Performance of the ECMWF IFS T799 L91 Spectral Model on the NEC SX-8 Vector Supercomputer

Mike O'Neill
マイク・オニール

Working with NEC High Performance Computing Europe GmbH
NECハイパフォーマンスコンピューティングヨーロッパと共に働いています。

Email moneill@hpce.nec.com UK Mobile Phone: +44 7725656575

For the last 5 years the ECMWF Forecasts Have been running on large cluster systems.

IBM p690-Power 4+ and p575-Power 5+

High Performance Computing History at ECMWF before IBM

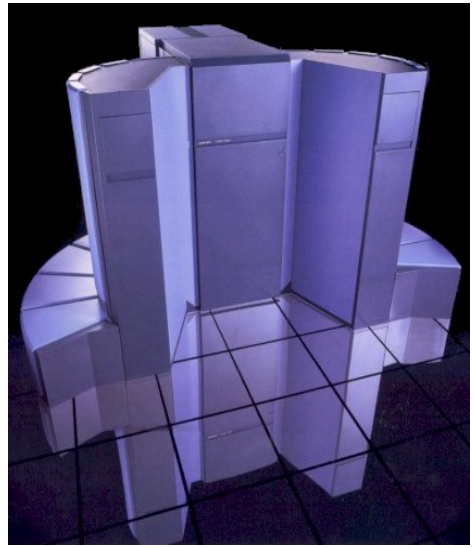
Year	Vendor	Model	Cpu type	Clock rate	# cpus	Peak Mflops	Sustained Mflops	% of Peak
1976	CRAY	Cray 1	Vector	80 MHz	1	160	100	65 %
1979	CRAY	Cray1A	Vector	80 MHz	1	160	100	65 %
1983	CRAY	XMP-2	Vector	105 MHz	2	420	235	56 %
1986	CRAY	XMP-4	Vector	118 MHz	4	944	453	48 %
1989	CRAY	YMP-8	Vector	167 MHz	8	2.7 Gf	1.1 Gf	42 %
1992	CRAY	C90-16	Vector	250 MHz	16	16 Gf	5.9 Gf	36 %
1994	CRAY	T3D	MPP	150 MHz	128	19.2 Gf	2.7 Gf	14 %
1996	Fujitsu	VPP 700	Vector	142 MHz	46	101 Gf	39.6 Gf	39 %
1997	Fujitsu	VPP 700	Vector	142 MHz	116	255 Gf	99.5 Gf	39 %
1999	Fujitsu	VPP5000	Vector	252 MHz	100	960 Gf	300 Gf	30%

From 1976 to 2003

CRAY 1 to VPP 5000

27 Years of Vector Computers at ECMWF

Clock rate increase	3:1
No of CPUs	100:1
Theoretical Peak Flop Rate	6000:1
Sustained Performance Increase	3000:1



The Old Cray Research Machines were more Photogenic

Parallelising the ECMWF Spectral model for the first time. 1983

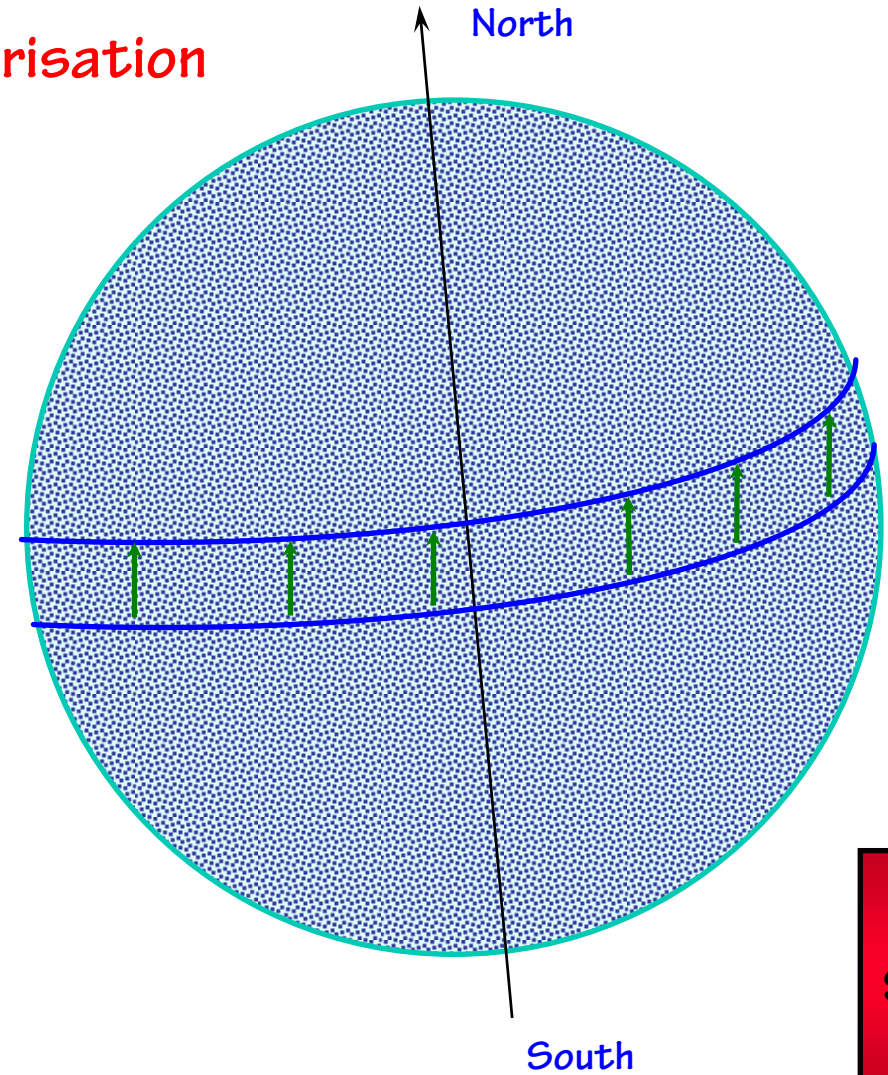
East West Vectorisation

Decomposition into 2 Hemi-spheres

Macro-Tasked work Distribution

ECMWF SPM Model T42 L15

Problem of Static Load balancing



CRAY XMP 2 Parallel Vector Machine

Solved by "moving" the equator Northwards

90% Parallel

Speed-up is 1.82 on 2 Processors.

David Dent ~ 1985

So the Question I address today is:-

Can Modern Vector Systems such as

The NEC SX-8 still have a performance

Advantage compared to Systems Built

With large Clusters of Scalar Processors

The T799 L91 model has been in Production from February 2006

Subsequently Tested on the UK Met. Office NEC SX-8 Systems

By Paul Burton and
Deborah Salmond of ECMWF

Deborah's Shopping List

T399 run	NEC SX-8	8 cpus	IBM P690+	32 cpus
Subroutine	Time secs	Mflops / cpu	Time secs	Mflops / cpu
CUADJTQ	146.6	1,038	62.1	613
CUBASEN	115.7	136	139	285
CPCUDDH	65.1	210	5.8	591
CUASCN	50.7	95	17.32	70
CLOUDSC	49.3	2,011	44.68	555
WVCOUPLE	31.5	4	4.68	8
RRTM-GASABSIA	26.6	111	2.16	408
CLOUDVAR	23.4	1376	20.8	387
FTINV	21.5	532	6.9	415
CUENTR	14.8	98	5.0	73
CUBASMCN	14.7	64	4.3	55
SLTEND	14.4	616	8.5	261
MXMAOP	13.5	12,875	14.0	3104
12 hour cpu time	259	2,375	282	563
System Performance	15 % of peak	19 G Flops	7.8% of peak	18 G Flops

Revised Shopping List After some basic Vectorisation

The T399 "Top Ten" on a short run

Subroutine Name	% of Cost	Mflops	V. OP Ratio	Average V length	Total Time	Vector Time
CLOUDVAR	8.1%	4,401.2	99.87	247.7	1333.85	1333.71
CUADJTQ	8.1%	8,089.2	99.59	198.6	1325.84	1253.84
CLOUDSC	6.8%	5,331.4	99.68	220.3	1110.38	1101.37
CUBASEN	5.2%	1,205.8	99.19	220.4	855.33	810.06
SGEMMX	4.4%	14,631.4	99.58	221.5	717.09	713.32
CUASCN	3.1%	3,356.5	99.61	208.7	505.57	492.76
VERINT	2.8%	14,249.8	99.61	239.5	457.53	456.07
LAITQM	2.8%	5287.6	99.97	247.7	453.23	453.10
LAITLI	2.5%	3902.2	99.89	247.7	409.04	408.87
LAITQMH	2.0%	5395.6	99.97	247.7	325.66	325.60
VDFMAIN	1.9%	7082.2	99.63	244.5	308.82	301.42

Flow Traces of cloudvar at T799 on 8 cpus

Pre-Optimisation cost 8.3%

Time in Seconds	Mflops	Cpu number
700	4335	0
702	4322	1
708	4282	2
706	4299	3
706	4299	4
705	4303	5
699	4341	6
699	4340	7

Flow Traces of cloudvar at T799 on 8 cpus

Pre-Optimisation cost 8.3%

Post-Optimisation cost 3.4 %

Time in Seconds	Mflops	Cpu number	Mflops	Time in Seconds
700	4335	0	1681	246
702	4322	1	1598	224
708	4282	2	1605	239
706	4299	3	1857	374
706	4299	4	1779	317
705	4303	5	1642	246
699	4341	6	1639	239
699	4340	7	1629	228

Flow Trace of cloudvar at T799 on 16 cpus

Cpu number	Mflops	Time in Seconds
0	1688	123
1	1651	125
2	1601	115
3	1578	111
4	1576	117
5	1609	123
6	1823	181
7	1863	196
8	1773	166
9	1740	155
10	1607	126
11	1625	124
12	1610	117
13	1663	123
14	1663	123
15	1588	105

max

min

Load-Imbalance Nearly 2:1

High percentage of peak speed
Is not always the most important
Attribute of Vector Architecture

Fastest Wall-clock time is the
only criteria that should be judged

NEC have a “TeraFlop Workbench “

The Challenge is to run Real Applications

At a sustained rate of One Teraflops

How well does IFS Run ?

HLRS

High Performance Computing Centre Stuttgart

Teraflop Workbench

NEC

Mike O'Neill

Stuttgart 17th October 2006

First set of runs on a large System October 13th 2006

ECMWF T799 L91 2 day Forecast

Nodes	Total Cpu s	Gflops per cpu	Total Gflops	WallClock Seconds	Node Speed-up
1	8	5.402	43.2	7,866	1.0
2	16	5.282	84.5	4,024	1.956

First set of runs on a large System October 13th 2006

ECMWF T799 L91 2 day Forecast

Nodes	Total Cpu s	Gflops per cpu	Total Gflops	WallClock Seconds	Node Speed-up
1	8	5.402	43.22	7,866	1.0
2	16	5.282	84.52	4,024	1.956
3	24	5.094	122.27	2,782	2.827
4	32	5.038	161.23	2,111	3.726
8	64	4.491	290.04	1,176	6.887
16	128	3.775	483.29	709.4	11.08

Problem, what is all this Imbalance?

Nodes	Total Cpu s	Gflops per cpu	Total Gflops	WallClock Seconds	Node Speed-up	Percentage Imbalance
1	8	5.402	43.22	7,866	1.0	14.68
2	16	5.282	84.52	4,024	1.956	17.1
3	24	5.094	122.27	2,782	2.827	21.1
4	32	5.038	161.23	2,111	3.726	23.3
8	64	4.491	290.04	1,176	6.887	33.2
16	128	3.775	483.29	709.4	11.08	48.5

So What Performance am I getting on the NEC SX-8 system Today ?

ECMWF T799 L91 10 day Forecast

# Nodes	Total Cpu s	Gflops per cpu	Total Gflops	WallClock Seconds	Node Speed-up
1	8	5.574	44.6	36,851	1.0
2	16	5.531	88.5	18,526	1.989
3	24	5.440	130.6	12,541	2.938
4	32	5.434	173.9	9,417	3.913
8	64	5,226	334.5	4,905	7.513
16	128	4.844	620.1	2,637	13.97
32	256	4,299	1,100,467	1,493	24.66

So What Performance am I getting on the NEC SX-8 system Today ?

ECMWF T799 L91 10 day Forecast

# Nodes	Total Cpu s	Gflops per cpu	Total Gflops	WallClock Seconds	Node Speed-up	Load Imbalance Percentage
1	8	5.574	44.6	36,851	1.0	6.9 %
2	16	5.531	88.5	18,526	1.989	7.1 %
3	24	5.440	130.6	12,541	2.938	10.8 %
4	32	5.434	173.9	9,415	3.913	12.5 %
8	64	5,226	334.5	4,905	7.513	14.6 %
16	128	4.844	620.1	2,634	13.97	19.3 %
32	256	4,299	1,100,467	1,486	24.66	21.5 %

ECMWF T799 L91 10 day Forecast

Comparison of NEC SX-8 Performance with Installed IBM Systems

System	Number Of Nodes	Number Of cpus	Wall-Clock Seconds	Sustained G flops
IBM p690+	32	1024	2534 (42)	644
IBM p690+	64	2048	1523 (25)	1,073
IBM p575+	48	768	2370 (39)	681
IBM p575+	96	1536	1341 (22)	~1,200
NEC SX-8	16	128	2637 (44)	620
NEC SX-8	32	256	1493 (25)	1,100

Raps8

Mike O'Neill

M & E O'Neill Consultants Ltd.

email m.eon@btopenworld.com

UK Mobile: +44 7725656575