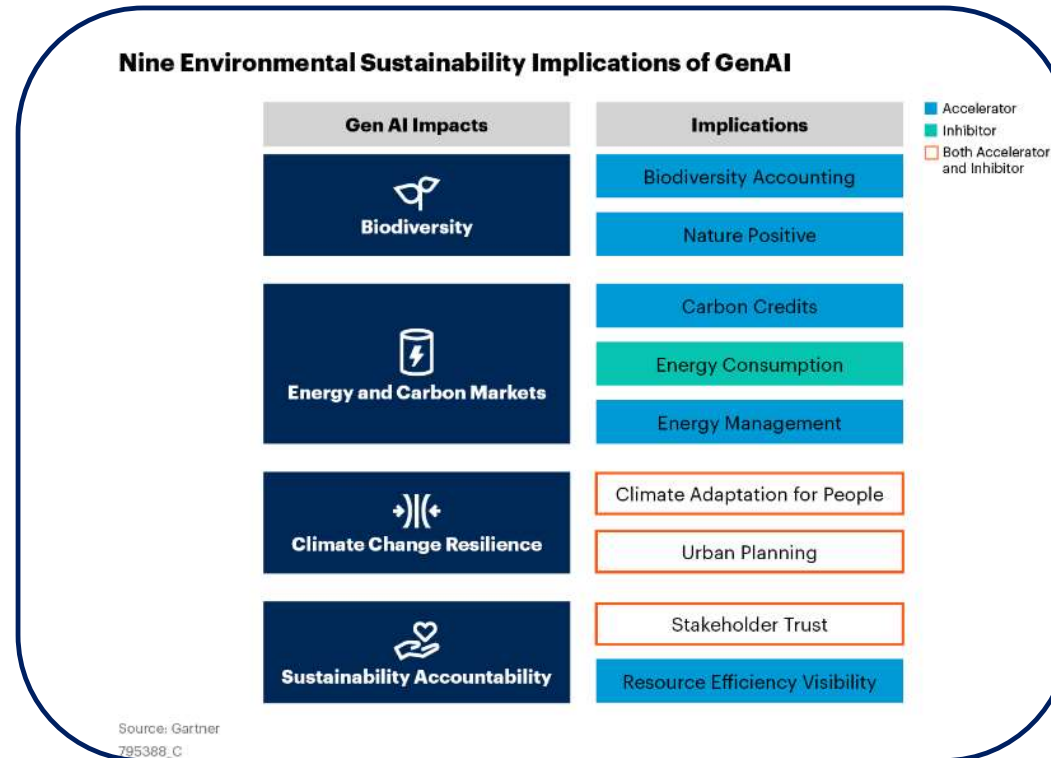


Energy Efficient Compute For A Sustainable Future

Gaurav Gupta, VP Analyst
Emerging Trends and Technologies
gaurav.gupta2@gartner.com



Larger training models → greater computing → higher energy demands



Compute-intensive and emerging applications based on AI will face **energy bottleneck**

RESTRICTED DISTRIBUTION

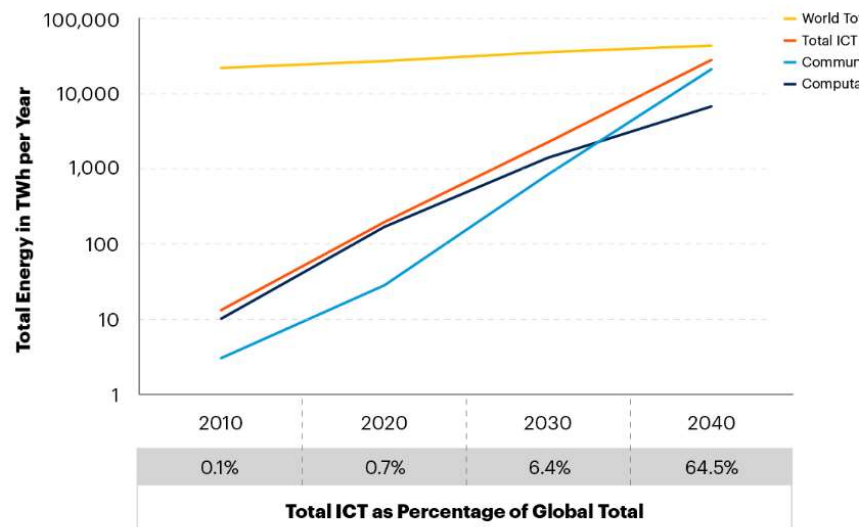
2 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

Power demands for high tech exceed global supply

Current Trends in ICT Global Electricity Requirements and Global Electricity Generation

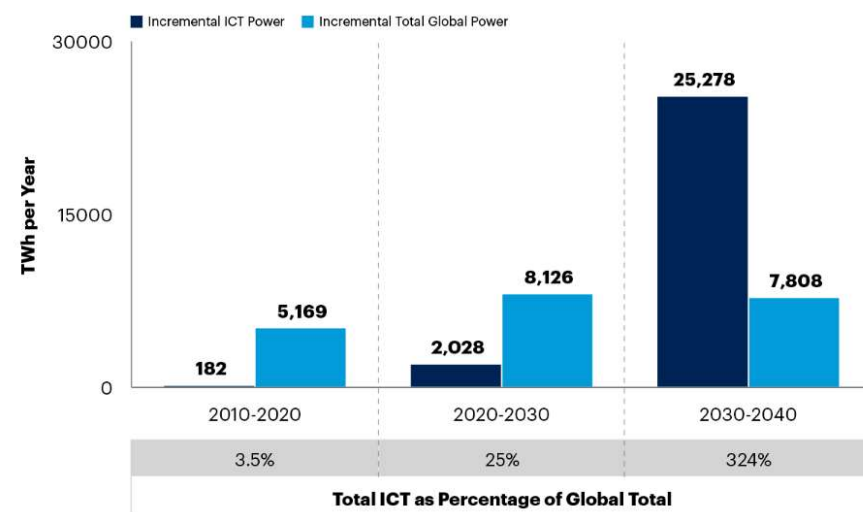
World Energy Production Versus ICT Energy Requirements



Source: SRC's Decadal Plan Update; International Energy Association World Energy Outlook 2022 — Stated Policies Scenario
782338_C

Current Trends: Incremental ICT Requirements and Global Electricity Generation

Incremental Power per Decade



Source: SRC's Decadal Plan Update; International Energy Association World Energy Outlook 2022 — Stated Policies Scenario
782338_C

RESTRICTED DISTRIBUTION

3 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

Power-efficiency isn't keeping up with performance

- ✓ Microsoft is hiring a nuclear energy expert to help power its AI and cloud data centers
- ✓ Unless things change, first zettaflop systems will need nuclear power, AMD's Su says



- The crux of challenge from Exascale to Zettascale will be power efficiency – server efficiency is starting to flatten
- I/O continues to be a limiting efficiency factor

RESTRICTED DISTRIBUTION

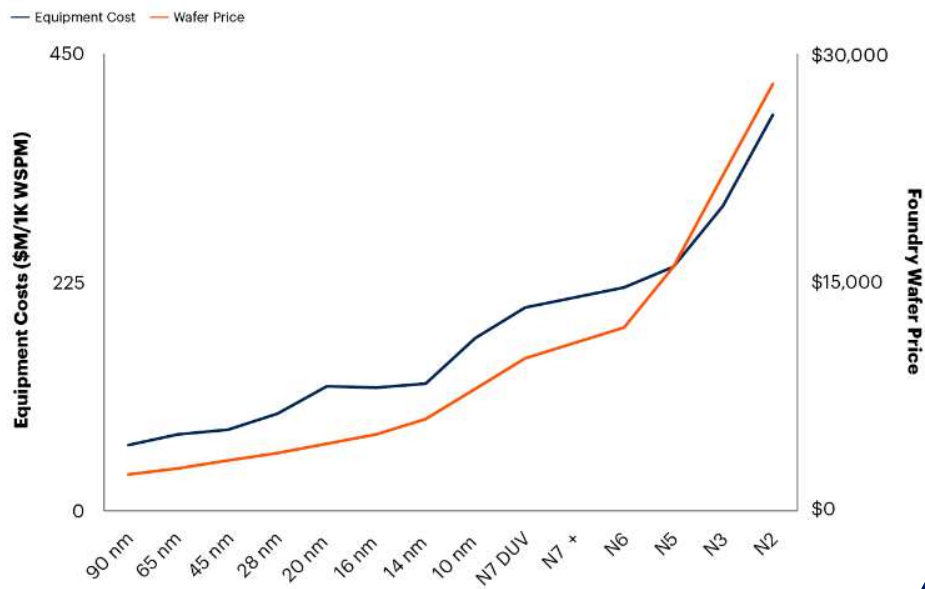
4 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Source: <https://jobs.careers.microsoft.com/global/en/job/1627555/Principal-Program-Manager-Nuclear-Technology>,
<https://www.cnn.com/2023/09/25/microsoft-is-hiring-a-nuclear-energy-expert-to-help-power-data-centers.html> ,
https://www.theregister.com/2023/02/23/amd_zettaflop_systems_nuclear/ , <https://www.nextplatform.com/2023/03/03/getting-to-zettascale-without-needing-multiple-nuclear-power-plants/>

Gartner®

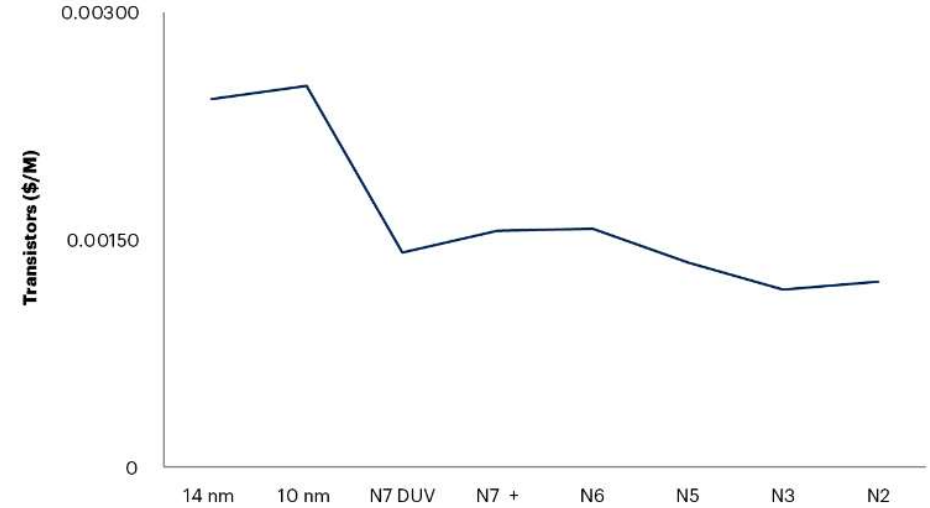
Moore's law can no longer be depended

Equipment Costs and Foundry Wafer Prices



Source: Gartner (September 2022)
767587_C

Wafer Price per Million Transistors



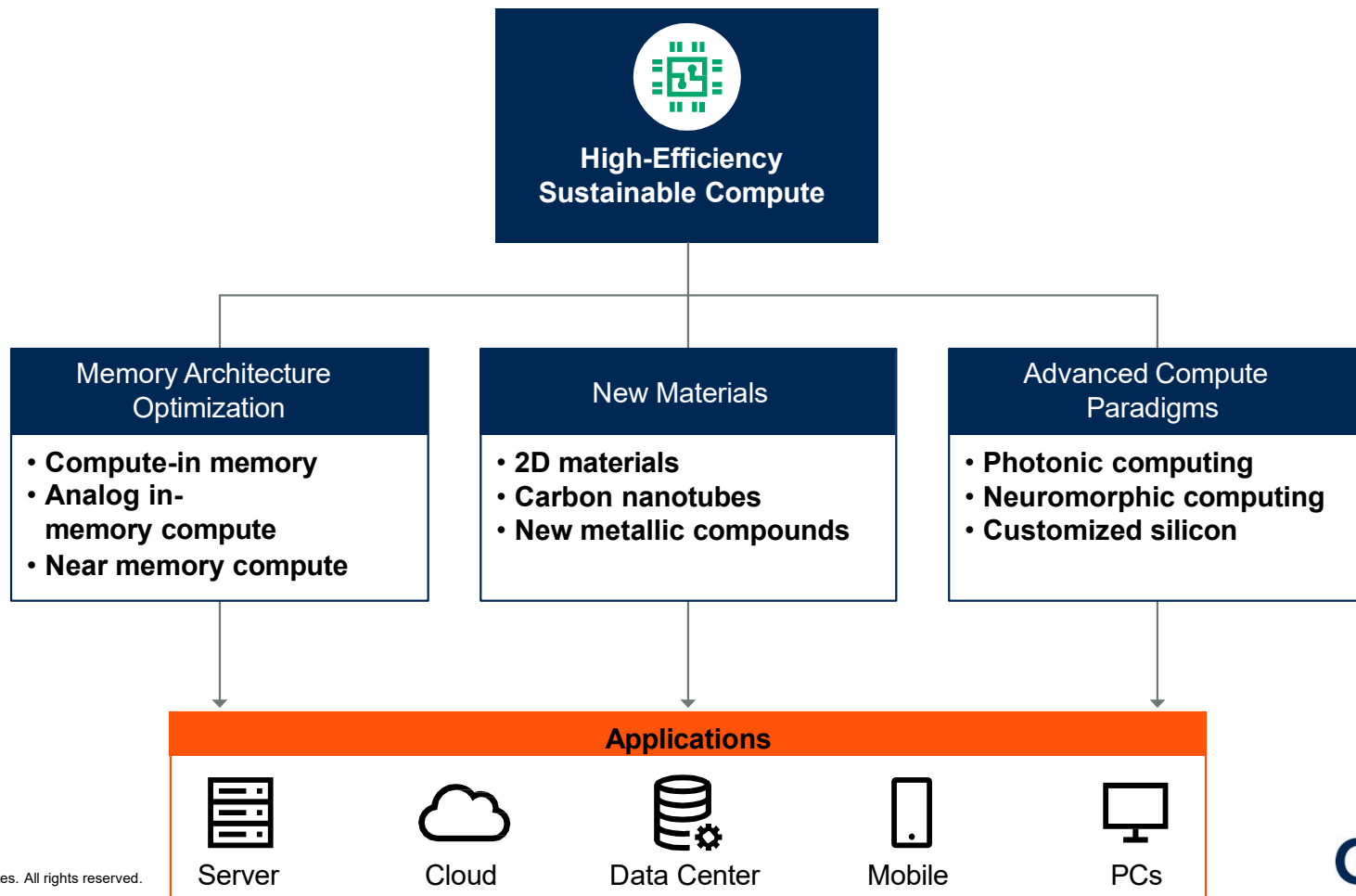
Source: Gartner (September 2022)
767587_C

RESTRICTED DISTRIBUTION

5 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

High efficiency compute for a sustainable future



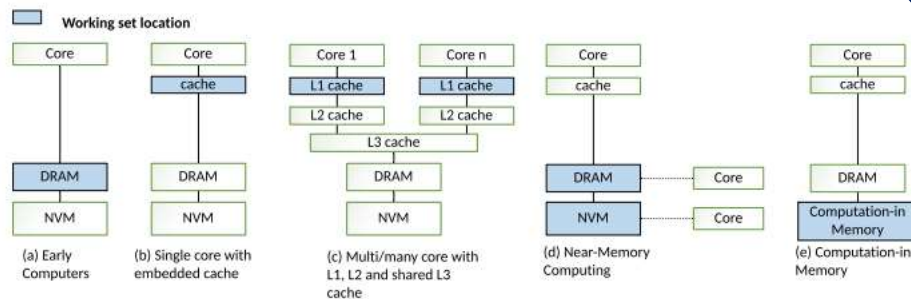
Source: Gartner 782737_C

RESTRICTED DISTRIBUTION

6 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

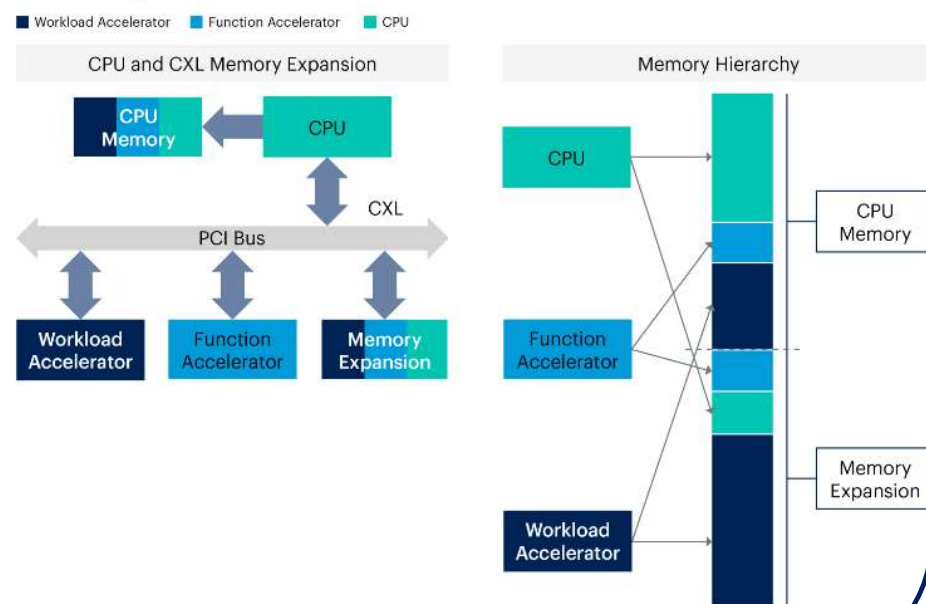
Gartner

Memory-centric architecture offers improvement



- Near-memory compute
- In-memory compute or compute in memory or processing in memory

Block Diagram of CXL-Enabled Server Architectures



Source: Gartner
780669_C

RESTRICTED DISTRIBUTION

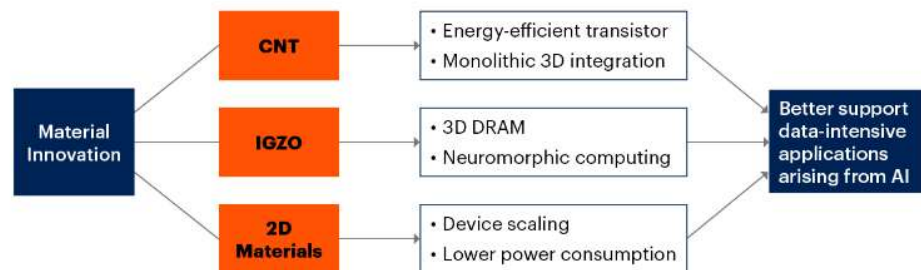
7 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Image source: <https://arxiv.org/pdf/1908.02640.pdf>

Gartner®

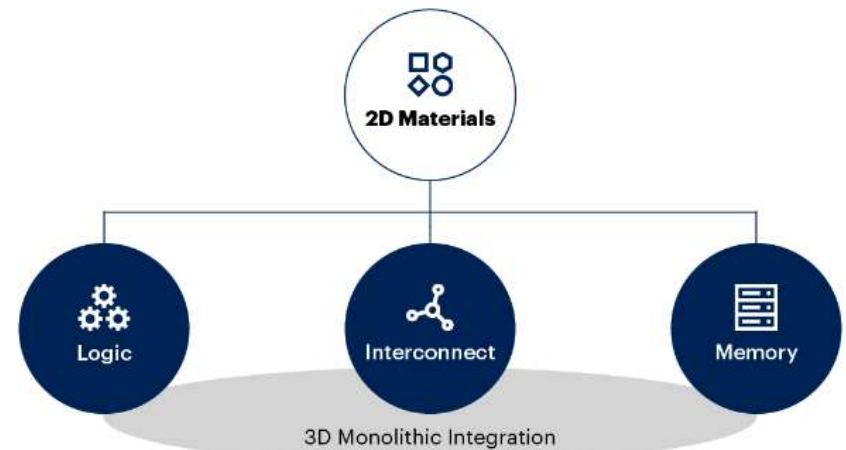
Material innovations enable further development

CNT, IGZO and 2D Materials



Source: Gartner
780563_C

2D Materials Enable Future Technology Development in Logic, Memory, and Interconnect



Source: Gartner
774346_C

RESTRICTED DISTRIBUTION

8 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

Current compute options for energy-efficiency

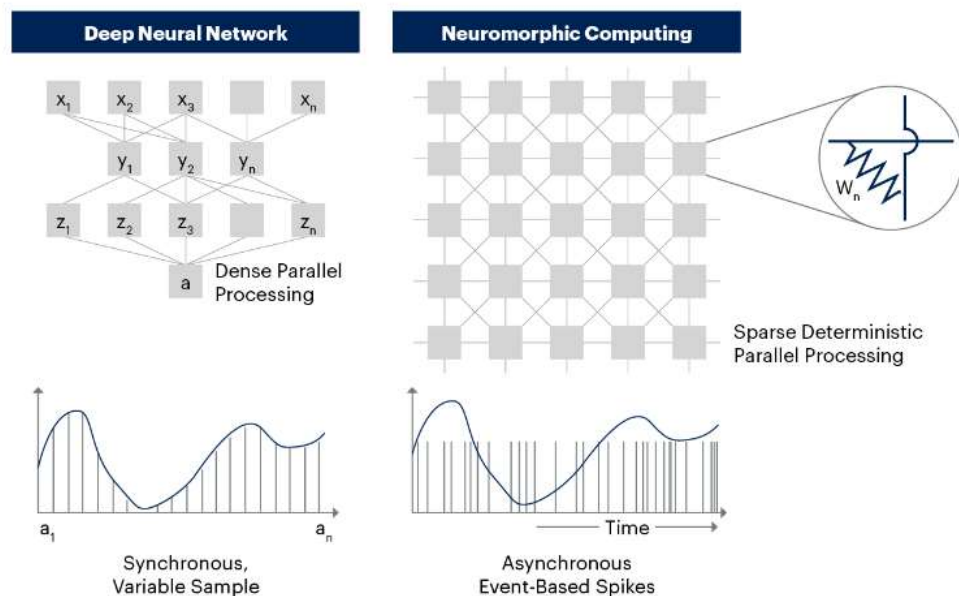
Energy efficiency is not an option; it is critical to achieving net-zero targets and sustainability.

Possibilities:

- Sub-threshold voltage designs
- Low-precision format arithmetic
- Machine virtualization
- Custom Silicon

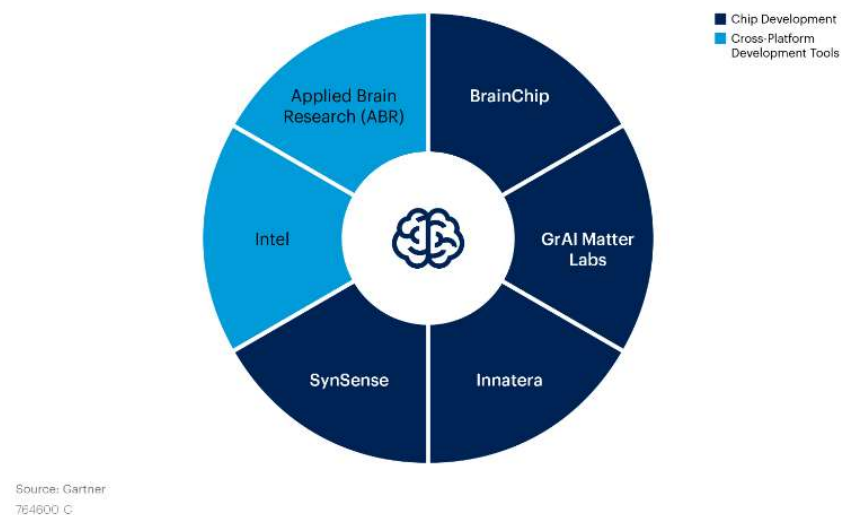
NC implementations are at an early stage

Comparison Between DNNs and Neuromorphic Computing Implementations



Source: Gartner
729376_C

Technology Innovators in Neuromorphic Computing



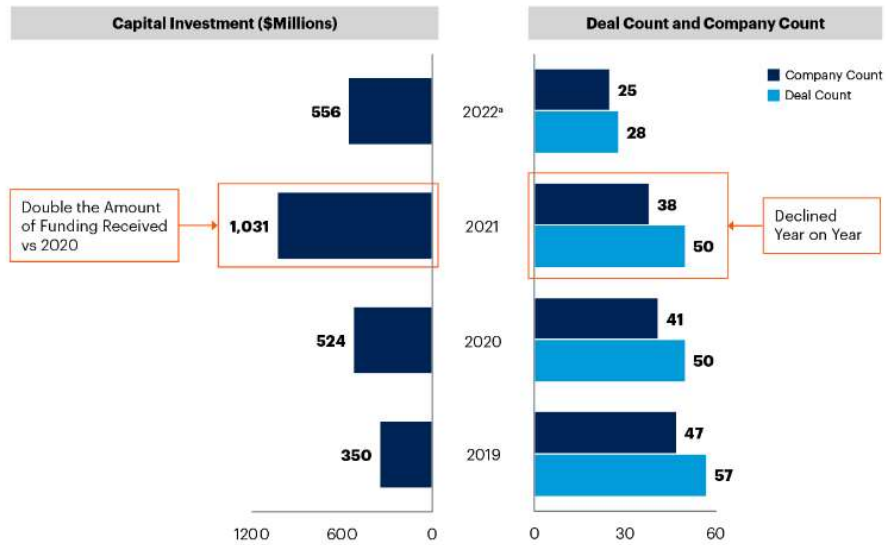
RESTRICTED DISTRIBUTION

10 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

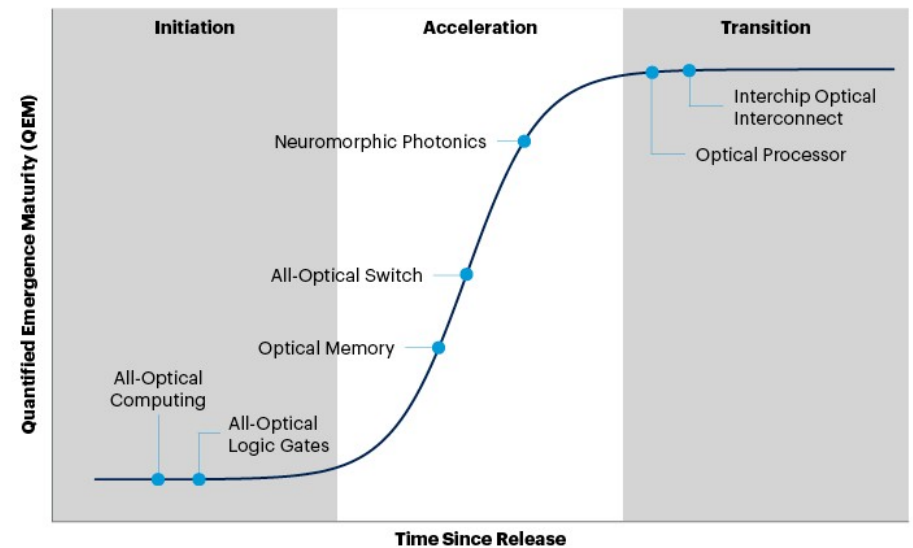
All-optical computing needs variety of components

Silicon Photonics Investment Trends



Source: Pitchbook Investment Data and Gartner Research
*2022 data is partial year only, until April
756832_C

Emergence Cycle for Photonic Computing



Source: Gartner
778572_C

RESTRICTED DISTRIBUTION

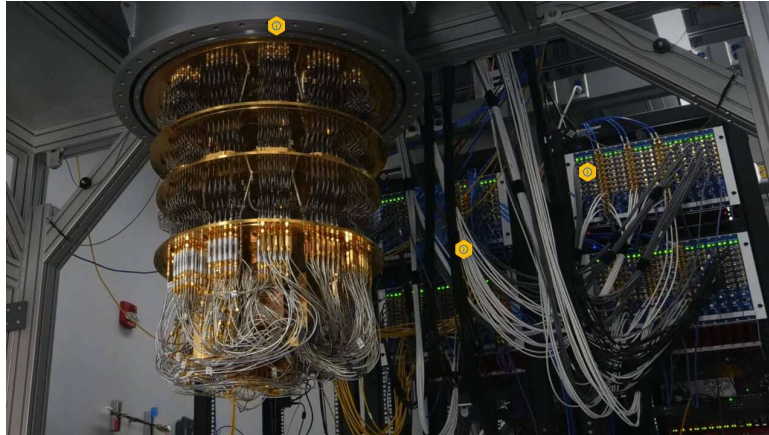
11 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Source: [Emerging Tech: Emergence Cycle for Photonic Computing](#), [Emerging Tech: Venture Capital Growth Insights in Silicon Photonics](#)

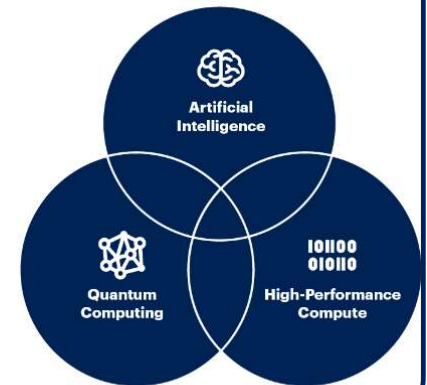
Gartner®

QC has several challenges to overcome

- Scalability
- Complex hybrid ecosystem
- System size
- Power consumption
- Error
- Hand-crafted



The Hybrid Workflow of the Future



Source: Gartner
801774_C

Timelines for Business Advantage From Quantum Computing



Source: Gartner
801774_C

Reference: <https://quantumai.google/learn/lab>

RESTRICTED DISTRIBUTION

12 © 2023 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

Product leader recommendations

- Target ROI for energy-efficient compute by prioritizing quick wins.
- Continue to watch-out for evolving compute paradigms:
 - ✓ Hyperdimensional computing
 - ✓ Fog computing
 - ✓ P-computing
 - ✓ Conservative computing
 - ✓ Energy-proportional computing



Energy Efficient Compute For A Sustainable Future

Gaurav Gupta, VP Analyst
Emerging Trends and Technologies
gaurav.gupta2@gartner.com

