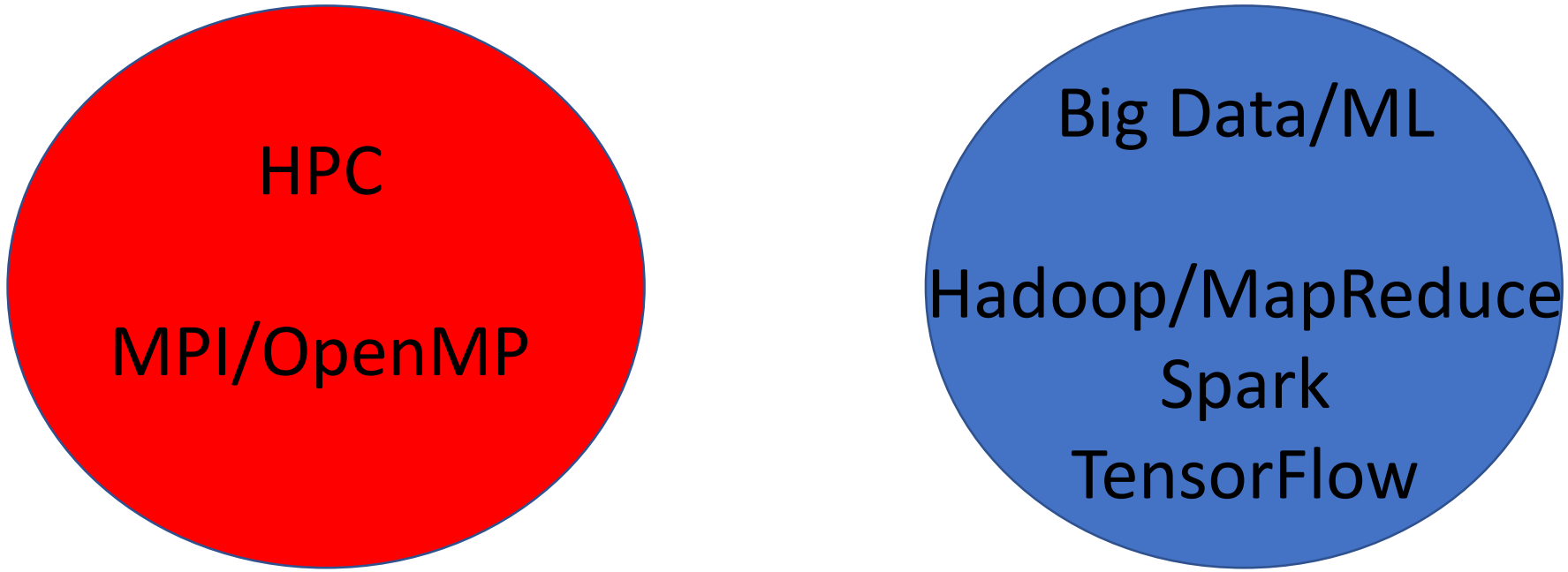
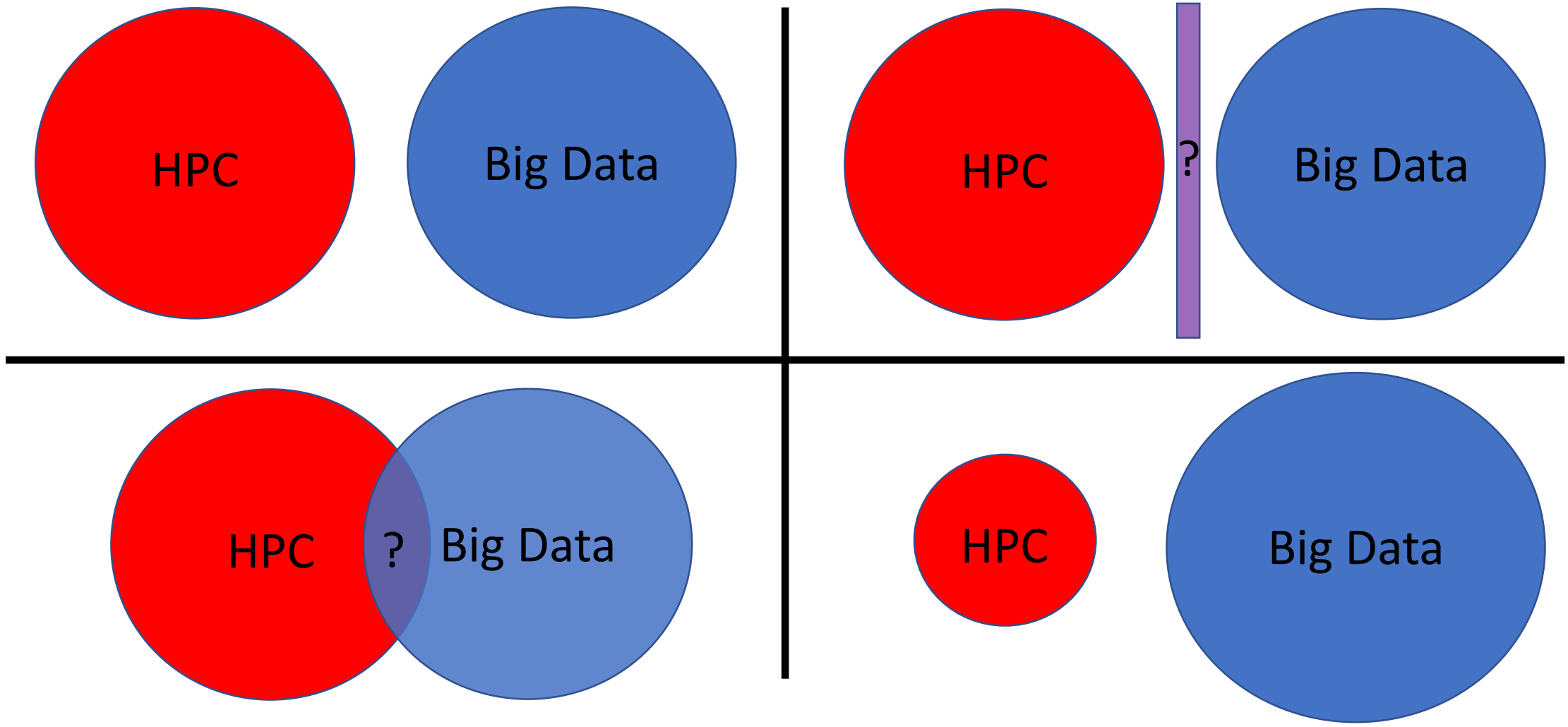
Convergence of HPC and Big Data : Software Panel

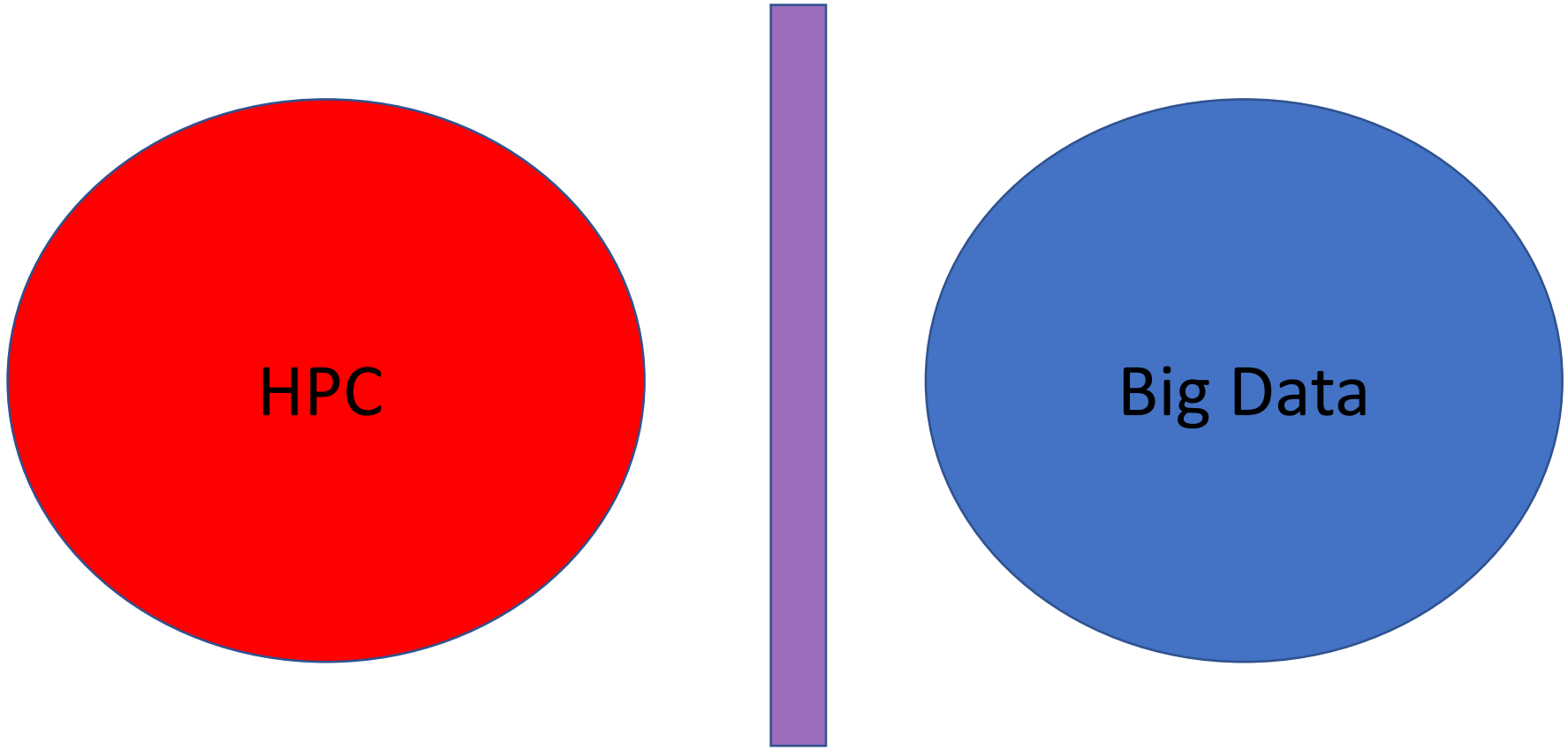
Alex Aiken
Stanford/SLAC

A Tale of Two Software Worlds




HPC
Big Data
HPC
?
Big Data
HPC
?
Big Data
HPC
Big Data

Are There Barriers to Convergence?



Priorities

HPC

- Performance
- Cost
- Correctness

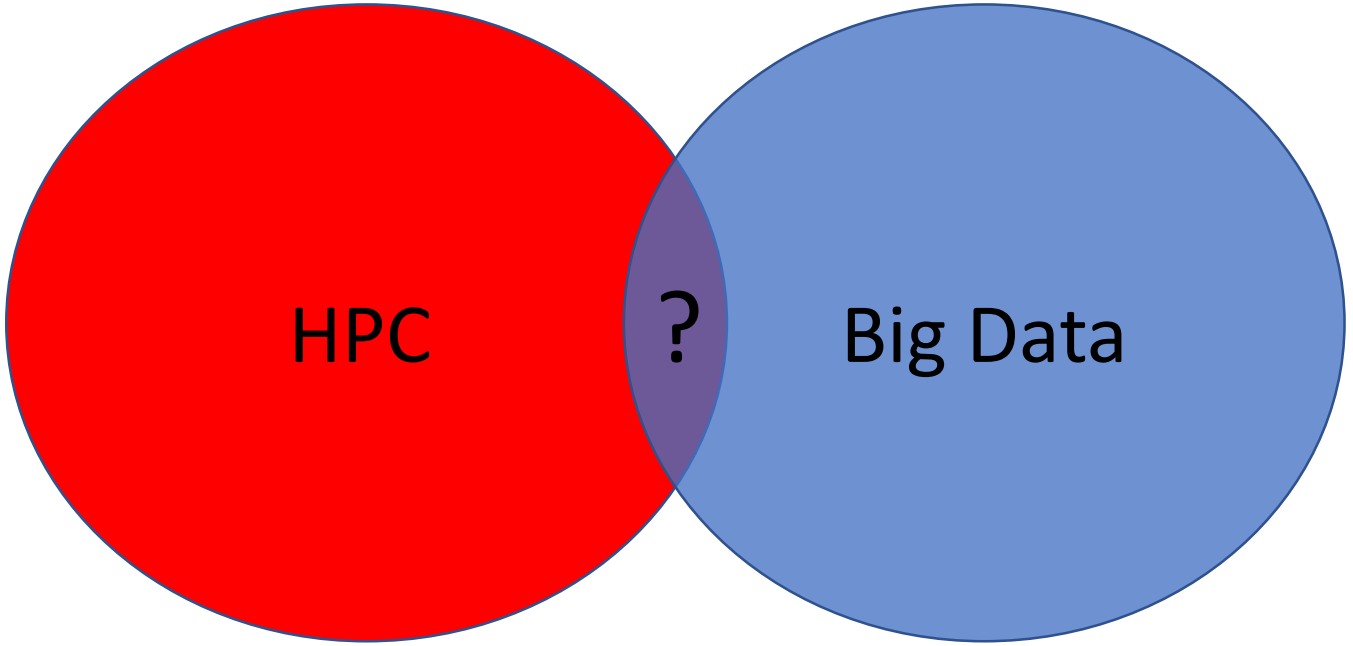
Big Data

- Cost
- Performance
- Correctness

Creates Significant Differences In …

- Platform performance
- Programmer productivity
- Scale of computations
- Economic model

Is There Overlap Today?



Who Would Switch from Big Data to HPC?

0%

Who Would Switch from HPC to Big Data?

- If performance improved by switching, everyone
- If performance were comparable or not overly harmed, some
- If performance is 10X worse, none

A Brief Digression: Hardware

- The hardware platform drives the software abstractions

- The current, slow-motion revolution: accelerators
  - GPUs today
  - Other specialized hardware tomorrow

A Key Point

- In new supercomputers, > 95% of performance is in the accelerators
- The tradeoff
  - Greatly complicates programming
  - But switching to GPUs can greatly increase performance
- This is the ground on which any convergence will happen

An Observation

- The HPC community values performance
  - Unless it is too hard
  - Many HPC systems perform far below their potential today

- The Big Data community values productivity
  - Until the code takes forever to run
  - Organizations spend inordinate amounts of time tweaking for performance

The Technical Issue

- The real problem in current and future systems is data movement
  - By far the most expensive part of any computation
  - And accelerators add multiple levels of memory hierarchy

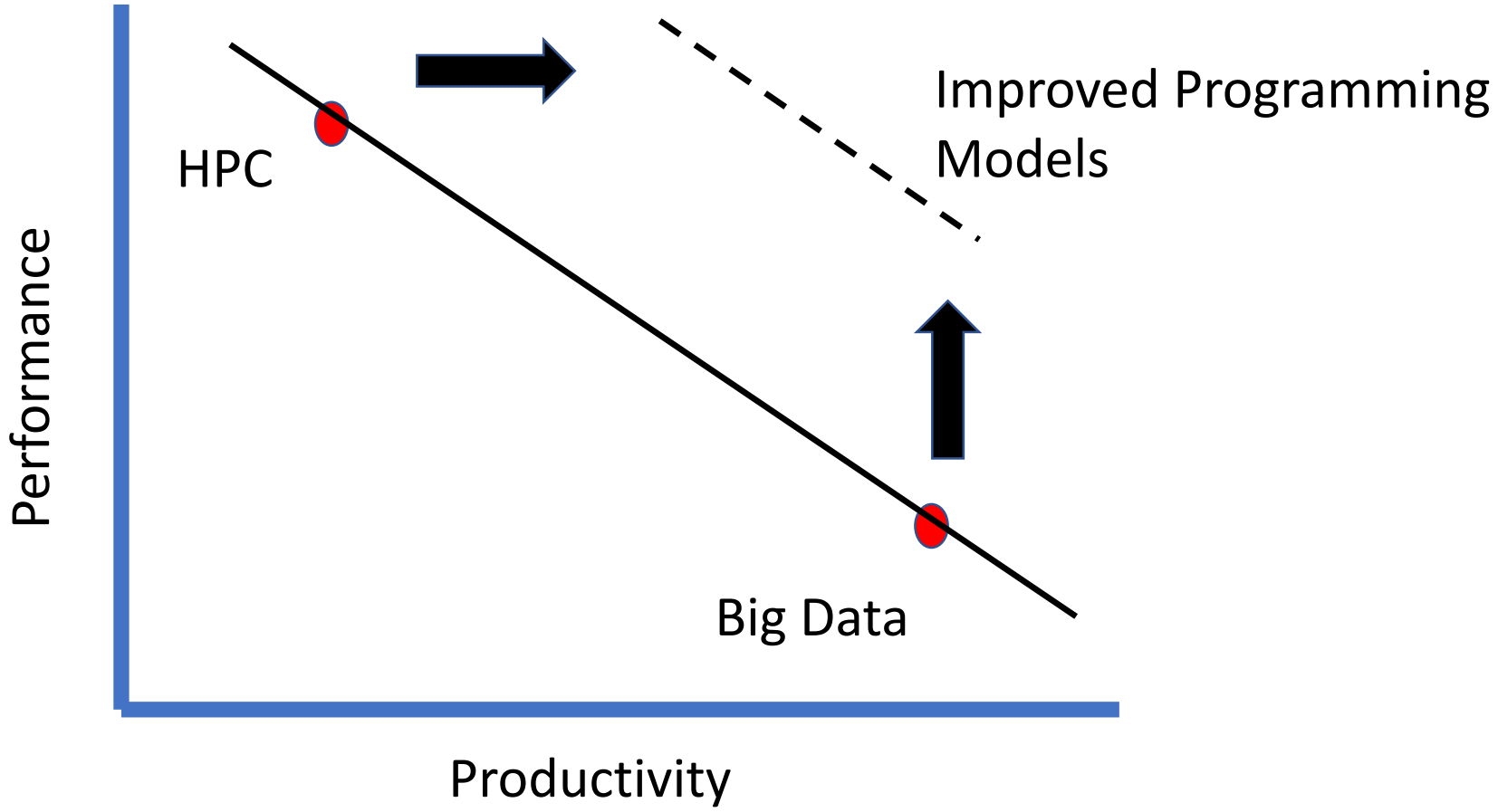
- Few programming abstractions in any programming model for
  - Locality
  - Partitioning
  - Mapping of compute/data into a machine

The Evidence

- S3D
  - Production chemistry combustion code
  - 7X off its potential
- Large graph analytics
  - Current state of the art ~10X off potential
- Machine Learning
  - 2-3X off potential

Switching to GPUs + good data partitioning & placement

Improved data partitioning


Improved Programming
Models
HPC
Performance
Big Data
Productivity

"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov