

# ML > The Internet

Summary of Ark Invest Analysis

# Source

- ARK identifies more innovation evolving today than ever before. We believe it is changing lives and businesses across the globe dramatically, creating opportunities to own the next big thing by investing in the future today.
- <https://ark-invest.com/analyst-research/ai-training/>

# Moore's Law

- The number of transistors in an IC doubles approximately every two years.
- Moore's law is an observation and projection of a historical trend.

# Let This Sink In...

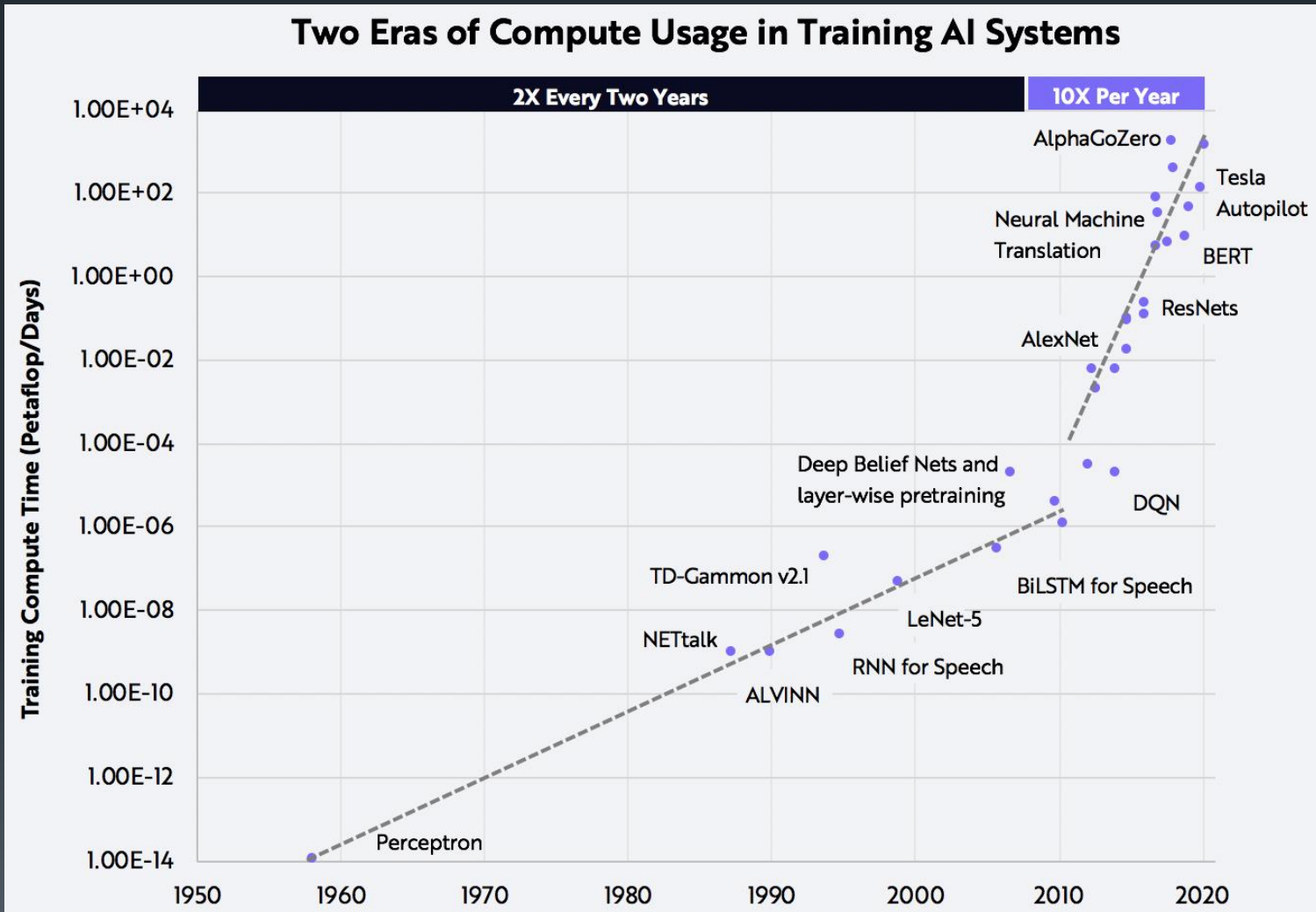
- The cost to train an artificial intelligence (AI) system is improving at 50x the pace of Moore's Law.

# Cost Approaches Zero

- For many use cases, the cost to run an AI inference system has collapsed to almost nil.
- After just five years of development, deep learning – the modern incarnation of AI – seems to have reached a tipping point in both cost and performance, paving the way for widespread adoption over the next decade.

# Compute

- For the past ten years, the computing resources devoted to AI training models have exploded.
- After doubling every two years from 1960 to 2010, AI compute complexity has soared 10x every year.



A “Petaflop-Day” refers to performing a quadrillion operations per second for a day.

# Incentive and Investment

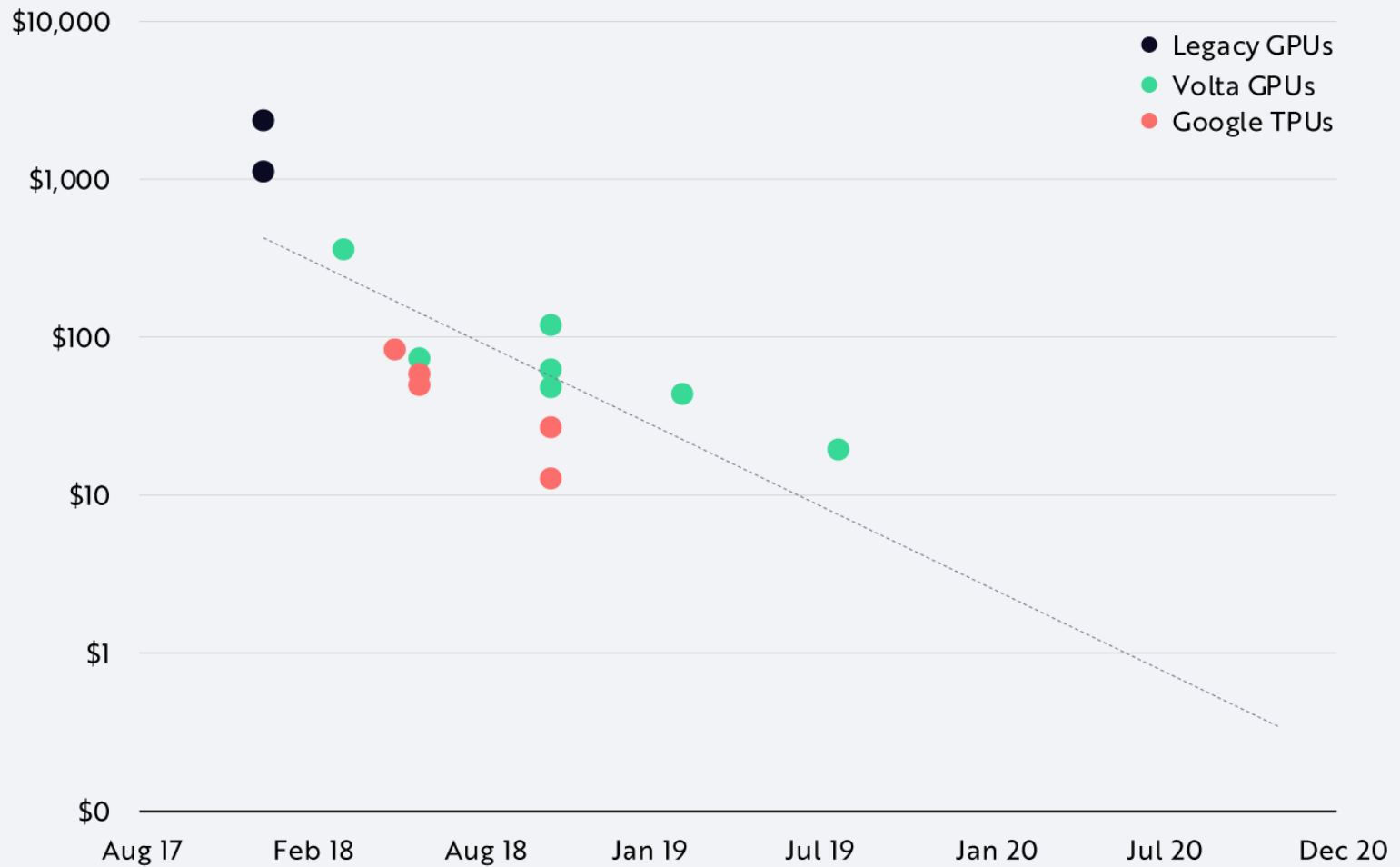
- It is believed that companies have had ample incentive to increase computing resources at five times the rate of Moore's Law: significant competitive advantages in revenue generation and hardware cost declines rapid enough to keep fueling the beast.
- As so-called hyperscale internet companies (Amazon, Microsoft, Google, etc.) have taken the reins from universities and trained deep learning networks on their data, they have budgeted hundreds of millions of dollars for AI hardware, expecting superior rates of return on investment over time.



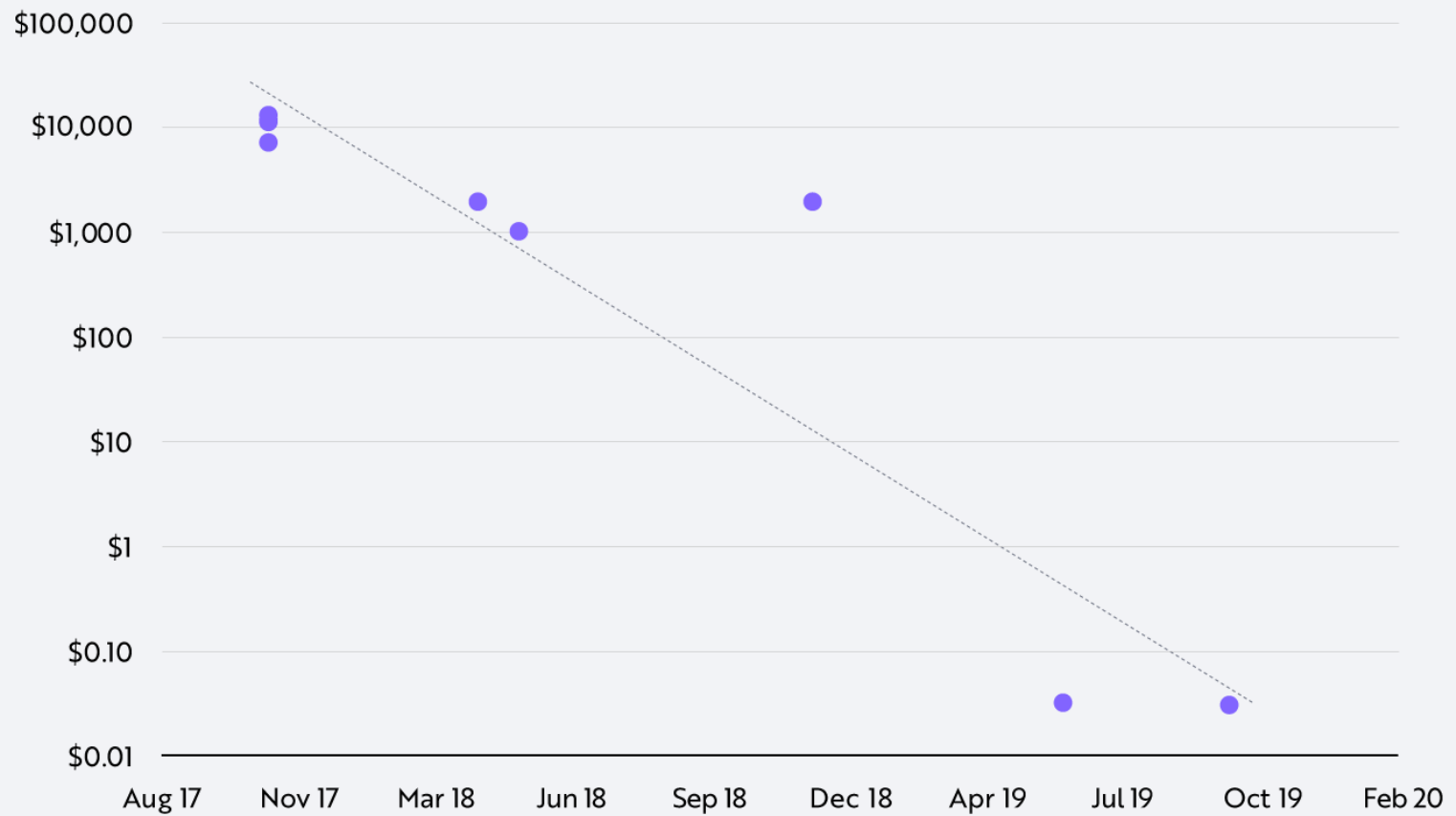
# Train and Inference

- AI training costs have dropped roughly 10x every year.
  - In 2017, for example, the cost to train an image recognition network like [ResNet-50](#) on a public cloud was ~\$1,000.
  - In 2019, the cost dropped to ~\$10, as shown below.
  - At the current rate of improvement, the cost should fall to \$1 by the end of this year.
- The cost of inference—running a trained neural network in production—has dropped even more precipitously. During the past two years, for example, the cost to classify one billion images has fallen from \$10,000 to just \$0.03, as shown below.

## Cost to Train a Neural Network (ResNet-50)



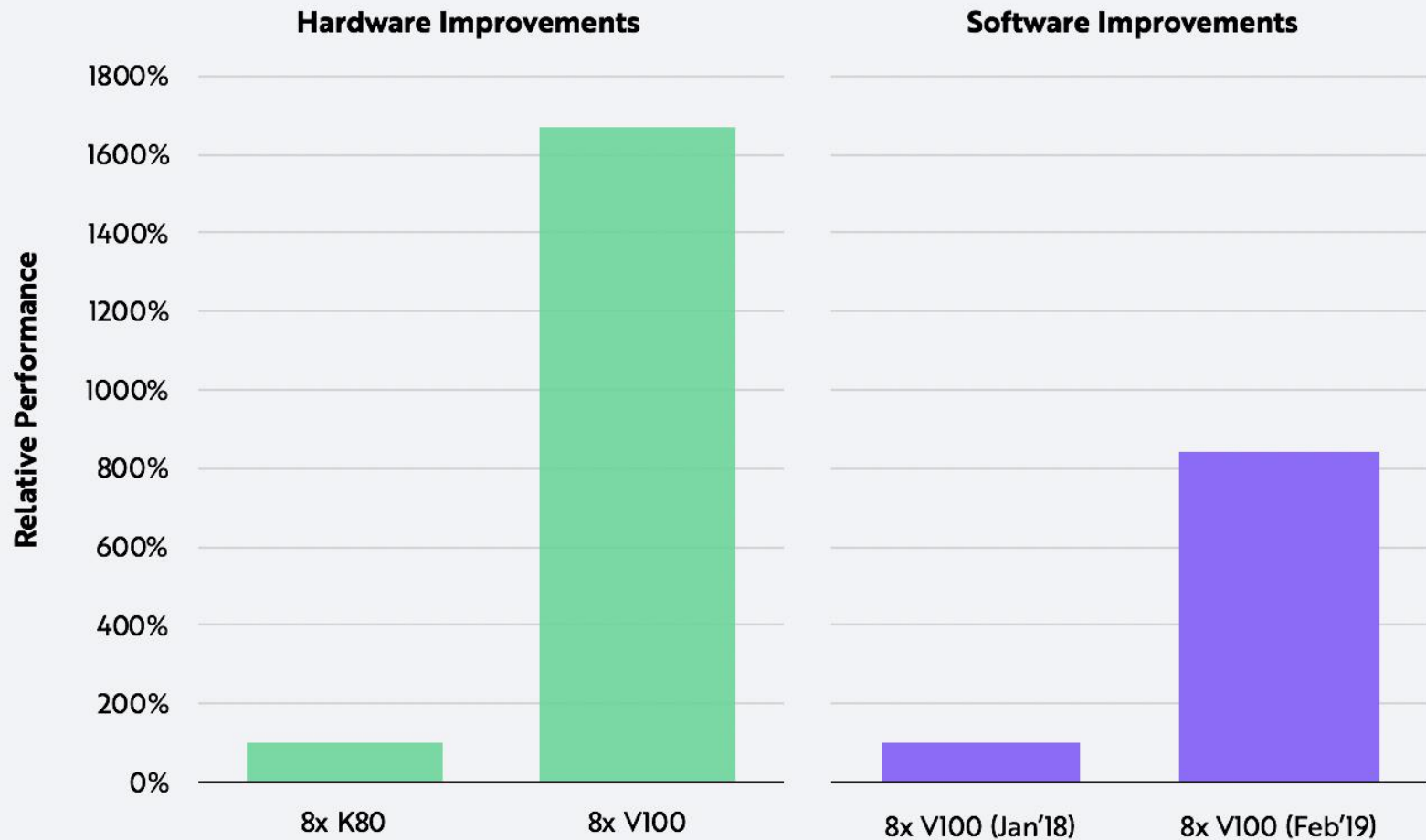
## Cost to Perform Inference on 1B Images



# Hardware and Software

- Breakthroughs in both hardware and software have enabled these cost declines.
- In the past three years, chip and system design have evolved to add dedicated hardware for deep learning, resulting in a 16x performance improvement.
- Holding hardware improvements constant, newer versions of TensorFlow and PyTorch AI frameworks in concert with novel training methods combine to generate an 8x performance gain.

## AI Training Performance Gains



# AI Chip Costs

- Curiously, AI training chip costs have not dropped in tandem with unit hardware prices. The price of Nvidia's data center GPU, for example, has tripled over the [last three generations](#). In fact, Amazon Web Services has not lowered the price of Nvidia's V100 GPU instances since it introduced them [in 2017](#). Competition from independent and hyperscale AI chip designs does have the potential to erode Nvidia's pricing power, but so far, no company has been able to field a comparable chip to Nvidia's V100 GPU with the same breadth of software and developer support.

# Early Days of AI

- Based on the pace of its cost decline, AI is in very early days. During the first decade of Moore's Law, transistor count doubled every year—or at twice the rate of change seen during decades thereafter.
- The 10-100x cost declines we are witnessing in both AI training and AI inference suggest that AI is nascent in its development, perhaps with decades of slower but sustained growth ahead. Based on ARK's research, while thus far AI has added roughly \$1 trillion to the global equity market cap, it is poised to scale to \$30 trillion by 2037, becoming the first foundational technology to dwarf the internet.

## Market Cap Creation Internet vs. Deep Learning

