WHAT'S GOING ON IN HIGH PERFORMANCE COMPUTING

PHONON-PHONON INTERACTION WORKSHOP

Dan Stanzione
Executive Director, TACC
Associate Vice President for Research, UT-Austin

EPW Summer School Seminar

June 2024

TACC - 2024

LEADERSHIP-CLASS COMPUTING FACILITY

TACC
TEXAS ADVANCED COMPUTING CENTER

WHAT WE DO

- Provide researchers with:
  - Computing, Data, AI , Software capabilities to support their research
  - The expert help to be able to use it!
  - In the ways they want to consume it
  - Help with grants/strategy
- Computation, AI, Data almost ubiquitous across the sciences.

OUR MISSION

- Mission: To enable discoveries that advance science and society through the application of advanced computing technologies.

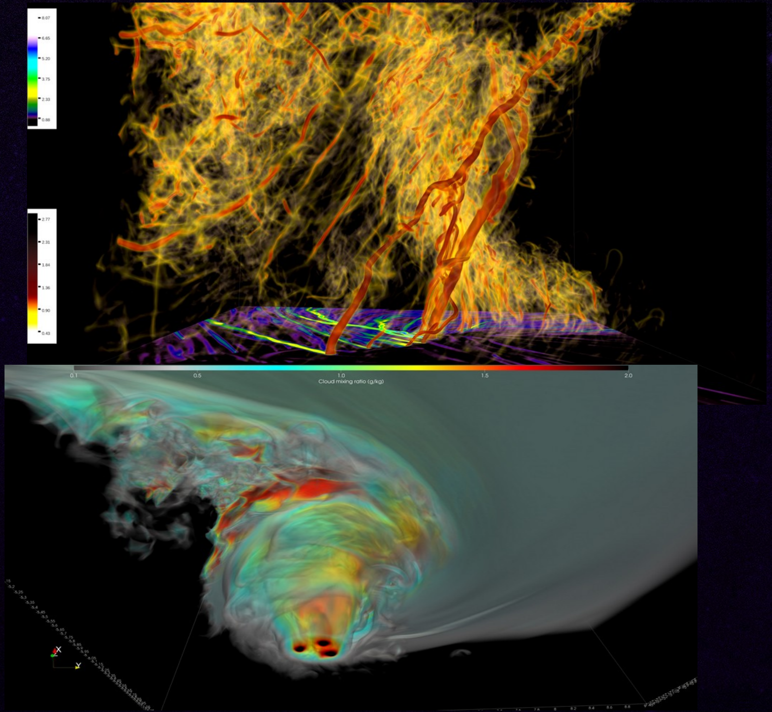
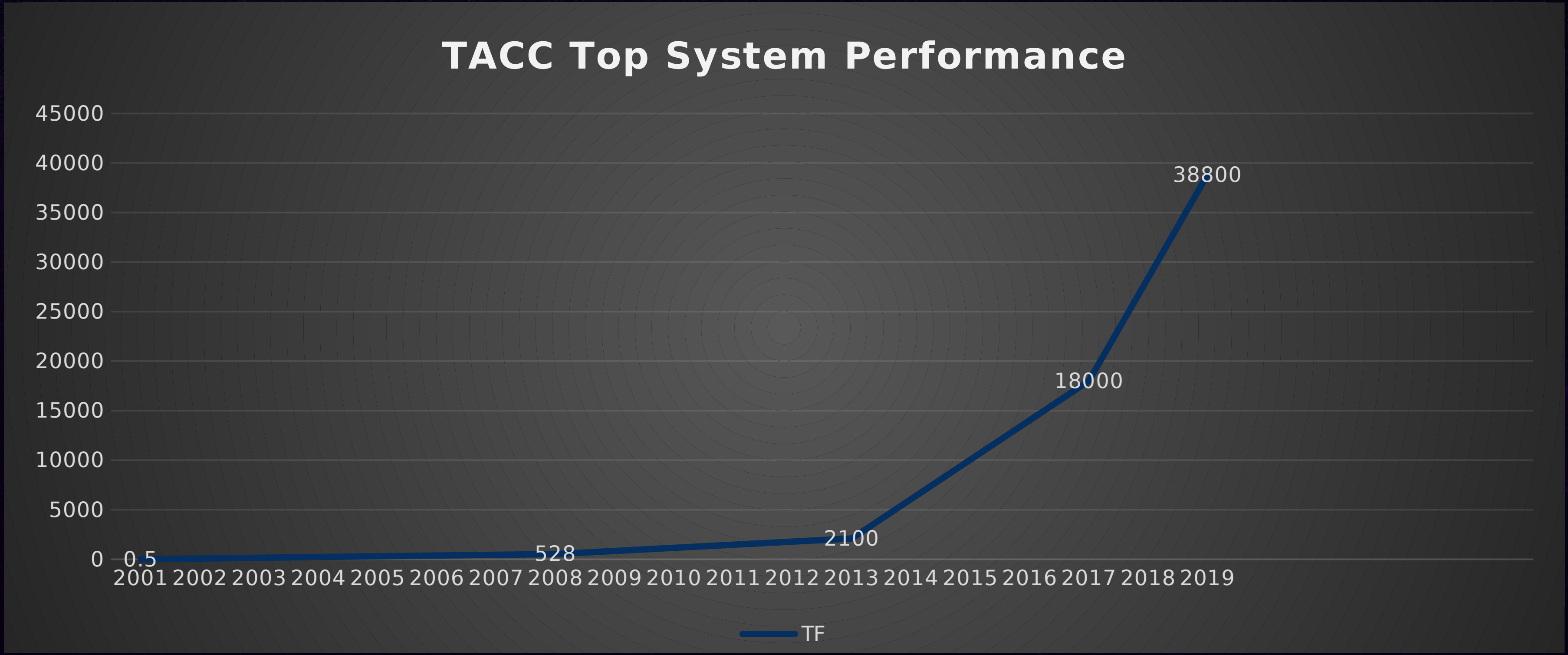


Figure: "World's Most Detailed Tornado Simulation", Leigh Orf, Wisconsin – used more than 200,000 cores on Frontera

SYSTEMS UPDATES

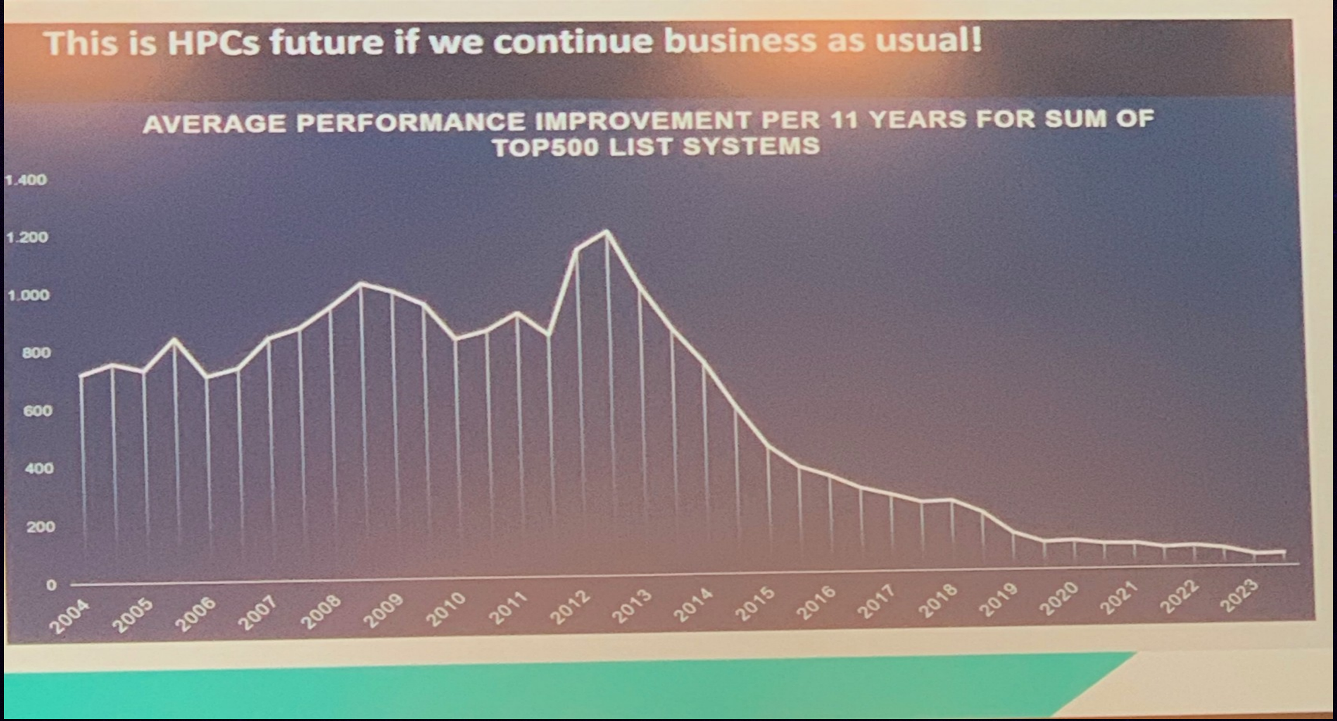
- Frontera, NSF Capability System, 2019-2026 (started #5, still in Top 30)
- ~~Stampede2, NSF Capacity System, 2017-2023~~
- Stampede3, NSF Capacity System (with HBM) 2024-2029
- Lonestar-6, Texas/Local System 2022-2027(ish)
- **Vista – AI/DL GPU System, 2024-2029**
- Jetstream2 - NSF "Cloud" System 2022-2027
- ~~Chameleon3~~ –Chameleon4 NSF CS Testbed 2024-2028 (multiple HW upgrades)
- Corral, Ranch, Stockyard – Storage Platforms
- *Aggregate: ~75PF, ~16,000 compute nodes, ~350PB*

TACC PERFORMANCE OVER TIME

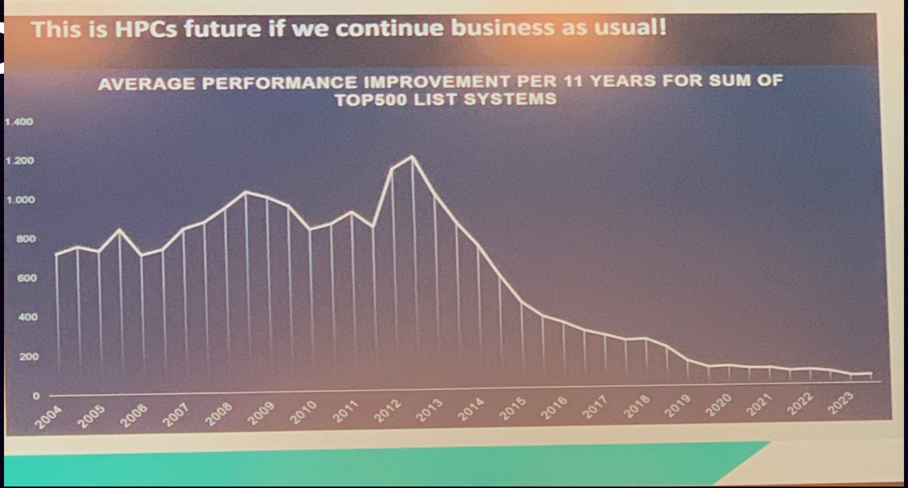


Our Oversubscription rate is still ~5x, Despite the Growth

PERFORMANCE GAINS HAVE SLOWED. .



MOORE’S LAW ENDED

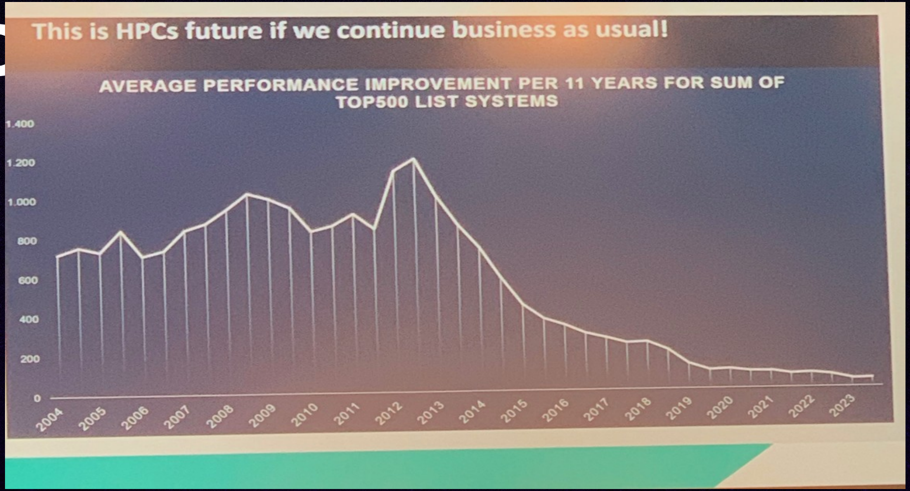


OK, so clearly, this looks like the fabled “End of Moore’s Law” that’s been forecast repeatedly for 30 years.

Is it? Sort of, but it’s important to remember what Moore’s Law actually said.

The observation is named for Gordon Moore, the co-founder of Fairchild Semiconductor and Intel, who in 1965 posited a doubling every year **in the number of components per integrated circuit**, and projected this rate of growth would continue for at least another decade. In 1975, looking forward to the next decade, he revised the forecast to doubling every two years. While Moore did not use empirical evidence in forecasting that the historical trend would continue, his prediction has held since 1975 and has since become known as a "law".

MOORE'S LAW ENDED



What people perceived as Moore's Law was actually *performance* improving, which was never promised.

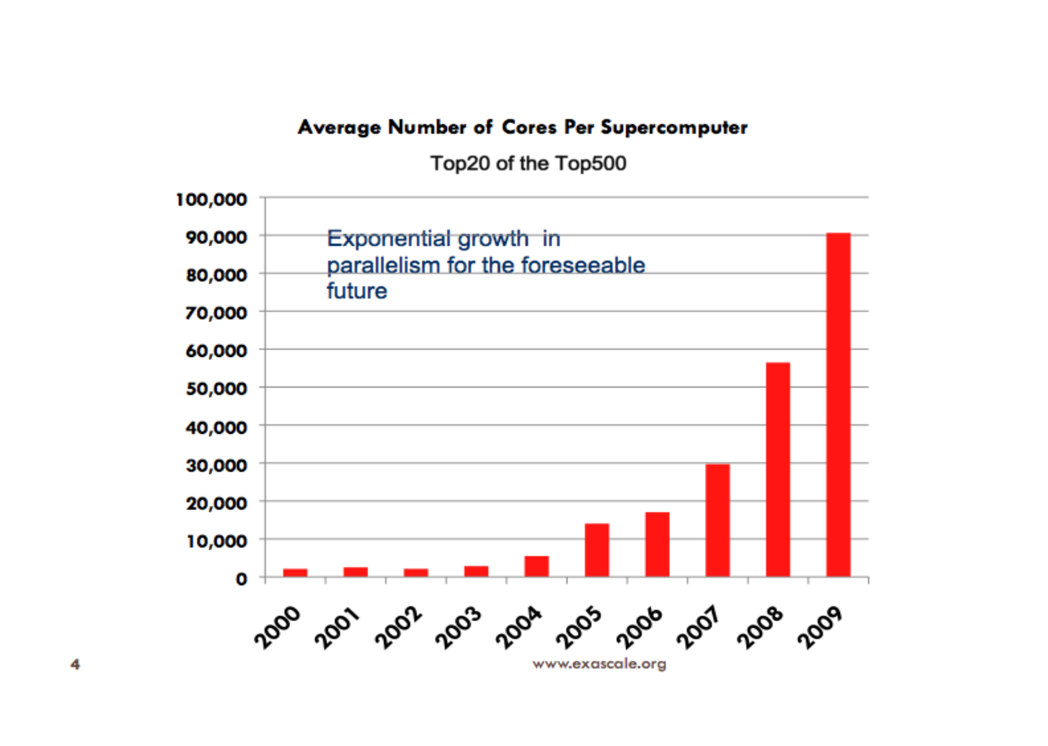
It's also confused with Dennard Scaling, which definitely ended a while ago.

Dennard scaling … is a scaling law which states roughly that, as transistors get smaller, their power density stays constant, so that the power use stays in proportion with area; both voltage and current scale (downward) with length

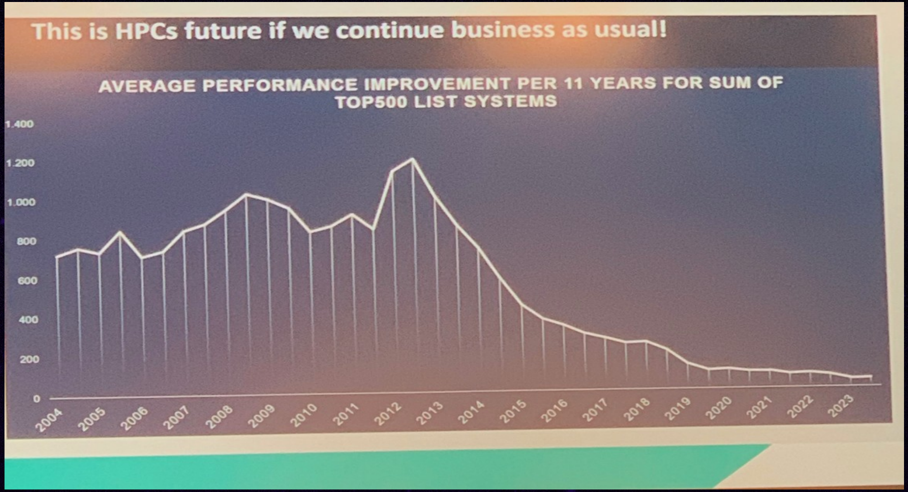
When Dennard scaling stopped, we stopped gaining frequency, which leads to this plot...

DENNARD SCALING FAIRLY DEFINITIVELY ENDED AROUND 2005.

- 1 GF – 1988 – 8 cores
- 1 TF - 1998 – 1024 cores
- 1 PF – 2008 - 150,000 cores
- 1 EF – 2022. - ~9M cores.



WHY THIS IS BAD



So, core counts stopped yielding power improvements in ~2018, meaning double performance was now double the power.

Consider:
Stampede Xeon (Sandy Bridge) – 8 cores, 130 Watts
Stampede 2 Xeon (Skylake) – 24 cores, 145 Watts
Frontera Xeon (Cascade Lake) – 28 cores, 210 Watts
Stampede 3 Xeon (Sapphire Rapids) – 56 cores, 350 Watts

So, less improvement on a fixed power (and dollar) budget since ~2018.

LIES, DAMN LIES, AND HPC BENCHMARKS

- OK, so the situation isn’t really as bad as the doomsayers forecast. . .
  - Consider:
    - Frontier, ~1 EF for 22MW
    - Aurora, ~1EF for 38.5 MW
  - But, Frontier supports 64 bit matrix operations (think tensor cores for AI), and Aurora does not (vector instructions at 64 bit).
  - The only known use of 64 bit matrix instructions is to run the benchmark above. . .
- HPL only measures, roughly, one dimension of performance –
  - For instance, the new Stampede3 processors get *practical* performance improvement over the last generation by improving memory performance, which doesn’t change your Top 500 (HPL, or Linpack) performance *at all*.

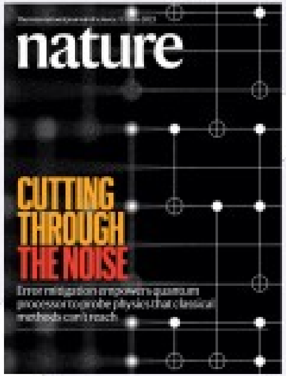
A WORD ABOUT QUANTUM

- I have been cautious for a long time about promising *anything* from Quantum Computing.
- But recently, there is reason to believe we are on a real path to useful hardware within 6 years. . .

Developments since last IS&T Capability Review

Indicate more optimism on QC maturity

- June 14 2023 – Evidence for the utility of quantum computing before fault tolerance (due to error mitigation) – IBM Quantum, Riken, UCB and LBNL



- April 3, 2024 - By applying an innovative qubit-virtualization system to ion-trap hardware, Microsoft and Quantinuum created four highly reliable logical qubits from 30 physical qubits, while demonstrating an 800x improvement in error rate
  - 14,000 experiments executed without a single error

Fault-Tolerant Quantum Computing is Coming Faster Than Expected

Trapped Ion (2022)

[5] C. Ryan-Anderson et al. **Implementing Fault-tolerant Entangling Gates on the Five-qubit Code and the Color Code. 2022**. arXiv: 2208.01863 [quant-ph].

Superconducting (2023)

[4] VV Sivak, Alec Eickbusch, Baptiste Royer, Shraddha Singh, Ioannis Tsioutsios, Suhas Ganjam, Alessandro Miano, BL Brock, AZ Ding, Luigi Frunzio, et al. "**Real-time quantum error correction beyond break-even**". In: Nature 616.7955 (2023), pp. 50–55.

Photonic (2023-2024)

[10] Reuters. PsiQuantum targets first commercial quantum computer in under six years. url: https://www.reuters.com/technology/psiquantum-targets-first-commercial-quantum-computer-under-six-years-2023-09-28/ (visited on 04/17/2024).

PsiQuantum to Build World's First Utility-Scale, Fault-Tolerant Quantum Computer in Australia - *The Australian and Queensland Governments Will Invest $940M AUD ($620M USD) into PsiQuantum* (announced on 04/17/2024)

Neutral Atom (2023)

[3] Dolev Bluvstein et al. "**Logical quantum processor based on reconfigurable atom arrays**". In: Nature 626.7997 (Dec. **2023**), pp. 58–65. issn: 1476-4687.

QuEra, Infleqtion (2024)

[8] QuEra. QuEra Computing Releases a Groundbreaking Roadmap for Advanced Error-Corrected Quantum Computers, Pioneering the Next Frontier in Quantum Innovation. url: https://www.quera.com/press-releases/quera- computing-releases-a-groundbreaking-roadmap-for-advanced-error-corrected-quantum-computers-pioneering-the-next-frontier-in-quantum-innovation (visited on 04/11/2024).

[9] Infleqtion. Infleqtion Unveils 5-year Quantum Computing Roadmap, Advancing Plans to Commercialize Quantum at Scale. url: https://www.infleqtion.com/news/infleqtion-unveils-5-year-quantum-computing-roadmap-advancing-plans-to-commercialize-quantum-at-scale (visited on 04/17/2024).

Superconducting (2024)

[2] Riddhi S Gupta, Neereja Sundaresan, Thomas Alexander, Christopher J Wood, Seth T Merkel, Michael B Healy, Marius Hillenbrand, Tomas Jochym-O'Connor, James R Wootton, Theodore J Yoder, et al. "**Encoding a magic state with beyond break-even fidelity**". In: Nature 625.7994 (**2024**), pp. 259–263.

Trapped Ion (2024)

[1] MP da Silva, C Ryan-Anderson, JM Bello-Rivas, A Chernoguzov, JM Dreiling, C Foltz, JP Gaebler, TM Gatterman, D Hayes, N Hewitt, et al. "**Demonstration of logical qubits and repeated error correction with better-than-physical error rates**". In: arXiv preprint arXiv:2404.02280 (**2024**).

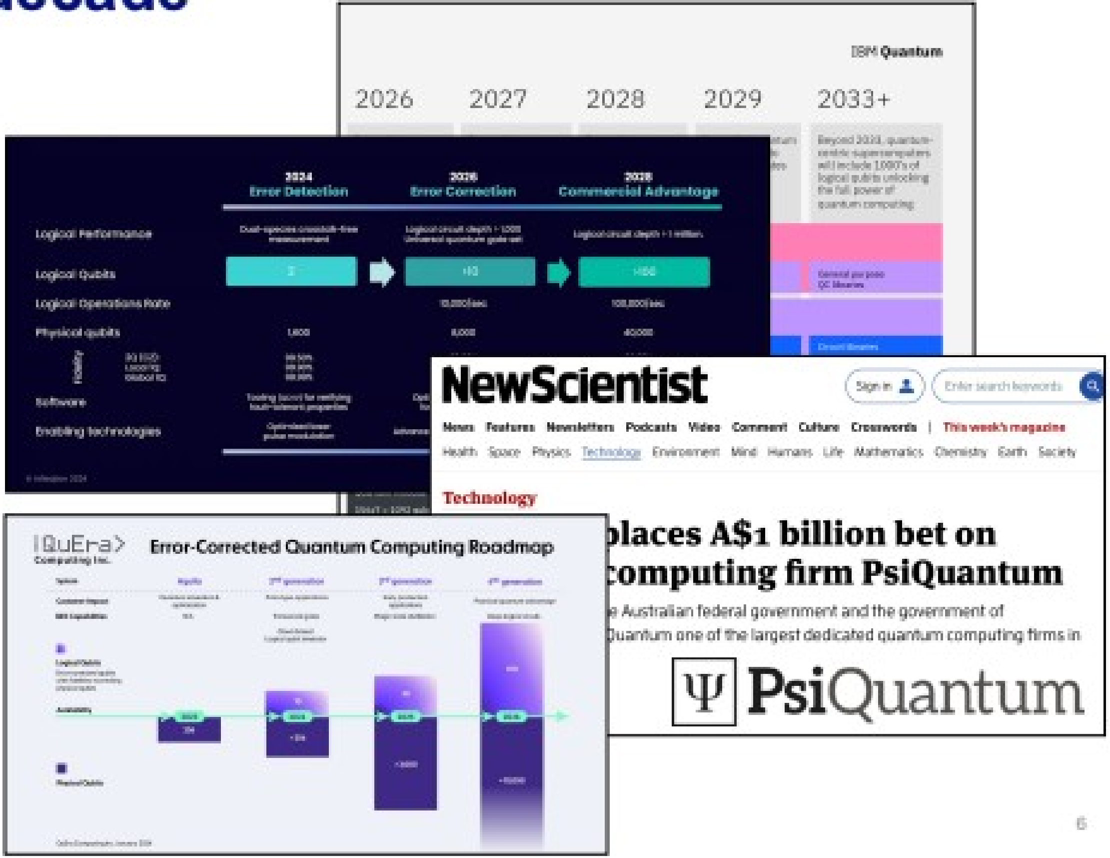
Entangled logical qubits encoded in a [[12, 2, 4]] code with error rates 4.7× to 800× lower than at the physical level, depending on the judicious use of post-selection

IBM (2024)

[11] IBM. IBM Quantum Roadmap. url: https://www.ibm.com/roadmaps/quantum/ (visited on 04/11/2024).

Public Roadmaps that aim for Fault-Tolerant Quantum Computing within this decade

- **IBM**
  - 200 logical qubits by 2029
  - 100M Gates
- **Infleqtion**
  - 100 logical qubits by 2028
  - 1-100M Gates
- **PsiQuanutm**
  - “utility-scale” by 2027
- **QuEra**
  - 100 logical qubits by *2026!*
  - Gates?



A WORD ABOUT QUANTUM

- So I’m still skeptical
- Even if the roadmaps hold, these machines will cover a very limited class of problems (and will likely be supercomputer-adjacent accelerators).
- However… quantum chemistry is where they will shine. . .
  - If you can write a Hamiltonian, they may be useful by 2030-2032.
  - By “useful”, I mean problems you won’t be able to solve anywhere else.

AND NOW, AI

MODERN AI EXISTS DUE TO HPC

- First, computers needed to be much faster.
  - Thanks to Physics(silicon process), Architecture (e.g. GPU), Algorithms (BLAS,FFT).
- The most important enabler for ”modern” AI through Deep Learning (e.g. Large Language Models, Foundation Models, and friends) came with an algorithmic discovery.
  - The definition of the “Transformer” Architecture in Google’s seminal paper “Attention is all You Need”
  - This broke the data dependencies in putting very large neural nets onto multiple compute servers.
  - And made AI/DL and HPC architectures virtually indistinguishable. . .

HPC AND MODERN AI

- Large Scale AI *is* HPC.
  - Three key to building large language models:
    - High Performance I/O
    - High Performance Computing
    - High Performance Communication

This is what we do.

- Large models are parallel applications
  - Data Parallelism
  - Operator Parallelism
  - Pipeline Parallelism
- AI in it’s current form would *not have happened* without HPC innovations.

HPC AND MODERN AI

- Though AI doesn't happen without HPC, and there is still much to learn from HPC…
  - Sparsity
  - Performance tuning in MPI applications
- HPC doesn't control it. To borrow from Torsten Hoeffler, ETH Zurich:
  - AI is a Gravity well.
    - We won't get any more 64-bit performance in the future, it will all be lower precision
    - We will need mixed precision methods in scientific computing
  - AI/Data Science is a Gravity Well
    - Everything will be in Python
  - Cloud is a Gravity Well
    - Low latency networks will converge – bet on low latency "Ultra Ethernet".
- We will have more AI users than HPC users in less than 3 years.

IN MANY WAYS, AI VINDICATES THE “HPC WAY”

- AI needs fast interconnects. We had them, the cloud and the enterprise did not.
- AI needs message passing; MPI was built for HPC, but is now the standard library for transformer-based generative AI wave (e.g. ChatGPT, DeepSpeed, etc.).
- AI needs heterogeneity – GPUs for general purpose compute came out of the HPC world.

- This means AI needs HPC hardware (probably good) and HPC programmers (good if you are one, bad if you need to hire one).

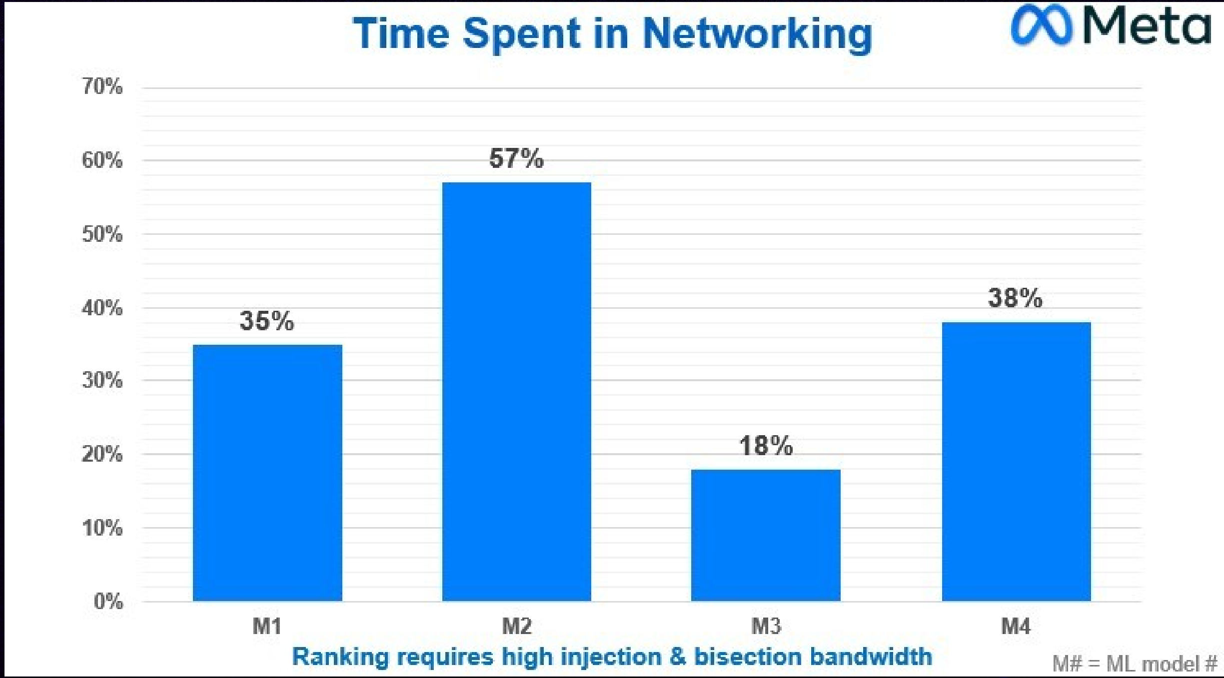
THE PAST 18 MONTHS HAVE BEEN A WATERSHED FOR AI

- ChatGPT was the “Sputnik moment” in an already building wave.
  - AI has been capturing headlines for the last 7-8 years.
  - The release of Transformers (from Google) unleashed the ability to scale to enormous sizes.
  - But ChatGPT changed everything, especially public perception.
- There is now a global ”AI Arms Race”, leading to a scramble (in both public and private sectors) for:
  - Funding
  - Expertise
  - Regulation/Policy
- I’m regularly hearing about billion-dollar machine orders paid for entirely by venture money to train products that don’t yet exist.
- AI and HPC are deeply intertwined – so academic HPC can’t pretend this is business as usual.

INTERCONNECTS ARE ONLY GROWING IN IMPORTANCE

- Interconnects have *always* been critical for HPC.
  - Mostly latency, but also bandwidth.
- The long time cloud rallying cry was “you don’t need all that expensive interconnect bandwidth if it’s not HPC”.
- Then AI came along. . .

INTERCONNECTS ARE ONLY GROWING IN IMPORTANCE – AI



- Often, one network rail per GPU
- Both latency *and* bandwidth seems to matter.
- The need for good interconnect is even *more* important than in HPC.
- And AI is the 800lb gorilla to HPC's modest sized chimp.
- This is unleashing new investments in networking.

AI HARDWARE WILL DOMINATE

- Per Hyperion:
  - The market for AI-driven hardware will be $300B/year in 2025.
  - The market for “pure” HPC hardware will be $10B/year in 2025
  - *Guess which will get more vendor attention?*

ADAPTING TO THE MARKET

- This isn’t actually a new problem in supercomputing.
  - And academics tend to lead the market on this.
- In 1991, the cold war was ending, which was killing the unlimited government budgets for vector-based custom silicon supercomputers. Cray, SGI, Thinking Machines, Convex, Raytheon Supercomputing, many other companies were falling apart – most didn’t survive.
- At NASA Goddard, Thomas Sterling and Don Becker started the “Beowulf” project exactly 30 years ago.
  - In Thomas’ exact words, those of us doing scientific computing needed to be “bottom feeding scumsuckers” - words I’ve built me career around ;-).
- The gist – silicon is expensive, use the commodity parts.
  - Step 1 – Don wrote network drivers for this thing called “Linux”. First time it talked via Ethernet. That worked out.
  - Step 2 – Come up with ways to use commodity processors.
  - Almost all Top 500 machines since have used this.
  - Even the addition of GPUs to HPC was about riding the commodity (gaming) markets.
- Universities led, agencies followed kicking and screaming (DOE still makes NRE investments with vendors).
- **WE CAN DO THIS AGAIN – and this time we have more to offer in the other directions.**

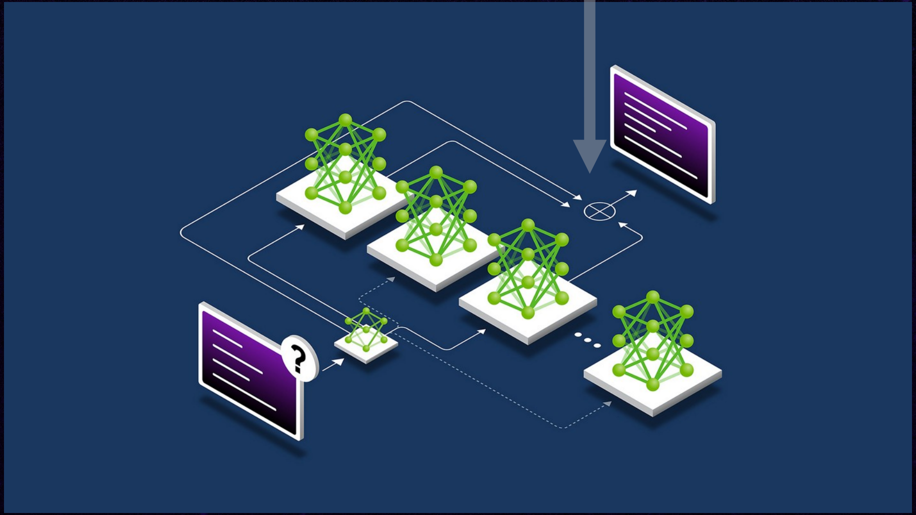
AI IS A YOUNG AND FAST CHANGING FIELD

- While AI research goes back more than 60 years, and Neural Nets have been around more than 30 years, the “modern” Deep Learning methodology has mostly been invented in the last five years.
- But it’s rich with Jargon. . .
  - RAG (Retrieval Augmented Generation)
  - MoE (Mixture of Experts)
  - Foundation Model, LLM, Hypothesis-Driven AI, etc., etc.
- It’s not possible for you to be *that* far behind in AI.
  - There are far more people working in it than have degrees in it, and many couldn’t spell AI ~7 years ago.
- However, once you get through the jargon, AI has both a great deal of cleverness, but also a lot of recycling of previous algorithms under new names.

YOUR AI TERM OF THE WEEK

They are about to discover this is a bottleneck

- MoE = “Mixture of Experts”
- An AI technique where a set of pre-trained “experts” are combined via a router to provide better answers than a single giant network.
- MoEs:
  - Are pretrained much faster vs. dense models
  - Have faster inference compared to a single model with the same number of parameters.
  - Require a lot of memory, as all models must be loaded at once.



This figure could have come from any 80s/90s Parallel Computing Textbook (though the graphics would be worse)

HPC, AI HARDWARE, AND SUSTAINABILITY

- To borrow from my friend David Keyes:
- *As computational infrastructure demands a growing sector of research budgets and global energy expenditure, we must enhance utilization efficiency.*
- As a community, we have excelled at this historically in three aspects:
  - architectures
  - applications
  - algorithms
- Among other opportunities, algorithmic opportunities abound:
  - reduced rank representations/ reduced precision representations

AI *HARDWARE* FOR SCIENCE
H100 PERFORMANCE ACROSS PRECISIONS

- *Source: NVIDIA*
- For Vector units, SP is unsurprisingly 2x DP.
- For Matrix units, it.s 15-1!!!
- At FP16, 2PF *Per socket*
- Maybe we need to spend a bit more time on using mixed precision Matrix ops, given **the 30X advantage**

FP64	34 teraFLOPS
FP64 Tensor Core	67 teraFLOPS
FP32	67 teraFLOPS
TF32 Tensor Core	989 teraFLOPS*
BFLOAT16 Tensor Core	1,979 teraFLOPS*
FP16 Tensor Core	1,979 teraFLOPS*
FP8 Tensor Core	3,958 teraFLOPS*

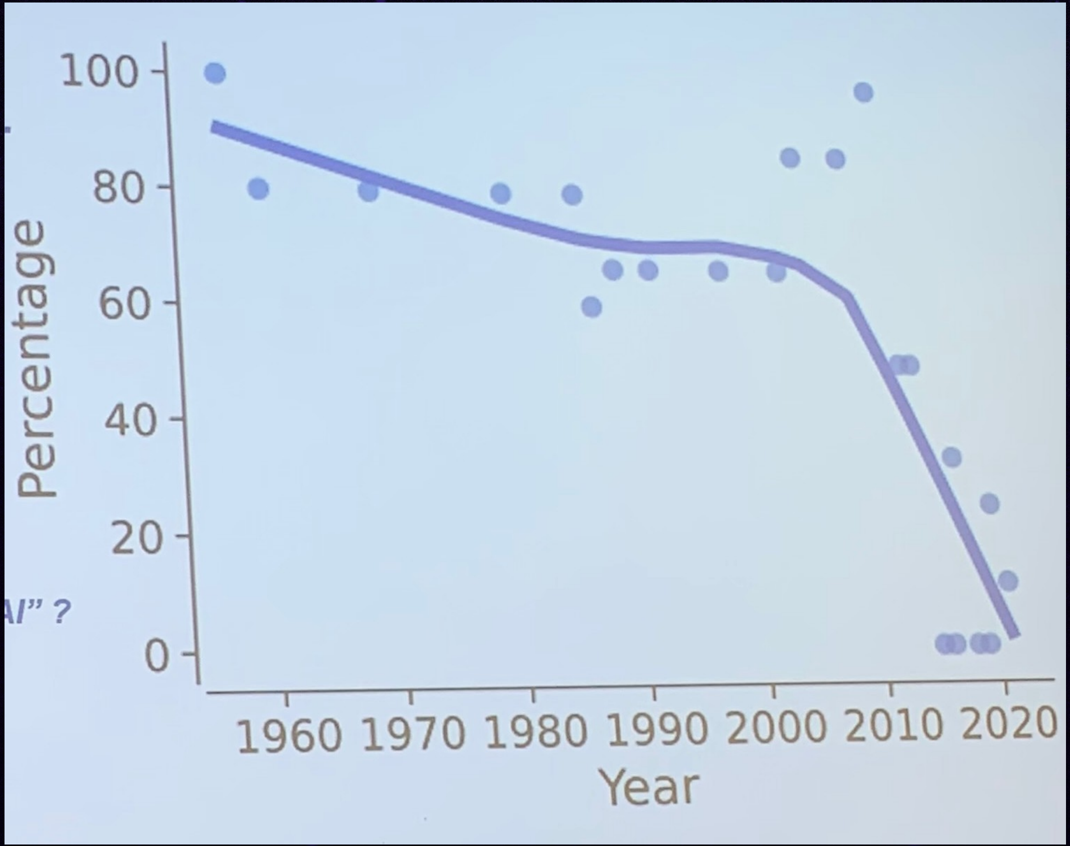
GPU ADVANTAGE – NAÏVE FIRST CUT

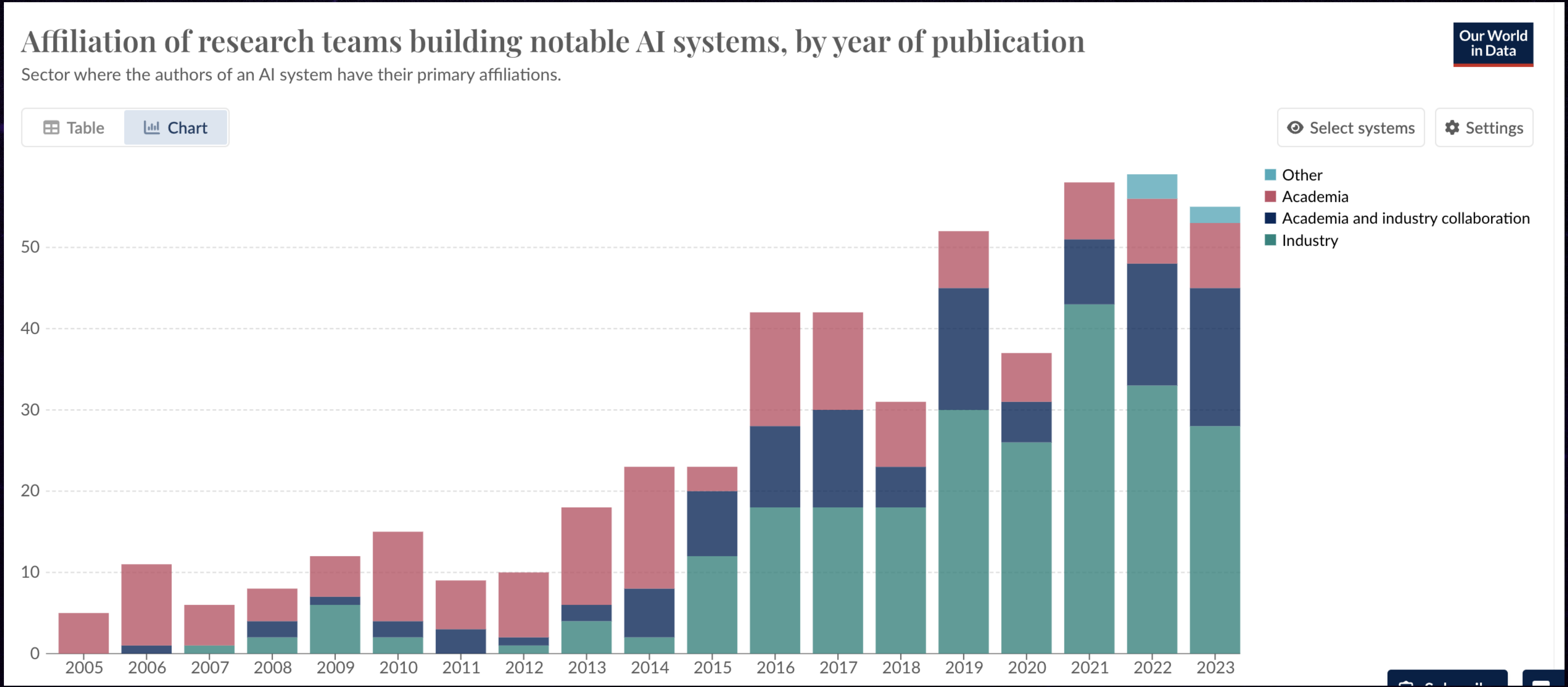
	TFlops	Watts	**Gflops/ Watt**	BW	**Flops/ Byte**
Intel ICX (Dual-Socket)	5.9	540	**10.93**	300	**20**
AMD Milan (Dual-Socket)	5.1	560	**9.11**	300	**17**
AMD MI250x	47.9	560	**85.54**	3277	**15**
NVIDIA A100	9.7	400	**24.25**	1600	**6**
NVIDIA A100 (Tensor)	19.5	400	**48.75**	1600	**12**

In terms of FLOPS/Watt, GPUs clearly win right now!

Even at this level, the GPU cost/TF advantage isn't that clear cut (Assume a node with two A100 cards cost 3x a node with no GPUs).

% OF LARGE SCALE AI RESULTS FROM US ACADEMIA




Affiliation of research teams building notable AI systems, by year of publication
Sector where the authors of an AI system have their primary affiliations.
Our World in Data
Table
Chart
Select systems
Settings
Other
Academia
Academia and industry collaboration
Industry
0
10
20
30
40
50
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023

AI WILL BE INCREDIBLY USEFUL, BUT LIKELY NOT UBIQUITOUS IN LIFE SCIENCES

WHERE WILL AI BE USEFUL?

- More places than you suspect. And probably half the talks at this conference will cover this.
- But the “AI Assistants” may become more useful than you think. . .

AI ASSISTANTS TO SYNTHESIZE, DISTILL AND EXTRACT KNOWLEDGE

- **A real example (courtesy Rick Stevens, Argonne) of a model trained on a ~quarter million genomics papers (in under half an hour):**
- **Seed Input:** CHD4 Chromodomain-helicase-DNA-binding protein 4

"Please tell me (using the knowledge you have been trained on) what you know about this Human gene and gene product " + line + ". In particular, we want to know the following information:

Is this gene well studied or is less well studied with some open questions about its function? Is the gene essential for survival?,

If this gene is mutated what impact does it have on phenotype of the individual? Is the gene or gene product a known drug target?, If it is a known drug target, what disease or diseases are associated with it? What other genes/proteins does this gene/protein interact with?, Is this gene or protein associated with Cancer?, Is this gene/protein involved in transcriptional regulation?,

Is it known what gene regulates this gene's expression? If you were starting out as a research oncologist, what might be a hypothesis you could explore related to this protein that would have significant scientific impact? Where possible give concise answers to these questions as well as describe the function of the gene more generally if it is known."

Model Output: CHD4

CHD4 (Chromodomain-helicate-DNA-binding protein 4) is a moderately well-studied gene involved in various biological processes. It encodes a multi-domain chromatin remodeling protein that plays a role in gene expression and chromatin organization.

While CHD4 is not absolutely essential for survival, it does have critical functions in cells. Mutations in this gene have been associated with several disorders and cancers. For example, mutations or deletions of the CHD4 gene lead to a rare syndrome called CHARGE syndrome, charactenzed by developmental and physical abnormalites. In cancers such as glioblastoma, CHD4 loss results in an altered cellular phenotype contributing to tumor progression and resistance to therapy.

Currently, the CHD4 gene or its encoded protein is no directly targeted by any approved drugs. However, researchers have identified small molecule inhibitors of its chromatin-remodeling function as potential therapeutic agents for various cancers.

As mentioned earlier, CHD4 mutations and loss are linked to glioblastoma and other types of cancer. By understanding the roie CHD4 plays in cancer progression and identifying mechanisms by which it interacts with other genes/proteins, new potential targets for therapeutic intervention may emerge.

- How well studied and role?
- Essential for survival?
- Any Drugs targeting this protein?
- Impact of Mutations?

CHD4 has indeed been found to interact with multiple proteins involved in chromatin remodeling, DNA repair, and transcriptional regulation. Some of its known interaction partners include HP1a (heterochromatin protein 1 alpha), MTA3 (metastasis-associated protein 2), and BRG1 (brahma-related gene 1).

In terms of cancer, CHD4 loss has been associated with the development and progression of glioblastoma, as well as other cancers such as lung ademocarcinomas. Since CHD4 is a components of the nucleosome remodeling deacetylase (NuRD) complex, its solr in chromatin organization and gene expression regulation could potentially contribute to those cancerous processes.

CHD4 plays an important role in transcriptional regulation through its involvement in the NuRD complex. The NuRD complex mediates histone deacetylation, which ultimately affects the accessibility of DNA for transcription factors. While it is not known which specific genes regulate CHD4 expression, epigenetic mechanisms and interactions with other chromatin-remodeling proteins are involved in its regulation.

As a research oncologist, an interesting hypothesis to explore could be the investigation of CHD4 as a potential biomarker or therapeutic agent for glioblastoma and other cancers associated with its loss or mutations. By understanding the downstream effects of CHD4 dysregulation in cancer cells and identifying how it interacts with other genes/proteins involved in tumorigenesis, new insights into targeted therapies might be gained. Additionally, exploring the role of CHD4 in normal cellular functions could provide further context for its abnormal behavior in mutated cells.

- What other proteins does it interact with?
- Associations with Cancer?
- Connections to transcriptional regulation?
- Hypothesis to pursue as a “research oncologist”?

WE CAN EXTRACT, CAN WE DISTILL?

- gpt_assistant_prompt = "You are a super smart AI that knows about science. You follow directions and you are always truthful and concise in your reponses.*

- gpt user_prompt = "Please summarize in approximately 800 words this paper. [## BEGIN PAPER "+ chunk1 + END PAPER##] Your summary should contain the TITLE of the paper. the YEAR the paper was published, the KEY FINDINGS. the MAIN RESULT, any HYPOTHESES the paper proposes or that you can infer from the text when the hypothesis is not explicit.

Please propose EXPERIMENTS that would validate the hypothesis. In addition to generating the summary. please generate a list of up to ten KEYWORDS that are relevant to the paper.

Please include these UPPER CASE words as headings in your response."

TITLE: CHD4 Coordinates Signaling and Repair After DNA Damage

YEAR: 2010

KEY FINDINGS:

- The adenosine triphosphate (ATP)-dependent chromatin-remodeling protein CHD4 (chromodomain helicase DNA-binding protein 4) becomes transiently immobllzed on chromatin after ionizing radiation (IR).
- Knockdown of CHD4 triggers enhanced Cdc25A degradation and p21Cip 1 accumulation, leading to more pronounced cyclin-dependent kinase inhibition and extended cell cycle delay.
- At DNA double-strand breaks (DSSs), depletion of CHD4 disrupts the chromatin response at the level of the RNF168 ubiquitin ligase, impalring local ubiquity lation and BRCA1 assembly
- These cell cycle and chromatin defects are accompanied by elevated spontaneous and IR-induced DNA breakings, reduced efficiency of repair, and decreased clonogenic survival.

MAIN RESULT:

- CHD4 is a novel genome caretaker that facilitates both checkpoint signaling and repair events after DNA damage.

HYPOTHESIS:

- The study suggests that CHD4 plays a crucial role in the coordination of DNA repair and cell cycle checkpoint signaling after DNA damage.

EXPERIMENTS TO VALIDATE THE HYPOTHESIS:

- Overexpression of CHD4 in cells undergoing DNA damage to repair and cell cycle checkpoint signaling.
- Knockdown of CHD4 in cells with detective DNA repair pathways to observe it it further impairs DNA repair cell cycle checkpoint signaling.
- In Vivo studies to observe the effects of CHD4 deletion or overexpression on tumor development.

WHERE WILL AI NOT BE USEFUL? (MEDICINE)

- AI (Deep Learning, really) needs a *lot* of data to be effective.
- In many ways “Personalized Medicine” will be the antithesis of AI (which will be really good at population-type things).
  - Preaching to the choir, but…
    - There are *lots* of kinds of tumors.
    - There are *lots* of variations between humans.
    - Can we build a big enough database to extrapolate a specific person’s specific tumor?
- While AI can scan the literature for treatments, personalized cancer treatment, for instance, could really benefit from first-principles mechanistic models of tumor growth – and AI may *never* solve that piece (though it might run the model for you and make recommendations).

AI RISKS AND CHALLENGES

- **Trust – Verification and Validation**
  - Seriously, we have to tell it not to lie…
  - No methodology for associating an output with a “how”, or which input.
- **Keeping private data private**
  - The recent “Company Company Company” example shows it may blurt out private training data at any time.
  - This *completely* defeats federated learning
  - We have no way to know what it *remembers*, any more than asking a human to write down everything they knew as useful ... there are still "oh yeah" moments when something triggers a memory -- LLMs have this behaviour too.
- **The data is mostly terrible, so the results will have issues too.**

LLM BEHAVIORS

- All an LLM can (should) do is predict the next word in a sequence, based on past patterns.   That's it.
- And yet, we have to tell it not to lie, we have to tell it to be helpful, we have to tell is to be concise.
- They are fascinatingly human sometimes, despite being built from simple, simple mechanisms…
  - Perhaps humans are too -- we just have a lot of scale of neurons.
- Emerging behavior of systems at ***scale*** is a remarkable phenomenon…

TACC IS HERE TO BRING YOU SCALE. . .

LET'S SEE WHAT EMERGES!

THANKS!
DAN@TACC.UTEXAS.EDU

THE VENDOR SPACE

- I was reading the release notes on a common, well-known open source code recently.   It noted that it supported most large scale supercomputers, including:
  - “IBM Blue Gene, SGI, SUN, and Cray”
  - What all those things have in common is that they **no longer exist**.
- For a long time, you would buy processors from Intel or AMD, Infiniband from Mellanox or Qlogic, and accelerators (if you had them) from NVIDIA.
- (BTW, there are dozens of new AI chips – they still have zero penetration in HPC).

IN THE END, THERE CAN BE ONLY ONE.

- NVIDIA bought Mellanox, and started making CPUs (and, you might notice, systems).
- Intel bought up Qlogic, made IB into Omnipath, killed it, then spun it off as Cornelis again, but still pretty tightly coupled.   Intel also has started making HPC GPUs, and pushing DAOS for filesystems.
- AMD is now making CPUs and GPUs.   As no interconnect is aligned, at large scale they are exclusively using HP-E Slingshot. Curiously, HP-E seems to be *only* delivering AMD systems at scale this year.
- See the trend? Good luck buying Infiniband for your AMD GPU cluster. (You can get it, but since delivered price<<1/5th list price, there are. . .complications).
- This is probably. . . Not good for innovation.

THEN THERE IS THE CLOUD, WHICH MEANS AMAZON, GOOGLE, OR MICROSOFT JUST MAGICALLY DELIVER YOU SERVICES…


FRONTERA
TACC
NSF
TEXAS

NVIDIA HPC SDK-PROGRAMMING MODEL

Proven, Productive, Portable and Performant

```
std::transform(par, x, x+n, y, y,
    [=](float x, float y){
        return y + a*x;
});
```

```
do concurrent (i = 1:n)
   y(i) = y(i) + a*x(i)
enddo
```

GPU Accelerated C++ and Fortran

```
#pragma acc data copy(x,y)
{

...

std::transform(par, x, x+n, y, y,
    [=](float x, float y){
        return y + a*x;
});

...

}
```

Incremental Performance Optimization with Directives

```
__global__
void saxpy(int n, float a,
           float *x, float *y) {
  int i = blockIdx.x*blockDim.x +
          threadIdx.x;
  if (i < n) y[i] += a*x[i];
}

int main(void) {
  ...
  cudaMemcpy(d_x, x, ...);
  cudaMemcpy(d_y, y, ...);

  saxpy<<<(N+255)/256,256>>>(...);

  cudaMemcpy(y, d_y, ...);
```

Maximize GPU Performance with CUDA C++/Fortran

GPU Accelerated Math Libraries

SO, WHAT DOES THIS MEAN FOR YOU, THE RESEARCHER/USER?

- ”Ready for Future” – use new parallel constructs in C++/Fortran/Python
  - (Expect this not to be “most performant” in the near term).
- Otherwise
  - Use Directives for thread-level parallelism (OpenMP/CUDA/OpenACC)
  - Use MPI for task-level parallelism, plus one of the above thread systems.
- Align MPI tasks with layout of chips (one AMD chiplet per task, one per GPU. . .), try to pin to sockets so memory doesn’t move.
- Think about scale – the checkpoint interval on your laptop is not the same as on 100k cores.

- Codes can work at this scale… but many still don’t.

MY TURN. . .

- (Programming) Language of choice?
- Incorporating AI?
- Ever think about I/O strategies?
- Do you use Jupyter?

- How much of your time goes to code/data (and related issues) vs. science/engineering?

A FEW QUICK COMMENTS ON THE CLOUD.

- Let’s assume:
  - Cloud storage if free *(it’s not).*
  - Cloud Parallel Filesystem is free *(it’s very much not)*
  - Cloud networking/transfer is free *(it’s not)*
  - Tightly coupled interconnects don’t matter *(they do).*
  - You would get 1-1 performance for large jobs on clouds vs. Frontera *(you won’t).*
  - There is zero effort to get diverse technical workloads to run on the cloud *(there is a lot).*

A FEW QUICK COMMENTS ON THE CLOUD.

- Even with those ridiculously optimistic assumptions:
  - The equivalent cloud cost for delivered cycles on Frontera for FY2021 is ***$291,000,000***.

- Turns out, $291M > $24M.   Any questions?