AlphaFold3 Workload Characterization:
A Comprehensive Analysis of Bottlenecks and Performance Scaling

Jinpyo Kim, Mingi Kwon, Jishen Zhao
University of California, San Diego
SK Hynix

The Game Changer: **AlphaFold**

- Predict **3D Protein Structures**
- **Revolutionizing** Structural Biology

200M Million
Predicted Protein Structures

Years → Minutes
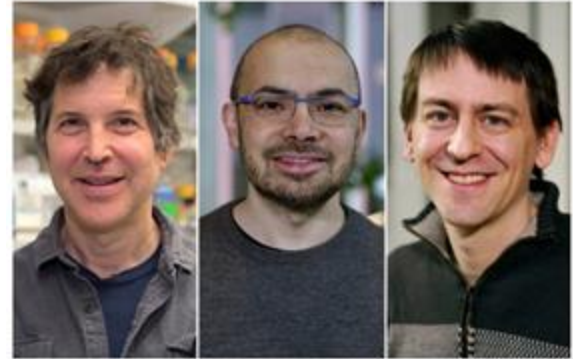Time per Structure
0:03:47

NEWS | 09 October 2024

Chemistry Nobel goes to developers of AlphaFold AI that predicts protein structures

This year's prize celebrates computational tools that have transformed biology and have the potential to revolutionize drug discovery.

By Ewen Callaway

David Baker, Demis Hassabis and John Jumper (left to right) won the chemistry Nobel for developing computational tools that can predict and design protein structures. Credit: BBVA Foundation

AlphaFold 3: Expanded Capabilities

- **Expanded Input Domain → Proteins, DNA, RNA, ligands**
- Predict **complex interactions** across **diverse biomolecules**

7PNM
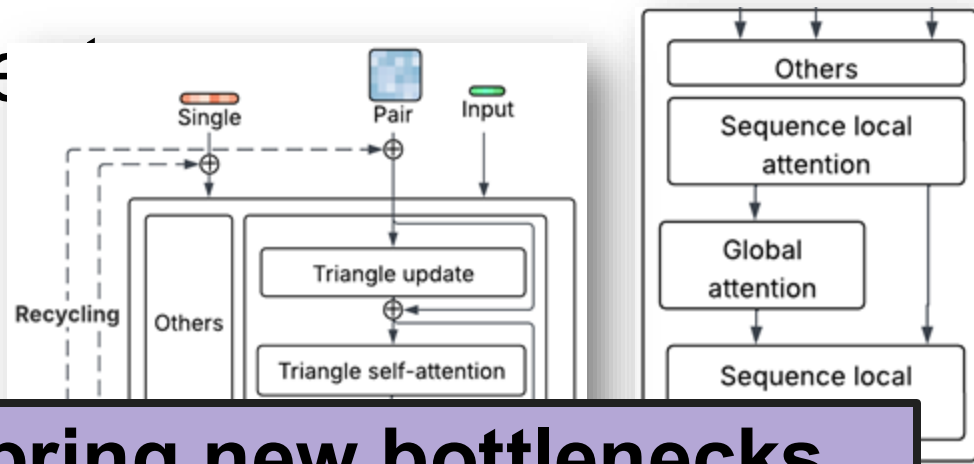
Handling diverse data changes performance behavior.

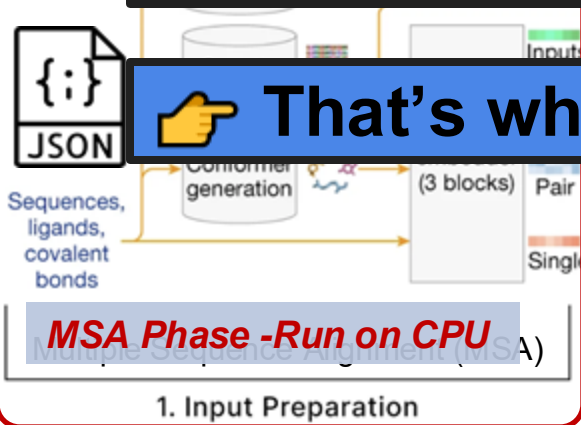👉 **That's why we need to profile AlphaFold 3**

Ground truth shown in grey

7PNM - Spike protein of a common cold virus (Coronavirus OC43)
Image credit: **Google DeepMind** (May 2024 blog post)

AlphaFold 3 Architecture

New Architecture bring new bottlenecks

👉 **That's why we need to profile AlphaFold 3**

MSA Phase -Run on CPU

Multiple Sequence Alignment (MSA)

1. Input Preparation

More details at https://elanapearl.github.io/blog/2024/the-illustrated-alphafold/

Hardware Baselines: Server and Desktop Platforms

● Server: Recommended by DeepMind.

Configuration	Server
CPU	Intel Xeon Gold 5416S
Core/Thread	16/32
Clock	2.0GHz (4.0GHz max)
L1/L2 Cache	80KB / 2MB per core
Last-level Cache	30MB shared
Memory Type	DDR5-4400
Memory Size	512GiB (32GiB x 16)
Mem. Expander	CXL (256GiB)
GPU	NVIDIA H100 80GB
Storage	PCIe 4.0 SSD

AFSysBench: A Reproducible Benchmark Suite

- **Automated Execution**: Automated testing of multiple, diverse input samples.
- **Thread Scaling**: Performance measurement across various thread counts (e.g., 1 - 8).
- **Integrated Profiling Tools**: Built-in support for detailed analysis using
 `Linux perf, AMD uProf, NVIDIA Nsight Systems`, etc.

TABLE II: Summary of Input Samples Used in AF3 Experiments.

Sample	Structure	Complexity	Seq. Length	Primary Benchmark Target / Workload Characteristic
2PV7	Protein (2 chains)	Low	484	Symmetric multi-chain processing
7RCE	Protein (1) + DNA (2)	Low-Mid	306	Baseline for mixed-type input
1YY9	Protein (3 chains)	Mid	881	Asymmetric multi-chain complex
Promo	Protein (3) + DNA (2)	Mid-High	857	MSA pipeline stress with low-complexity sequence
6QNR	Protein (9) + RNA (1)	High	1,395	High chain-count assembly with mixed input types

AFSysBench **enables us measure, profile, and compare performance of AF3 across diverse inputs and systems.**

Overall **Benchmark Results** and **Key Observations**

1. **The MSA Phase is the bottleneck.**
2. **Desktop** often **outperforms Server.**
3. **Inference bottlenecks shift** between platforms.
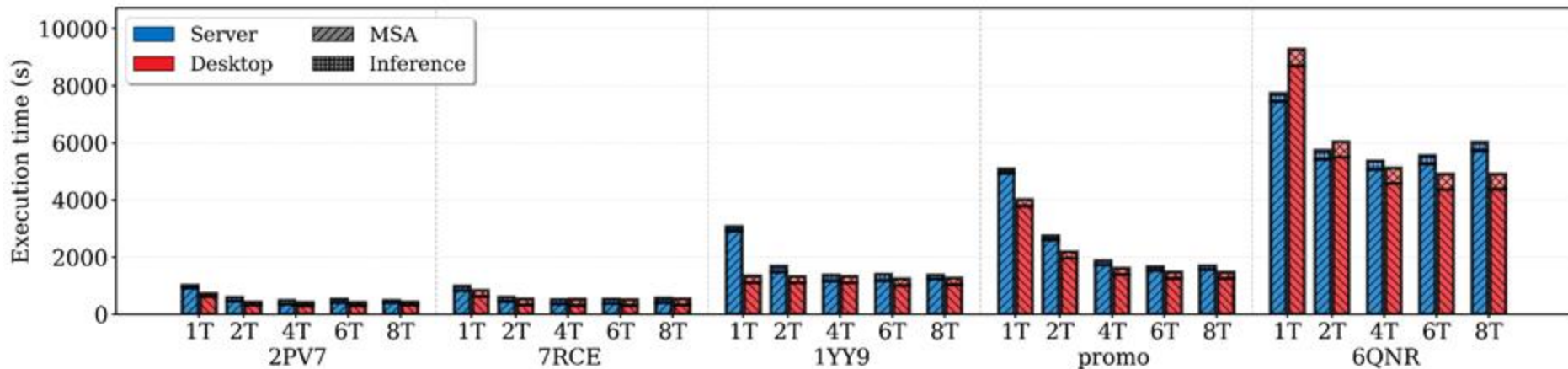

Total AF3 execution time across five input samples and two hardware platforms. Stacked bars indicate the relative contributions of MSA and inference phases under varying thread counts.

Observation 1: **The MSA Phase is the Bottleneck**

MSA dominates the **total execution time**

— taking from **75%** to over **94%.**

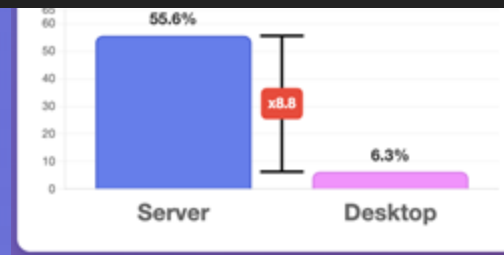The **MSA** phase is **essential** to **optimize** the overall **AlphaFold 3** process.
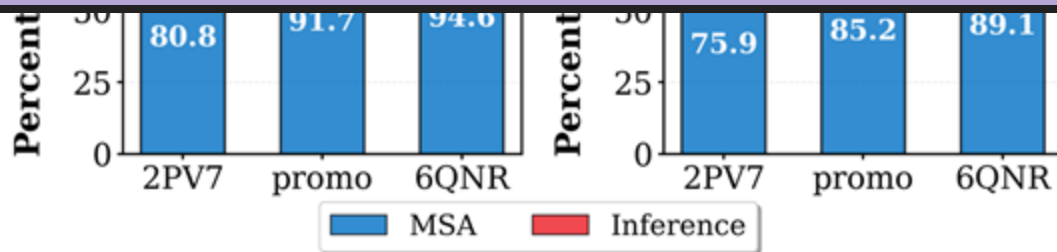
Observation 2: **Desktop Outperforms Server**

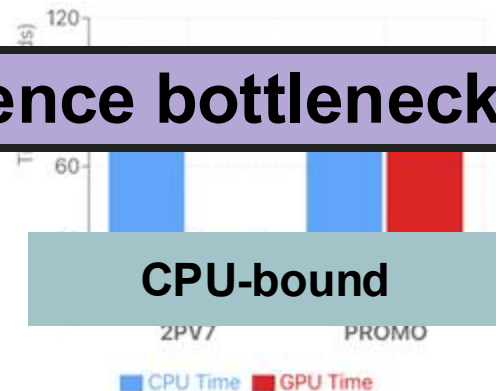**The desktop** can be a **cost-effective option** for typical protein inputs

Observation 3: **Inference Bottlenecks Shift between platforms**
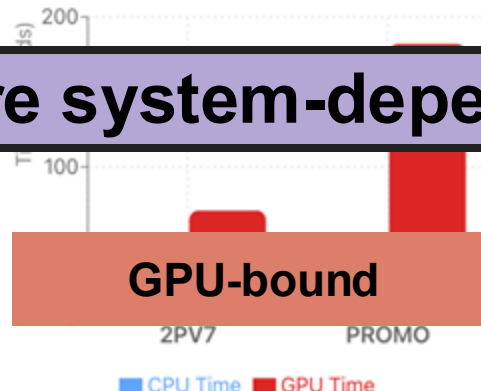
CPU vs GPU Compute Time

Server Platform

Intel Xeon + H100

Desktop Platform

AMD Ryzen + RTX 4080

Inference bottlenecks are system-dependent.

CPU-bound

GPU-bound

2PV7   PROMO

■ CPU Time ■ GPU Time

2PV7   PROMO

■ CPU Time ■ GPU Time

Inference time breakdown

Potential of CXL Memory in **AlphaFold3**

- **CXL** is a **high-speed interface** that connects **CPUs** with **devices** such as **memory** or **accelerators**.

1335-residue

1135-residue

935-residue

CXL memory unlocks new scalability for AlphaFold 3 in fixed-memory HPC environments.

~~memory from 512 GiB to 768 GiB to~~ support **RNA-heavy sequences.**
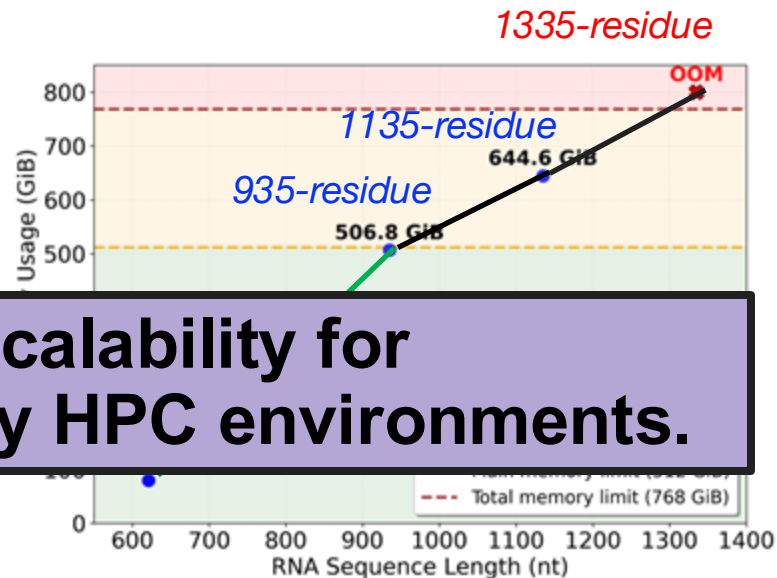
*Peak memory consumption as a function of **RNA sequence length.** Horizontal lines indicate the **main memory** and **total memory with CXL expansion**. ("Residues" – fundamental units that make up the RNA structure.)*

Conclusion

This **work** provides the **first in-depth, system-level performance characterization of AlphaFold3**, enabled by **AFSysBench,** a reproducible benchmark suite.

Key Takeaways:

- The **MSA phase** is the **dominant bottleneck.**
- **Desktop** shows **high performance efficiency** relative to cost.
- **Inference bottlenecks** are system-dependent.

Q&A

Code
Available

https://github.com/stable-lab/AFSysBench

Contact: jik066@ucsd.edu

STA
BLE