



HYPERION RESEARCH

Considerations for Cloud

ISC24 Market Update Briefing
May 2024

Mark Nossokoff

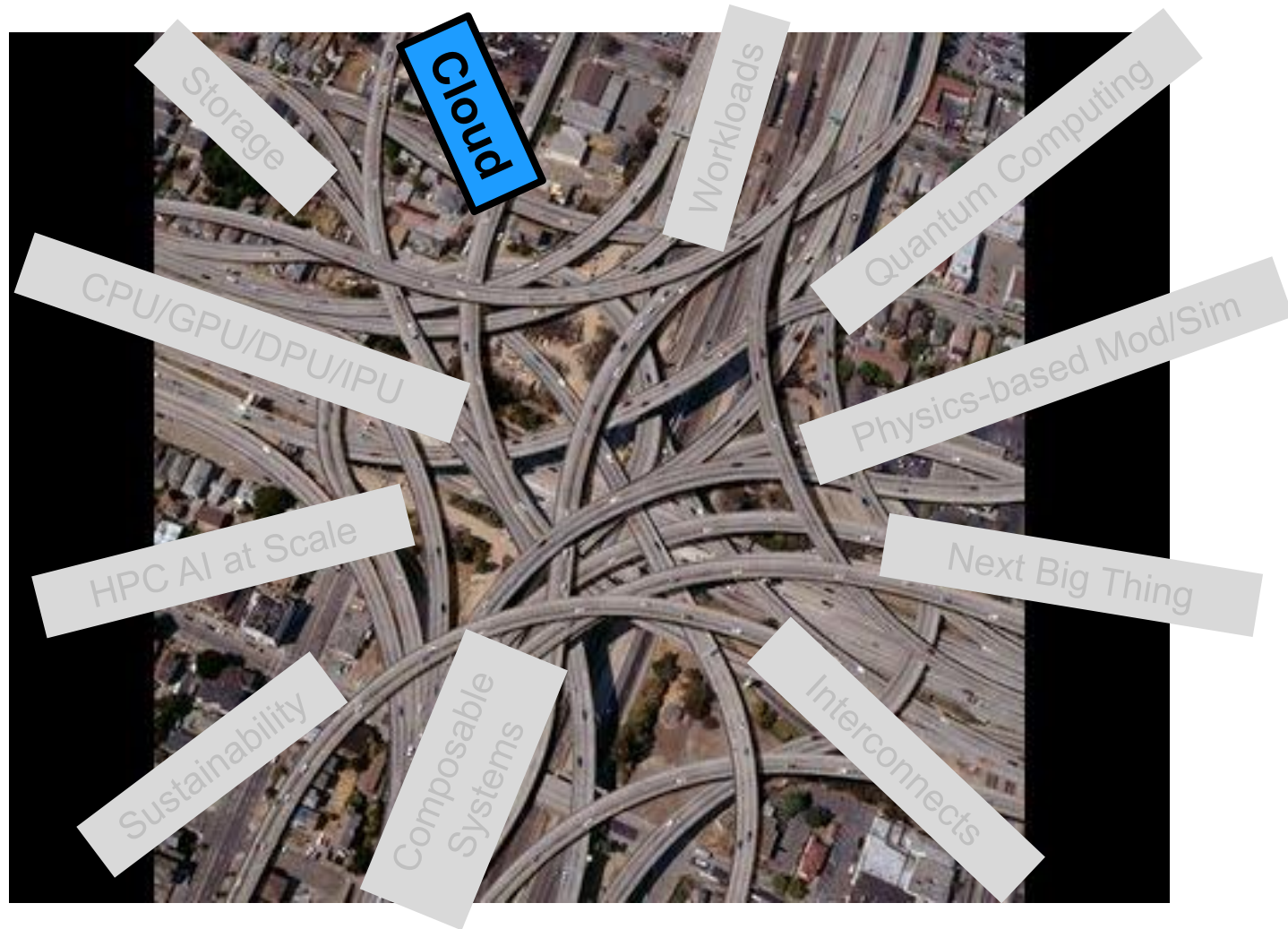
www.HyperionResearch.com
www.hpcuserforum.com

Not Your Father's HPC... ..Now Add AI

An Almost Untenable Increase in Complexity

HPC + AI

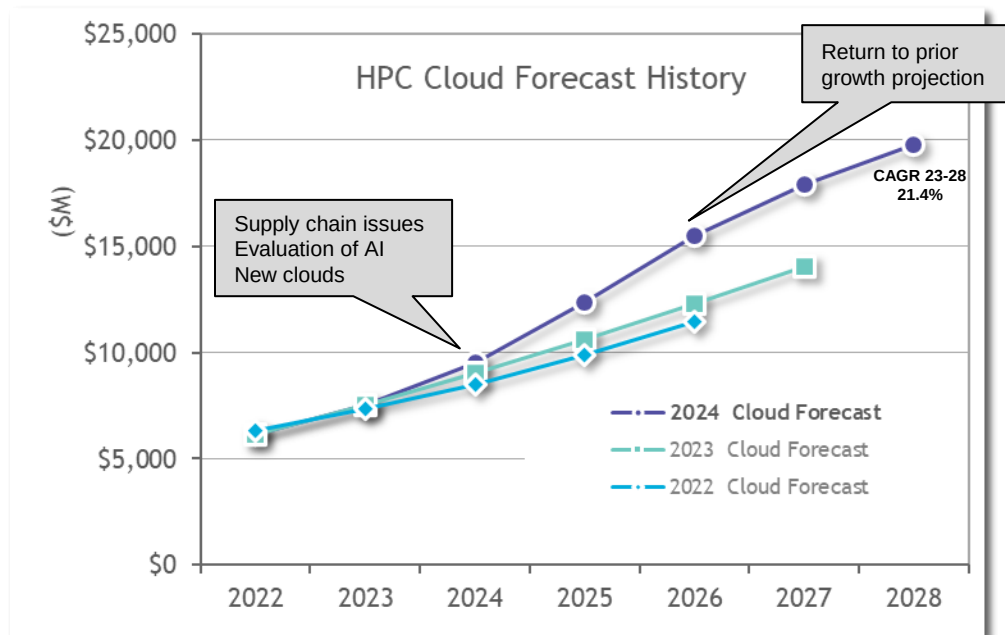
Research
Engineering
Scientific
Discovery
Business
Analytics
LLMs
SLMs
Multimodal
GenAI



HPC/AI Cloud Forecast

Cloud revenue expected to approach \$20B by 2028

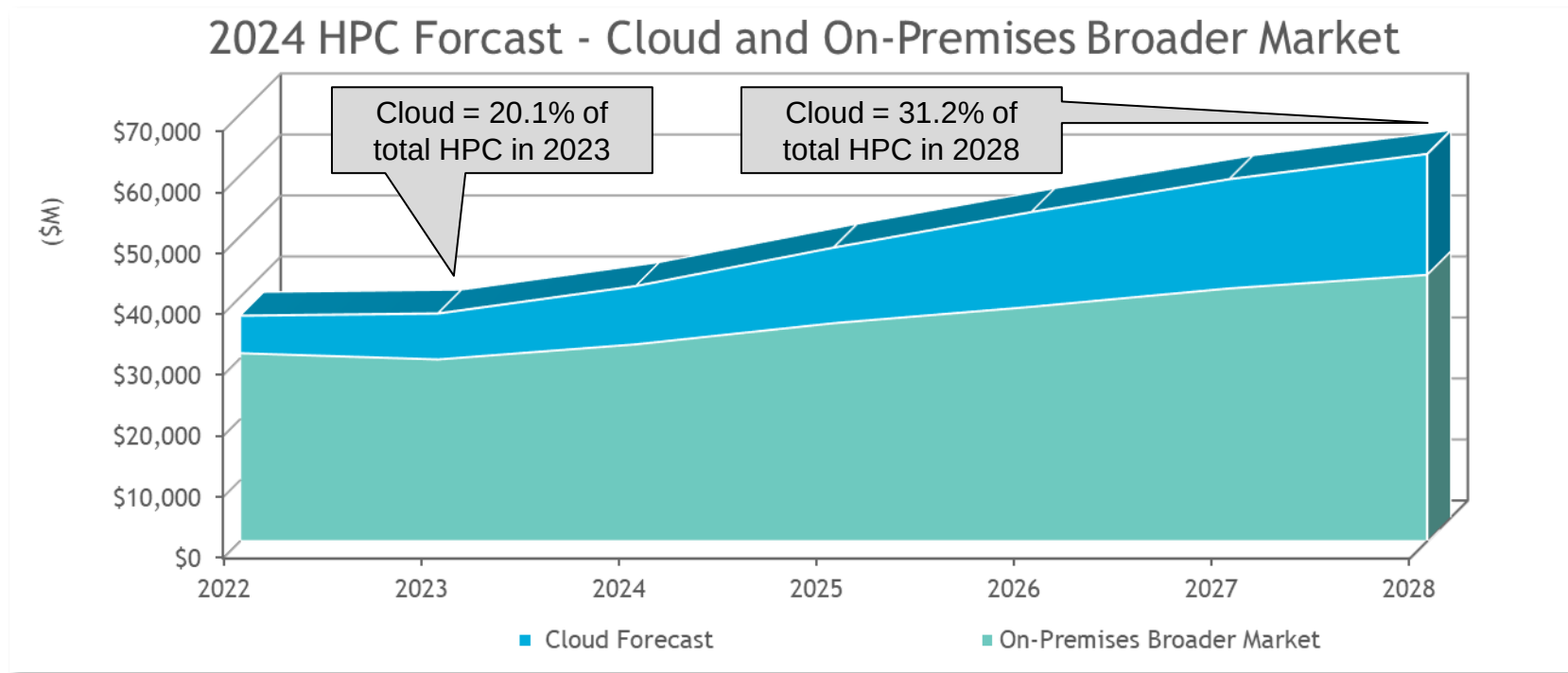
- **Global HPC/AI buyers around the world continue shifting portions of their on-premises budgets to spending in the cloud**
- **Increasing growth rate for 2024, returning to prior projected growth in 2026**
- **Primary growth drivers:**
 - Supply chain issues
 - Access to latest GPU technology
 - Experimentation and evaluation of AI workflow and infrastructure needs
 - Emergence of sovereign clouds



- **Forecast has steadily increased over previous forecasts**

The Total HPC/AI Market: On-Prem and Cloud Computing

The cloud market approaching 1/3 of total market in 2028



	2022	2023	2024	2025	2026	2027	2028	23-28 CAGR
Cloud	\$6,132	\$7,516	\$9,540	\$12,376	\$15,519	\$17,892	\$19,804	21.4%
On-Premises	\$30,813	\$29,791	\$32,265	\$35,735	\$38,420	\$41,433	\$43,656	7.9%
Total	\$36,945	\$37,308	\$41,805	\$48,112	\$53,939	\$59,326	\$63,460	11.2%

Cloud Providers' Responses to Demand

Rising cloud utilization driven almost entirely by AI adoption

- **Extensive investments by cloud providers**
 - Application-specific processors and accelerators
 - New data center buildouts
- **AI and GPU clouds**
- **Sovereign clouds**
 - Driven by local legal requirements, cost constraints, and tax & privacy policies, a shift to smaller, more adaptable data centers could result.
 - CSP's ability to dynamically shift load demand between geographic regions and national borders may be impacted
 - Local provisioning to become more critical
 - Potentially detrimental performance impact in counties where they get it wrong
 - What is the correct metric to key on to retain “sovereignty”?
 - Where data is created and where it can be moved, or...
 - ...security and governance of users who can access it, wherever it may be stored?

What's next in clouds?

Dynamic environment anticipated for the foreseeable future

- **Continued growth of user spending for HPC/AI advanced computing resources in the cloud from both new users and migration of workloads from current users**
- **More “specialization”**
 - Focused CSP accelerator and system designs for optimized performance and energy utilization
 - Capabilities (e.g, AI/GPU clouds) and outcomes (e.g., scope and complexity of workloads)
- **Build-out of new co-located hyperscaler data centers to support:**
 - New liquid cooling requirements of advanced architectures
 - Increasing number of “mid tier” service providers

Questions?



**We welcome questions,
comments, and suggestions**

**Please contact us at:
mnosokoff@hyperionres.com**