HPCG: Architectures and Performance

Peter M. Kogge
McCourtney Prof. of CSE
Univ. of Notre Dame
IBM Fellow (retired)

Goal

- Match each report with actual hardware
- Correlate performance with hardware & system parameters
  - Hardware: Core type, Peak flops, bandwidth, …
  - System: System architecture, …
- Look at results thru lens of architectural parameters
- Do so in way that allows apples-apples across benchmarks
- Note: not all current reports fully correlated

2D Architectural Classification

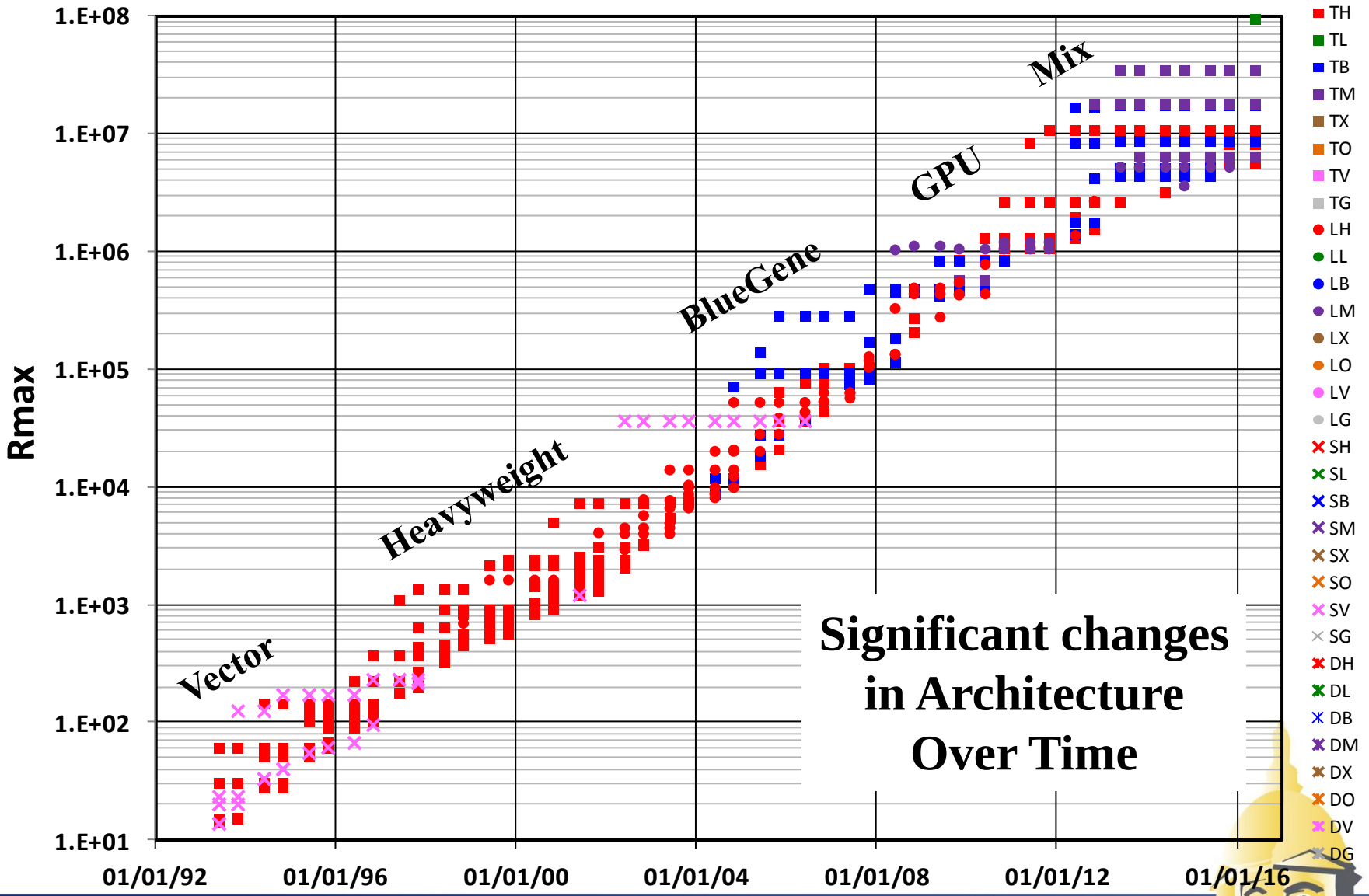
System Interconnect

- **L**: Loosely coupled distributed memory
  - Commodity networking with software I/F
- **T**: Tightly coupled distributed memory
  - Specialized NICs & some H/W RDMA ops
- **S**: Shared Memory
  - Single domain in H/W
- **D**: Distributed Shared Memory
  - Single domain but S/W assist for remote references (typically via traps)

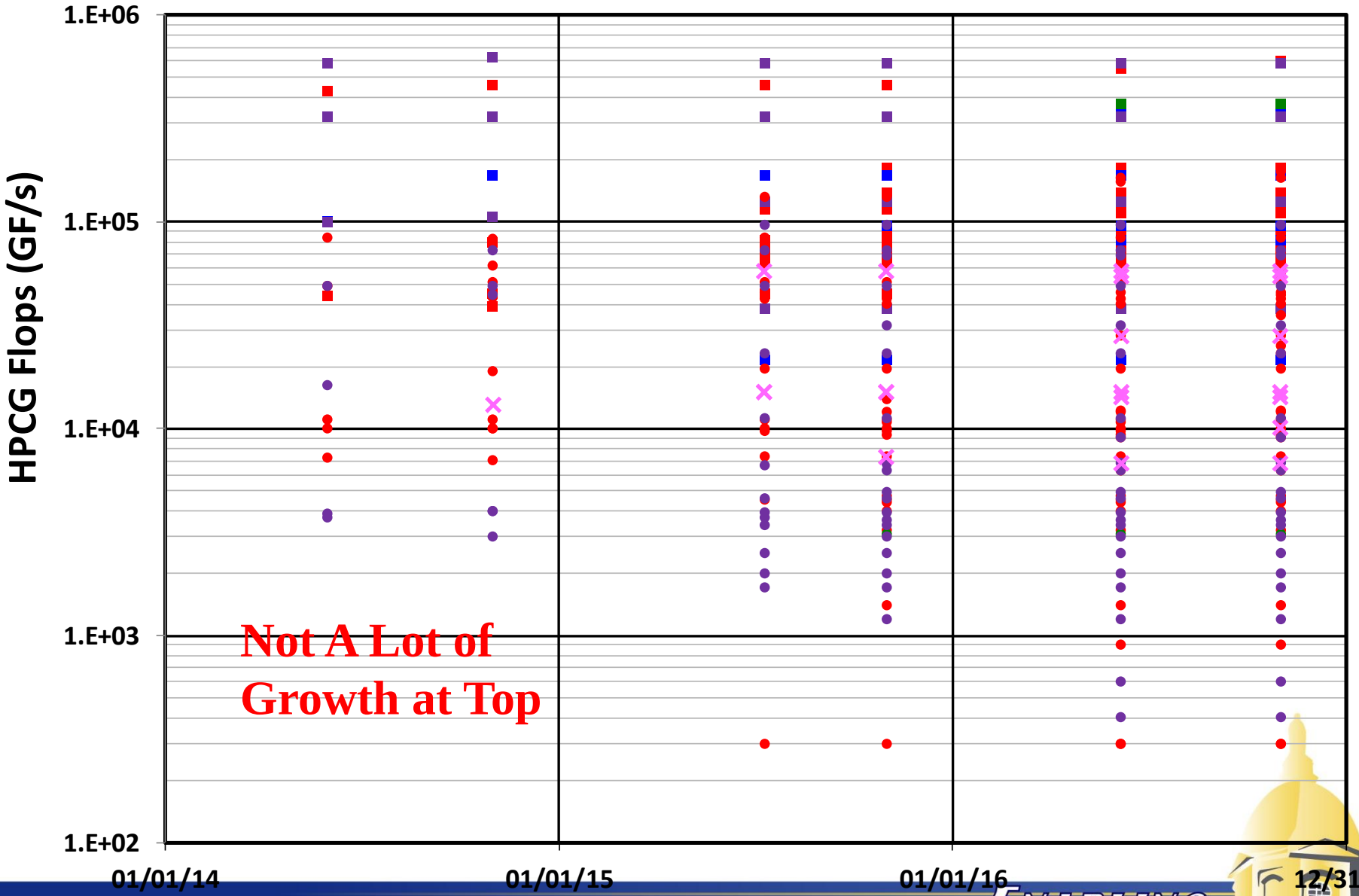
Core Architecture

- **H**: Heavyweight
- **L**: Lightweight
- **B**: BlueGene
- **X**: Multi-threaded
- **V**: Vector
- **O**: Other
- **G**: GPU-like
- **M**: a mix

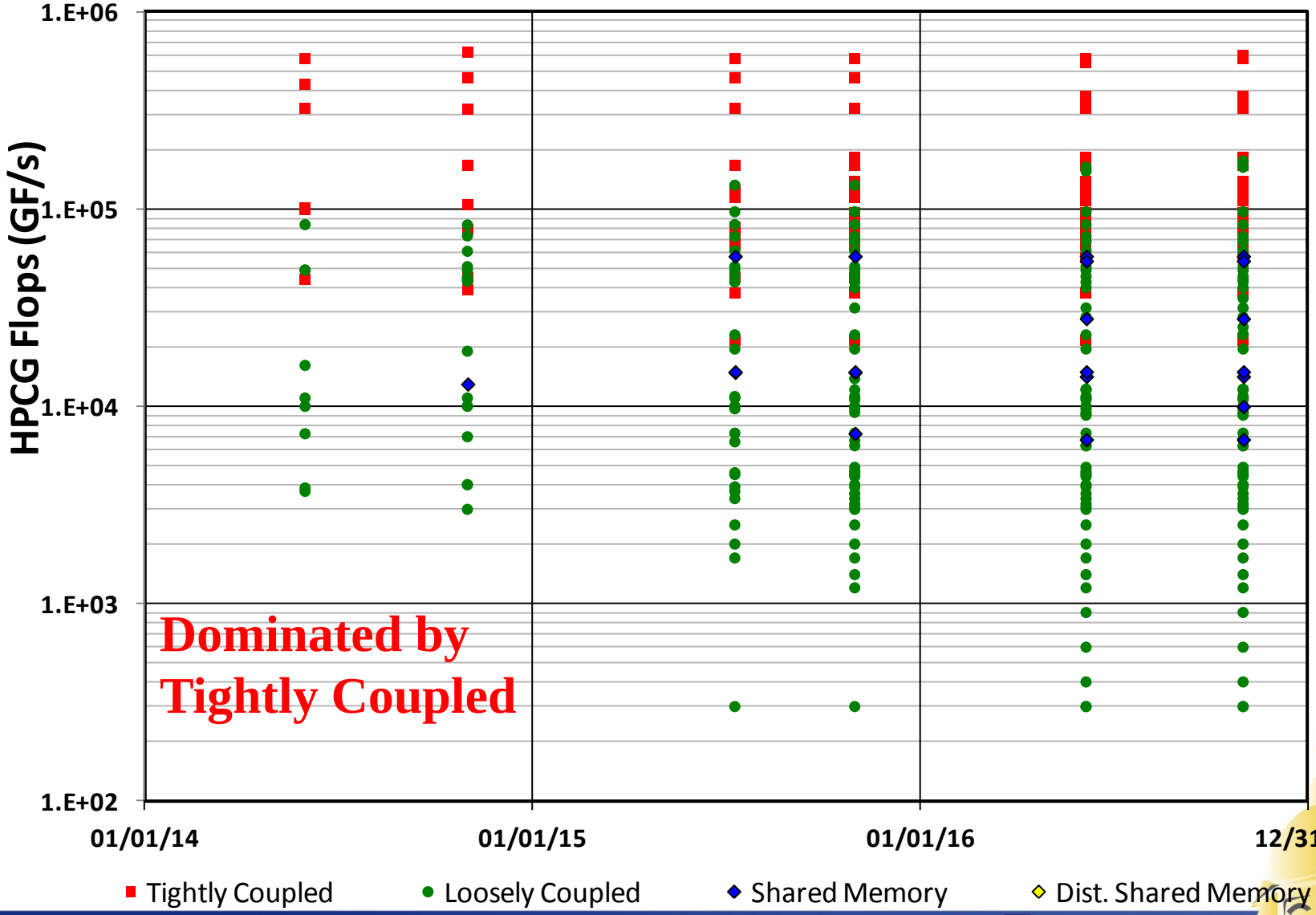
HPL Architectural Change



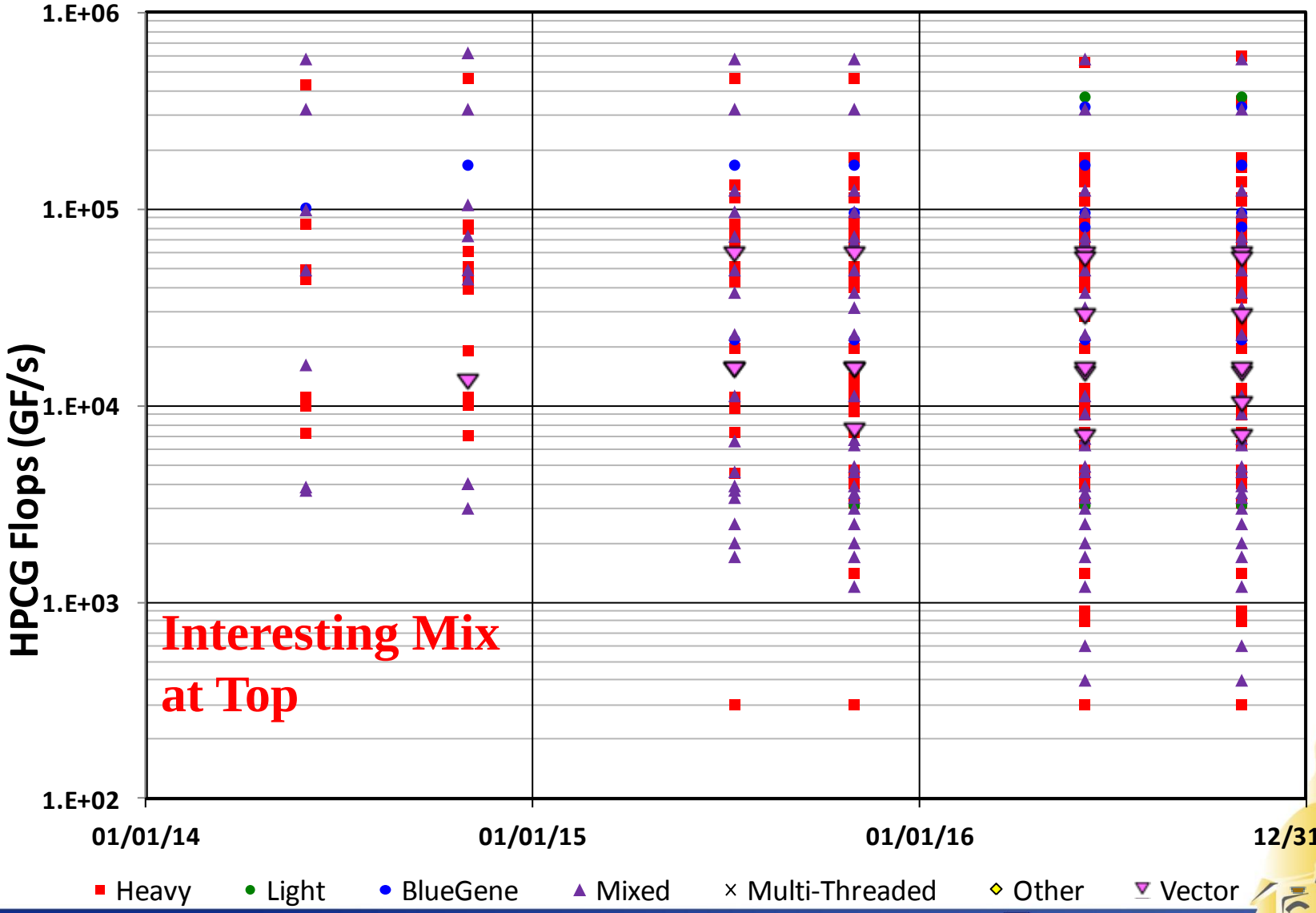
HPCG Over Time



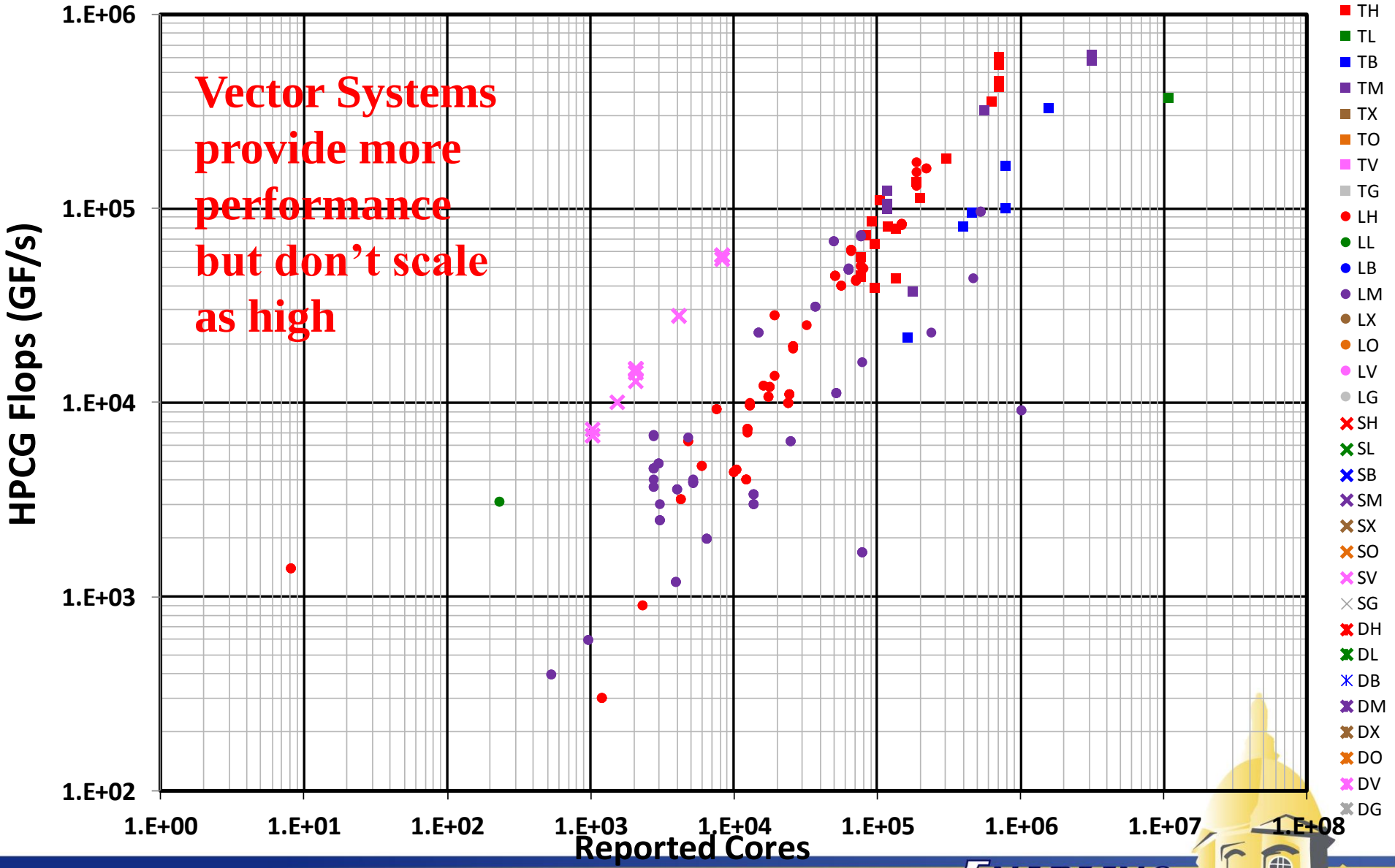
Trends by System Class



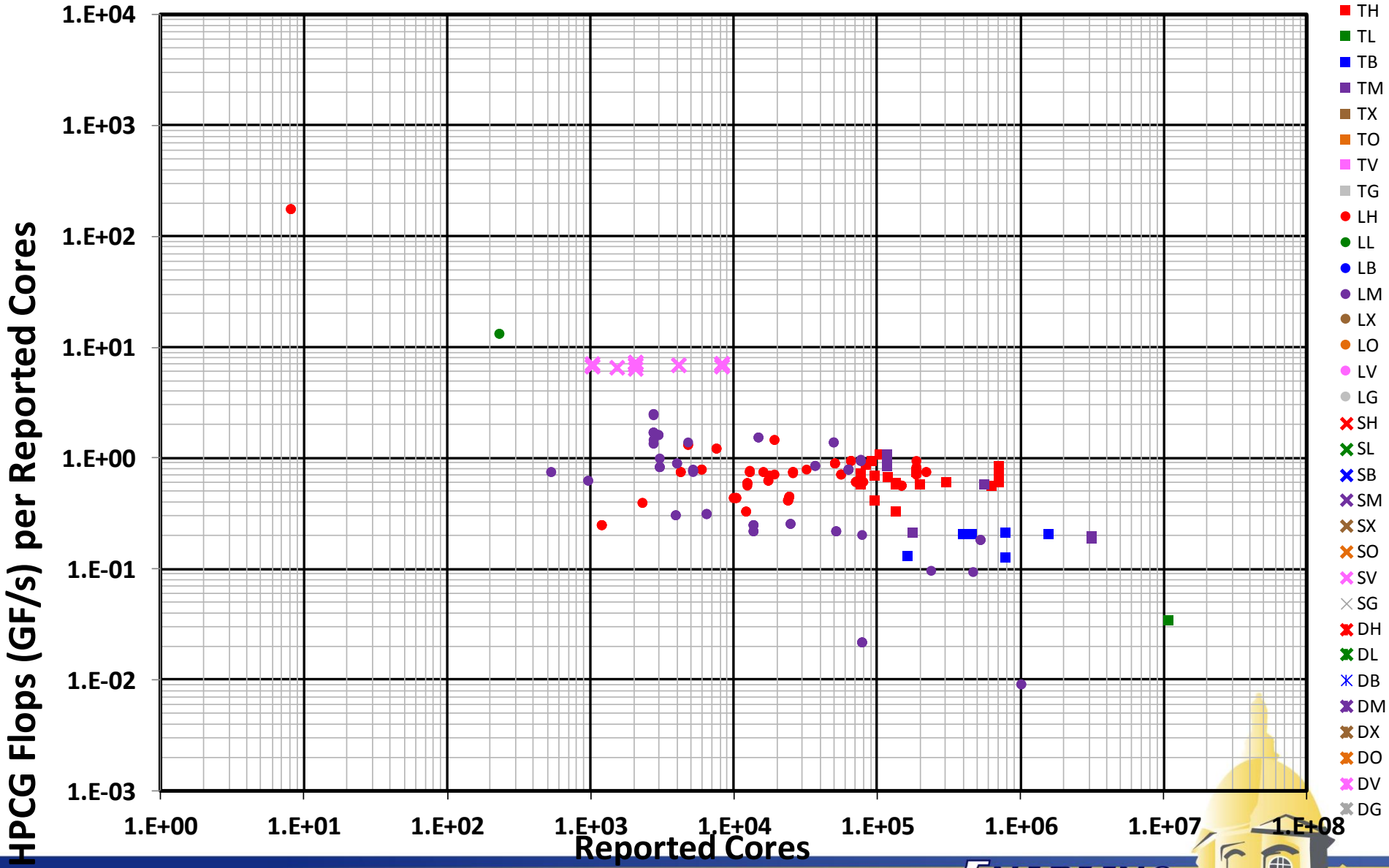
Trends by Core Architecture



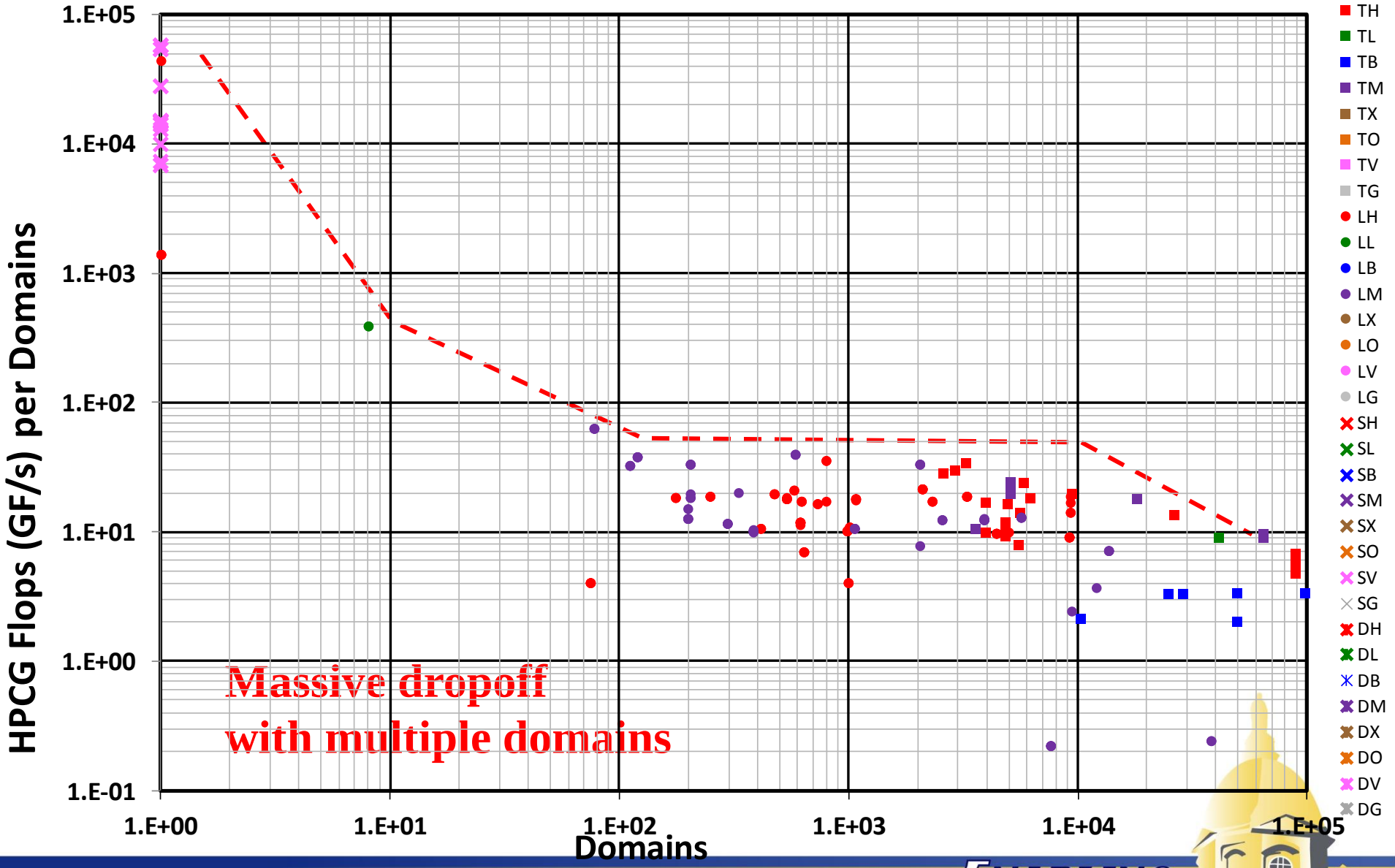
Performance vs # Cores



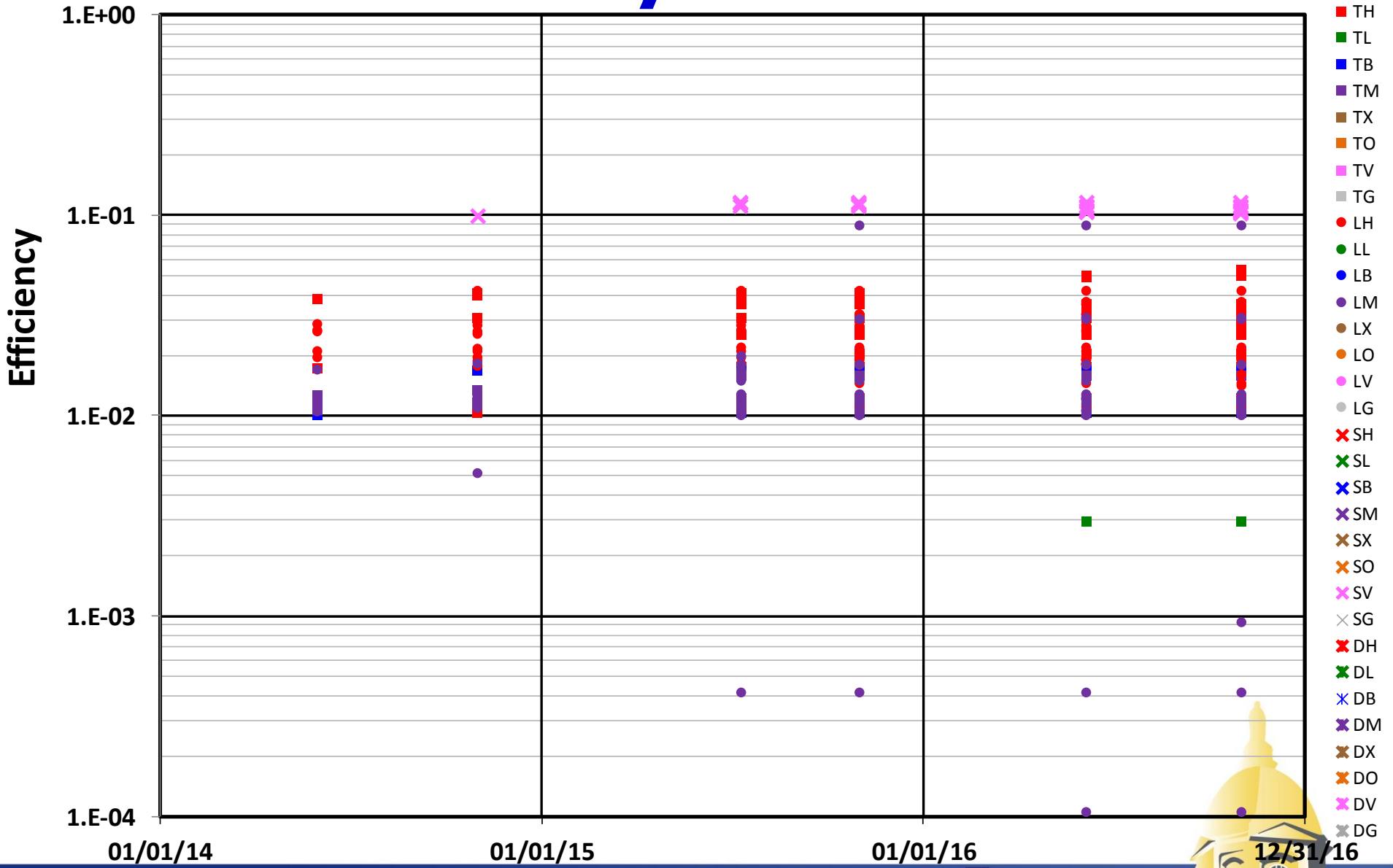
Performance Per Core



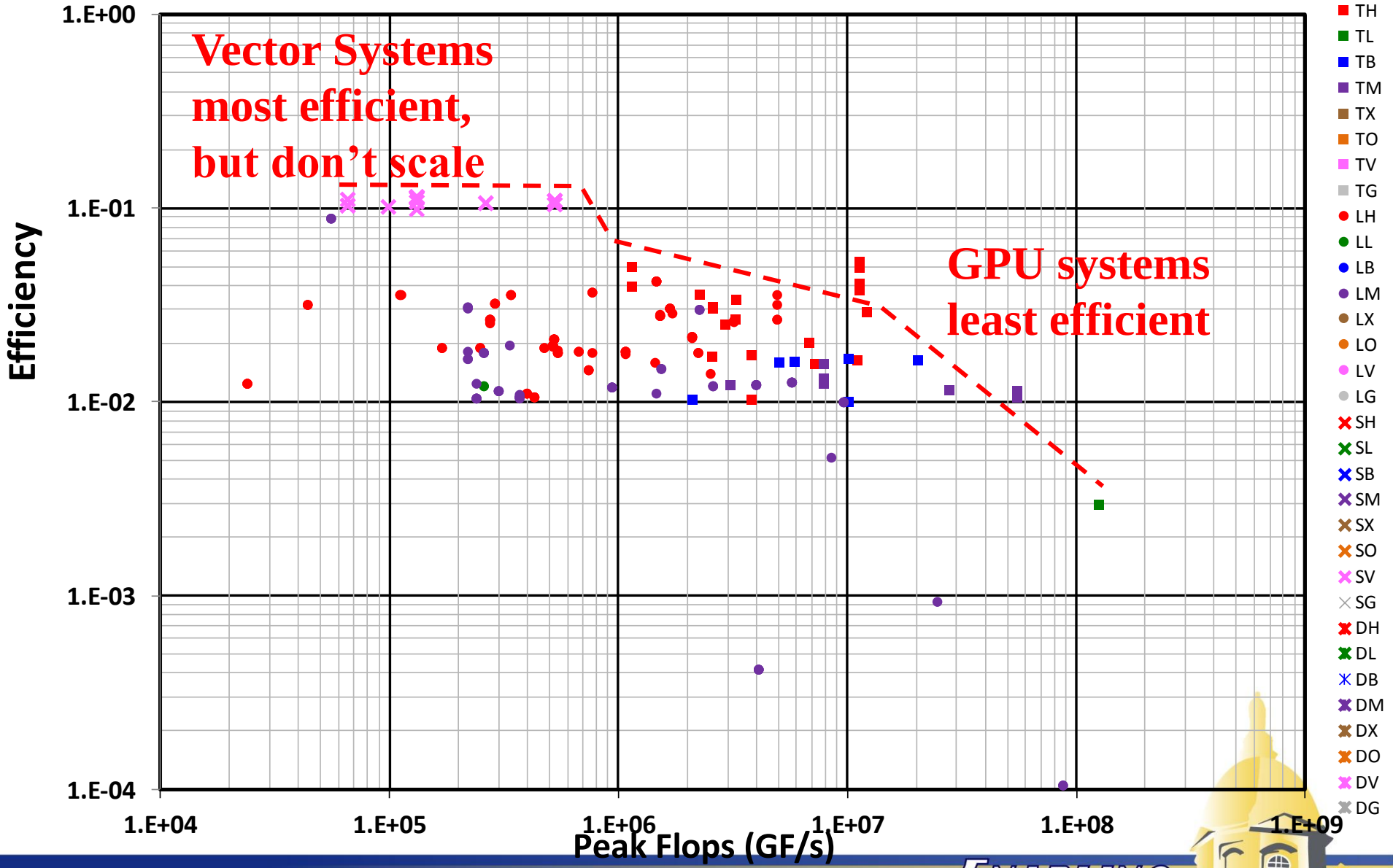
Performance per Domain



Efficiency over Time

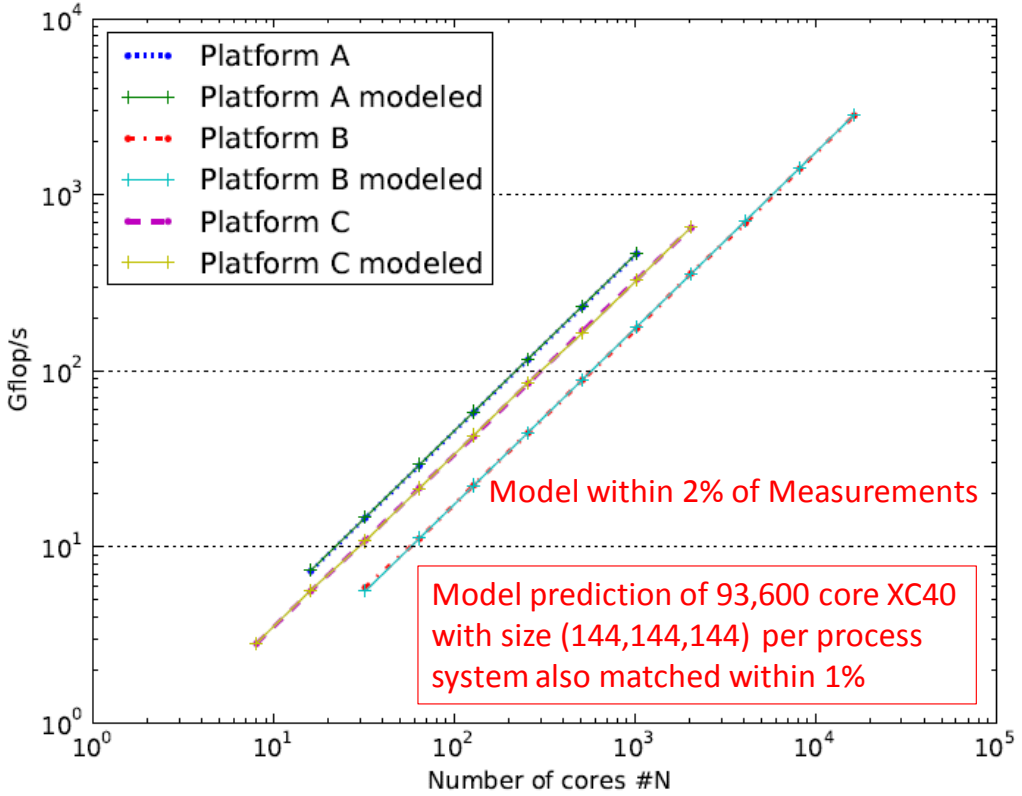


Efficiency vs Rpeak

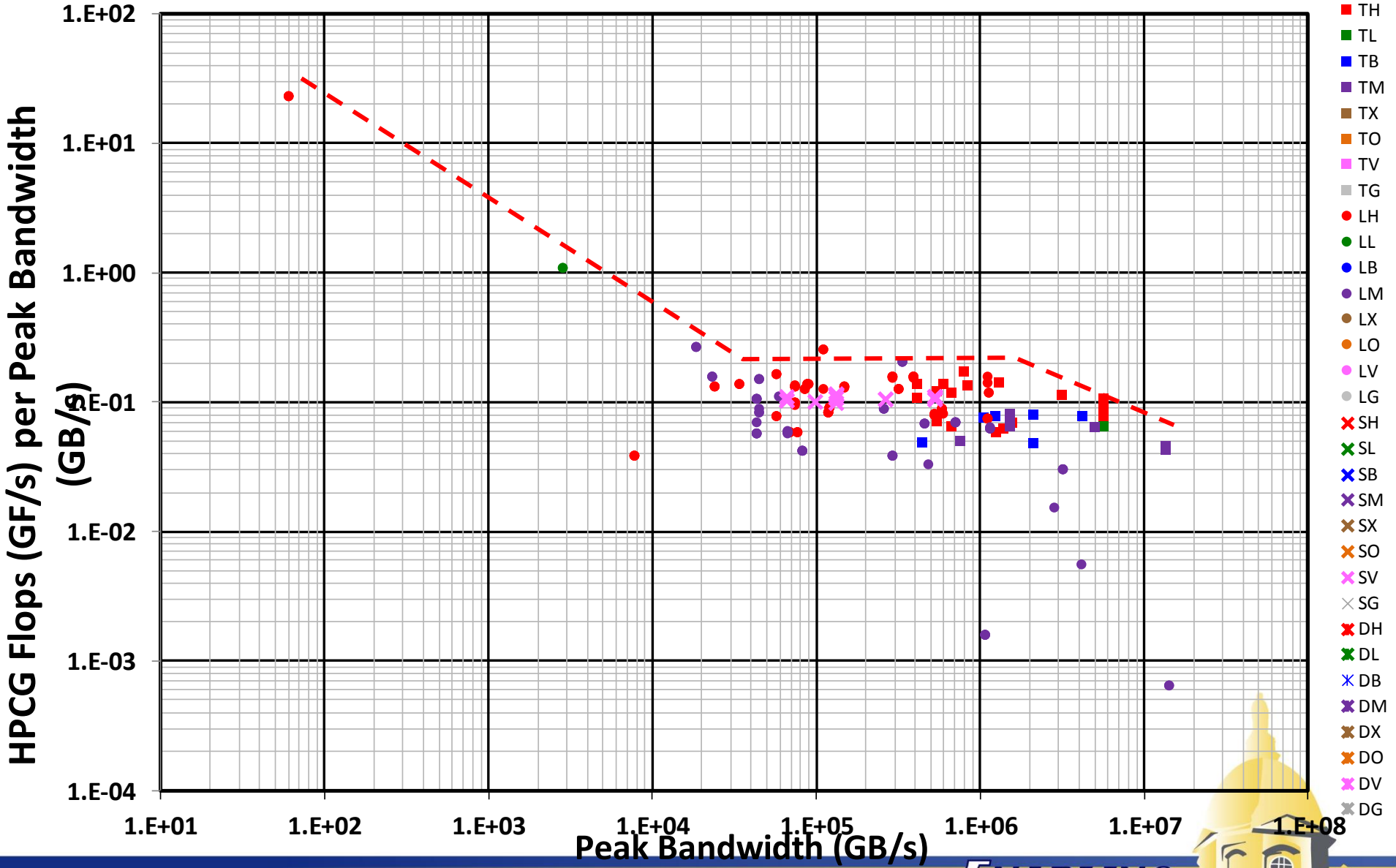


An HPCG Performance Model

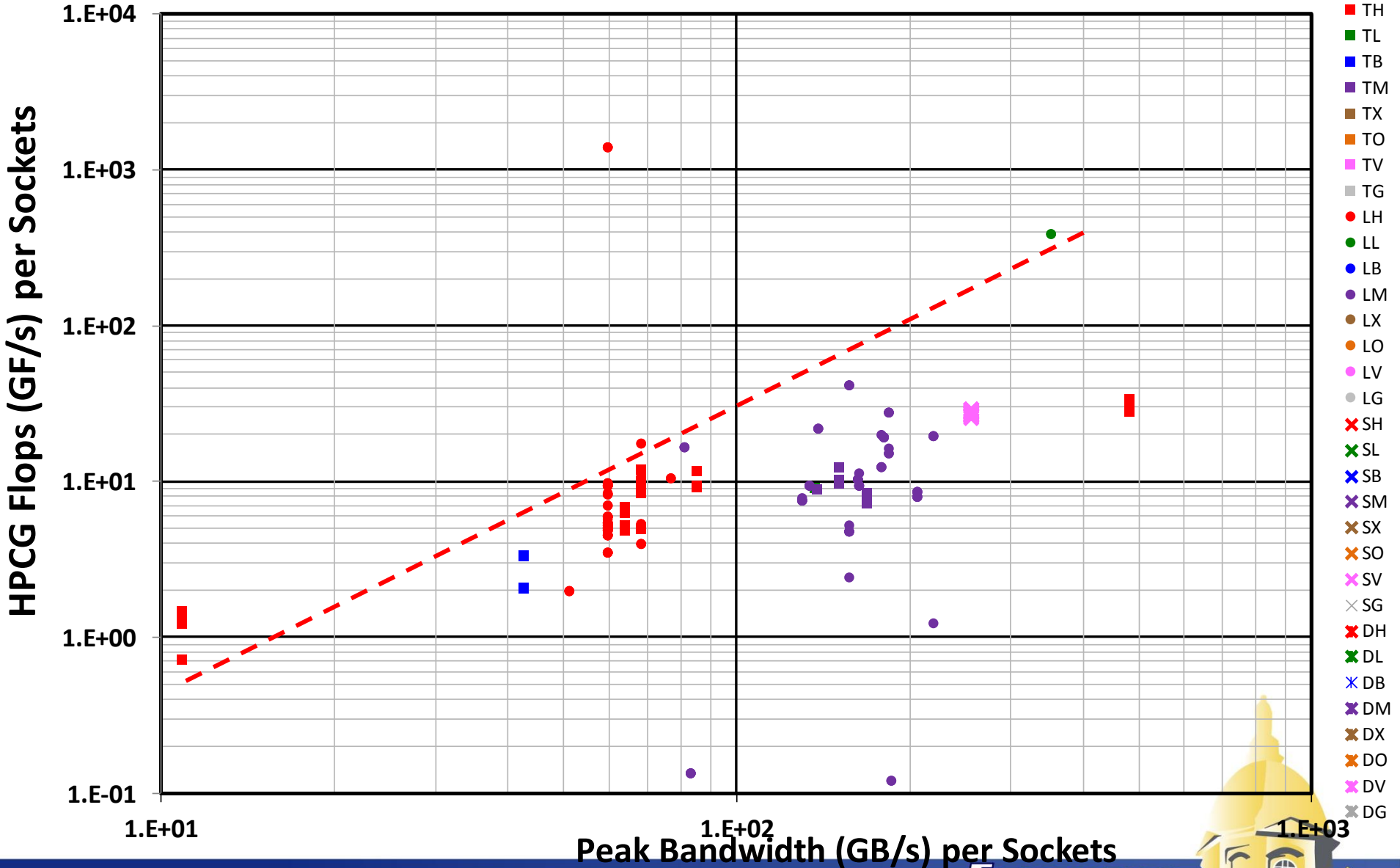
- from Marjanovic et al, "Performance modeling of the HPCG benchmark"
- Model based on:
  - # Non-zeros/row
  - Memory Bandwidth
    - Stream Metric
  - (Lesser) Network Bandwidth



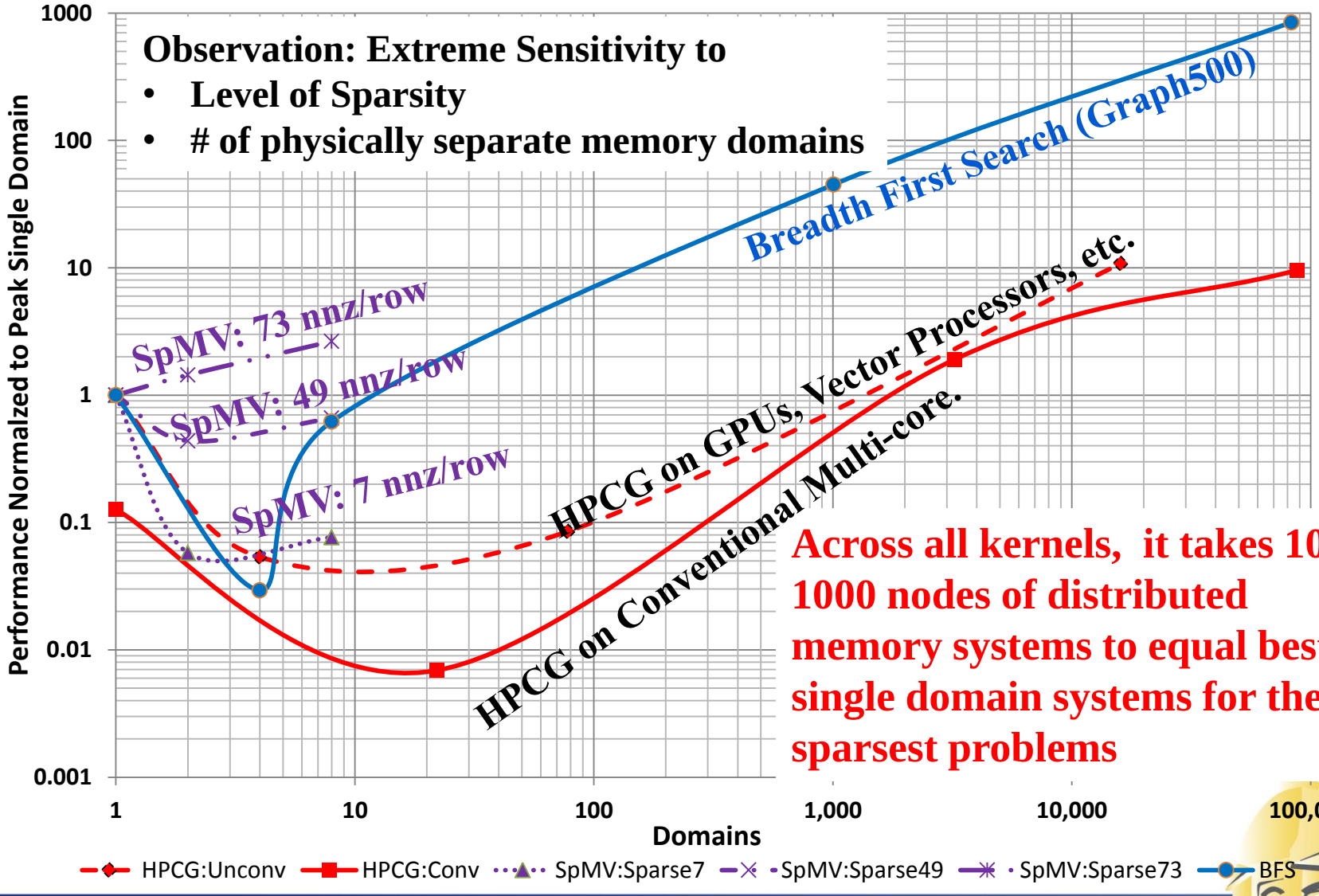
Per Memory Bandwidth



Per Bandwidth & Socket



Sparsity & Parallelism



Across all kernels, it takes 10-1000 nodes of distributed memory systems to equal best of single domain systems for the sparsest problems

Conclusions

- 3 Performance regions
  - Single Domain: highest performance per core – by far
  - < 1 Rack
    - Significant drop-off from single domain
    - But excellent weak scaling
    - Especially shared memory vector machines
  - >1 Rack
    - Some dropoff from single rack
    - But again good scaling up to about 1 million cores
- Strong correlation with memory bandwidth
  - Will be looking at recasting as memory access rate
- Plea: PLEASE REPORT <Max Parallelism