Optimization of SV1 Application Codes in a Production Environment

David Turner and Tina Butler

NERSC, Lawrence Berkeley National Laboratory

Mike Stewart and Bob Thurman

Cray Inc., Lawrence Berkeley National Laboratory

Background

- Cray customer for over 20 years
- Many long-time users
- Currently three SV1s and one T3E
- Resources allocated by DOE Office of Science

Objectives

- The impact on 24x7 production on optimization
- The feasability of MSP in production environment

NERSC PVP Cluster

Name	Processors	Memory	Purpose
Killeen	16	1 GW	interactive
Seymour	24	1 GW	batch
Bhaskara	24	1 GW	batch

- Each machine has about 350GB of fast RAID disk
- Killeen exports about 140GB of home directories

Execution Environment

- Large interactive limits (80MW/10 CPU-hours)
- NQE, with simple queue structure
  - 80MW, 256MW, and 512MW
  - Time limit: 120 CPU-hours, disk limit: 80GB
- Class of service batch priority system
  - Determines how long a job remains pending
  - Determines charging

    Premium: 2.0, Regular: 1.0, Low: 0.5
  - Example: Average pending time in February 2001

    Premium: 2.5 hours, Regular: 14 hours, Low: 78.5 hours
- All CPUs oversubscribed

The Benchmarks

- Users contacted based on allocations and/or usage
- t743lin1

  Toroidal nonlinear 3D-MHD equations
- xqcd_hot

  Lattice QCD
- classic

  Core collapse supernova simulation

Initial Observations

Machine	User	Sys	Total	nts	Ratio
Killeen	815.6	9.8	825.4	250	
Seymour	1013.4	13.6	1027.0	250	1.244
Killeen	2334.7	111.8	2446.5	750	
Seymour	2870.7	53.9	2924.6	750	1.195

- Killeen 0.25 conflicts/reference
- Seymour 0.41 conflicts/reference
- Killeen 20% faster!

Investigating Memory Conflicts

"Pounder": eight-way autotasked, vectorized, 150MW

	Killeen		Seymour	
Pounders	User	Con/Ref	User	Con/Ref
none	253.7	0.16	254.0	0.16
8/150	273.3	0.20	273.0	0.20
16/300	342.4	0.46	320.8	0.31
24/450	335.6	0.43	429.7	0.66

HPM Group 2

Killeen: 0.4 conflicts/memory refernce

Seymour and Bhaskara: 0.6 – 0.7 conflicts/memory reference

Impact of Memory Conflicts

- Timing results masked by memory contention
  - Consistent when run closely
  - Wide variation day-to-day
  - Wide variation machine-to-machine

The Multi Streaming Processor

- Four SV1 processors combined to form a single MSP
  - Similar to autotasking
  - More efficient synchronization and communication
- Originally required reboot; only execute MSP code
  - Too much idle time
- Both restriction now gone
- Cannot mix MSP and autotasking

Initial MSP Results

N	User	Sys	Total	Elapsed	Opt
1	1928.7	17.7	1946.4	2980.5	task3
4	2659.5	102.6	2762.1	1285.7	task3
4	2767.2	2.9	2770.1	696.6	stream3

N	User	Sys	Total	Elapsed	Opt
4	3135.5	3.8	3139.3	1558.4	stream3
4	3083.0	4.6	3087.6	1544.8	stream3
4	3103.3	3.6	3106.9	1557.4	stream3
4	3123.6	3.6	3127.2	1560.4	stream3

Additional MSP Results

N	User	Sys	Total	Elapsed	Opt
1	4594.8	54.8	4649.6	9637.8	task3
4	5141.4	352.4	5493.8	3204.3	task3
4	5787.3	7.1	5794.4	1778.9	stream3

N	User	Sys	Total	Elapsed	Opt
1	710.6	13.2	723.8	1561.1	task3
4	1064.1	51.2	1115.3	726.3	task3
4	2944.7	5.3	2950.0	1122.5	stream3

SV1 vs. SV1e

SV1

N	User	Sys	Total	Elapsed	Opt
1	924.6	13.0	937.6	2531.8	task3
2	1212.4	14.4	1226.8	1202.0	task3
4	1252.9	50.0	1302.9	1002.9	task3
4	1339.4	2.1	1341.5	341.1	stream3

SV1e (dedicated)

N	User	Sys	Total	Elapsed	Opt
1	529.8	0.1	529.9	530.3	task3
2	628.0	0.1	628.1	314.4	task3
4	765.1	0.2	765.3	191.8	task3
4	752.1	0.3	752.4	188.3	stream3

Conclusions

- Variability of CPU charges based on system load
  - Code on Killeen 20% faster than on Seymour or Bhaskara
  - Wide variations on single machine
  - Complicates resource management
- MSP performance lacking
  - Better turnaround
  - Reproducibility
  - Lack of scaling
  - Incompatible with priority class system