SAND2008-1166C

Trends in parallel computing and their implications for extreme-scale parallel coupled cluster

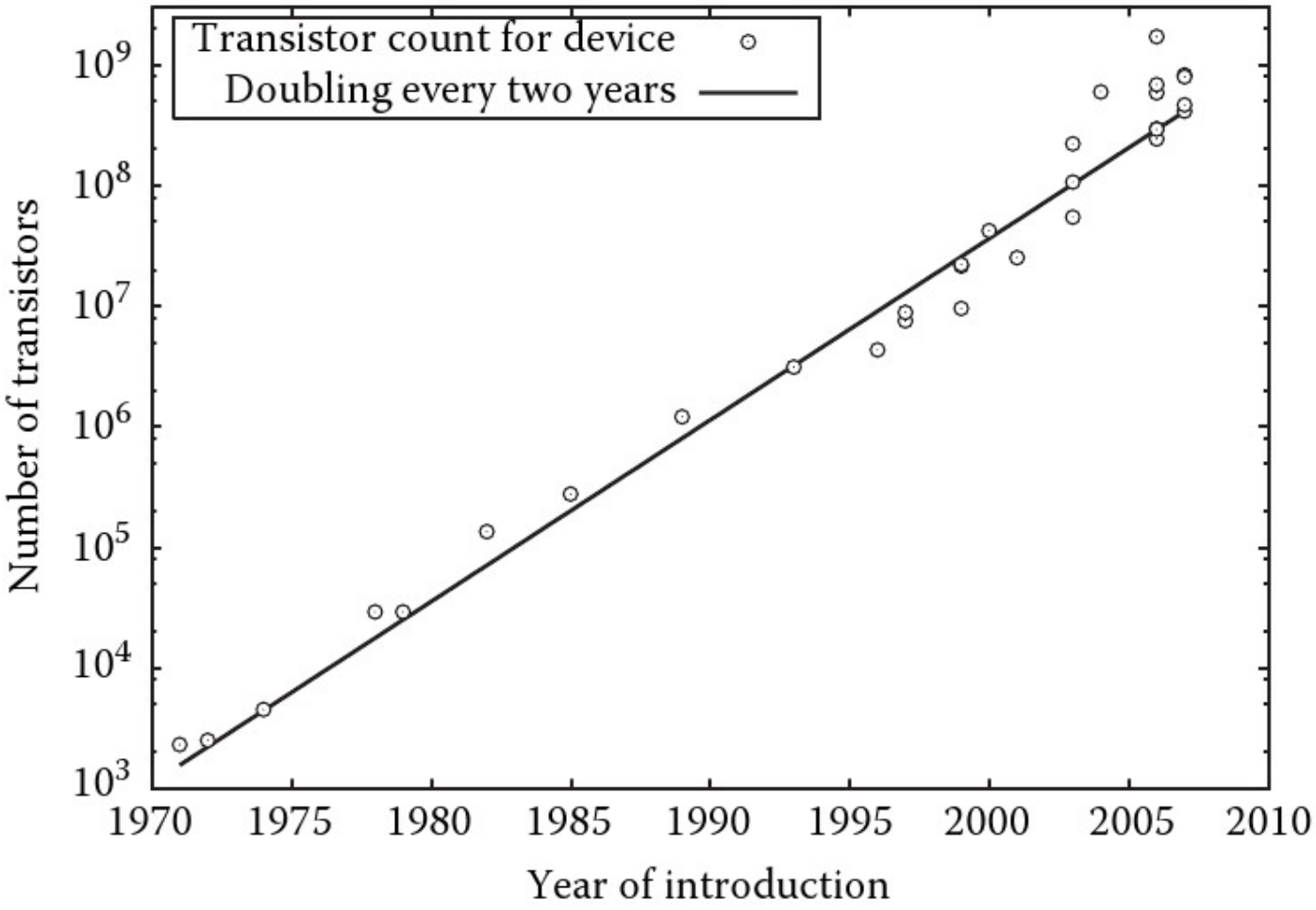
Workshop on Parallelization of Coupled Cluster Methods

February 23 to 24, 2008
St. Simons Island, GA

Curtis Janssen

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

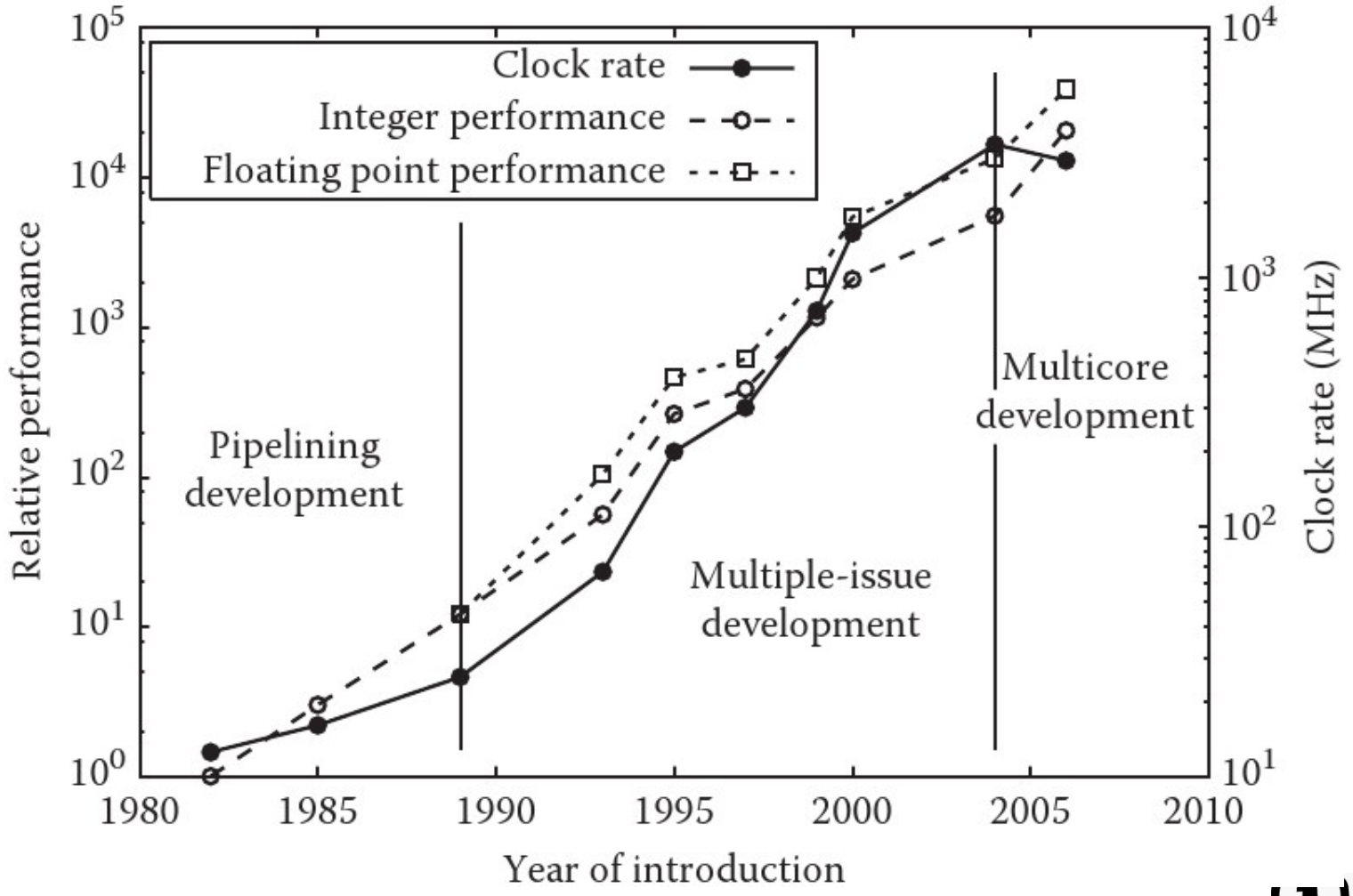
Cost of semiconductor production continues to follow Moore's law



Reproduced with permission from:
C. L. Janssen, I. B. Nielsen, *Parallel Computing in Quantum Chemistry*, CRC Press, May 2008.

Sandia National Laboratories

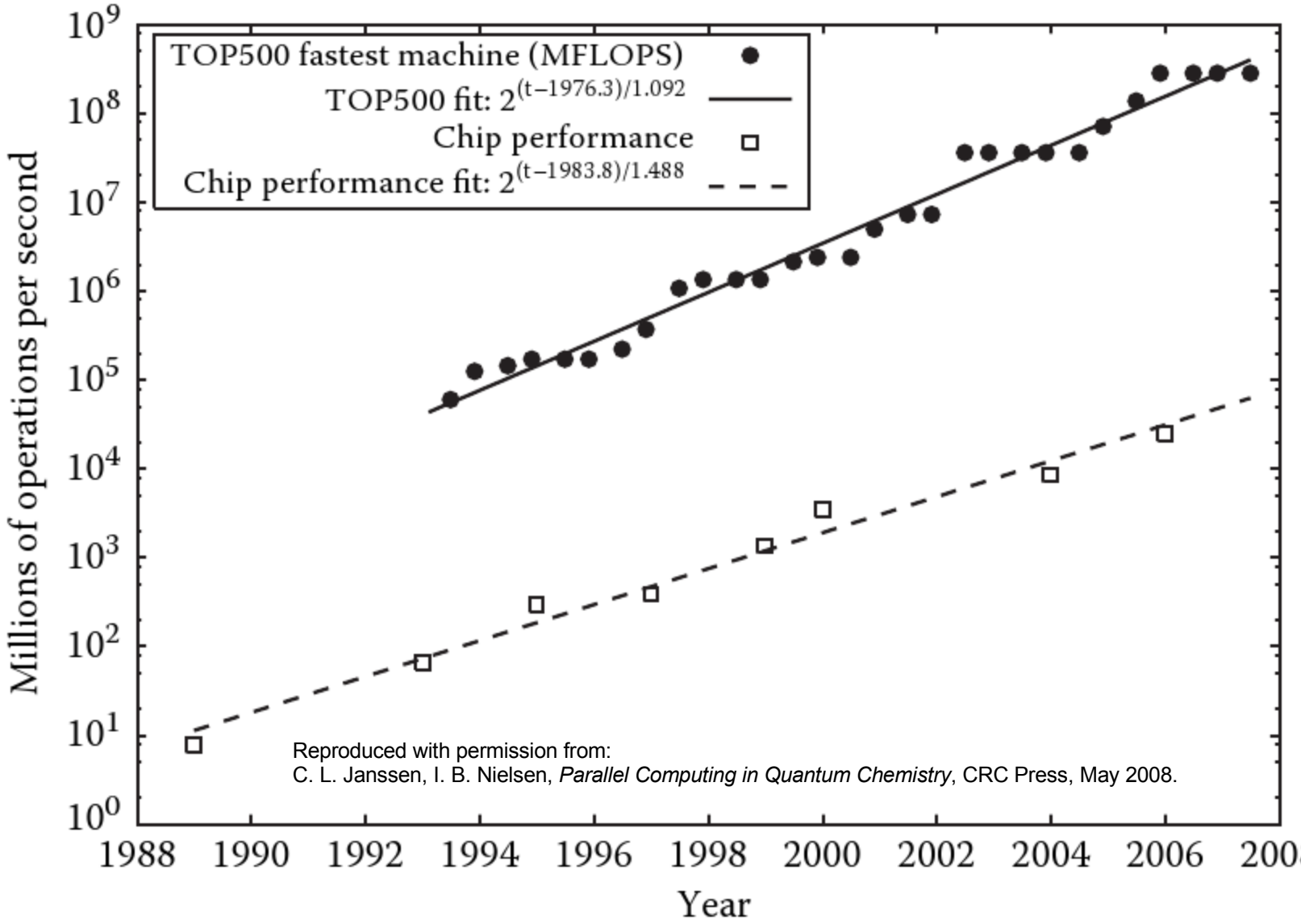
Moore's law affects realized performance in a variety of ways



Reproduced with permission from:
C. L. Janssen, I. B. Nielsen, *Parallel Computing in Quantum Chemistry*, CRC Press, May 2008.

Sandia National Laboratories

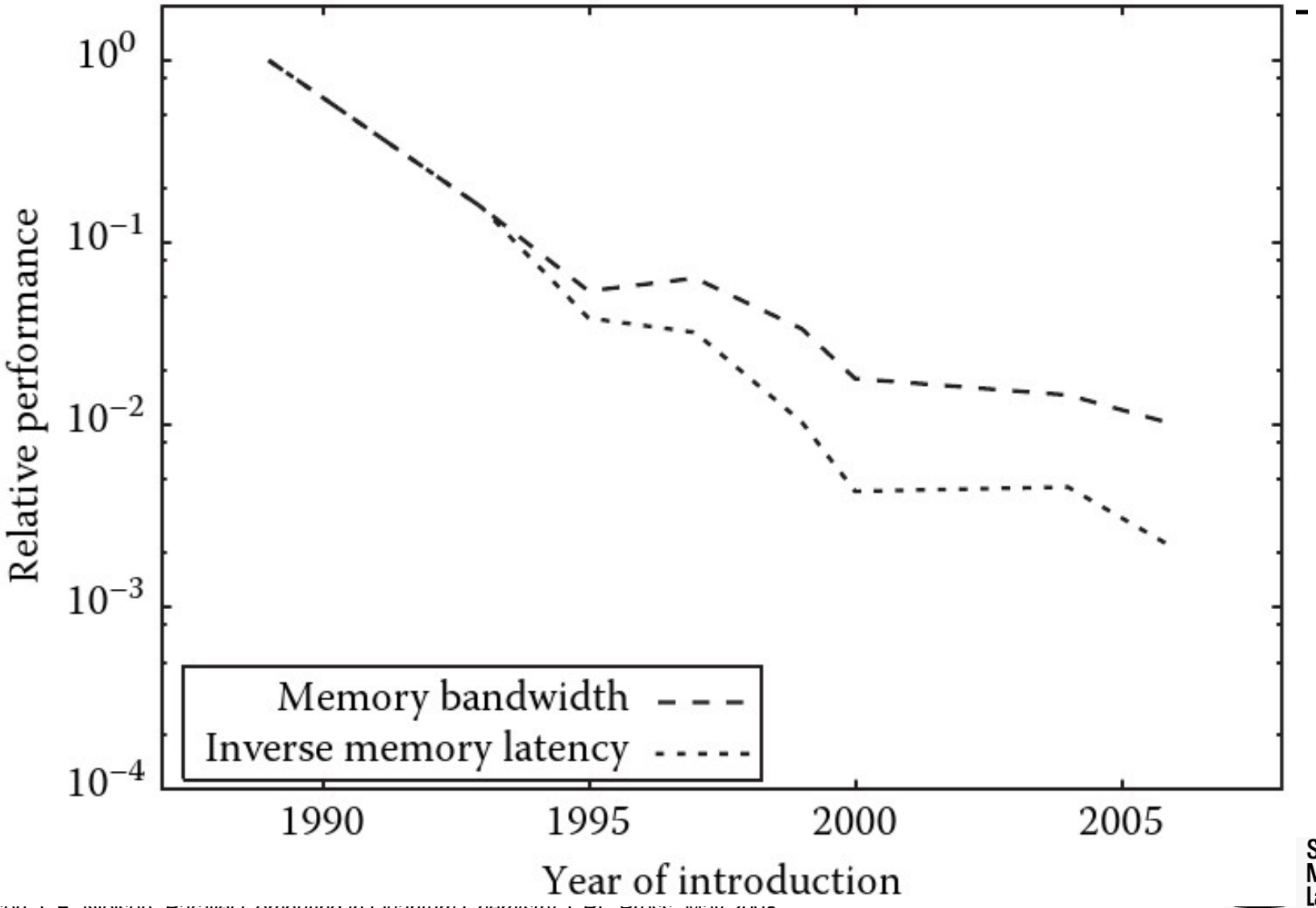
Parallel machines increase performance by faster chips and more chips



Reproduced with permission from:
C. L. Janssen, I. B. Nielsen, *Parallel Computing in Quantum Chemistry*, CRC Press, May 2008.

Sandia National Laboratories

These improvements in speed are not matched by latency and bandwidth



Reproduce
C. L. Janssen, I. B. Nielsen, *Parallel Computing in Quantum Chemistry*, CRC Press, May 2008.

Sandia National Laboratories

Net result: future parallel environments very different from today's.

- Multiple levels of memory (both local and remote)
- Failure rate proportional to number of components
  - Can we assume future machines are 100% reliable?
- Even if crashes avoided what about “brown outs”?
  - Thermal throttling, cores going offline, ECC recovery.
  - Nodes might provide heterogeneous performance.

Efficient utilization of future machines will be hard.

We also must consider how the application will be transformed.

- We all agree that high accuracy quantum methods are important, but how best employed at such scale?
  - Canonical CC methods?
  - Reduced scaling CC methods?
  - Periodic CC methods?
  - All of the above.
- Obligated to provide the most impactful science possible on such large and expensive machines.
  - With $100M machine, 5 yr lifetime, a few MW power, several FTEs support: cost to run a one week job on entire machine > $500,000.

We also must consider how the application will be transformed.

- 1 week of an exaflop machine is about 2500^7 flops
  - 2500 basis functions is about 22 CH_2 units
  - Is canonical CCSD(T) the best way to utilize this machine? —Probably not.

Reduced scaling methods have tremendous parallelization challenges

- **Canonical MP2, parallelizing only the *O(N^5)* step:**

$$t_{MP2}(N,p) \approx A\frac{N^5}{p} + BN^4 \qquad S_{max,MP2} = \frac{AN+B}{B}$$

- **Local MP2, parallelizing most of the *O(N)* steps:**

$$t_{LMP2}(N,p) \approx C\frac{N}{p} + DN \qquad S_{max,LMP2} \approx \frac{C+D}{D}$$

In a local method, data is more irregular, making load balancing more difficult—this is where current programming models break down.

Current programming models do not scale

- MPI+Remote Accumulate has worked up to now
- MPI+Remote Accumulate+threads will help use scale up — but there are problems
  - MPI is difficult, threads is difficult, hybrid is difficult[2]
  - Results in a fragile environment, expensive to develop, debug, and obtain portable performance
- Memory hierarchy is deep, but imperative programming languages encourage random access
  - Need to think of new models akin to dataflow

Key Points

- Getting good scaling to > 100,000 will be hard.
- Current methods must expand and adapt to the types of problems that will be solved at that scale.
- Must reexamine current programming models to find best way to allow human scalability—entire software life cycle must be considered.