Systems for Big Data

MSc Thesis Matchmaking Event, Th. May 14, 2020

H. Peter Hofstee

h.p.hofstee@tudelft.nl

TU Delft

Things to know

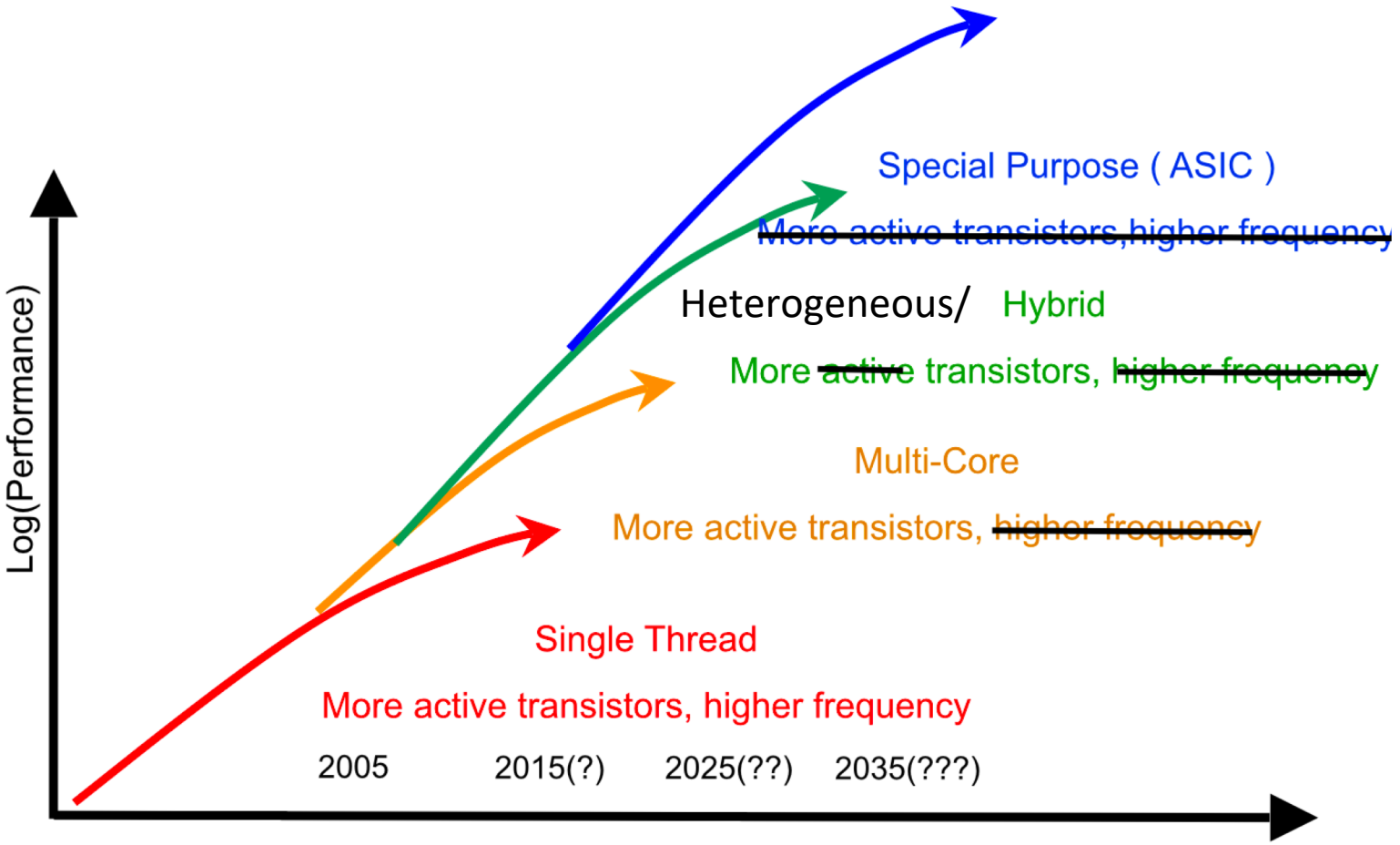
- About me: https://en.wikipedia.org/wiki/Peter_Hofstee
- I live in Austin TX, USA (day job at IBM) work on IBM POWER Systems
- Do a lot of work with Zaid Al-Ars and his MSc/PhD students: Accelerated Big Data Systems group
  - Please look at slides from his session if you have not attended it!

Things that interest me

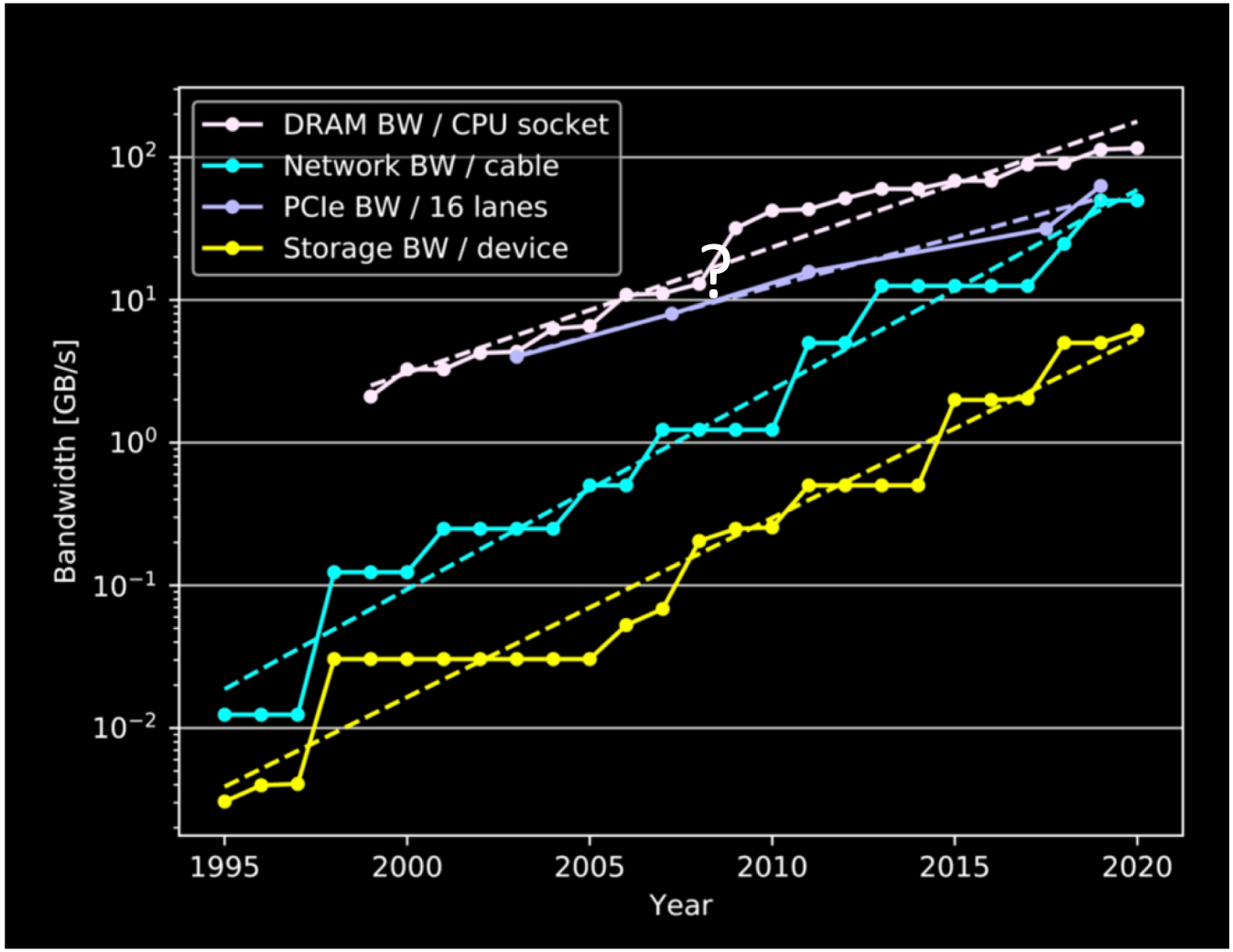
- Processor and system architecture
  - Especially in response to fundamental changes in underlying technology
- Heterogeneous Processing & Accelerators
  - Compression, sorting, data movers …
  - New floating-point formats
- Odds and ends
  - Synchronization
  - Security
  - Reversible computing

And many other things …

Microprocessor Trends



System (Bandwidth) Trends



Based on Sandisk Blog

Example Past MSc Thesis Projects

- Leveraging early form of nonvolatile memory for Apache SPARK
  - (Also done similar work for Cassandra etc.)
- A functional framework for Big Data
  - Replacement for SPARK
- Using FPGAs (with High Bandwidth Memory) for
  - Fast large sorts
  - Fast Database operations (Hash-Join, Filter, etc.)
  - Fast decompression and deserialization
  - New floating-point formats (Posits)
- Computer architecture
  - Raising level of abstraction: Object caches

Example MSc Thesis Projects Ideas

- Refining a concept for a Brain-Like computer
  - Sierra & Summit supercomputer are large enough to approximate human brain-like complexity (I have a high-level design in mind)
- Many ways to further leverage/improve FPGAs
  - Integrate w. Big Data frameworks (e.g. Dremio)
  - FPGA frameworks/tools (build on Fletcher)
  - Leverage our tools to build new solutions
  - Further improve on basic operations
- More computer architecture topics
  - Raise level of abstraction
  - Copy vs. cache
  - Reversible computing on FPGAs (???)

Note: not an exhaustive list … I am interested in many things, happy to have A one on one chat to brainstorm to find an idea that interest both you and me!