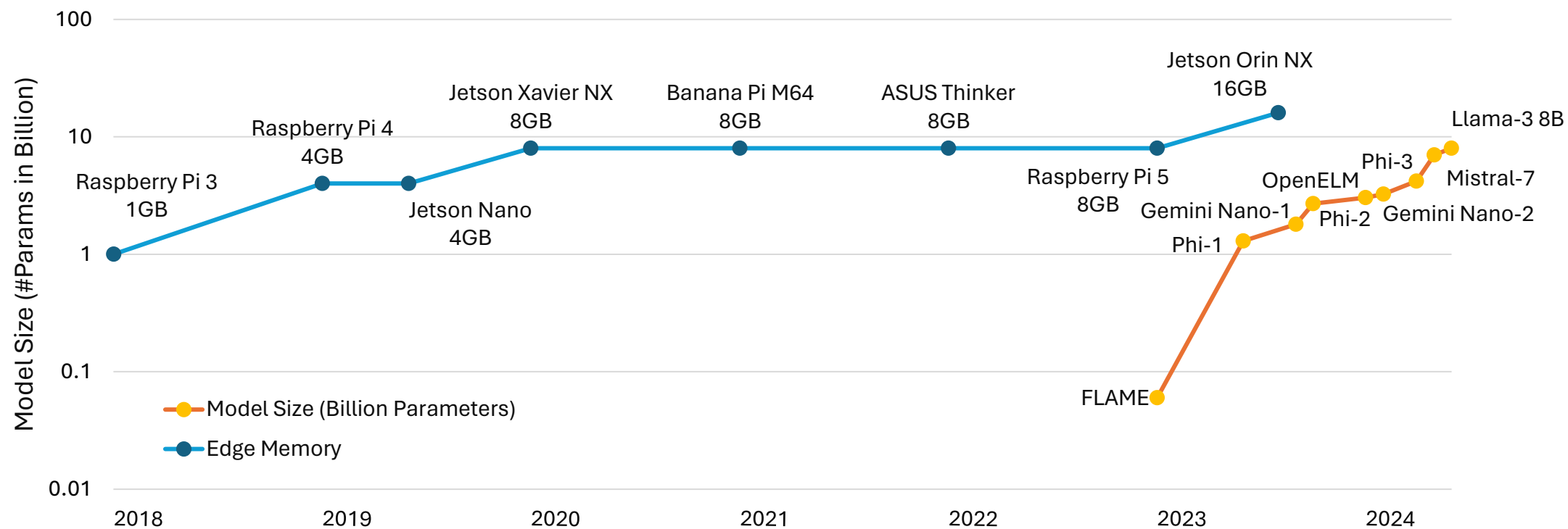# Edge number over the past 10 years

Source:

(a) Model Size of Large Language Models (LLM) and Cloud GPU Memory
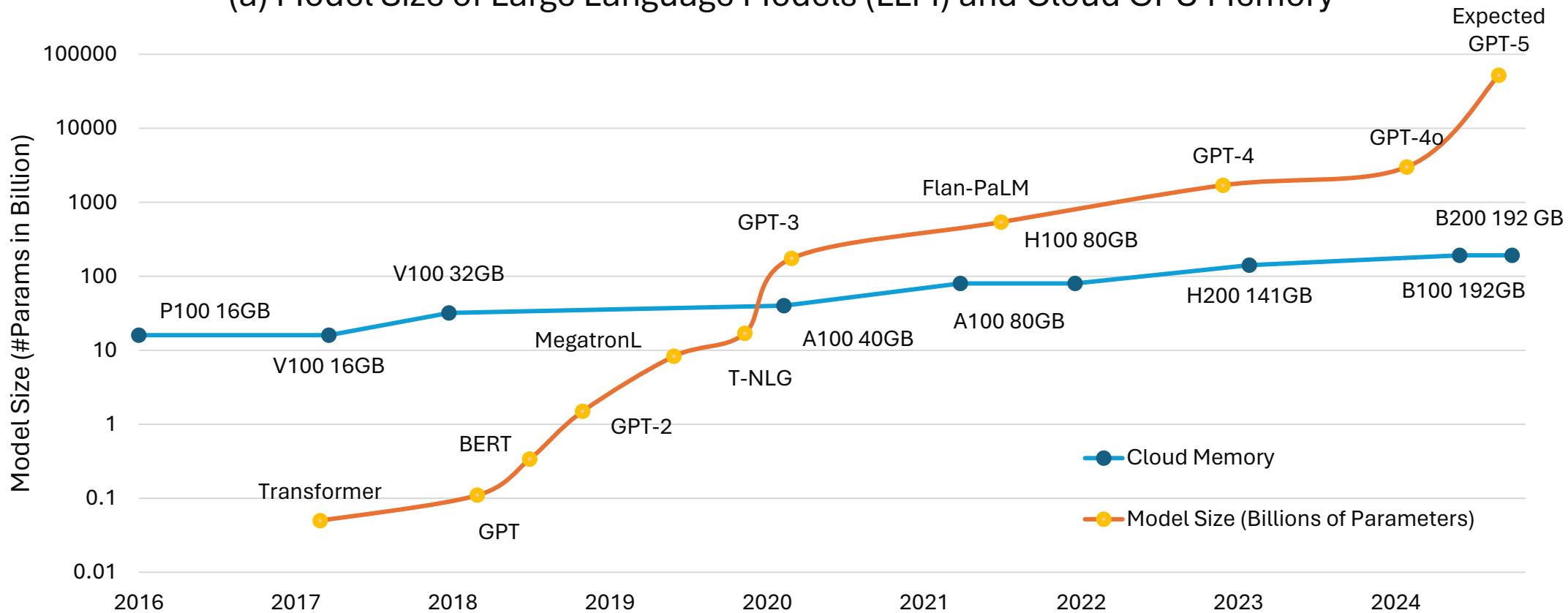
SLM vs Edge GPU Memory

Gap also grow exponentially

## (b) Model Size of Small Language Models (SLM) and Edge GPU Memory

Legend:
- Model Size (Billion Parameters)
- Edge Memory

Data points (Edge Memory): Raspberry Pi 3 1GB, Raspberry Pi 4 4GB, Jetson Nano 4GB, Jetson Xavier NX 8GB, Banana Pi M64 8GB, ASUS Thinker 8GB, Raspberry Pi 5 8GB, Jetson Orin NX 16GB

Data points (Model Size): FLAME, Phi-1, Gemini Nano-1, OpenELM, Phi-2, Phi-3, Gemini Nano-2, Mistral-7, Llama-3 8B

Y-axis: Model Size (#Params in Billion) — 0.01, 0.1, 1, 10, 100
X-axis: 2018, 2019, 2020, 2021, 2022, 2023, 2024

Raspberry Pi 1ˢᵗ gen to 5ᵗʰ gen

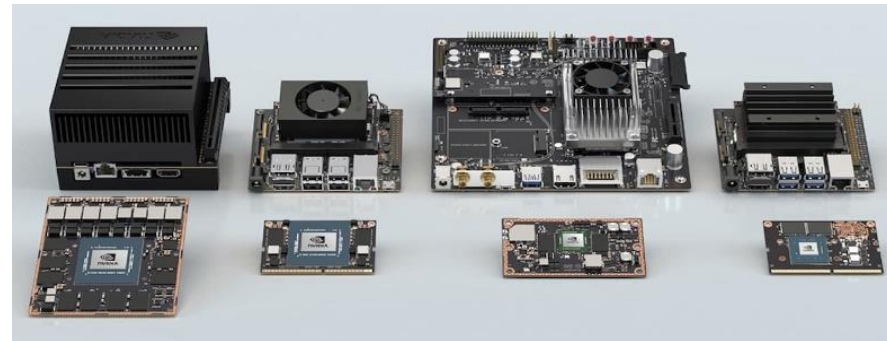
(a) Performance, Memory Capacity, and Power Consumption of Different Generations of Raspberry Pi
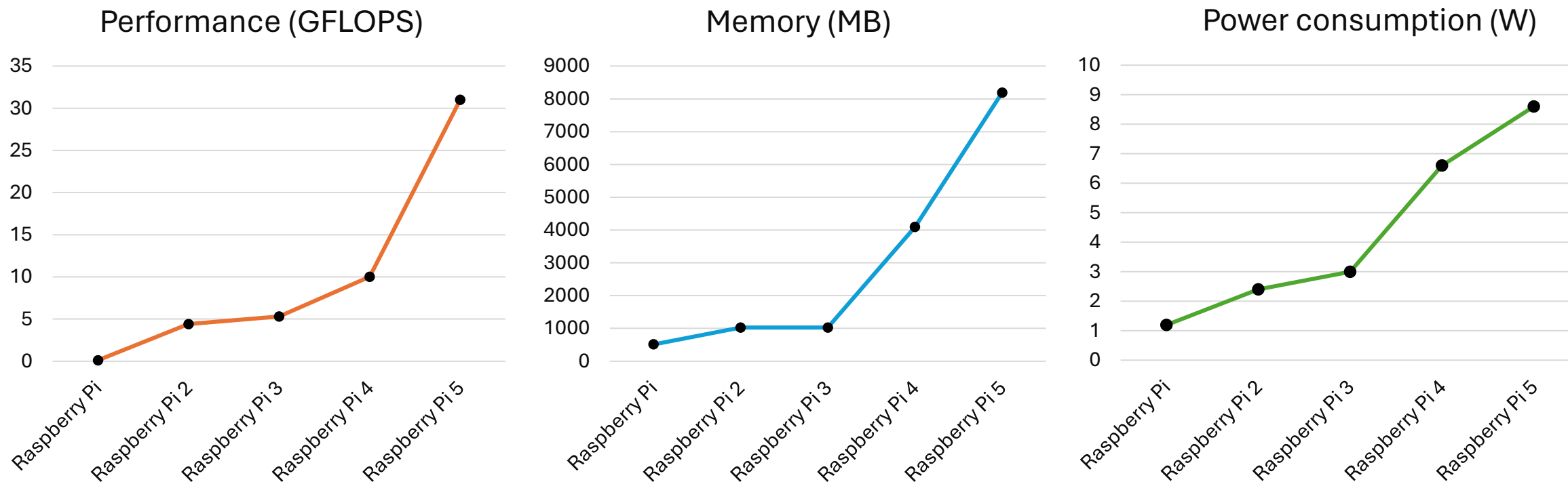
Nvidia Jetson series
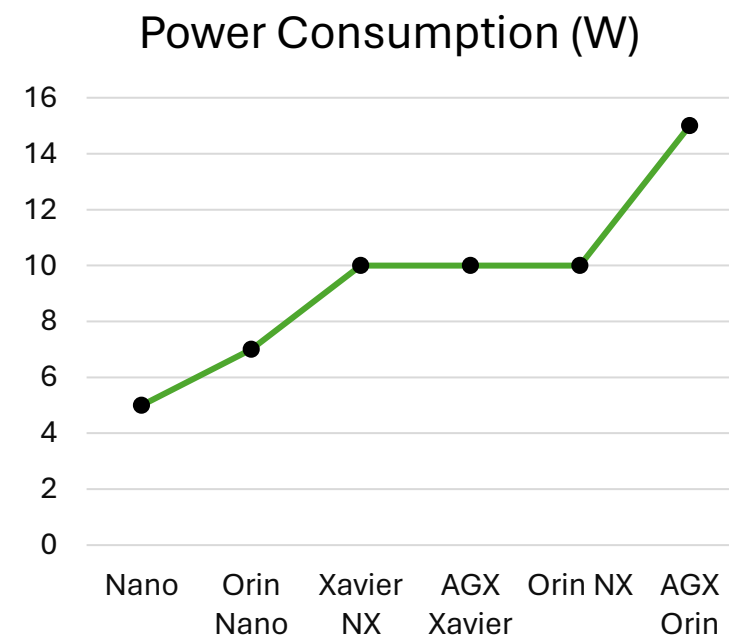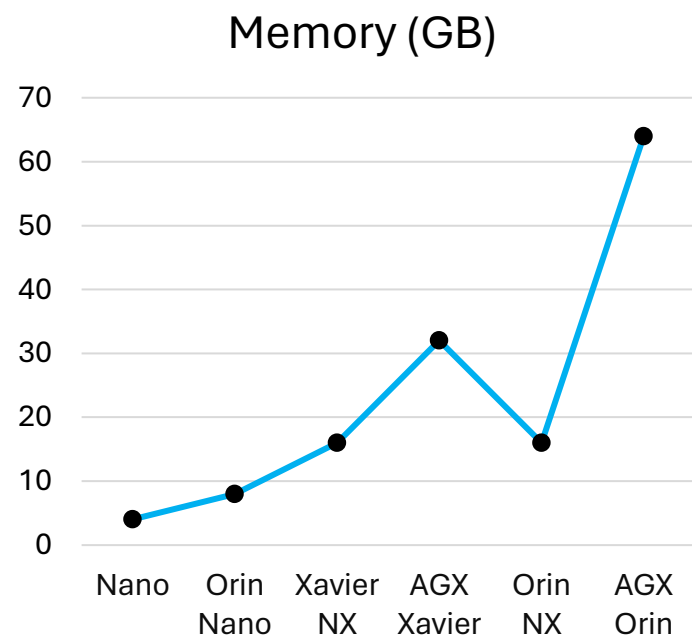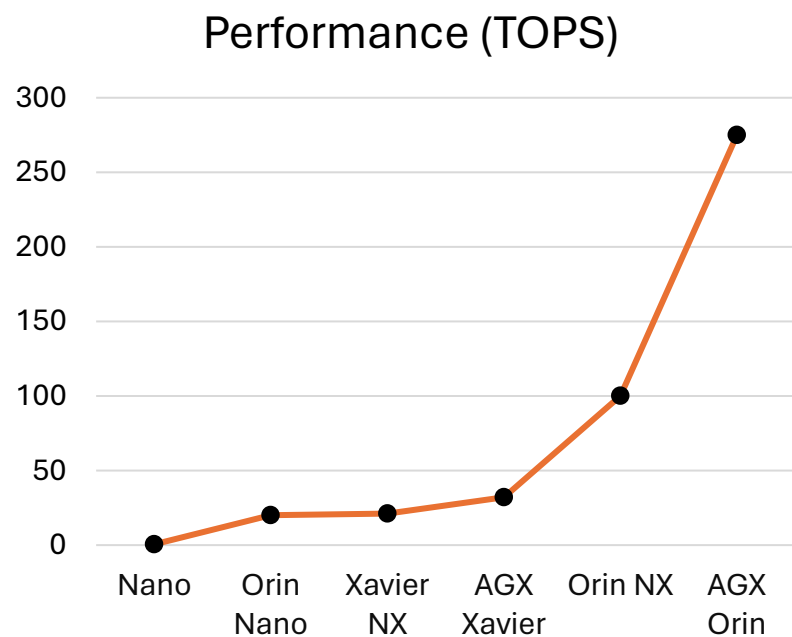
(b) Performance, Memory Capacity, and Power Consumption of Different Generation Nvidia Edge Devices

Performance (TOPS)

Memory (GB)

Power Consumption (W)

Apple Silicon from m1 to m4

(c) NPU Performance, Memory Bandwidth, and Relative Performance Per Watt for Apple M Series Chips

NPU AI Performance (TOPS)

Memory Bandwidth (GB/s)

Performance per Watt to M1