IBM **Research**

IBM: Pioneering Emerging Compute

Julian Warchall, Ph.D.
Technology Business Development Executive
IBM Research
Yorktown Heights, NY, USA
Julian.Warchall@ibm.com
(914) 945-3000

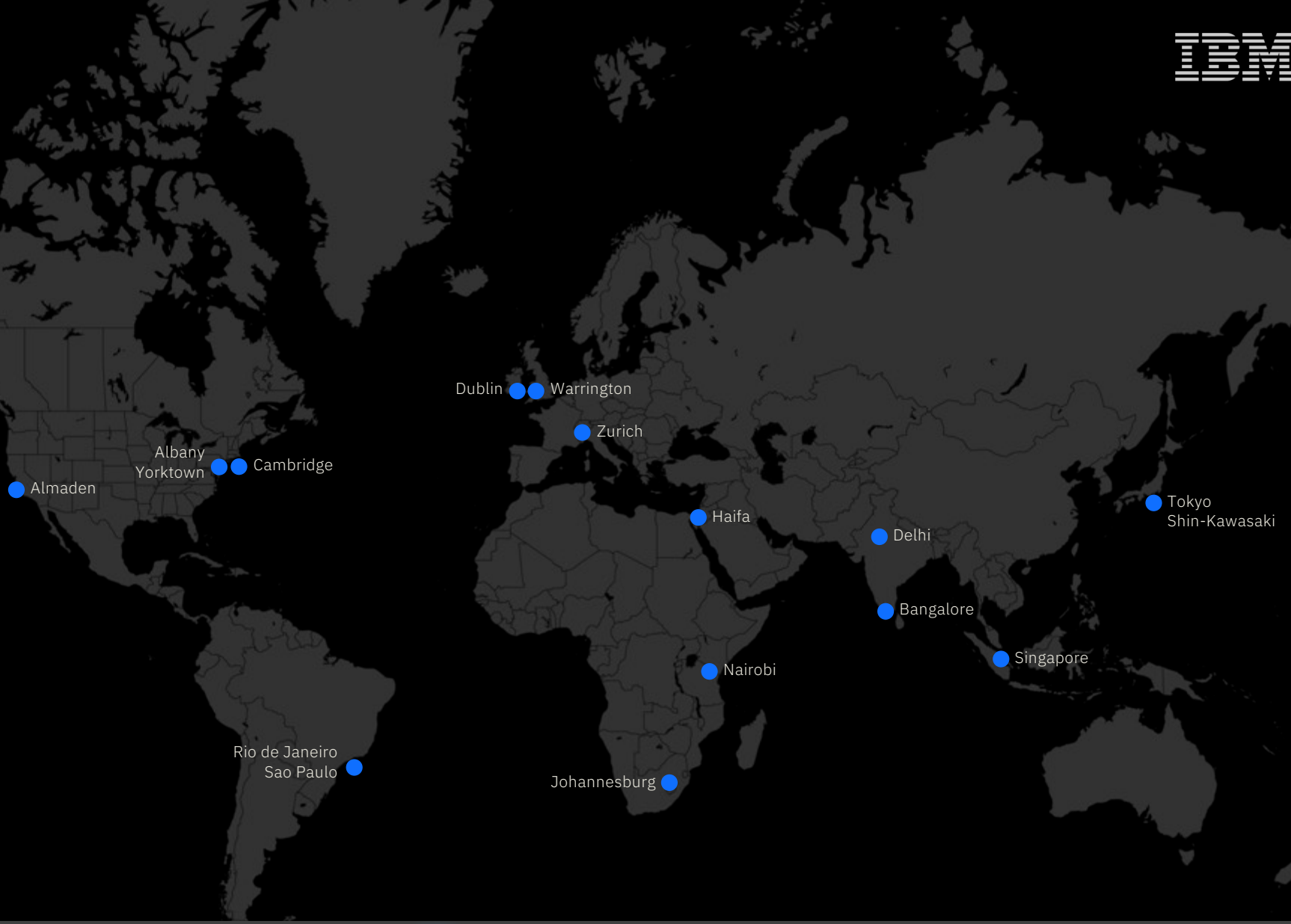
IBM Research / IBM Corporation © 2025

IBM Research Global Footprint

3,000
researchers

100s
of disciplines

>150,000
patents granted



6 Nobel Laureates

10 Medals of Technology

5 National Medals of Science

6 Turing Awards

IBM Research/© 2025 IBM Corporation

→ Yorktown Heights, NY

→ Cambridge, MA

→ Hursley, UK

→ Albany, NY

→ Zurich, CH

→ Tokyo, JP

IBM Research / IBM Corporation © 2025

The Future of Computing

Intersection of Bits, Neurons, & Qubits

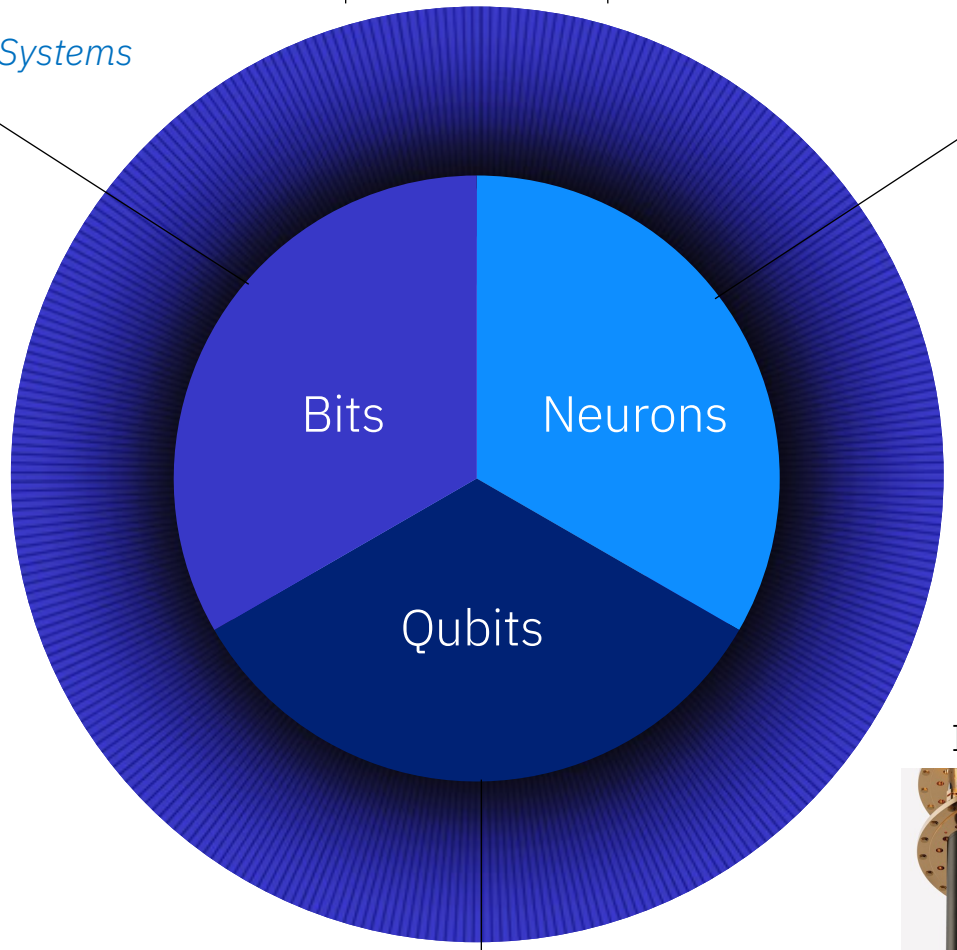
Hybrid Cloud

Secure, accessible, heterogeneous compute

Mathematics + Information

Ultra-Reliable High-Performance Systems

IBM z16



Cognition + Information

AI Systems

IBM AI Hardware

Physics + Information

Quantum Systems

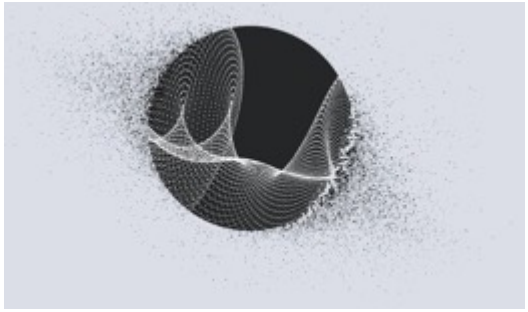
IBM Quantum

IBM Research / IBM Corporation © 2025

IBM Research Focus Areas

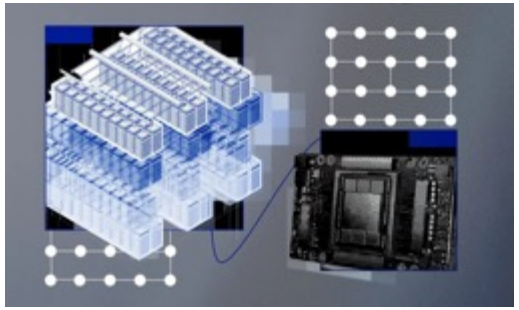
AI & Machine Learning

We're developing software, middleware, and hardware to bring frictionless, cloud-native development and use of foundation models to enterprise AI.



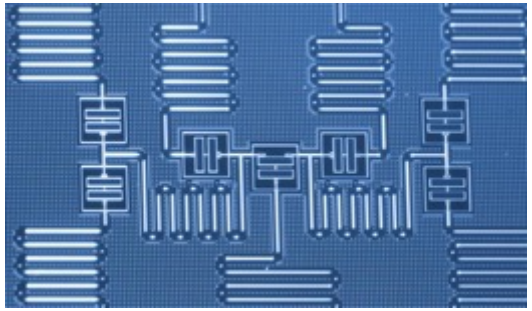
Hybrid Cloud

At IBM Research, we're designing new systems that provide flexible, secure computing environments — from bits to neurons and qubits.



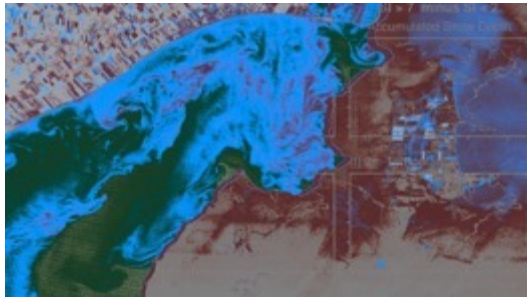
Quantum Computing

We combine quantum communication and computation to increase system capacity, and uses a hybrid cloud middleware to seamlessly integrate quantum and classical workflows.



Science

At IBM Research, we're tackling some of the most pressing challenges across computer science, materials discovery, climate change, drug discovery, physical sciences and sustainability.

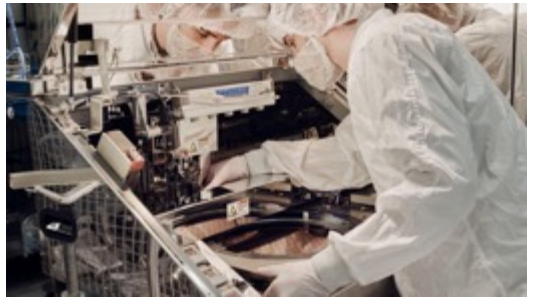


Security

Our pioneering technologies in confidential computing, decentralized trust and a secure supply chain will enable more secure, zero-trust infrastructures for all.

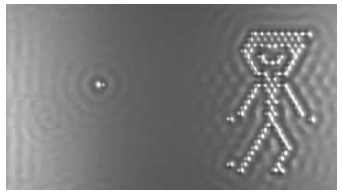
Semiconductors

We're pushing the boundaries of logic scaling as well as chiplet technology and design, and with an ecosystem of partners, we're moving innovations from our labs to the manufacturing line.



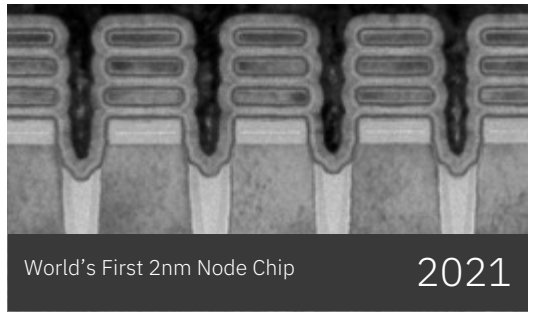
IBM Research / IBM Corporation © 2025

A Durable Legacy of World-Class Research

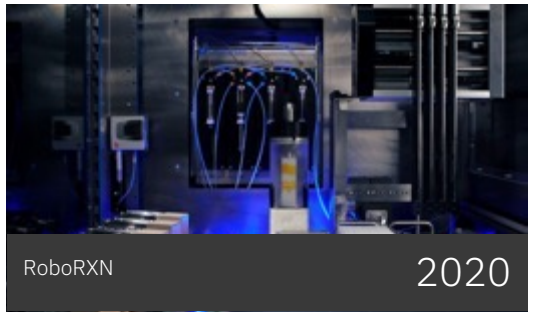


Artificial Intelligence Unit (AIU) 2022

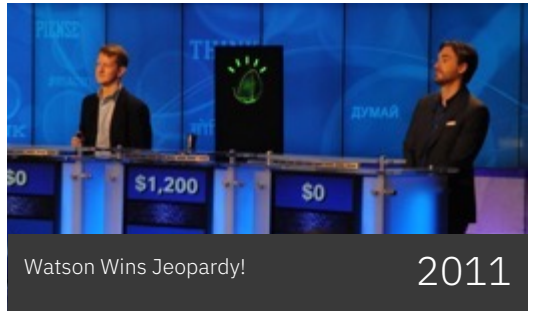
433-Qubit Osprey Processor 2022



World's First 2nm Node Chip 2021



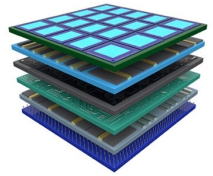
RoboRXN 2020

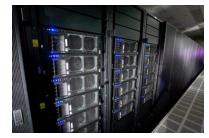


Watson Wins Jeopardy! 2011

Scanning Tunneling Microscope 1986

2023	**3D Heterogenous Integration**
2022	Artificial Intelligence Unit (AIU)
2021	World's First 2-nm Node Chip
2016	Quantum Computing in the Cloud
2012	**Atomic Imaging and Manipulation**
2011	Watson System for Jeopardy
2009	Nanoscale Magnetic Resonance Imaging (MRI)
2008	**World's First Petaflop Supercomputer (@ Los Alamos)**
2007	Web-scale Data Mining
2005	**Cell Processor (Sony Playstation)**
2004	Blue Gene/L
2003	5 Stage Carbo Nanotube Ring Oscillator
2000	Performance Java
1998	Silicon on Insulator (SOI)
1997	**Copper Interconnect Wiring**
1994	Silicon Germanium (SiGe)
1990	Chemically Amplified Photoresist
1987	**High-Temperature Superconductivity (Nobel Prize)**
1986	**Scanning Tunneling Microscope (Nobel Prize)**
1980	Reduced Instruction Set Computing (RISC)
1979	Thin Film Recording Heads
1973	**Modern Winchester Hard Disk Drive**
1971	Speech Recognition
1970	Relational Database
1967	Fractals
1966	**One-Device Memory Cell (DRAM)**
1957	FORTRAN
1956	Random Access Memory Accounting Machine (RAMAC)



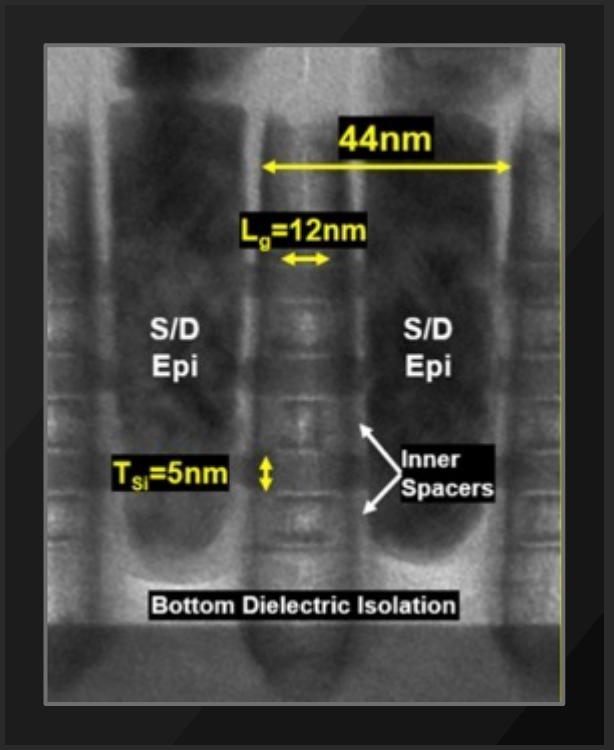


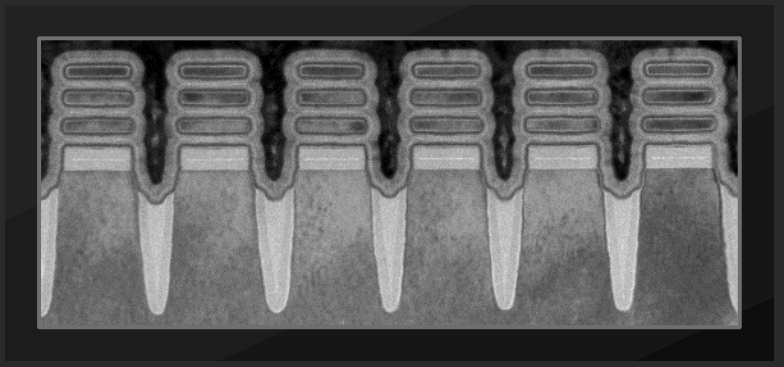
IBM Research / IBM Corporation © 2025

IBM Research produces the world's first 2 nm technology node.

45% better performance or

75% less power consumption compared to 7 nm technology.





Big Blue Goes Tiny With World's First 2nm Chip Tech — Forbes

WIRED
To Make These Chips More Powerful, IBM Is Growing Them Taller
The company reveals a process that it says can cram two-thirds more transistors on a semiconductor, heralding faster and more efficient electronic devices.

IBM Unveils World's First 2 nm Chip — EE Times

The New York Times
IBM on Thursday announced another leap in miniaturization, a sign of continued U.S. prowess in the technology race.

Rapidus – IBM Partnership

- Strategic partnership to build advanced semiconductor technology and ecosystem in Japan
- Deploy IBM's 2nm node technology into market-leading offering
- Leverage IBM's long history of successful joint development partnerships in semiconductors
- Rapidus engineers working alongside IBM at Albany Nanotech, at IBM Japan, and in Chitose

SONY NEC
DENSO TOYOTA NTT
KIOXIA SoftBank MUFG

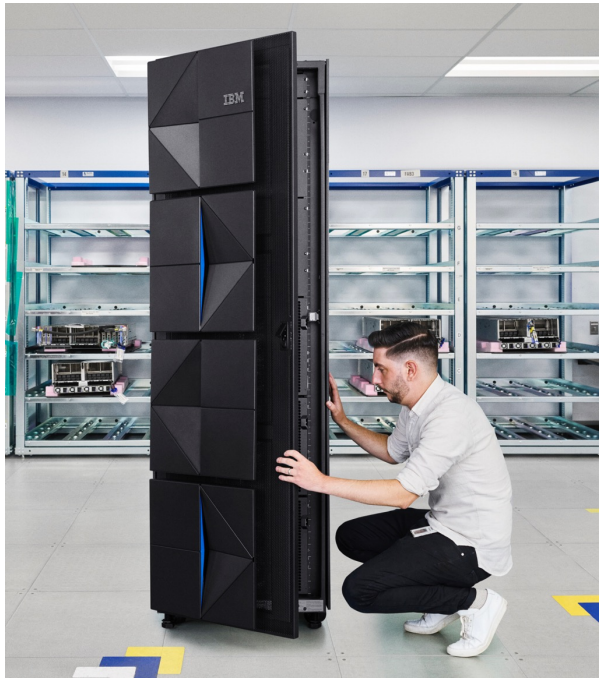
IBM Semiconductors

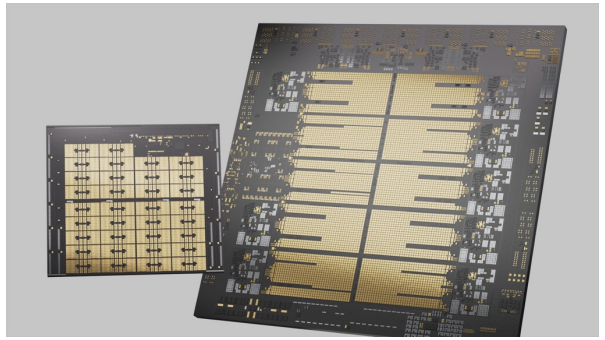
Semiconductor technology is central to IBM's core infrastructure business:
The most reliable, scalable, and secure computing system on the planet.

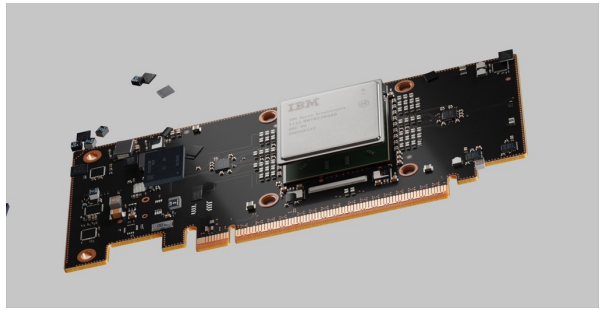
2000	2003	2005	2008	2010	2012	2015	2017	2019	2022	2025
z900	z990	z9 EC	z10 EC	z196	zEC12	z13	z14	z15	z16	z17
180 nm SOI	130 nm SOI	90 nm SOI	65 nm SOI	45 nm SOI	32 nm SOI	22 nm SOI	14nm FinFET	14 nm FinFET	7 nm FinFET	5 nm FinFET

50+ year track record of leading-edge performance and reliability

AI Accelerator Integration for IBM Z Systems







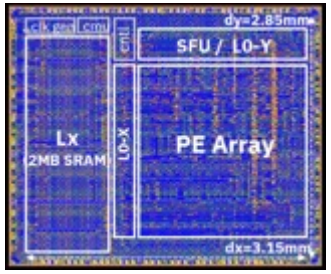
- **67 of the Fortune 100**
- **45 of the world's top 50 banks**
- **8 of the top 10 insurers**
- **4 of the top 5 airlines**
- **7 of the top 10 global retailers**
- **8 of the top 10 telcos**

- Focus: Secure reliable on-prem inference requirements
- AI accelerator (zAIU) integrated into Telum II processor plus AIU Spyre cards for large models
- Datacenter-class inference performance at 1ms response time
- Enabling real-time data inference for applications such as fraud detection

IBM Research AIU Family

2018

Gen 1 Core
for DARPA Seedling



14nm Technology
Proof of Concept

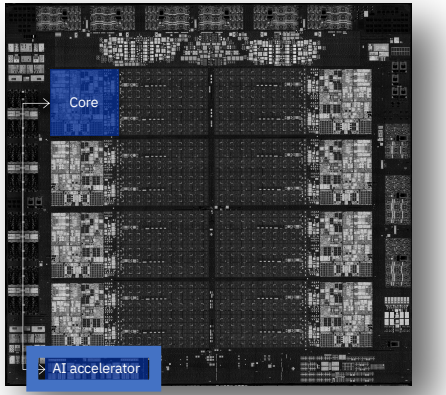
2019

AI HW Center
Launch

Full Stack
Ecosystem

2021

zAIU in IBM's
System Z
Telum processor



7nm Technology
Custom Productized SoC

2022

IBM Research
AIU Spyre

5nm Technology
PCIe Form Factor
DARPA-funded

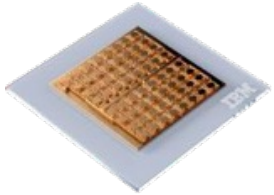
2024

IBM Research
AIU NorthPole

12nm Technology
VPX Form Factor
AFRL-funded

2023

IBM Research
Analog AI
DARPA-funded

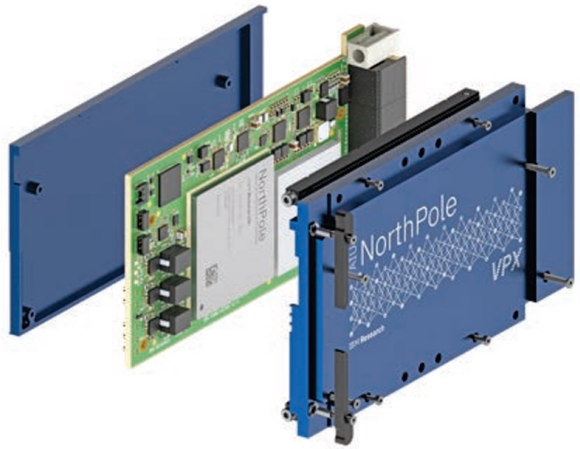


Innovation: Computation with less bit movement

Value: More efficient AI hardware

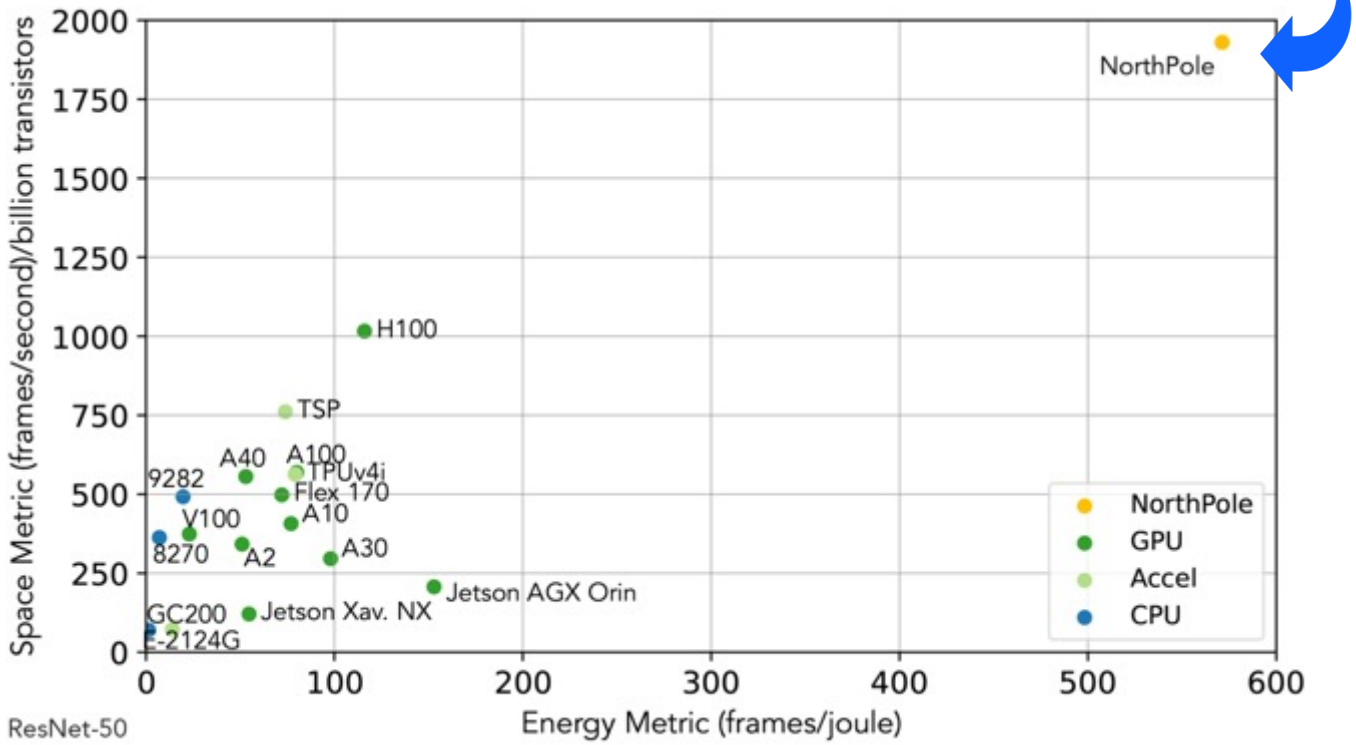
IBM Semiconductors | © 2025 IBM Corporation

AIU NorthPole



Brain-inspired accelerator chip
that supercharges edge AI by
working faster with far less power

Neural Inference at the Frontier of Energy, Space, & Time



IBM Semiconductors | © 2025 IBM Corporation

IBM AIU: Packaging Toward the Next AI Accelerators

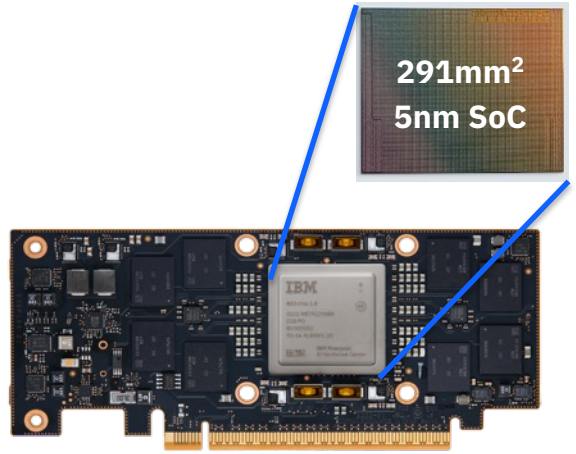
Key technology enablement:

- State-of-the-art foundry CMOS
- State-of-the-art silicon-verified IP blocks for support functions (memory controllers, I/O interfaces)
- **State-of-the-art advanced packaging including chiplets**

AIU SoC

Optimized for FM Inference

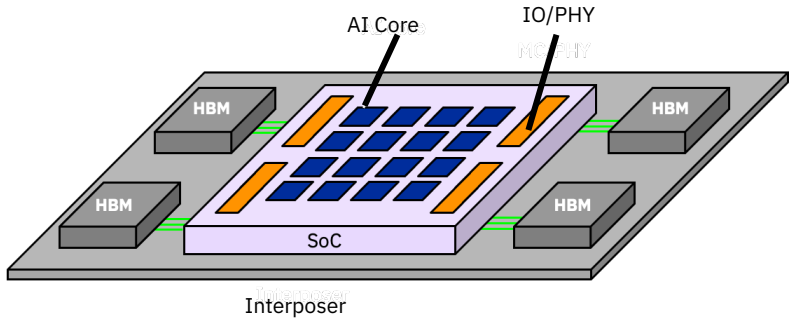
Leverages novel multi-precision inference architecture for energy efficiency and low latency



2.5D Vision

Optimized for FM Inference, Fine-Tuning, & Training

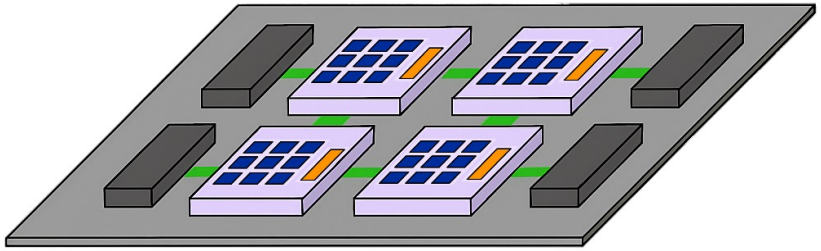
Adds HBM for increased performance



Chiplet Vision

Optimized for future very large FM Inference + Fine-Tuning + Training

Adds chiplet technologies for rapid development cycles and cost competitiveness



IBM Semiconductors | © 2025 IBM Corporation

Engaging with IBM AIU

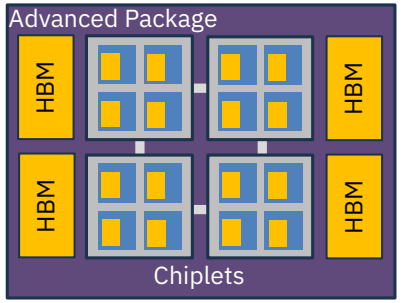
Today

On-Ramp with AIU Spyre

Begin development of models and workflows to handle business-specific workloads using proven product

Tomorrow

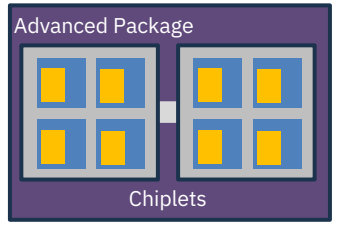
Access to AIU Roadmap



Leverage and benchmark enhanced performance from next-generation chiplet-based accelerators

Path Forward

Custom Systems in Package (SiPs)

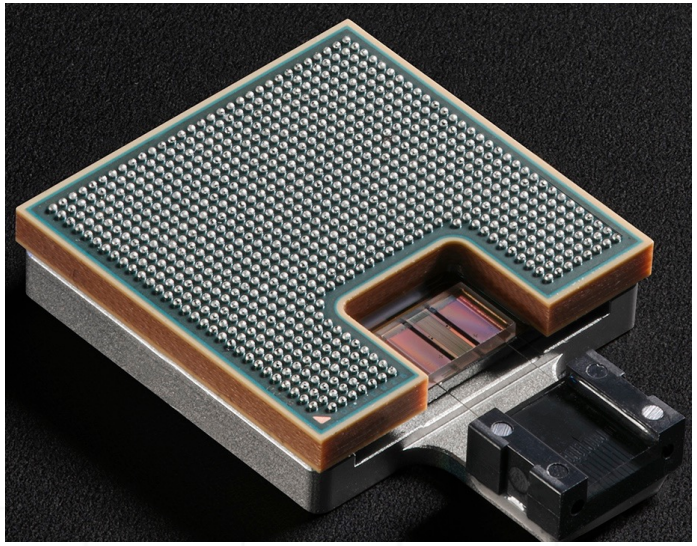


License AIU and build out bespoke application-specific hardware with minimal NRE

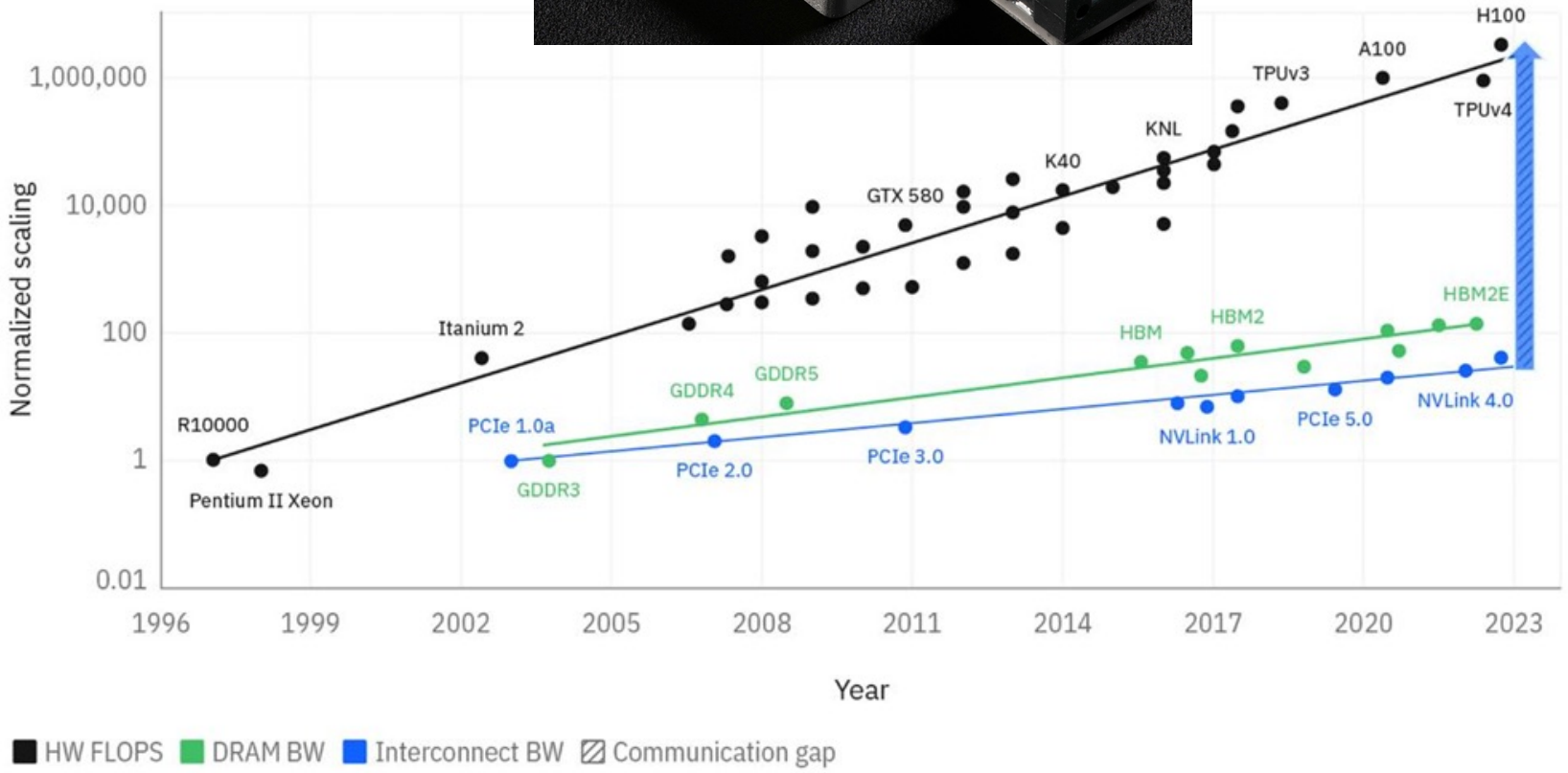
 | © 2025 IBM Corporation

Co-Packaged Optics

Closing the gap between large-scale compute nodes at the speed of light



- **80x higher bandwidth** than today's chip-to-chip communication
- **Lowers costs** for scaling generative AI:
  - Extends length of data center interconnect cables from ~1 to **100s of meters**.
  - **5x power reduction** over mid-range electrical interconnects
- Results in **5x faster AI training**

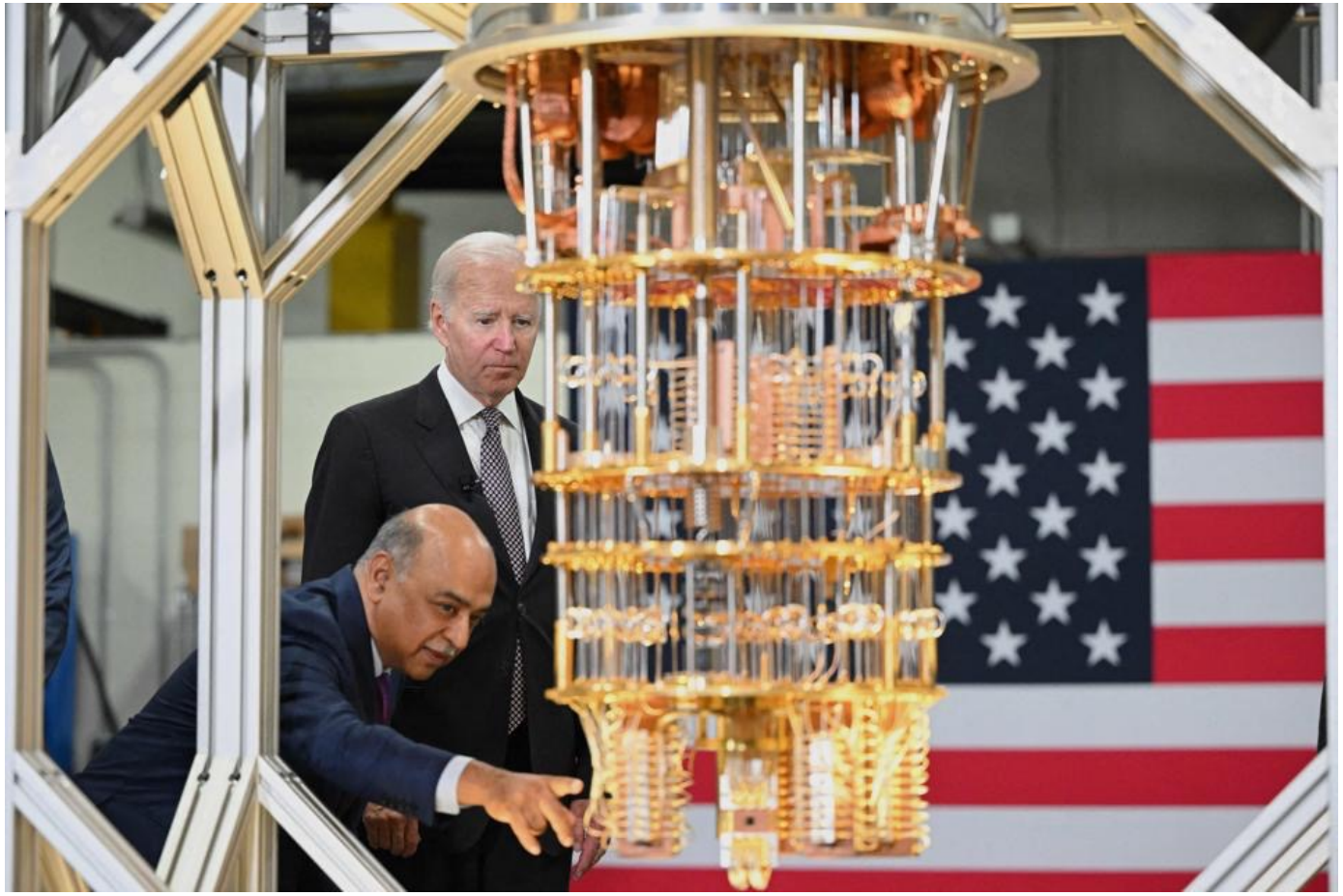


IBM Semiconductors | © 2025 IBM Corporation

IBM Quantum

Advances beyond classical computing to solve problems too complex for current computers.

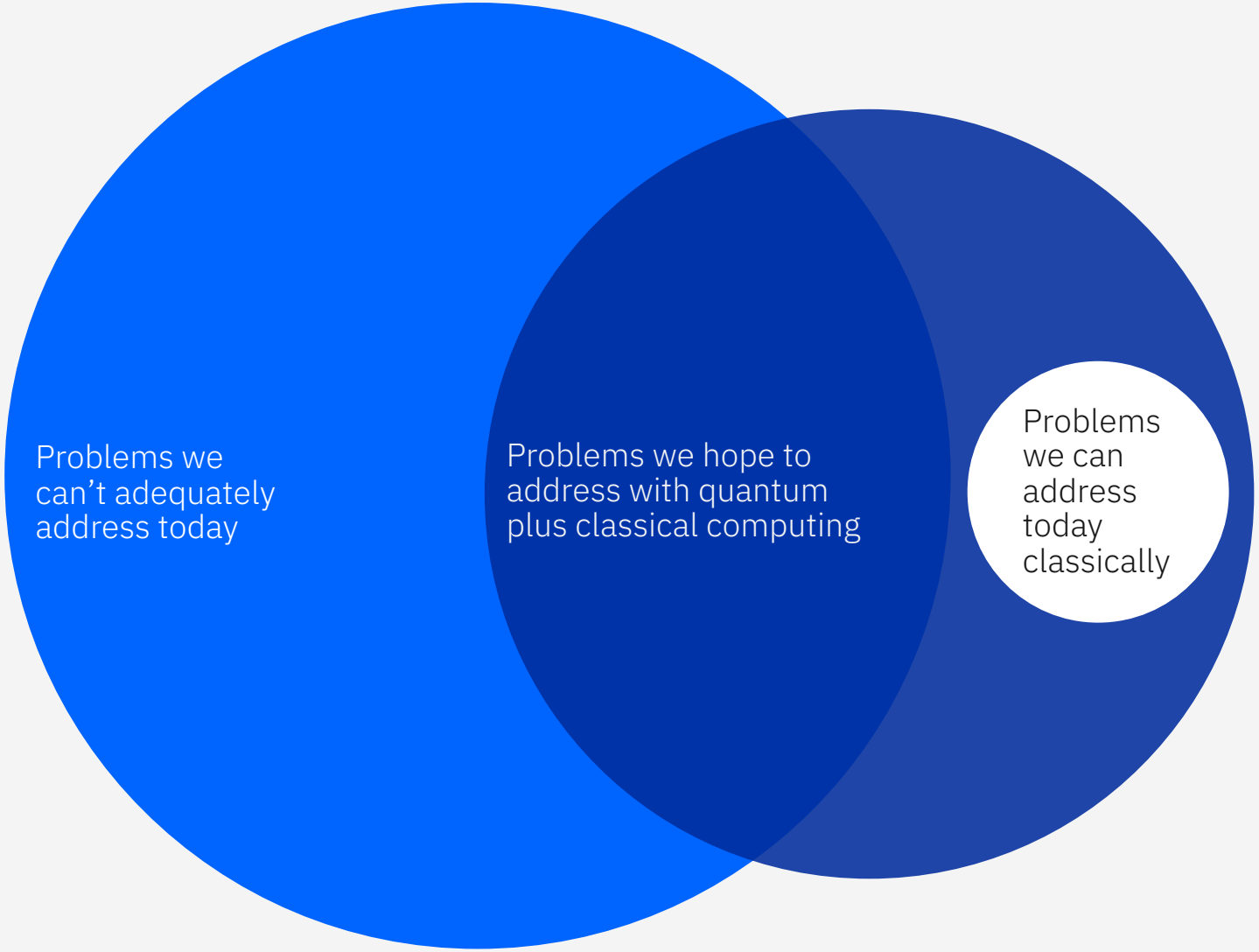
A critical technology area that will transform nearly every industry dependent on speed and processing power, from agriculture and financial services to health care and defense.



US President Joe Biden listens to IBM CEO Arvind Krishna as he tours the IBM facility in Poughkeepsie, New York, on October 6, 2022. (Photo by MANDEL NGAN/AFP via Getty Images)

IBM Research / IBM Corporation © 2025

Why Quantum?



Known Applications

Simulating Nature
- Physics
- Chemistry
- Materials Science

Data with Complex Structure
- Machine Learning
- Ranking in groups
- Factoring

Other (non-exponential)
- Sampling and Monte-Carlo problems
- Optimization
- Risk analysis and option pricing

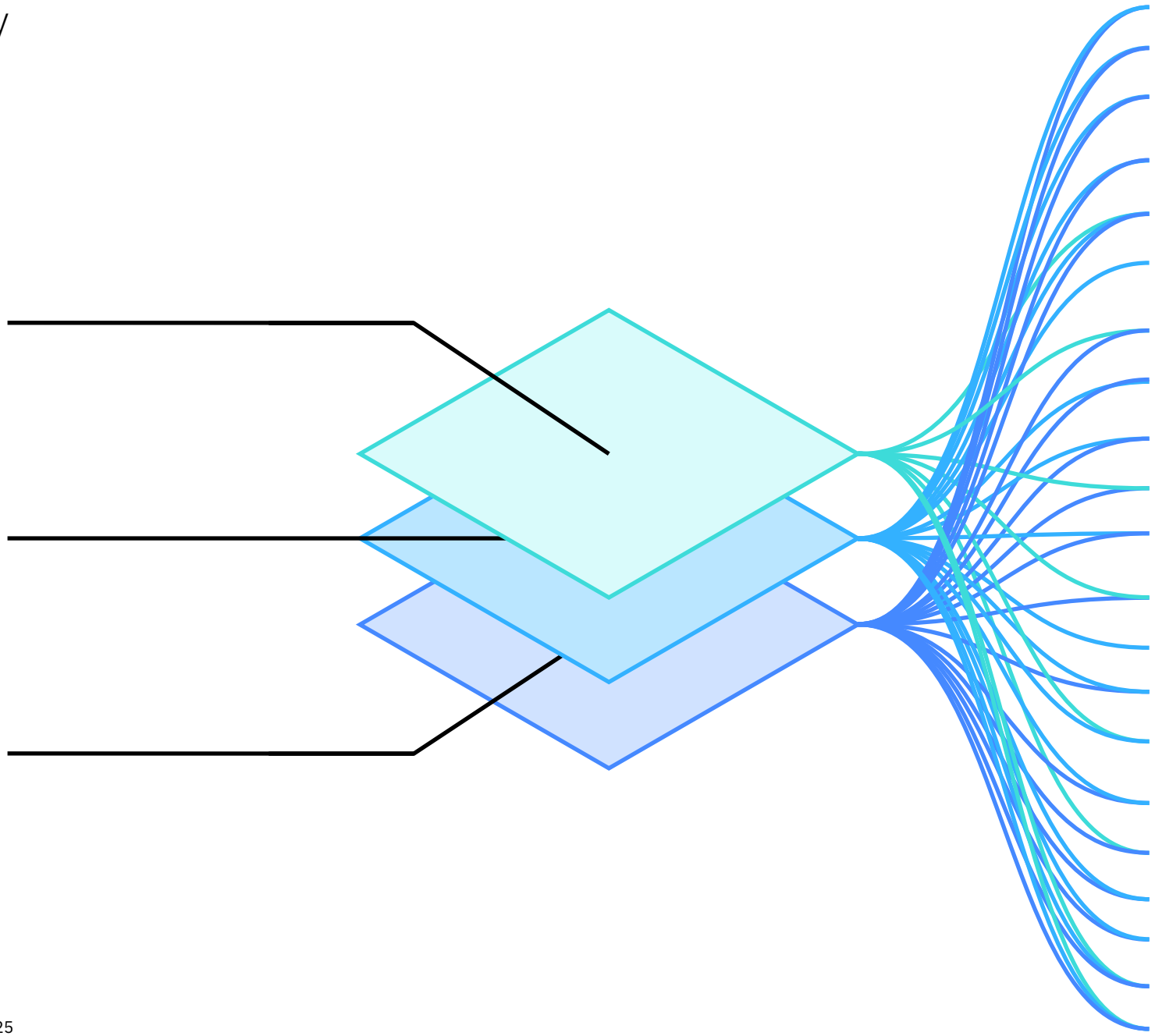
IBM Research / IBM Corporation © 2025

Quantum industry use cases

Simulating nature

Data with complex structure

Search and optimization



Crédit Mutuel

TEL

ERSTE Bank

vodafone

BOSCH

HSBC

JSR

Goldman Sachs

Woodside Energy

BOEING

JPMorgan Chase & Co.

bp

A leading insurance company

e·on

ExxonMobil

WELLS FARGO

A Leading Global Technology & Services Company

AMGEN

A Leading Global Consumer Products Company

SAMSUNG

LG

IBM Research / IBM Corporation © 2025

Development Roadmap

IBM **Quantum**

	2016–2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2033+
	Ran quantum circuits on the IBM Quantum Platform	Released multi-dimensional roadmap publicly with initial aim focused on scaling	Enhanced quantum execution speed by 100x with Qiskit Runtime	Brought dynamic circuits to unlock more computations	Enhanced quantum execution speed by 5x with Quantum Serverless and execution modes	Improve quantum circuit quality and speed to allow 5K gates with parametric circuits	Enhance quantum execution speed and parallelization with partitioning and quantum modularity	Improve quantum circuit quality to allow 7.5K gates	Improve quantum circuit quality to allow 10K gates	Improve quantum circuit quality to allow 15K gates	Improve quantum circuit quality to allow 100M gates	Beyond 2033, quantum-centric supercomputers will include 1000's of logical qubits unlocking the full power of quantum computing
Data scientists						Platform						
						Qiskit Code Assistant	Qiskit Functions Service	Mapping collections	Specific libraries			General purpose QC libraries
Researchers					Middleware							
					Qiskit Serverless	Qiskit Transpiler Service	Resource management	Circuit knitting x p	Intelligent orchestration			Circuit libraries
Quantum physicists			Qiskit Runtime Service									
	IBM Quantum Experience		QASM 3	Dynamic circuits	Execution modes	Heron (5K)	Flamingo (5K)	Flamingo (7.5K)	Flamingo (10K)	Flamingo (15K)	Starling (100M)	Blue Jay (1B)
	Early Canary 5 qubits Albatross 16 qubits Penguin 20 qubits Prototype 53 qubits	Falcon Benchmarking 27 qubits		Eagle Benchmarking 127 qubits		Error mitigation 5k gates 133 qubits Classical modular 133x3 = 399 qubits	Error mitigation 5k gates 156 qubits Quantum modular 156x7 = 1092 qubits	Error mitigation 7.5k gates 156 qubits Quantum modular 156x7 = 1092 qubits	Error mitigation 10k gates 156 qubits Quantum modular 156x7 = 1092 qubits	Error mitigation 15k gates 156 qubits Quantum modular 156x7 = 1092 qubits	Error correction 100M gates 200 qubits Error corrected modularity	Error correction 1B gates 2000 qubits Error corrected modularity

Innovation Roadmap

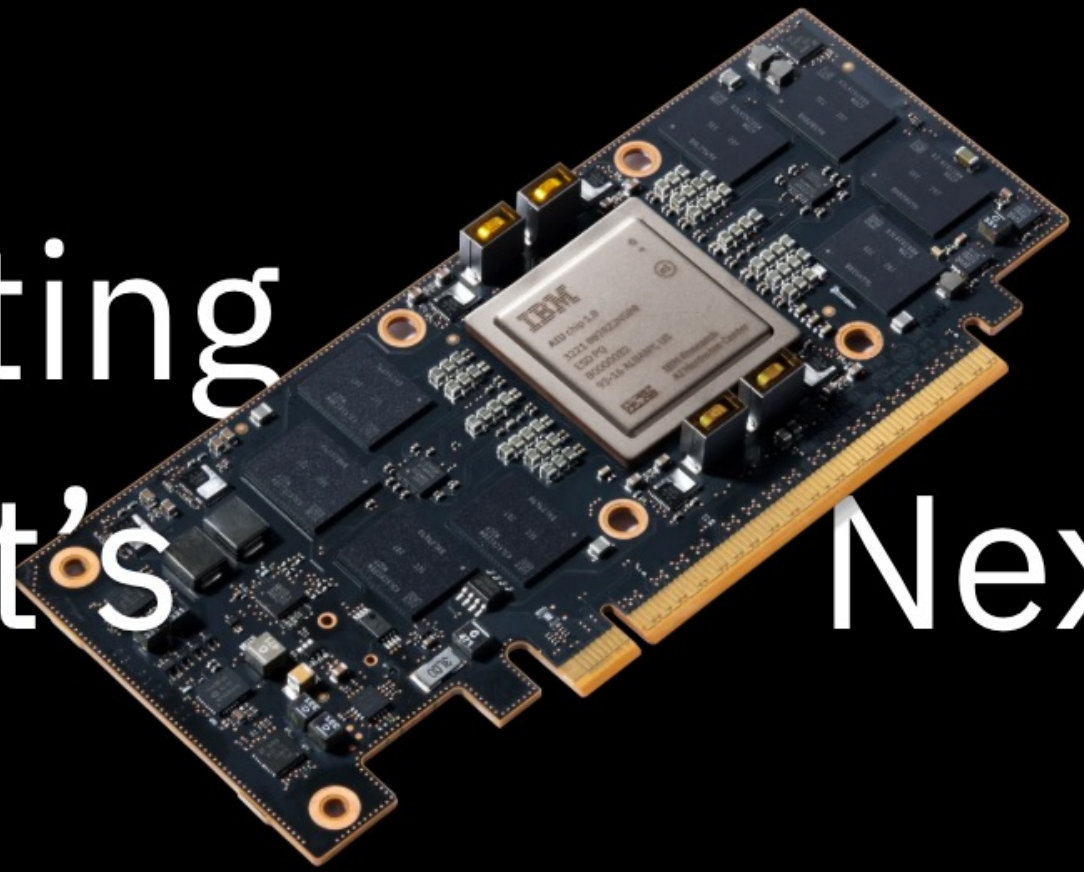
Software innovation	IBM Quantum Experience	Qiskit Circuit and operator API with compilation to multiple targets	Application modules Modules for domain specific application and algorithm workflows	Qiskit Runtime Performance and abstraction through primitives	Quantum Serverless Demonstrate concepts of quantum-centric supercomputing	AI-enhanced quantum Prototype demonstrations of AI-enhanced circuit transpilation	Resource management System partitioning to enable parallel execution	Scalable circuit knitting Circuit partitioning with classical reconstruction at HPC scale	Error correction decoder Demonstration of a quantum system with real-time error correction decoder			
Hardware innovation	Early Canary 5 qubits Penguin 20 qubits Albatross 16 qubits Prototype 53 qubits	Falcon Demonstrate scaling with I/O routing with bump bonds	Hummingbird Demonstrate scaling with multiplexing readout	Eagle Demonstrate scaling with MLW and TSV	Osprey Enabling scaling with high density signal delivery	Condor Single system scaling and fridge capacity	Flamingo Demonstrate scaling with modular connectors	Kookaburra Demonstrate scaling with nonlocal c-coupler Demonstrate path to improved quality with logical memory		Cockatoo Demonstrate path to improved quality with logical communication	Starling Demonstrate path to improved quality with logical gates	
				Egret Tunable coupler demonstration		Heron Architecture based on tunable-couplers	Crossbill Demonstrate m-couplers					

Executed by IBM

On target

IBM Quantum / © 2024 IBM Corporation

Inventing What's Next.



IBM Semiconductors | © 2025 IBM Corporation