



Gelato

UNSW

Performance and
Scalability on Itanium

www.gelato.unsw.edu.au



Linux Kernel Limits to Performance and Scalability

Peter Chubb

Gelato Project

peterc@gelato.unsw.edu.au

Computer Science and Engineering

UNSW

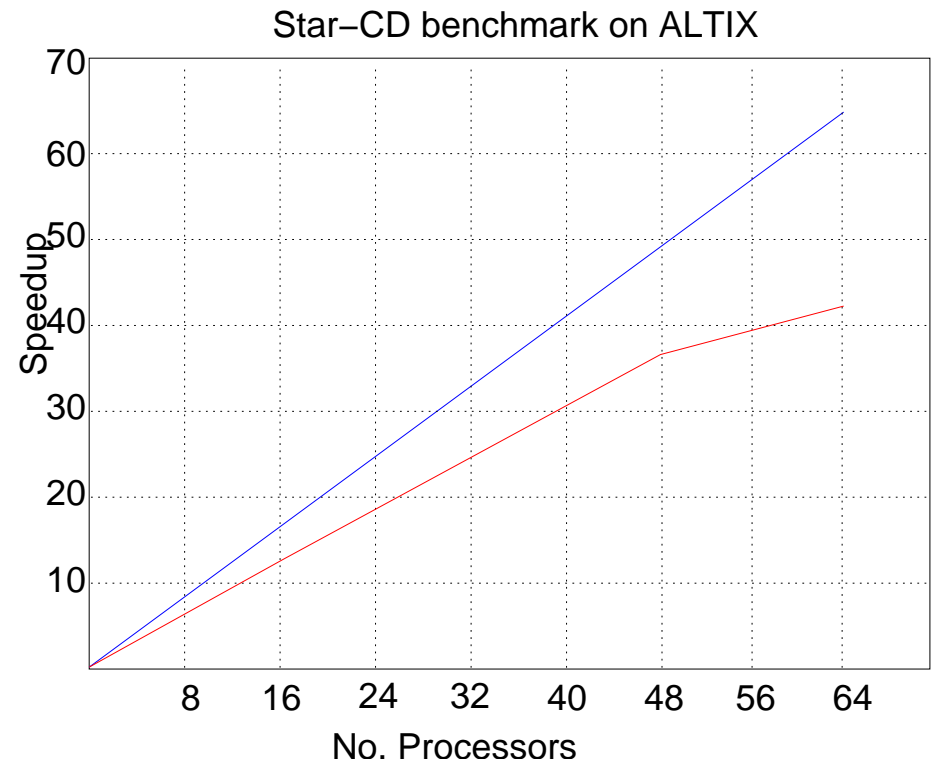
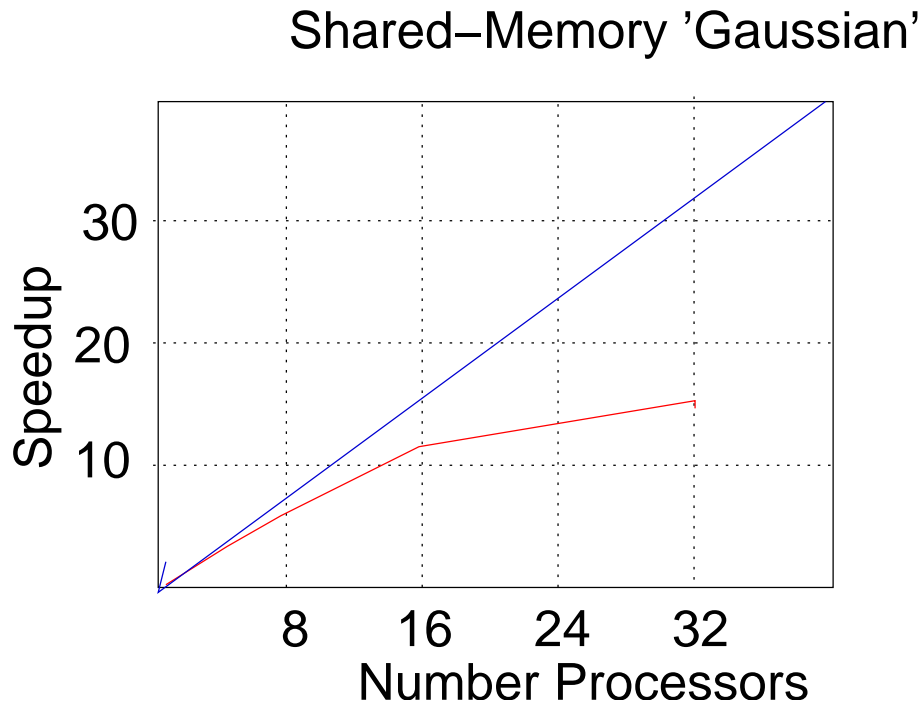
Monday 13th of October



Linux Kernel Scalability

- Core kernel reasonably good
 - system time small in most benchmarks
 - SGI Altix scales to > 64 processors
- However ...

Carrier Grade Linux work at OSDL: Linux works fine until the workload gets heavy.



(from OLS presentation by SGI, June 2003, 2.4 kernels)



Improvements

- Better use of hardware resources
- Disk I/O
- Network I/O
- Performance measurement
- Accounting
- Job isolation
- Reliability



- Memory
 - Superpages
 - More sharing (TLBs, pagetables, etc)
- Bus/interconnect bandwidth
 - Use zero-copy as far as possible
 - Replicate read-mostly memory segments
- Processor
 - reduced overhead — system calls, context switch, etc
 - Better Compilers Needed!!!



Disk I/O

- Scales to three spindles (IA32)
- Max throughput reported on Ise-tech as 30% of thoretical.
- Need to validate on IA64 — may be a HIGHMEM effect



Network I/O

- Unaligned accesses
- TCP/IP problems?



Reliability

- Failure modes currently ungraceful (OOM killer etc)
- Bug fixes



Profiling and Performance measurement

- Many disparate tools and patches (qprof, oprofile, gprof, &c)
- Need specialised kernel tools
 - μ -state accounting
 - Lockmeter
 - Semaphore meter
- Unification desirable — collect data from more than one source in one run; present together.



Accounting & Job Isolation

- Virtual machines
- Class schedulers (IBM's work)
- CPUsets



Working Together

- Communicate
 - Use the Wishlist
 - Use the WiKi
 - Use the Fora



What are *your* problems?