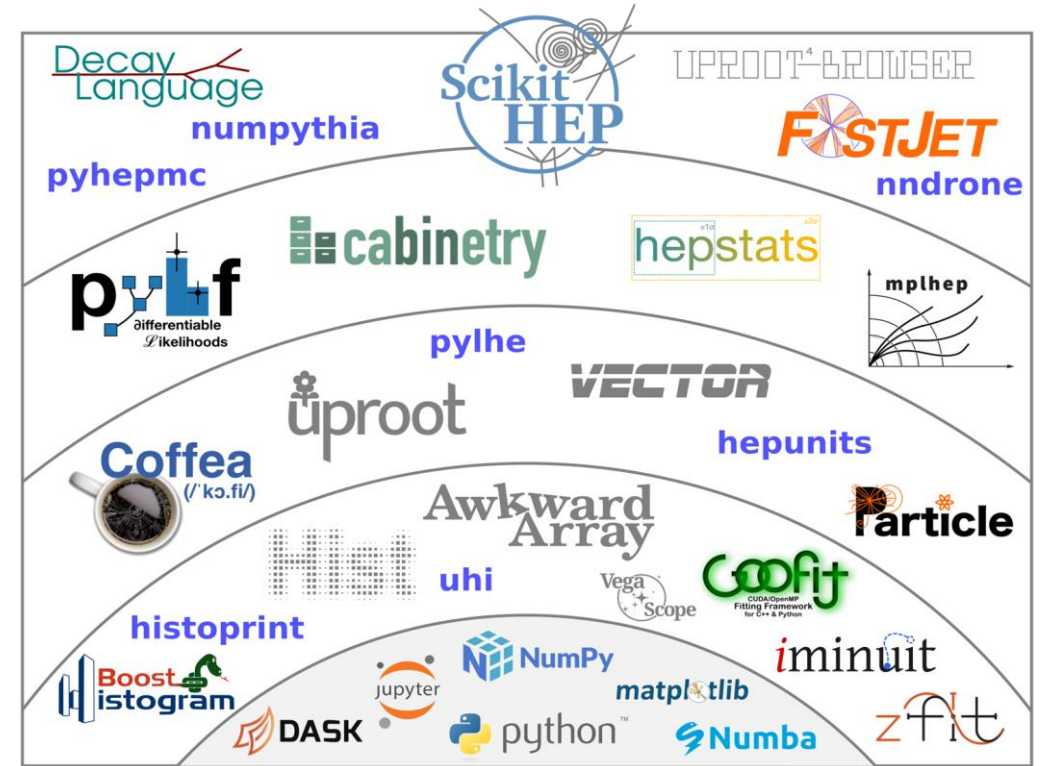


Application Specific

Domain Specific

Technique specific

Foundational



The full HEP ecosystem is of course wider, ROOT being prominent,

Landscape of research software projects

Basics
Awkward(array)
Vector
hepunits

HEP specific libraries and interfaces to HEP libraries
Particle
DecayLanguage
fastjet
pylhe
pyhepmc

Data manipulation & interoperability
uproot
Coffea
uproot-browser
hepconvert
formulate

Fitting and Statistics
iminuit
pyhf
cabinetry
resample
hepstats

Full list at <https://scikit-hep.org/packages>

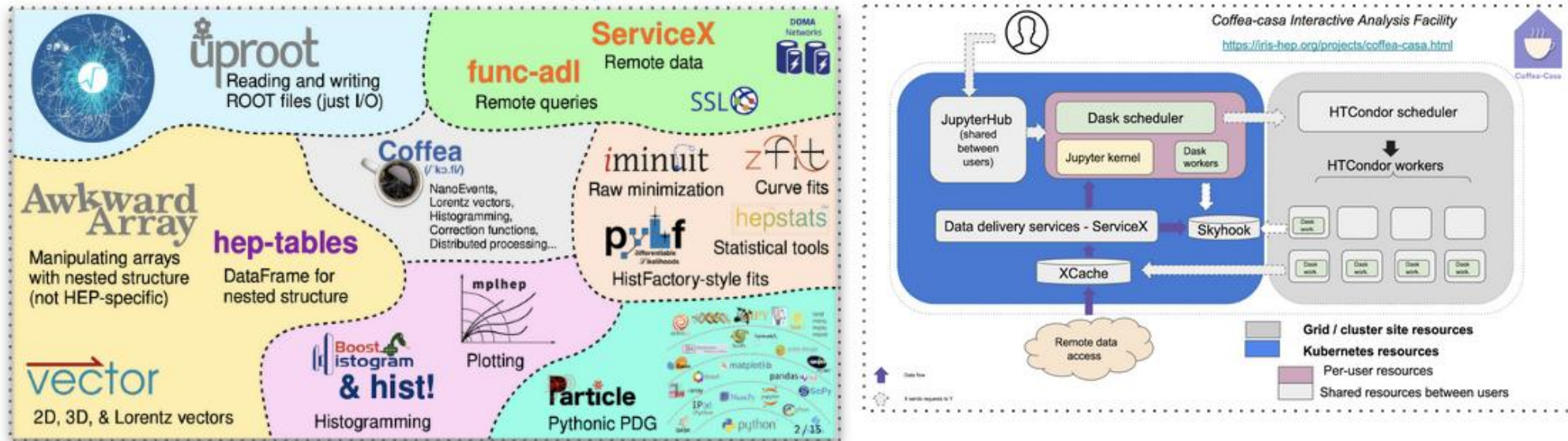
Histogramming
boost-histogram
Hist
histoprint
UHI

Visualisation
mplhep

Analysis Tools

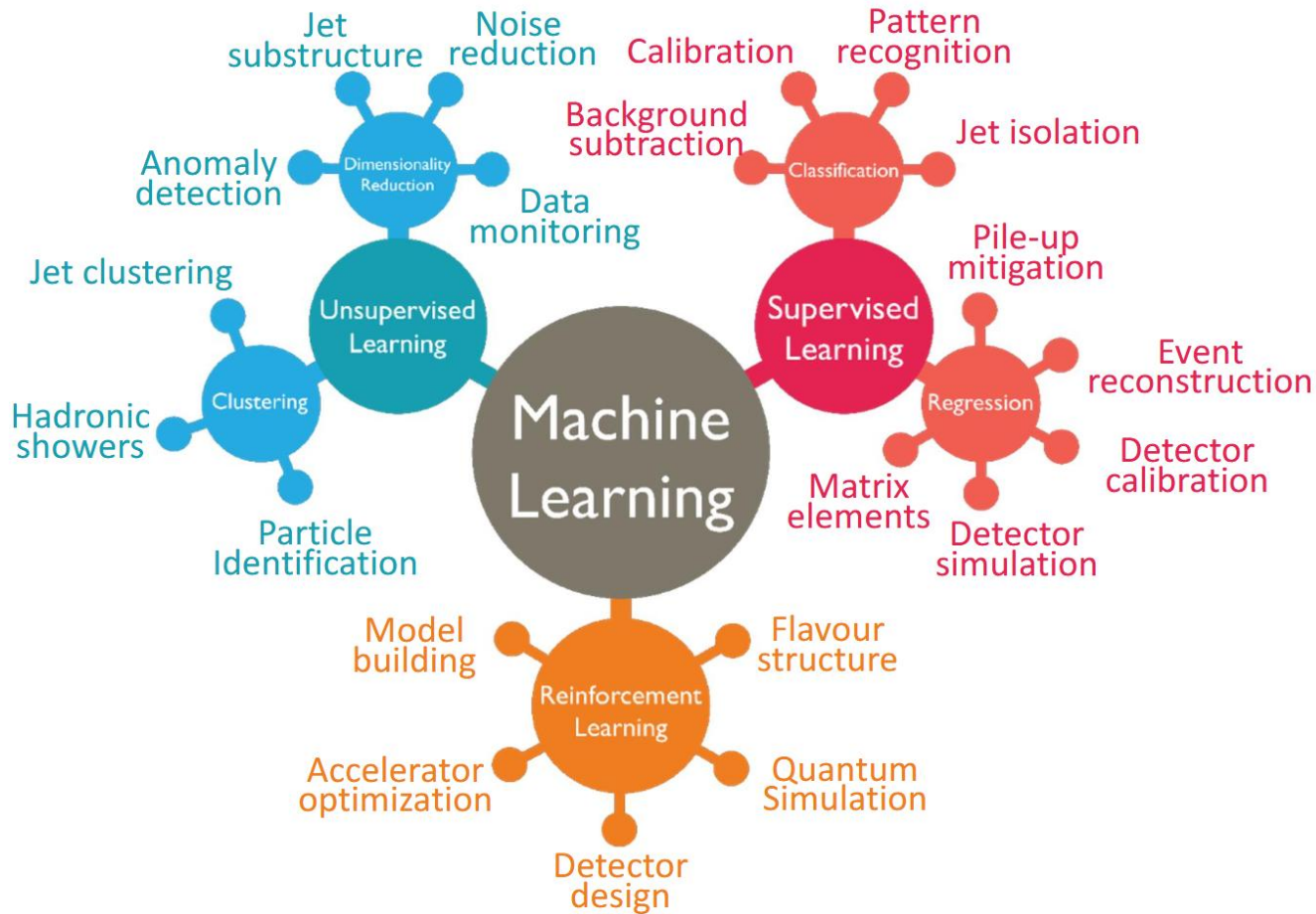
Analysis Facilities

(coffea-casa AF or any other facility matching tech.requirements)

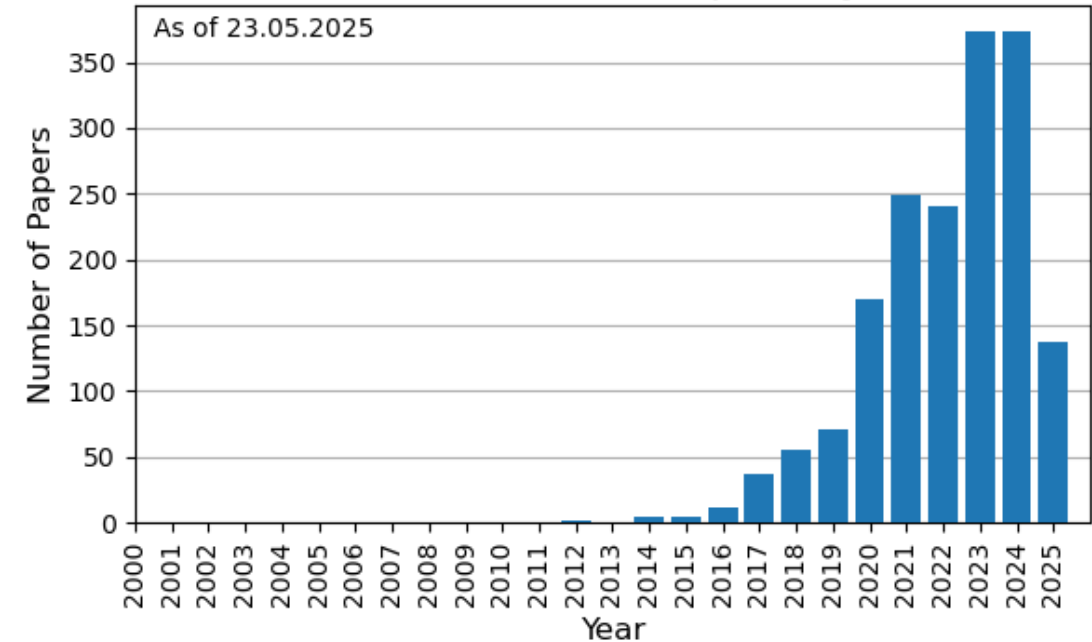


Execution of AGC analysis benchmark

Artificial Intelligence



Number of HEP-ML Papers by Year



Advancements in platforms and architectures.

Advancements in platforms and architectures.

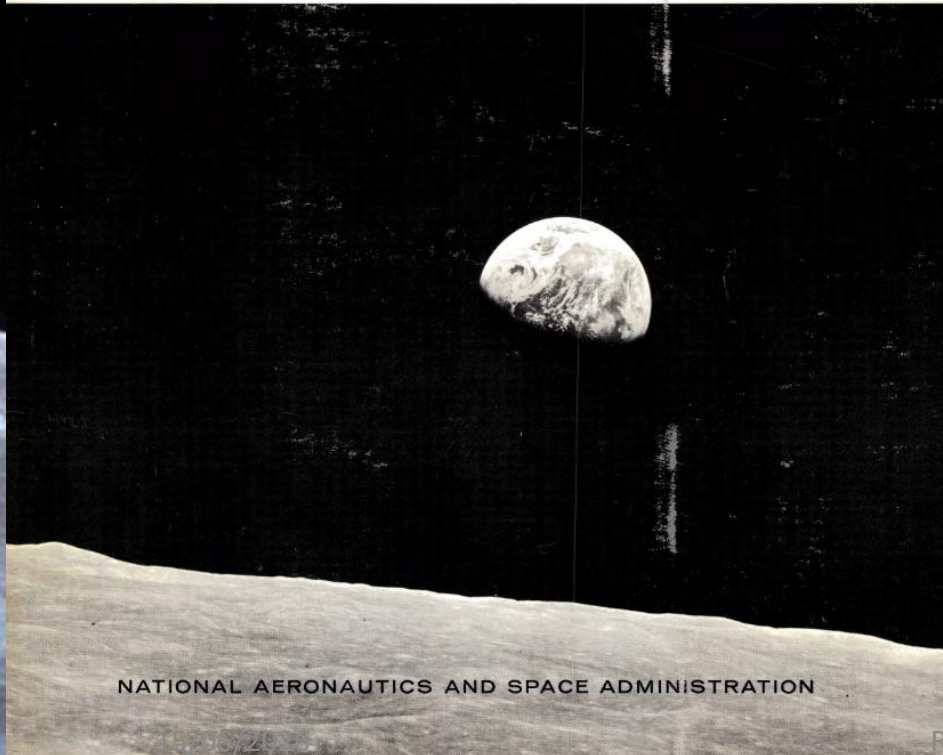
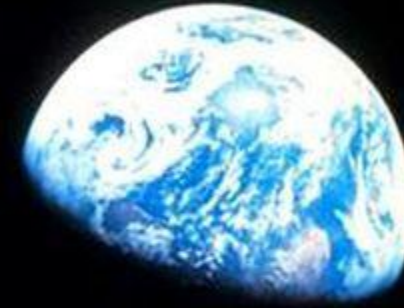
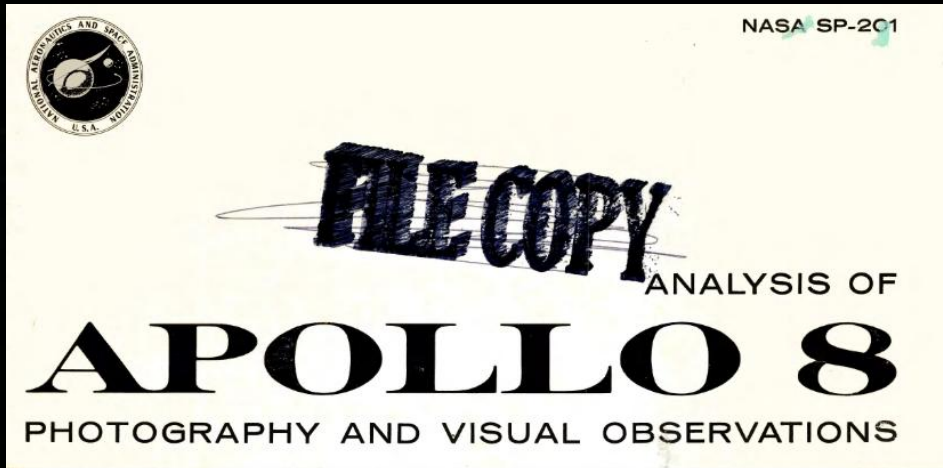
"New directions in science are launched by new tools much more often than by new concepts.

The effect of a concept-driven revolution is to explain old things in new ways.

The effect of a tool-driven revolution is to discover new things that have to be explained."

"Imagined Worlds" (1997) by Freeman Dyson

Advancements in platforms and architectures.

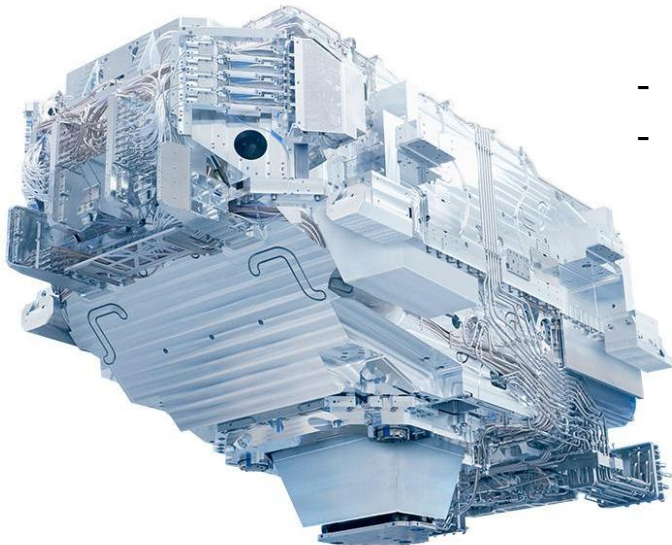
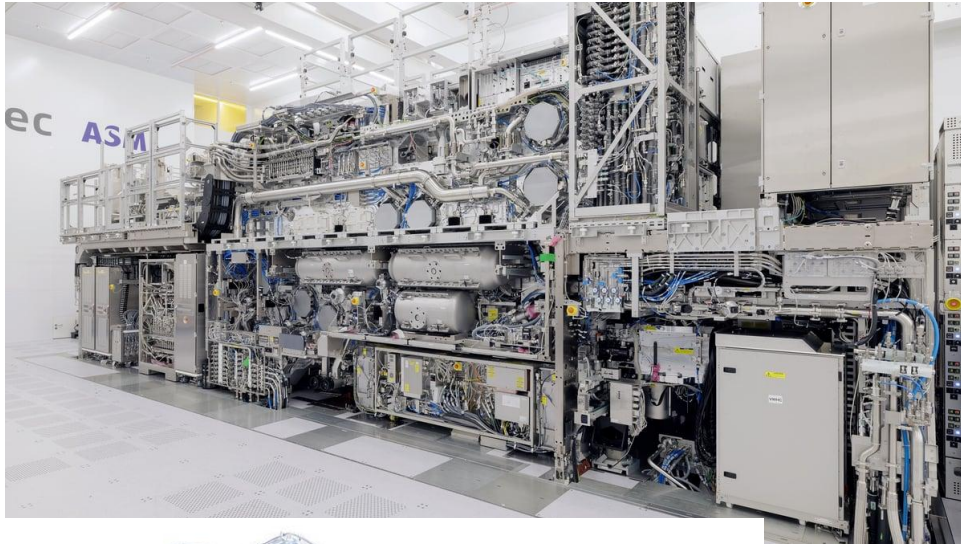


HASSELBLAD ON THE MOON

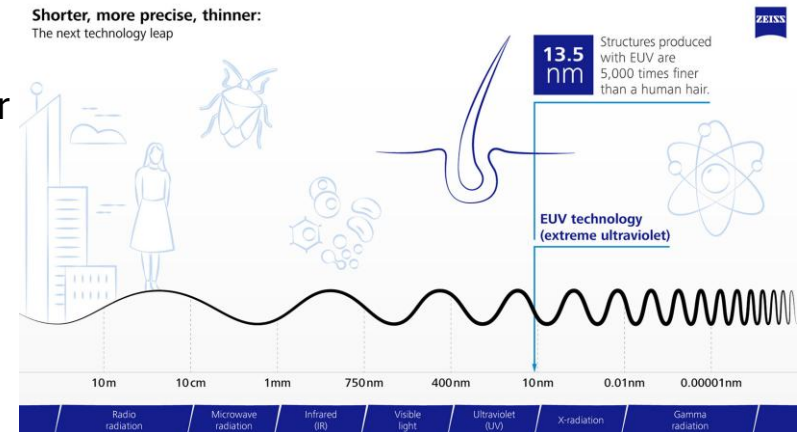
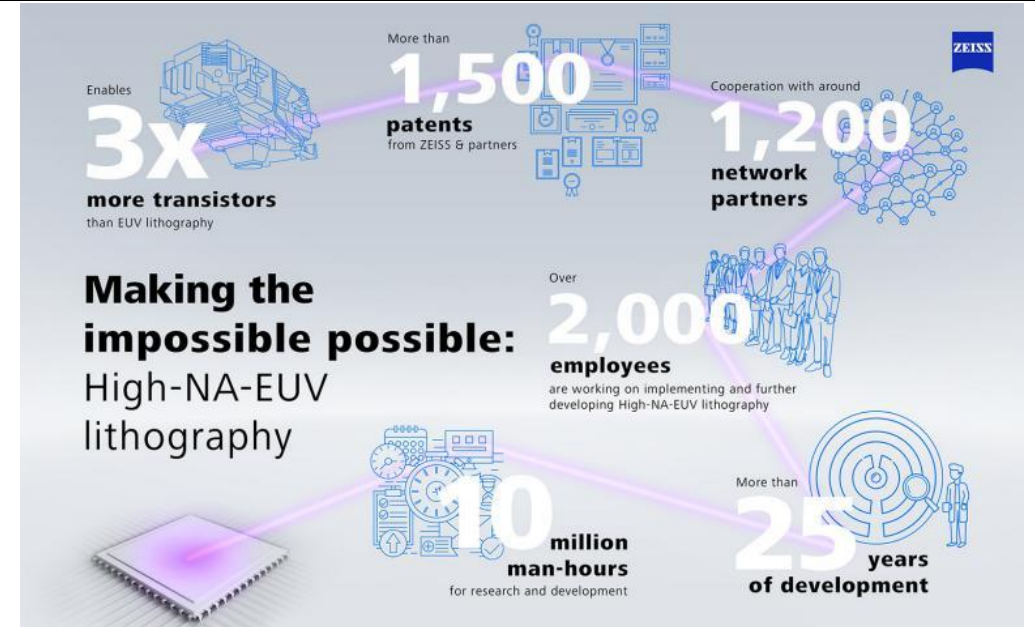
What could be deemed as one of the most iconic moments of Hasselblad in space was when the Apollo 11 mission successfully landed the Eagle on the Moon on 20 July 1969, signifying humanity's first steps off our own planet. A silver Hasselblad Data Camera (HDC) with Réseau plate, fitted with a Zeiss Biogon 60mm $f/5.6$ lens, was chosen to document the lunar surface and attached to astronaut Armstrong's chest. A second black Hasselblad Electric Camera (HEC) with a Zeiss Planar 80mm $f/2.8$ lens was used to shoot from inside the Eagle lunar module. The HDC had never been tested in space before, adding to the pressure of this once in a lifetime moment. Would the one Hasselblad camera used to shoot on the lunar surface capture the results everyone was hoping for? Working perfectly under the extreme conditions of the lunar surface, the HDC produced some of history's most iconic photographs. After the successful shooting on 21 July 1969, the Hasselblad was hoisted up to the lunar lander with a line. Securely removing the film magazines, both cameras with lenses were left behind on the Moon in order to meet narrow weight margins for successful return. The journeys home from the Moon made very special demands on what could return regarding weight; from Apollo 11 to the final Apollo 17 mission, a total of twelve camera bodies were left behind on the lunar surface. Only the film magazines containing the momentous images were brought back. The resulting photographs captured the history of humanity in the making.

Advancements in platforms and architectures.

High-NA-EUV lithography at a glance

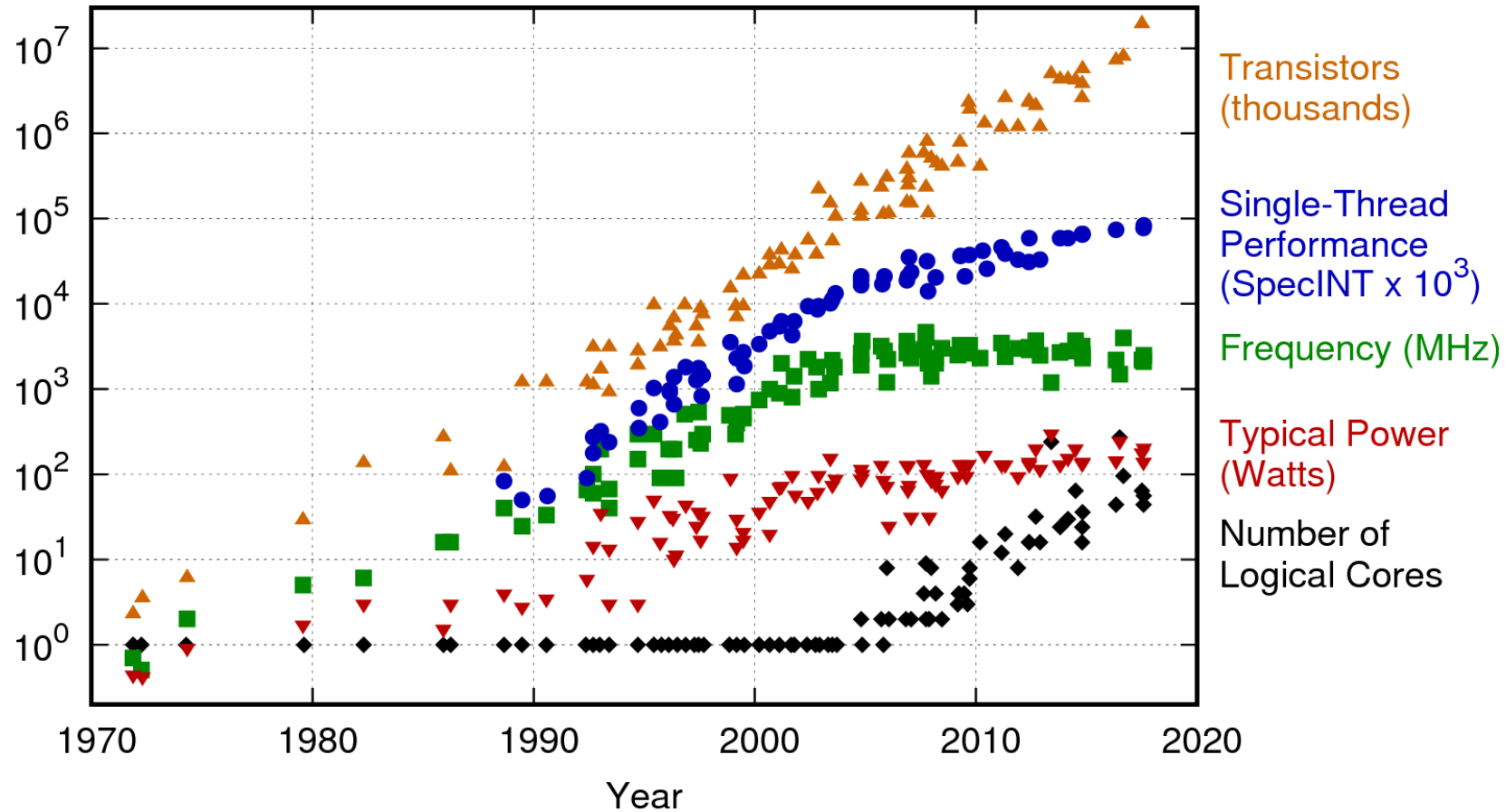


- ASML and Zeiss
- The world's most powerful pulsed industrial laser
Will enable next generation of chips.
 - Intel 18A and 14A
 - 2.9x more density
 - Similar plans by TSMC and



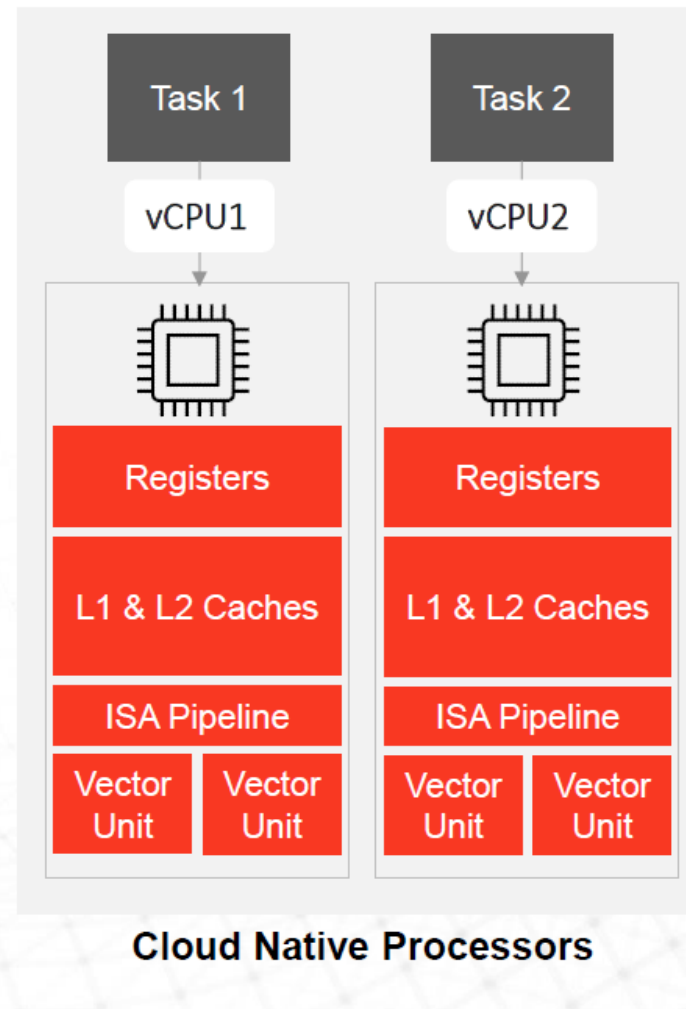
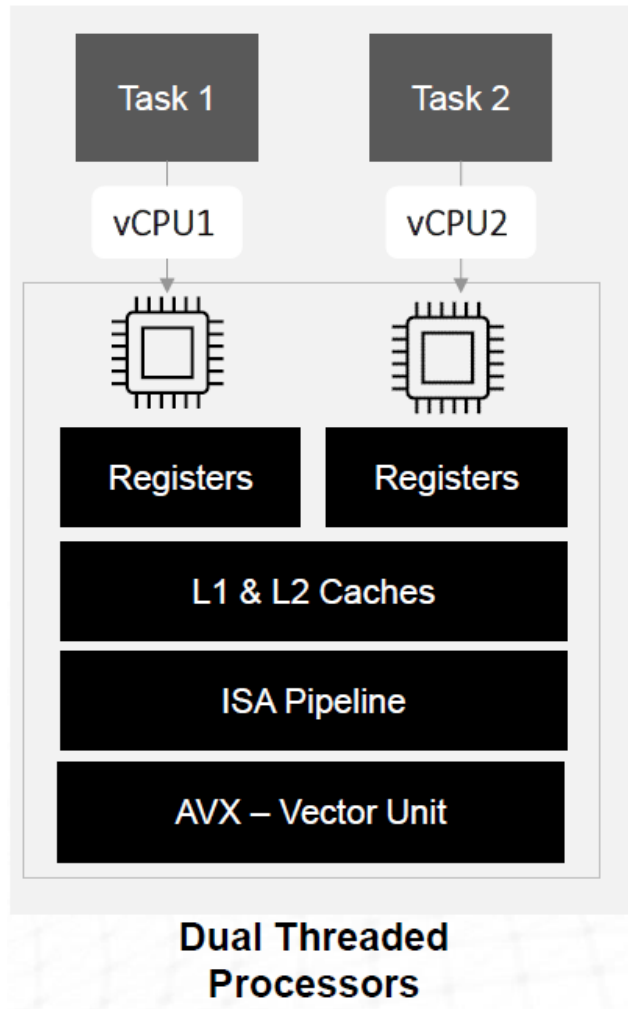
Advancements in platforms and architectures.

42 Years of Microprocessor Trend Data



Original data up to the year 2010 collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond, and C. Batten
New plot and data collected for 2010-2017 by K. Rupp

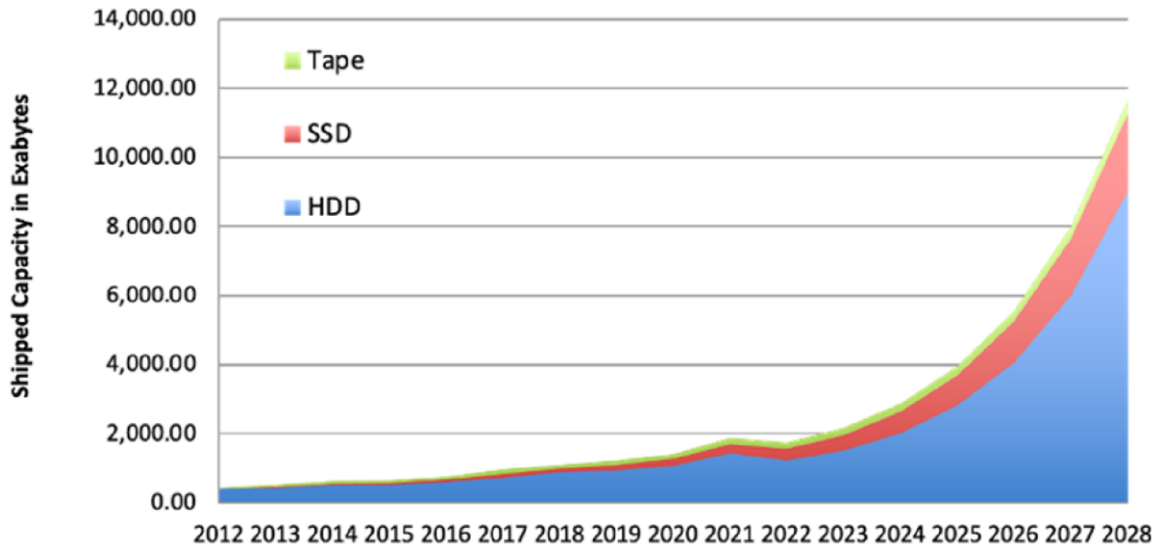
Advancements in platforms and architectures.



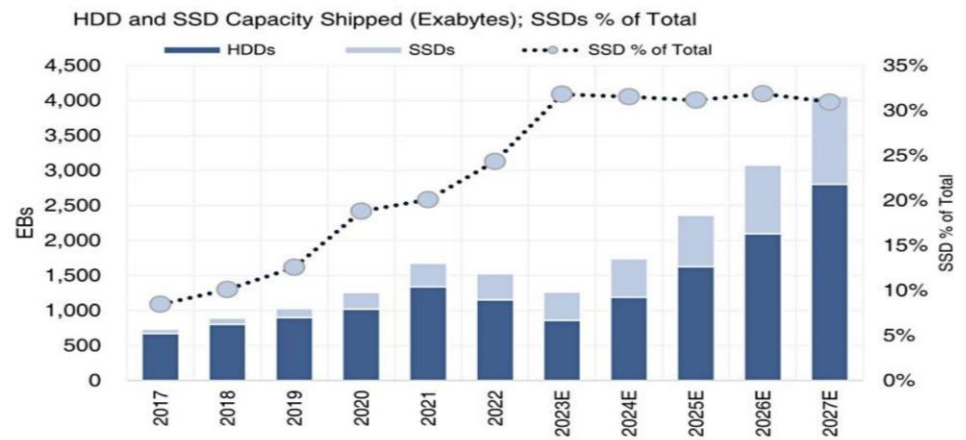
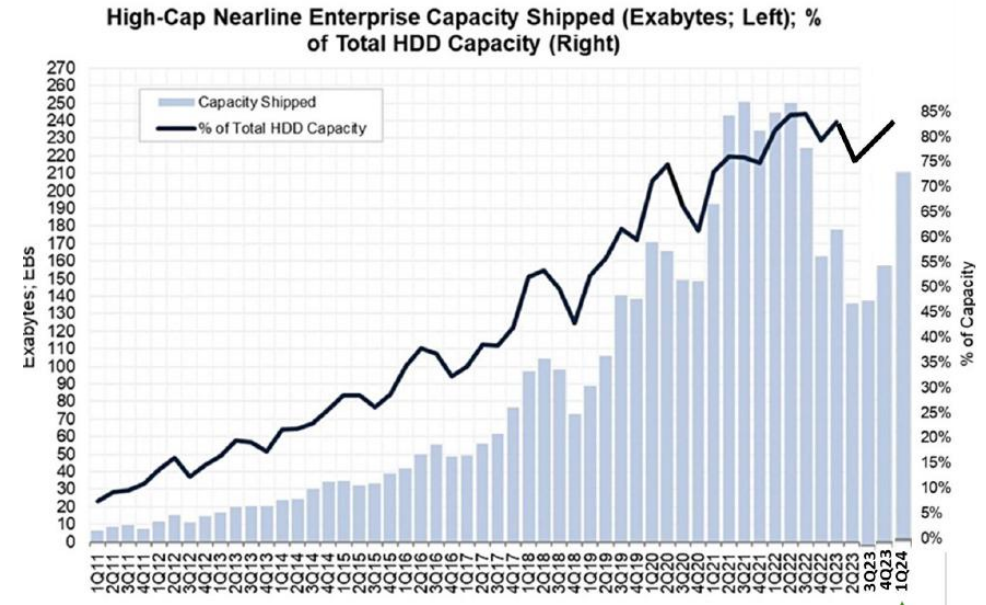
For all workloads

- Private resources in each core
- Resists noisy neighbour influence
- Predictable latency
- Linear Scaling
- Up to 384 Vector Engines
- Scale out requires SW optimisation
- Containerization of services and

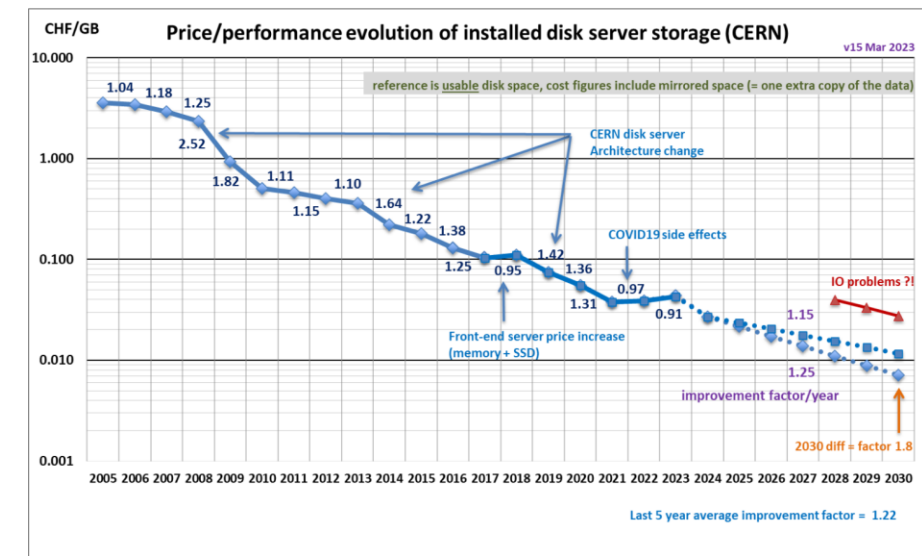
Advancements in platforms and architectures.



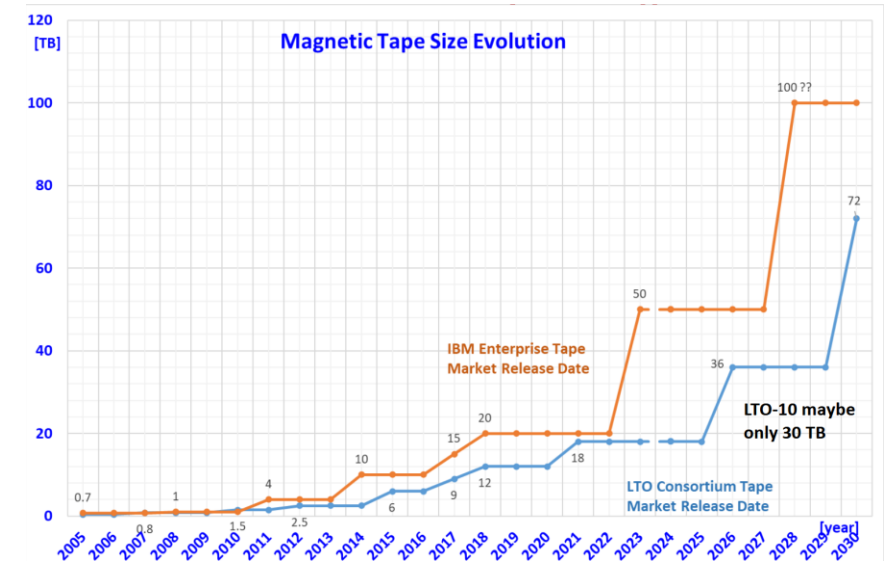
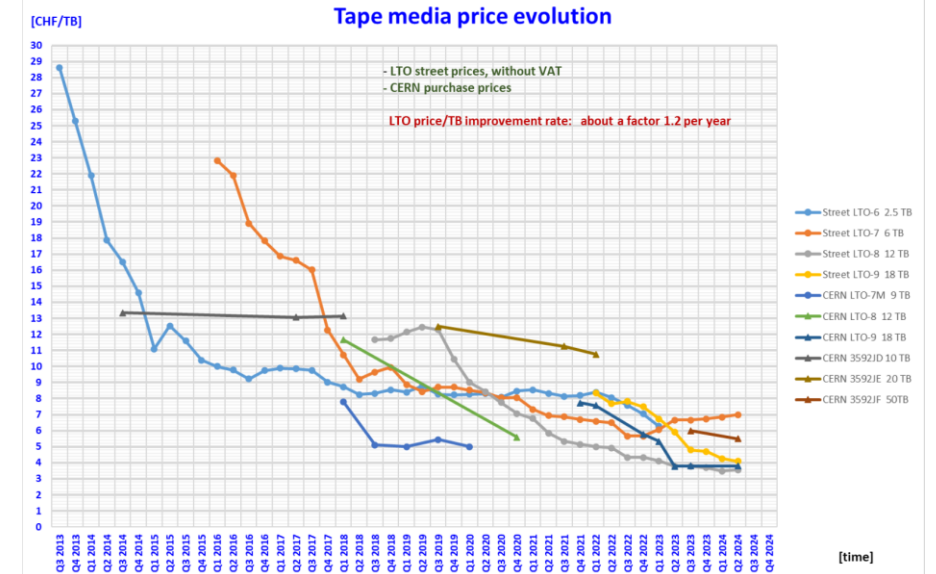
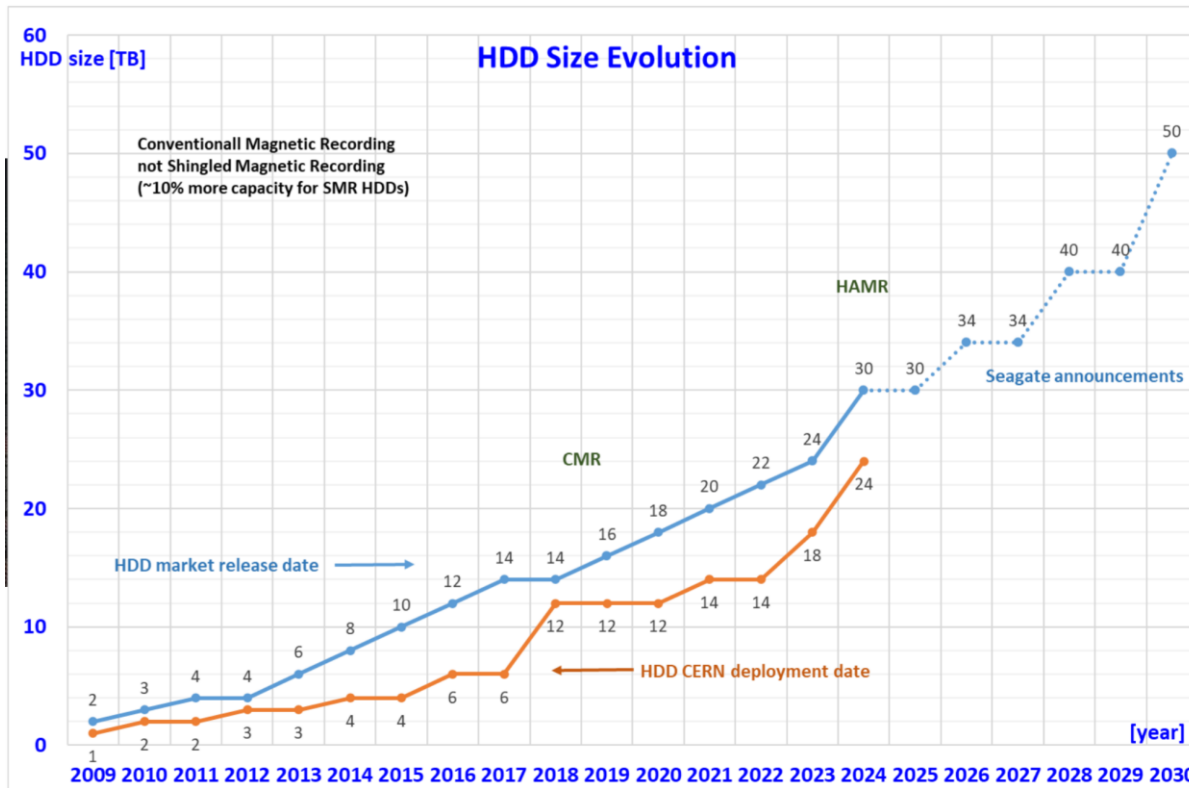
History and Projections for Digital Storage Capacity Shipments of HDDs, SSDs and Magnetic Tape
COUGHLIN ASSOCIATES CHART



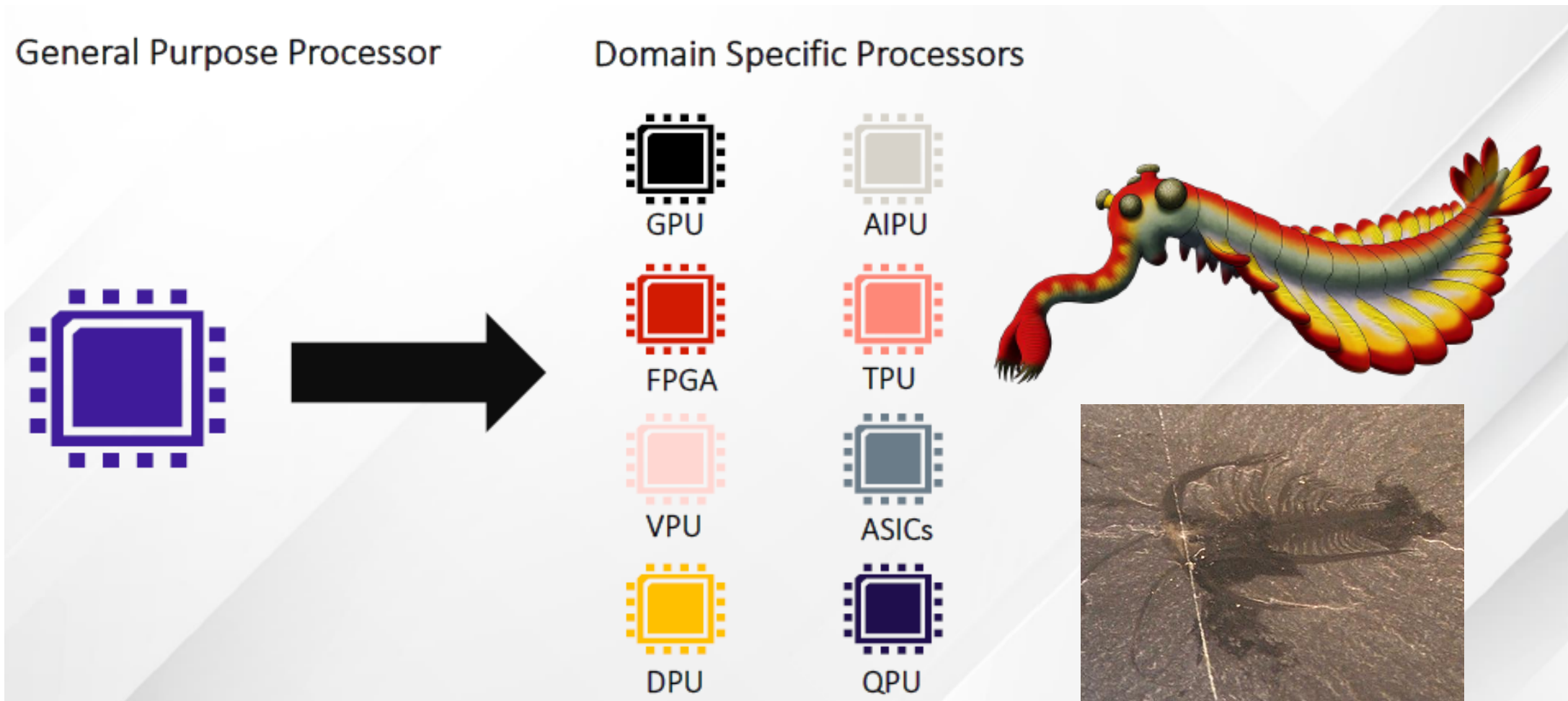
Source: Gartner; Wells Fargo Securities, LLC.



Advancements in platforms and architectures.



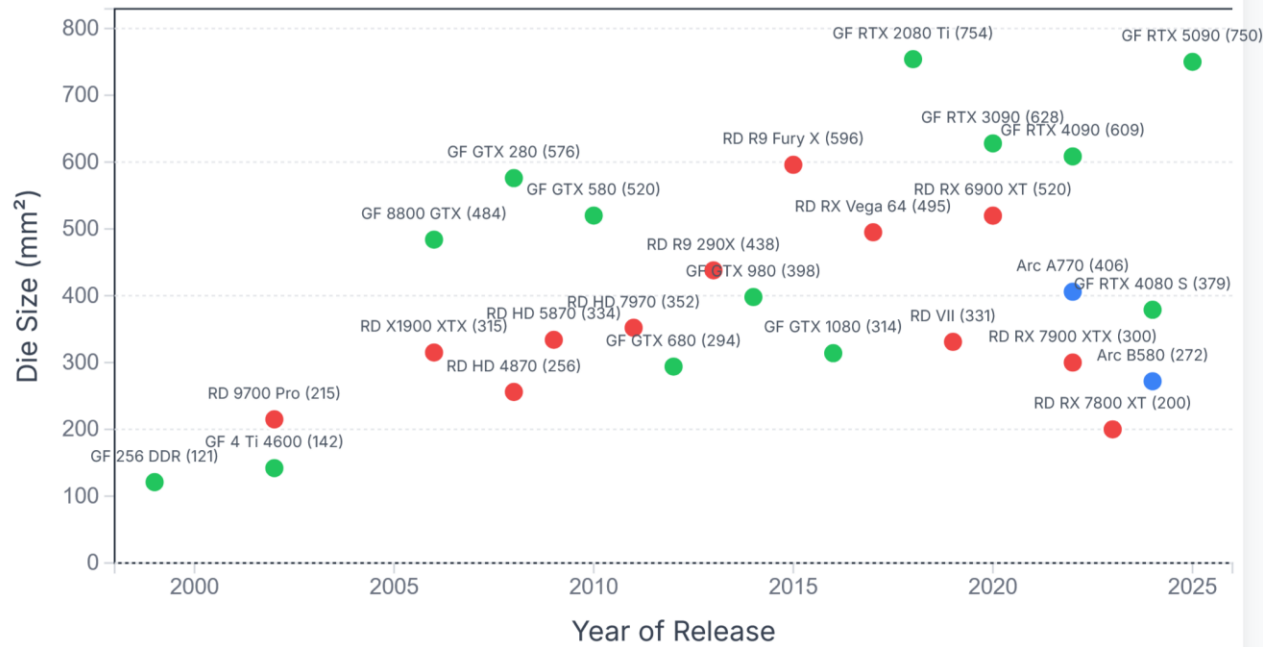
Advancements in platforms and architectures.



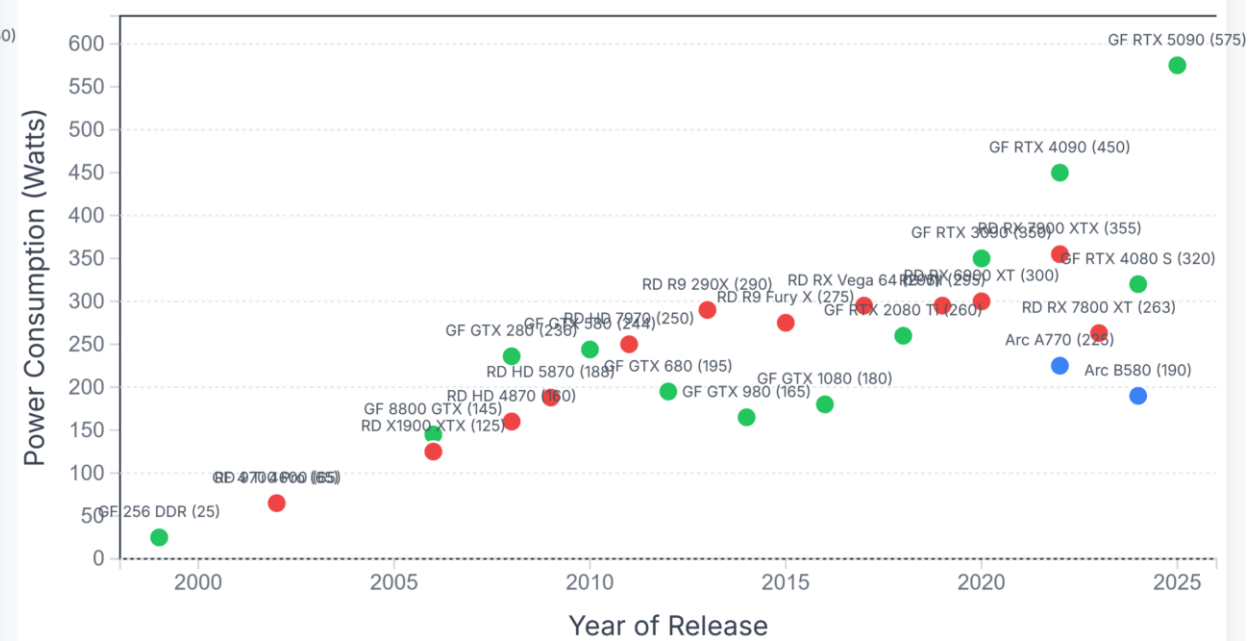
Architecture-aware optimizations in hot sections of our code can yield huge gains overall

Advancements in platforms and architectures.

GPU Die Size Evolution



GPU Power Consumption (TDP) Evolution



Advancements in platforms and architectures.

Dielectric-fibre surface waveguides for optical frequencies Kao&Hockham 1966

K.C. Kao, B.Sc.(Eng.), Ph.D., A M I E E and G. A. Hockham, B.Sc.(Eng.), Graduate I.E.E.

Synopsis

A dielectric fibre with a refractive index higher than its surrounding region is a form of dielectric waveguide which represents a possible medium for the guided transmission of energy at optical frequencies. The particular type of dielectric-fibre waveguide discussed is one with a circular cross-section. The choice of the mode of propagation for a fibre waveguide used for communication purposes is governed by consideration of loss characteristics and information capacity. Dielectric loss, bending loss and radiation loss are discussed, and mode stability, dispersion and power handling are examined with respect to information capacity. Physical-realisation aspects are also discussed. Experimental investigations at both optical and microwave wavelengths are included.

List of principal symbols

J_n = nth-order Bessel function of the first kind
 K_n = nth-order modified Bessel function of the second kind
 β = phase coefficient of the waveguide
 β_0 = first derivative of β
 K'_n = first derivative of K_n
 h_{2m} = radial wavenumber or decay coefficient
 ϵ_1 = relative permittivity
 k_0 = free-space propagation coefficient
 a = radius of the fibre
 γ = longitudinal propagation coefficient
 k = Boltzman's constant
 T = absolute temperature, deg K
 β_c = isothermal compressibility
 λ = wavelength
 n = refractive index
 $H_n^{(i)}$ = vth-order Hankel function of the ith type
 H'_v = derivation of H_v
 v = azimuthal propagation coefficient = $v_1 - jv_2$
 L = modulation period
 Subscript n is an integer and subscript m refers to the m th root of $J_n = 0$

for HE_{nm} modes

for E_{nm} modes

$$\frac{v^2 J'_v \left(\frac{u_1}{a} \right)}{u_1 J'_v \left(\frac{u_1}{a} \right)} \left[\frac{E_1 K'_v(u_2)}{u_2 K_v(u_2)} \right]$$

$$\left[\frac{1 - \frac{J'_v(u_1)}{u_1 J'_v(u_1)} \frac{u_2 K'_v(u_2)}{u_2 K_v(u_2)} \right] \circledast$$

$$\frac{E_1 J'_v(u_1)}{u_1 J'_v(u_1)} - \frac{E_2 K'_v(u_2)}{u_2 K_v(u_2)} \quad (2)$$

for H_{nm} modes

$$\frac{1 - \frac{J'_v(u_1)}{u_1 J'_v(u_1)} \frac{u_2 K'_v(u_2)}{u_2 K_v(u_2)}}{u_1 J'_v(u_1)} - \frac{1 - \frac{K'_v(u_2)}{u_2 K_v(u_2)}}{u_2 K_v(u_2)} \quad (3)$$

The auxiliary equations defining the relationship between u_1 and u_2 are

$$u_1^2 + u_2^2 = (k_0 a)^2 (\epsilon_1 - \epsilon_2)$$

$$h_1^2 = \gamma^2 + k_0^2 \epsilon_1$$

$$-h_2^2 = \gamma^2 + k_0^2 \epsilon_2$$

$$u_i = h_i a, i = 1 \text{ and } 2$$

where subscripts 1 and 2 refer to the fibre and the outer region, respectively.

All the modes exhibit cutoffs except the HE_{11} mode, which is the lowest-order hybrid mode. It can assume two orthogonal polarisations, and it propagates with an increasing percentage of energy outside the fibre as the dimensions of the structure decrease. Thus, when operating the waveguide in the HE_{11} mode, it is possible to achieve a single-mode operation by reducing the diameter of the fibre sufficiently. Under this condition, a significant proportion of the energy is carried outside the fibre. If the outside medium is of a lower loss than the inside dielectric medium, the attenuation of the waveguide is reduced. With these properties, HE_{11} mode operation is of particular interest.

The physical and electromagnetic aspects of the dielectric-fibre waveguide carrying the HE_{11} mode for use at optical frequencies will now be studied in detail. Conclusions are drawn as to the feasibility and the expected performance of such a waveguide for long-distance-communication application.






1151

Paper 5033 E, first received 24th November 1965 and in revised form 15th February 1966.
 Dr Kao and Mr. Hockham are with Standard Telecommunication Laboratories Ltd., Harlow, Essex, England
 PROC. IEE, Vol. 113, No. 7, JULY 1966

Nobel Prize in Physics in 2009

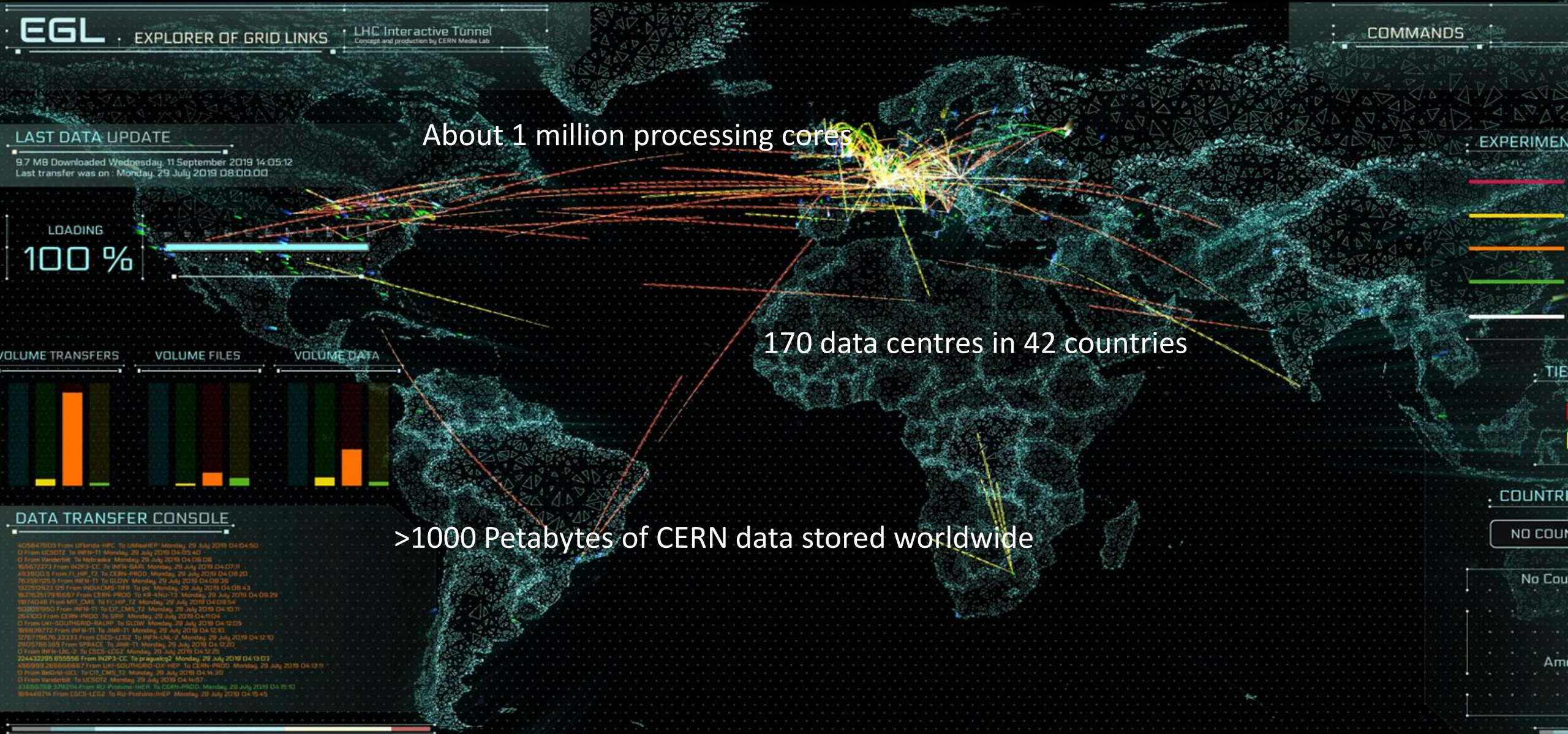
World Record Achieved in Transmission Capacity and Distance: With 19-core Optical Fiber with Standard Cladding Diameter 1,808 km Transmission of 1.02 Petabits per Second

- Expectation for Future Long-Distance High-Capacity Optical Communication Infrastructure -

Optical Fiber	Uncoupled 4-Core Fiber	Uncoupled 4-Core Fiber	15-Mode Fiber	Coupled 19-Core Fiber	Achievement New Coupled 19-Core Fiber
Cross-sectional View					
Achieved Records	Jun. 2021	Oct. 2023	Mar. 2023	Mar. 2023	Apr. 2025
Data Rate (petabit/s)	0.319	0.138	0.273	1.7	1.02
Distance(km)	3,001	12,345	1,001	63.5	1,808
Capacity Distance Product (exabit / s · km)	0.95	1.71	0.27	0.107	1.86 World Record
Wavelength Bands	S, C, L	S, C, L	C	C, L	C, L
MIMO Processing Load	None	None	Large	Moderate	Moderate

<https://www.nict.go.jp/en/press/2025/05/29-1.html>

The Worldwide LHC Computing Grid (WLCG): Our cyberinfrastructure



About 1 million processing cores

170 data centres in 42 countries

>1000 Petabytes of CERN data stored worldwide

Languages and software engineering.

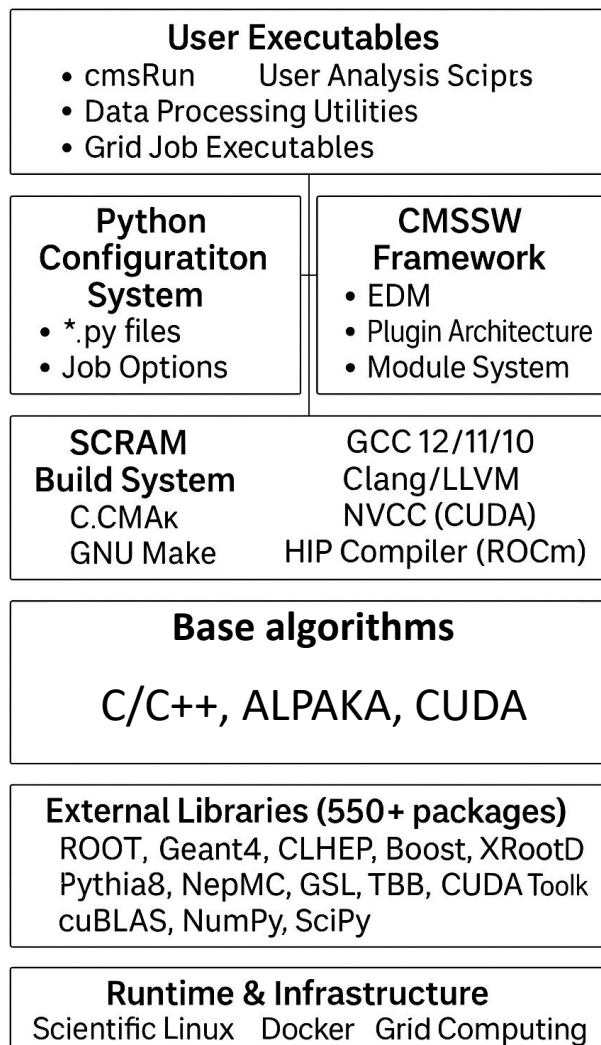
What if we could talk to a computer in a language closer to English

"The most dangerous phrase in the language is, 'We've always done it this way.'"

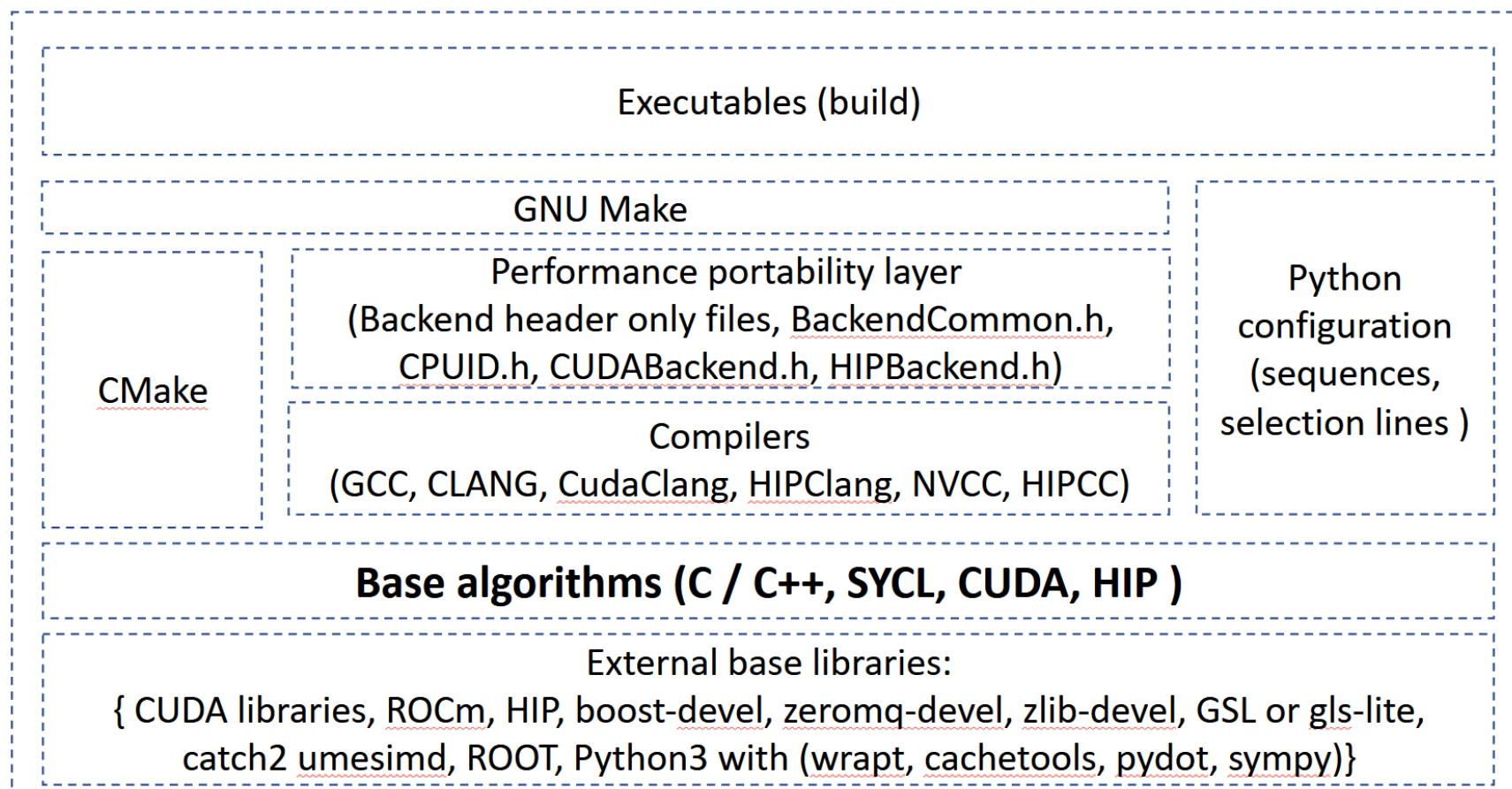
Grace Hopper,
1952, Developer of world's first Compiler (A-O)

Modern software project stack example

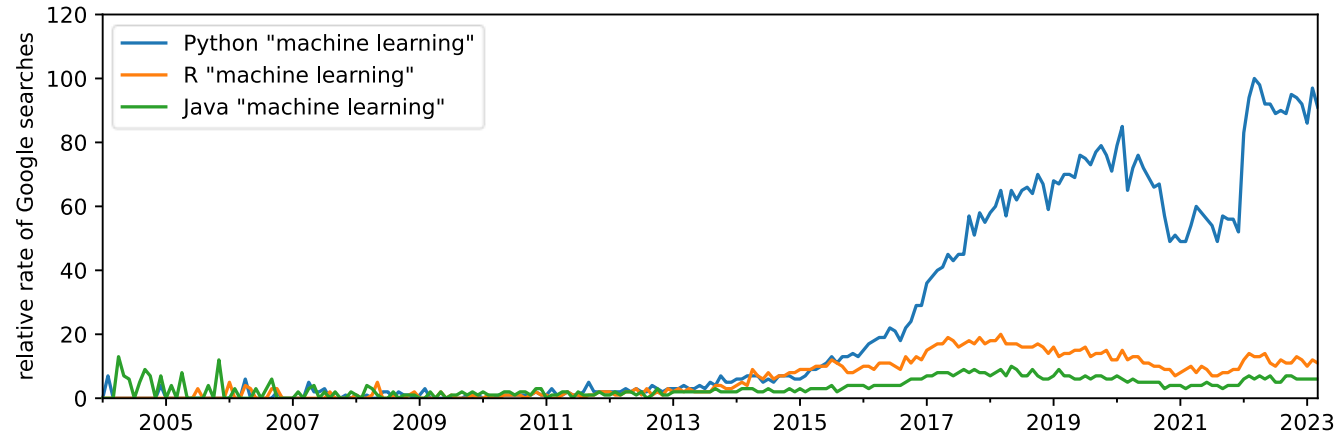
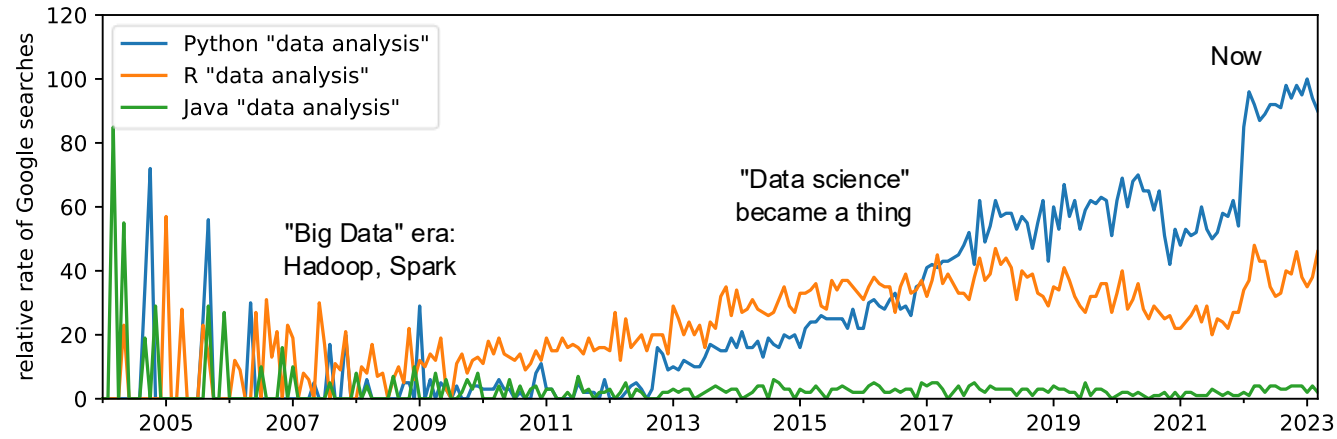
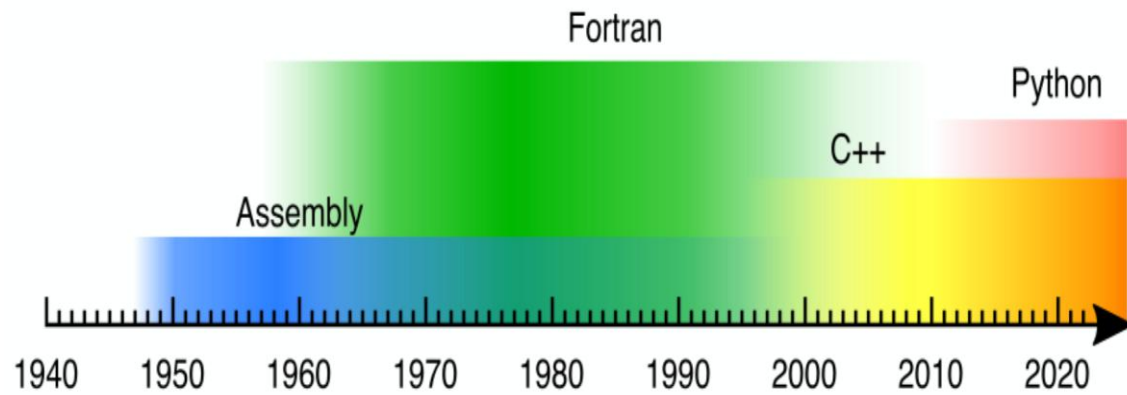
CMS CMSSW framework



LHCb HLT: Allen framework

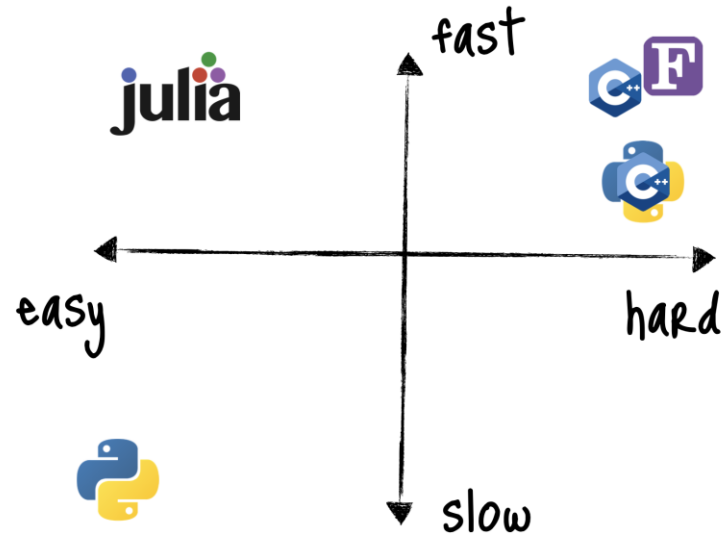


Languages and software engineering.



Languages and software engineering.

[CHEP24-JuieaHEP](#)



Metric	C++	Python	Julia
Performance	✓	✗	✓
Expressiveness	⚠	✓	✓
Learning Curve	✗	✓	✓
Safety (memory)	⚠	✓	✓
Composability	✗	⚠	✓

- Julia isn't perfect or magic
 - Startup time
 - Only LLVM backend
 - Static binaries and performance analysis a bit cumbersome
 - Pure Julia ML libraries not beating PyTorch
- But it does have clear advantages in many areas
- So its tradeoffs compare favourably

Core Features in C++23

- **Modules**
 - Simplify code imports with `import std;`—reduces compile times and avoids header complexities.
 - Great for large projects: improved encapsulation and clarity.
- **String Enhancements**
 - New `.contains()` method for substring search: `if (text.contains("word"))`.
- **`std::print` & `std::println`**
 - Clean, type-safe, and format-friendly output.
 - Example: `std::println("Value is: {}", value);`.
- **`std::flat_map` & `std::flat_set`**
 - Optimized for insertion-heavy workloads—use array-based storage for better cache performance.
- **Parallel Algorithms**
 - Built-in multi-core support via execution policies (`std::execution::par`): `std::for_each(std::execution::par, data.begin(), data.end(), func);`.
- **Debugging Tools**
 - `std::stacktrace` for error diagnosis and backtracing: efficient call stack inspection.

Languages and software engineering.

Software approaches beyond evolutionary baseline: HLT

Complexity challenges ahead: Despite clever simplifications, the complexity bounds highlight significant challenges for future Runs.

	Theoretical problem	Simplification	
Data sorting	$O(n^2)$	-	Quicksort or merge sort
Track seeding	$O(2^n)$	$O(n \cdot \log(n)^2)$	Geometry or physical constrains,
Track following	$O(2^n)$	$O(n \cdot \log(n))$	Kalman filter to most likelihood path
Likelihood minimisation	$O(2^n)$	$O(n^6)$	Gradient descent from exp to high-deg pol
Clustering	$O(n^2)$	$O(V + E)$	Graph based clustering
Selections	$O(2^n)$	$O(n^2)$	Exp to Quad

Need for advancements in algorithms:

- Similar challenges for MC simulations and offline processing.
- Development of more advanced and efficient data traversal algorithms is essential to manage exponentially growing data throughput,

Languages and software engineering. Python for HEP: Tutorials

- Array oriented programming for particle physicists
 - <https://hsf-training.github.io/array-oriented-programming/0-intro.html>
 - <https://github.com/hsf-training/array-oriented-programming>
- Python with Exercises
 - <https://research-software-collaborations.org/python-june2025/intro.html#>
 - [15 Sorting Algorithms in 6 Minutes](#)

Break: Questions and discussion

Thank you