HOW TO UTILIZE MULTI CORE CPUS
- Toward Sustained Petascale Computing -

Motoi Okuda[1]

[1] Technical Computing Solutions Unit
Fujitsu Limited
9-3, Nakase 1-chome, Mihamaku, Chiba City Chiba 261-8588, JAPAN
m.okuda@jp.fujitsu.com

The improvement of semiconductor technologies makes it possible to integrate several cores in one CPU chip. This type of CPU is called as multi core or many core CPU. This implementation can improve one CPU chip peak performance dramatically. However, it also brings up new problems, i.e. how to use multi/many core effectively and easily and how to balance core performance and memory bandwidth between core and memory?
Fujitsu has been developing new architecture called **Integrated Multi-core Parallel ArChiTecture** to respond these problems. In this presentation, I will explain the concept and the outline of Integrated Multi-core Parallel ArChiTecture and the performance of Fujitsu high-end technical computing server FX1 which implements Integrated Multi-core Parallel ArChiTecture. The outline of SPARC64™ VIIIfx, a Fujitsu's new high-end CPU for technical computing, and Fujitsu's future petascale computer which inherits Integrated Multi-core Parallel ArChiTecture will also be given in this presentation.

How to utilize multi core CPUs
- Toward Sustained Petascale Computing -

April 24th, 2009

Motoi Okuda

Fujitsu Limited

Agenda

- Outline of Fujitsu's HPC Solution Offerings
- High end Technical Computing Server FX1
- Fujitsu's Challenges for Petascale Computing
- Conclusion

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

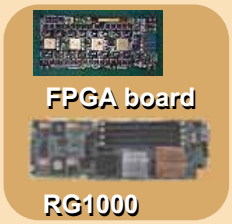
Fujitsu's Technical Computing Platform Solutions

Cluster Solutions

- Optimal price/performance for MPI-based applications
- Highly scalable
- InfiniBand interconnect

Solidware Solutions

- Ultra high performance for specific applications



High-end TC Solutions

- Scalability up to 100 TFlops class
- Remarkable real application performance
- High-end RISC CPU

Large-scale SMP System Solutions

- Up to 2TB memory space for TC applications
- High I/O bandwidth for I/O server
- High reliability based on mainframe technology
- High-end RISC CPU

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

Agenda

- Out line of Fujitsu's HPC Solution Offerings
- **High end Technical Computing Server FX1**
- Fujitsu's Challenges for Petascale Computing
- Conclusion

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

FX1 : New High-End TC Server - Outline -

- **Targeting highly efficient application performance**
  - High-performance CPU designed by Fujitsu
    - SPARC64™ VII : 4 cores by 65 nm technology
    - Performance : 40 GFlops (2.5 GHz)
  - New architecture for high-end TC server
    - **Integrated Multi-core Parallel ArChiTecture** by leading edge CPU and compiler technologies
    - Blade type node configuration for high memory bandwidth
  - High-speed intelligent interconnect
    - Combination of InfiniBand DDR interconnect and the highly-functional switch
    - Highly-functional switch realizes barrier synchronization and high-speed reduction between nodes by hardware
- **Petascale system inherits Integrated Multi-core Parallel ArChiTecture**
  - FX1 is a suitable platform to develop and evaluate Petascale applications

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

FX1 Specifications

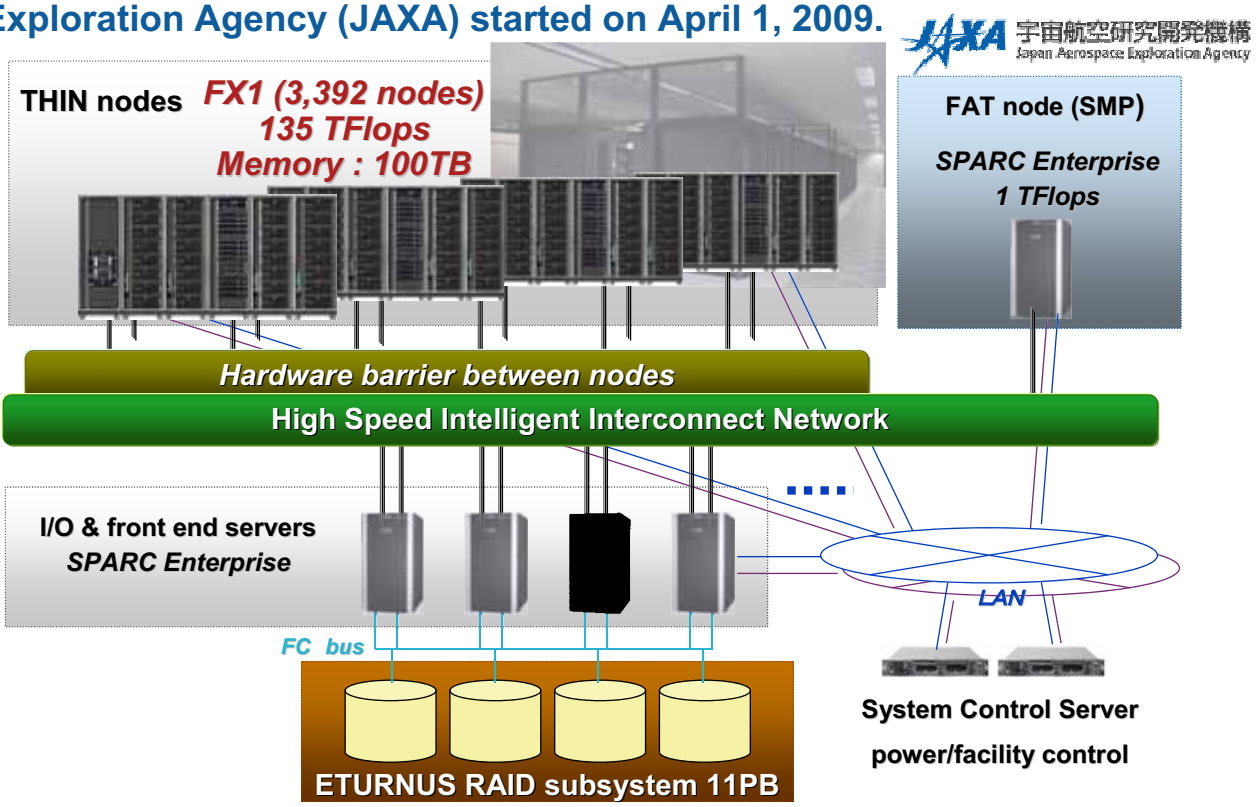
CPU	Processor	SPARC64™ VII @ 2.5 GHz
	Cache	L1: 64 KB instruction & 64 KB data / core L2: 6 MB/CPU, shared
	Cores	4
	Performance	40 GFlops
	Simultaneous multi-threading	2 threads/core
	Barrier synchronization	CPU-wide high-speed barrier mechanism between cores
Node	CPUs	1
	Memory capacity	Max 32 GB
	Memory error-checking	ECC, extended ECC
	Memory bandwidth	40 GB/s
	Interfaces	InfiniBand™ HCA (2 GBps) x 1; 1000baseT x 2
Inter-connect	Topology	Fat-tree
	Interface	InfiniBand™ DDR
	Additional functions	Intelligent SW with barrier synchronization and hardware assisted reduction capabilities

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

FX1 Launch Customer

- Operations of a new supercomputer system for the Japan Aerospace Exploration Agency (JAXA) started on April 1, 2009.



 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

FX1 LINPAC Benchmark Score on JAXA system

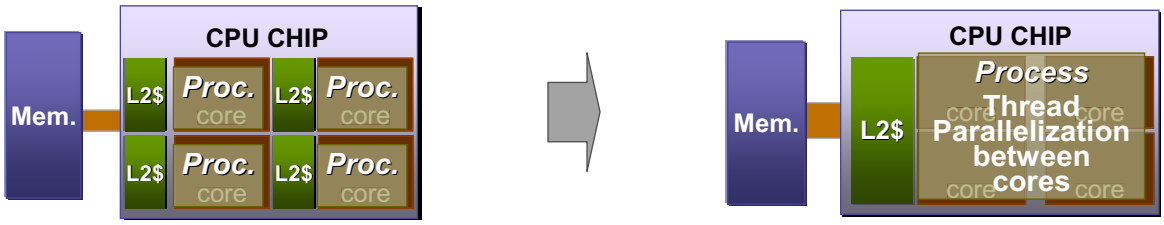
- FX1 LINPAC Benchmark on 130TFlops JAXA system (3,008 nodes = 3,008 CPUs = 12,032 cores)

	Results	Compared to November 2008 TOP500 list (latest)
Performance	110.6 TFlops	1st in Japan, 17th in world
Efficiency	91.19%	1st in world
Runtime	60 hours, 40 minutes	1st in world

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

Introduction

- **Concept**
  - Highly efficient thread level parallel processing technology for multi-core chip
  - Supports highly efficient hybrid parallel programming model (MPI + thread parallelization by OpenMP or automatic parallelization)

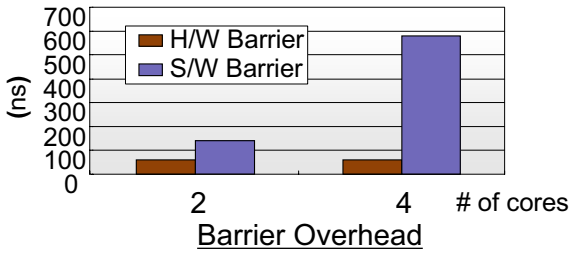


- **Advantage**
  - Handles the multi-core CPU as one equivalent faster CPU
    - Reduces number of MPI processes to $1/n_{core}$
      - ➔Increases parallel efficiency
      - ➔Reduce OS jitter effect
    - Reduces memory access and increase cache usage
- **Challenge**
  - How to decrease the thread level parallelization overhead?
  - How to decrease the cost for application implementation?

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

Key Technologies

- **CPU technologies**
  - Hardware barrier synchronization between cores
    - ➔Reduces overhead for parallel execution, 10 times faster than software emulation
    - ➔Start up time is comparable to that of the vector unit
    - ➔Barrier overhead remains constant regardless of number of cores

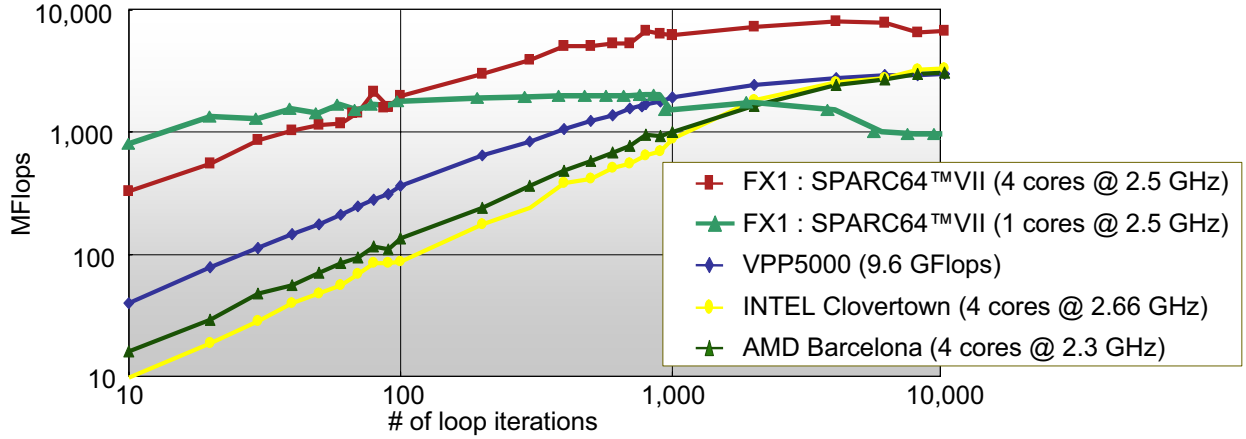


  - Shared L2 cache memory (6 MB)
    - ➔Reduces the number of cache to cache data transfers
    - ➔Efficient cache memory usage
- **Compiler technologies**
  - Highly efficient thread parallelization (automatic parallelization or OpenMP) by vectorization technology

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FX1 High Thread parallelization Performance

- **LINPACK performance on 1 CPU (4 cores, thread parallelization)**
  - 37.02 GFlops (91.82%)
- **Performance comparison of DAXPY (EuroBen Kernel 8) on 1 CPU**
  - 4core with Integrated Multi-core Parallel ArChiTecture shows better performance than
    - → 1core performance with small number of loop iterations
    - → Other X86 servers
    - → Vector server

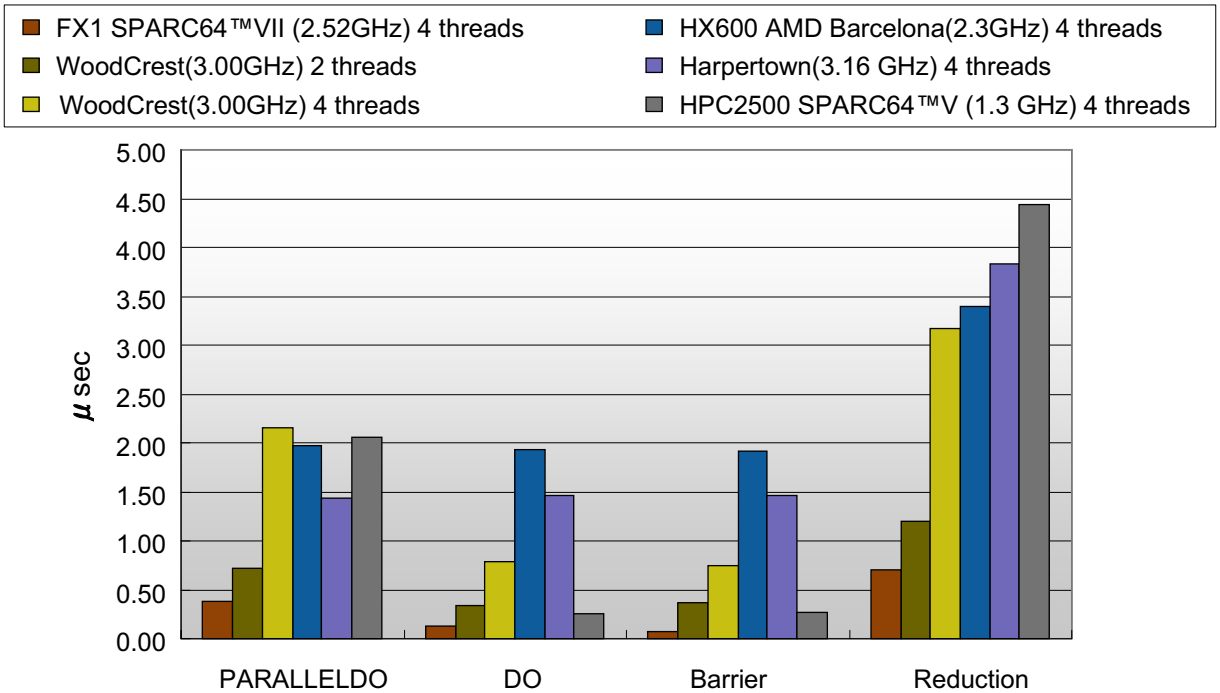


Performance comparison of DAXPY

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FX1 OpenMP Thread Parallelization Performance

- **Comparison of thread overhead on several OpenMP functions**

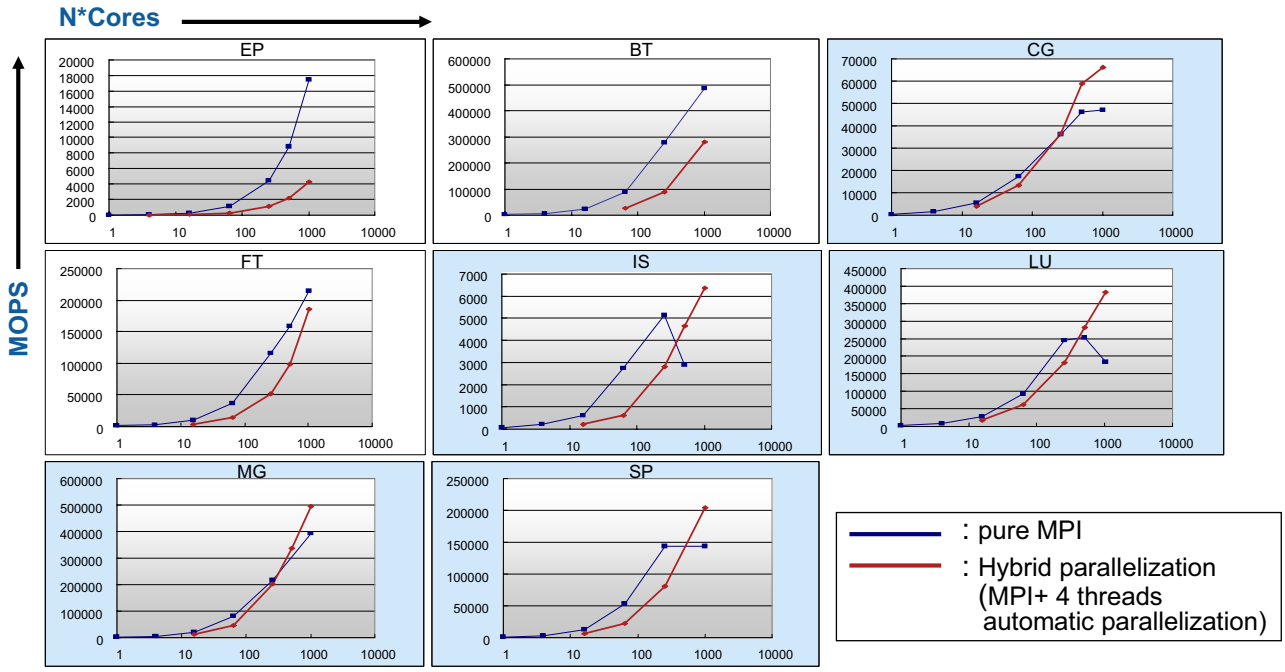


Overhead of OpenMP functions

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FX1 Hybrid Parallelization Performance

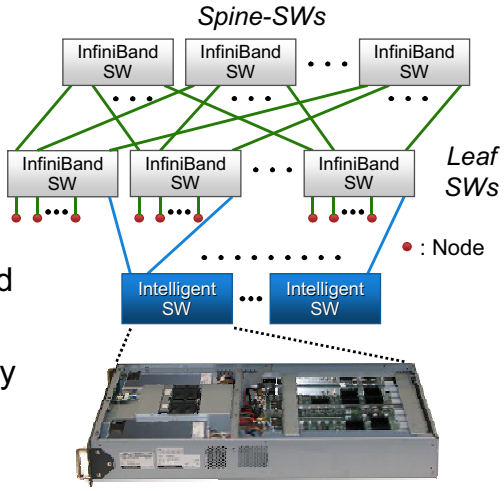
- **Performance comparison of NPB class C between pure MPI and Hybrid parallelization (automatic parallelization) on 256 CPUs (1,024 cores)**
  - Hybrid parallelization shows better performance than pure MPI with 5/8 programs



 All Rights Reserved, Copyright FUJITSU LIMITED 2009

Outline

- **Combination of fat tree topology InfiniBand DDR interconnect and the highly-functional switch (Intelligent switch)**
- **Intelligent switch (ISW)**
  - Result of the PSI (Petascale System Interconnect) national project
  - Functions
    - Hardware barrier function among nodes
    - Hardware assistance for MPI functions (synchronization and reduction)
    - Global ping for OS scheduling
  - Advantages
    - Faster HW barrier speeds up OpenMP and data parallel FORTRAN (XPF)
    - Fast collective operations accelerate highly parallel applications
    - Reduces OS jitter effect



Intelligent Switch & its connection

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FX1 Hybrid Parallelization Performance

```
C***********************************************************
      subroutine jacobi(nn,gosa)
C***********************************************************
      IMPLICIT REAL*4(a-h,o-z)
C
      include 'mpif.h'
      include 'param.h'
C
      DO loop=1,nn
         gosa=0.0
         wgosa=0.0
         DO K=2,kmax-1
            DO J=2,jmax-1
               DO I=2,imax-1
                  S0=a(I,J,K,1)*p(I+1,J,K)+a(I,J,K,2)*p(I,J+1,K)
     1                 +a(I,J,K,3)*p(I,J,K+1)
     2                 +b(I,J,K,1)*(p(I+1,J+1,K)-p(I+1,J-1,K)
     3                 -p(I-1,J+1,K)+p(I-1,J-1,K))
     4                 +b(I,J,K,2)*(p(I,J+1,K+1)-p(I,J-1,K+1)
     5                 -p(I,J+1,K-1)+p(I,J-1,K-1))
     6                 +b(I,J,K,3)*(p(I+1,J,K+1)-p(I-1,J,K+1)
     7                 -p(I+1,J,K-1)+p(I-1,J,K-1))
     8                 +c(I,J,K,1)*p(I-1,J,K)+c(I,J,K,2)*p(I,J-1,K)
     9                 +c(I,J,K,3)*p(I,J,K-1)+wrk1(I,J,K)
                  SS=(S0*a(I,J,K,4)-p(I,J,K))*bnd(I,J,K)
                  WGOSA=WGOSA+SS*SS
                  wrk2(I,J,K)=p(I,J,K)+OMEGA *SS
               enddo
            enddo
         enddo
C
         DO K=2,kmax-1
            DO J=2,jmax-1
               DO I=2,imax-1
                  p(I,J,K)=wrk2(I,J,K)
               enddo
            enddo
         enddo
C
         call sendp(ndx,ndy,ndz)
C
         call mpi_allreduce(wgosa, gosa, 1, mpi_real4, mpi_sum, mpi_comm_world,
     >                      ierr)
C
      enddo
CC End of iteration
      return
      end
```

Loop body of the HIMENO BMT

- **Performance measurement of HIMENO-BMT***
- **How to extract 4 cores performance on HIMENO-BMT**
  - Loop body is automatically parallelized
  - User only specifies the number of processes and its node assignment

Automatically parallelized

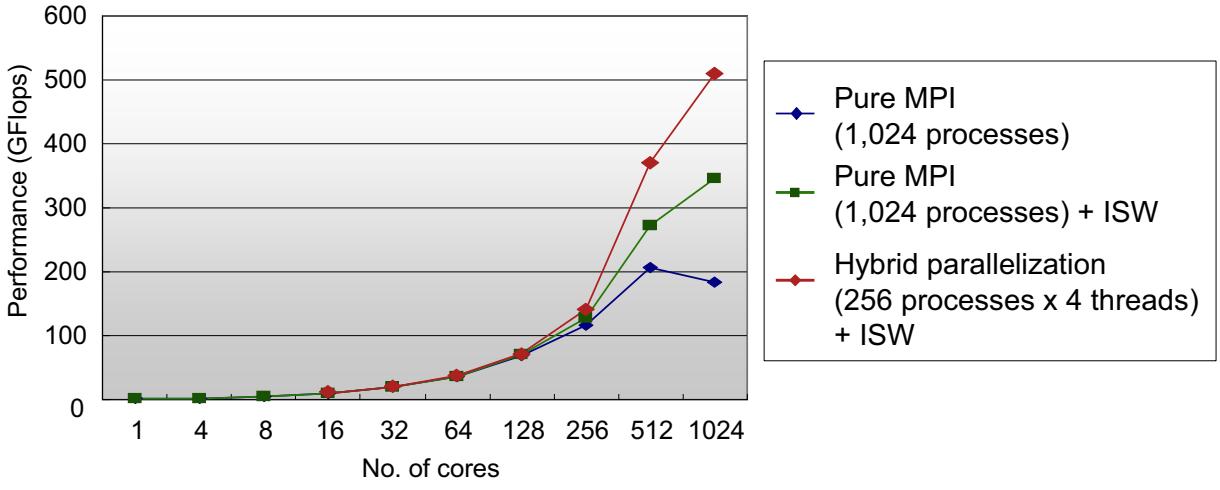
Uses ISW

* : Benchmark program which measures the speed of major loops to solve Poisson's equation solution using Jacobi iteration method.

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FX1 Hybrid Parallelization Performance

- **Performance comparison of HIMENO-BMT grid-M* between pure MPI, pure MPI + ISW and hybrid parallelization + ISW**
  - Hybrid parallelization (MPI + Automatic parallelization between four cores) assisted by Integrated Multi-core Parallel ArChiTecture and ISW achieves high parallel efficiency on FX1



Performance comparison by HIMENO BMT grid-M

* : Size M means that mesh size is 256 X 128 X 128.

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

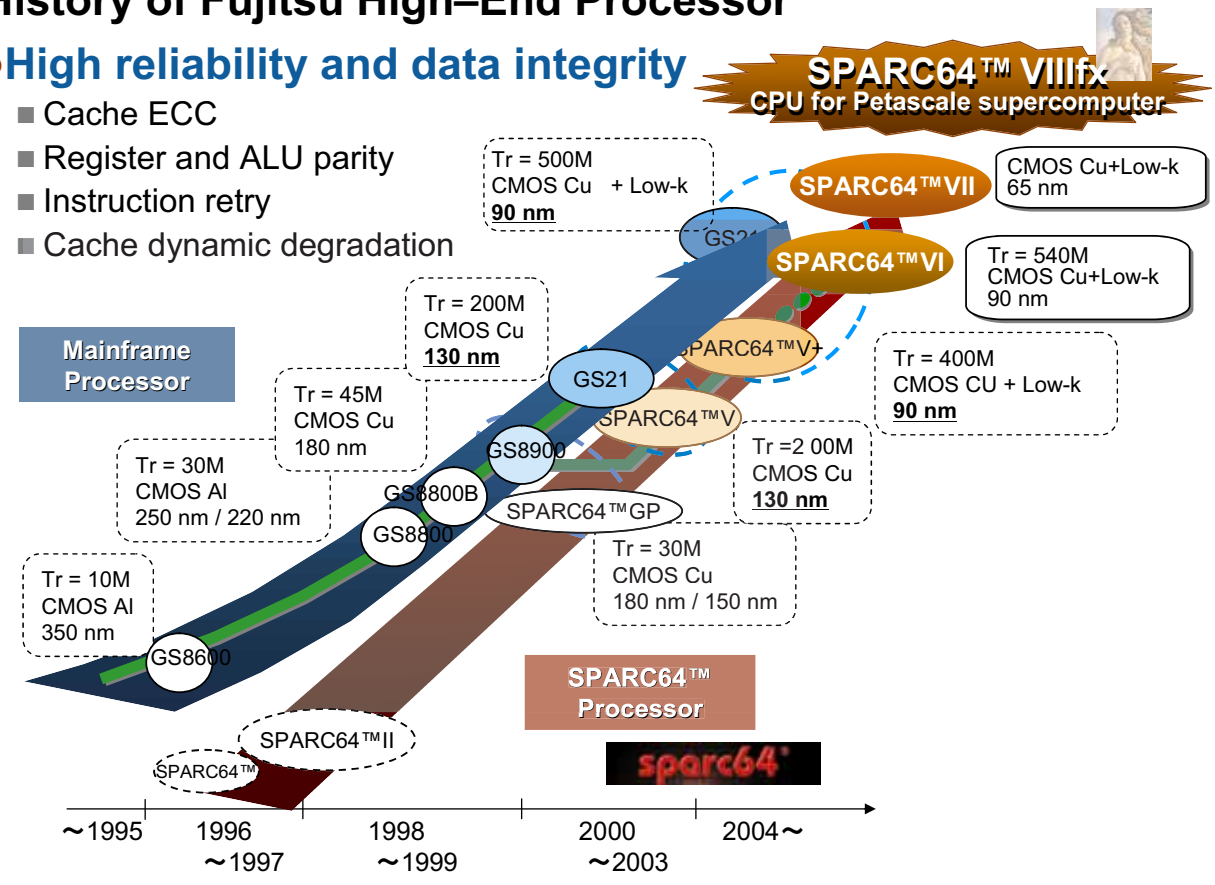
Agenda

- Out line of Fujitsu's HPC Solution Offerings
- High end Technical Computing Server FX1
- **Fujitsu's Challenges for Petascale Computing**
- Conclusion

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

History of Fujitsu High–End Processor

- **High reliability and data integrity**
  - Cache ECC
  - Register and ALU parity
  - Instruction retry
  - Cache dynamic degradation

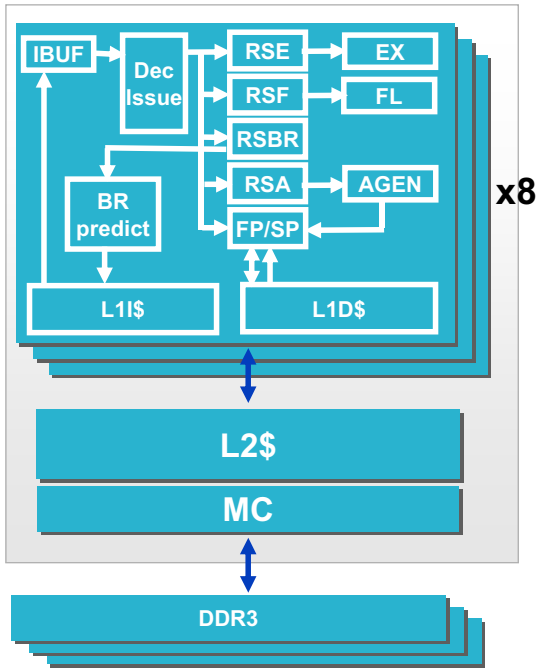


 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

SPARC64™ VIIIfx Overview

- **For Petascale computing**
  - 8 cores
  - Embedded memory controller
- **Architecture**
  - SPARC-V9 + extension (HPC-ACE)
    - SIMD
    - Hardware barrier
    - :
- **Semiconductor technologies**
  - Fujitsu 45 nm CMOS
- **Performance**
  - 128 GFlops@socket



Outline design

RSE : Reservation station for integer operation
RSF : Reservation station for floating operation
RSBR : Reservation station for branch operation
RSA : Reservation station for load /store
FP/SP : load/store queue
FL : Floating point pipeline
EX : Integer pipeline
IBUF : Instruction buffer
Dec Issue : decode & Issue
AGEN : Address generation

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

Agenda

- Out line of Fujitsu's HPC Solution Offerings
- High end Technical Computing Server FX1
- Fujitsu's Challenges for Petascale Computing
- **Conclusion**

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

Conclusion

- **Key Issues for sustained Petascale computing**
  - How to utilize multi-core CPU ?
  - How to handle a hundred thousand processes ?

- **Fujitsu's technical challenge**
  - New Integrated Multi-core Parallel ArChiTecture and innovative interconnect which provide a highly efficient hybrid parallel programming environment
- **Fujitsu's stepwise approach to product release ensures users to be ready for Petascale computing**

  - ***Step 1 :***
    - The new high end technical computing server **FX1** provides the environment for applications migration for Petascale system.
    - Design of Petascale system which inherits FX1 architecture
  - ***Step 2 :***
    - Petascale system with new high performance, highly reliable and low power consumption CPU and innovative interconnect

 All Rights Reserved, Copyright FUJITSU LIMITED 2009

FUJITSU

THE POSSIBILITIES ARE INFINITE