Dolphin software

SCI Software

Hugo Kohmann
hugo@dolphinics.no
Dolphin Interconnect Solutions
SISCI software Developers Kit

- Interconnect Resource Manager (IRM) Driver
- SISCI Driver and API Library
- SISCI Demo and Benchmark Programs
- Cluster Toolbox
Software Architecture

- **User Space**
  - SISCI API
  - Socket API
- **Kernel Space**
  - SISCI
  - TCP/UDP
  - IP
  - DLPI/NDIS
- **GENIF**
  - IRM
IRM Driver

- Managing Hardware Cluster Resources
- Adapter Card Initialization
- SCI Address Translation Table Manager
- Adapter DMA Engine Manager
- In-band Switch Configuration and Management
- Cluster Heartbeat Alive Checks
- Endpoint to Endpoint Session Control
- Error Recovery and Notification
- Kernel Programming Interface (GENIF)
- Highly portable
IRM Supported Hardware

- **Adapter Cards**
  - D307 - SBus
  - D310 - PCI32
  - D314 - PMC32
  - D320 - PCI64
  - D323 - PMC64
  - D330 - PCI 66

- **Switches**
  - D505 - 4 way (SBus)
  - D512 - 4 way (PCI)
  - D515 - 4 - 16 way (PCI)
  - D525 - 8 way switch
IRM Driver Architecture

- GENIF
  - OSIF Library
    - NT
    - Solaris
    - Linux
- IRM
- PAL Library PCI/SBus
- PCI or SBus Adapter
GENIF

- Kernel Level Programming Interface
- Memory Segment Allocation, Connection and Mapping
- DMA Engine queues and completion control
- Local Interrupt Routine Registration
- Remote Interrupt Triggering
- Error Detection Mechanisms
- Cluster and Connection Callbacks
- Hartbeat and Remote Node Alive Check
Main line Topologies

- SCI Ring Configurations 2+ Nodes

- Switched Configurations 4 - 16 nodes
Topologies

• 2D Topology
Cluster Toolbox

- Sciconfig - Cluster configuration
- Scidiag - Hardware diagnostic
- Scibench - Performance benchmark
- Isptools - Field firmware upgrade tools
• SCI Adapter configuration utility
  – Adapter NodeId
  – Adapter Number

• GUI
  – Windows NT

• Console Application
  – All system
  – Interactive
  – Command line option
• SCI Diagnostic utility

• Local Adapter diagnostic
  – Transaction Controller
  – On Board RAM/ROM
  – LinkController
  – Interrupt statistics

• SCI Cluster Diagnostic
  – Local Link Status
  – Switch Port Status
  – Remote Adapter Access
• **In System Programming**

• **Adapter Firmware Update**

• **Switch Firmware Update (D525)**
• Esprit Project 73124
  – Software Infrastructure for SCI

• Application Programming Interface (API)
  – Low Level SCI Programming

• User space Access to basic SCI and Adapter properties
  – High Bandwidth
  – Low Latency
  – Memory Mapped Remote Access
  – DMA Engine
  – Interrupts
  – Callbacks
SISCI Features

- Access to High Performance HW
- Highly Portable
- Cross Platform / Cross Operating system interoperable
- Simplified SCI Programming
- Flexible
- Reliable Data transfers
- Hostbridge / Adapter Optimization in libraries
SISCI Performance

- Highly dependent of the PC Chipsets
- Requires special attention
  - Write Combining
  - Write Posting
  - Burst Assembly
  - Cache line manipulation / flush

- Taken care of by
  - IRM
  - SCIMemCopy()
Bandwidth test results
PIO - D330

SUSTAINED ONE WAY DATARATE, D330
PIO on ServerWorks HE-III

Segment size

MBytes/s

With WC
Without WC

4 8 16 32 64 128 256 512 1024 2048 4096 8192 16384 32768 65536

0 20 40 60 80 100 120 140 160 180
## Bandwidth test results
### PIO - D330

<table>
<thead>
<tr>
<th>Size (byte)</th>
<th>One way (MB/s)</th>
<th>Bi directional (MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2.62</td>
<td>2.62</td>
</tr>
<tr>
<td>8</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>16</td>
<td>8.12</td>
<td>9.41</td>
</tr>
<tr>
<td>32</td>
<td>10.92</td>
<td>10.94</td>
</tr>
<tr>
<td>64</td>
<td>30.38</td>
<td>28.98</td>
</tr>
<tr>
<td>128</td>
<td>45.85</td>
<td>40.96</td>
</tr>
<tr>
<td>256</td>
<td>91.49</td>
<td>82.68</td>
</tr>
<tr>
<td>512</td>
<td>150.38</td>
<td>124.25</td>
</tr>
<tr>
<td>1024</td>
<td>154.96</td>
<td>129.44</td>
</tr>
<tr>
<td>2048</td>
<td>158.91</td>
<td>131.44</td>
</tr>
<tr>
<td>4096</td>
<td>160.35</td>
<td>133.35</td>
</tr>
<tr>
<td>8192</td>
<td>160.96</td>
<td>133.48</td>
</tr>
<tr>
<td>16384</td>
<td>161.38</td>
<td>133.52</td>
</tr>
<tr>
<td>32768</td>
<td>161.57</td>
<td>133.46</td>
</tr>
<tr>
<td>65536</td>
<td>161.66</td>
<td>133.51</td>
</tr>
</tbody>
</table>
Bandwidth test results
DMA - D330

SUSTAINED ONE WAY DATARATE, D330
DMA on ServerWorks HE-III

Segment size

MB/Bytes

0
50
100
150
200
250
300

64 128 256 512 1024 2048 4096 8192 16384 32768 65536 131072 262144 524288 1048576 2097152 4194304 8388608
<table>
<thead>
<tr>
<th>Segment (byte)</th>
<th>Latency (us)</th>
<th>Throughput (MB/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>71</td>
<td>0.9</td>
</tr>
<tr>
<td>128</td>
<td>78</td>
<td>1.64</td>
</tr>
<tr>
<td>256</td>
<td>78</td>
<td>3.28</td>
</tr>
<tr>
<td>512</td>
<td>78</td>
<td>6.56</td>
</tr>
<tr>
<td>1024</td>
<td>78</td>
<td>13.13</td>
</tr>
<tr>
<td>2048</td>
<td>81</td>
<td>25.28</td>
</tr>
<tr>
<td>4096</td>
<td>93</td>
<td>44.04</td>
</tr>
<tr>
<td>8192</td>
<td>110</td>
<td>74.47</td>
</tr>
<tr>
<td>16384</td>
<td>140</td>
<td>117.03</td>
</tr>
<tr>
<td>32768</td>
<td>207</td>
<td>158.3</td>
</tr>
<tr>
<td>65536</td>
<td>360</td>
<td>182.04</td>
</tr>
<tr>
<td>131072</td>
<td>620</td>
<td>211.41</td>
</tr>
<tr>
<td>262144</td>
<td>1160</td>
<td>225.99</td>
</tr>
<tr>
<td>524288</td>
<td>2250</td>
<td>233.02</td>
</tr>
<tr>
<td>1048576</td>
<td>4423</td>
<td>237.07</td>
</tr>
<tr>
<td>2097152</td>
<td>8781</td>
<td>238.83</td>
</tr>
<tr>
<td>4194304</td>
<td>17484</td>
<td>239.89</td>
</tr>
<tr>
<td>8388608</td>
<td>34906</td>
<td>240.32</td>
</tr>
</tbody>
</table>
SISCI software layers

Node A

Application (Performance tool)
SISCI library
SISCI Driver
IRM Driver
Hardware abstraction layer (PAL)
PCI-SCI adapter card

Node B

Application (Performance tool)
SISCI library
SISCI Driver
IRM Driver
Hardware abstraction layer (PAL)
PCI-SCI adapter card
*remotePtr = value;
*remotePtr = value;
The programs are made simple to demonstrate the use of the SISCI functions.

Example programs available:

- Shared memory
- DMA
- Interrupts
- Probe
- Connection to SCI Space
- Block transfers
- CSR
- Query
- Scibench2
Operating Systems supported

- Linux 2.0 / 2.2 x86
  - SISCI
  - Available since December 1998
  - TCP/IP Prototype available
- Linux 2.2 Alpha
  - SISCI
  - Running Prototype
- Linux 2.4
  - Port in progress
• Solaris 2.5.1 / 7 / 8 SPARC
  – SISCI and DLPI
  – Available since 1994

• Solaris 2.6 / 7 x86
  – SISCI
  – Available since 1998
Operating Systems supported

- **Windows NT 4.0 x86**
  - *SISCI and NDIS*
  - *Available since August 1996*

- **Windows 2000 x86**
  - *SISCI and NDIS*
  - *Available since March 2000*
• Dolphin is a Lynx Synergy partner
• Lynx 3.01 x86
  – SISCI
  – Available since December 1999
• Lynx 3.0.1 PowerPC
  – SISCI
  – Beta testing in progress.
• Lynx HA Package support is being added
• Lynx Messenger is being reviewed
• Lynx 3.1 Prototype available
• Dolphin is a WindLink Partner

• VxWorks PowerPC
  – SISCI
  – Beta available since September 2000

• VxWorks x86
  – SISCI
  – Beta planned mid October 2000

• cPCI backplane messaging API ported to SCI
Operating Systems supported

- Compaq Tru64 Unix Alpha
  - SISCI
  - Running Prototype
Software Distribution

- Distributed using web technology
  - www.dolphinics.com

- Free of Charge
  - Operating system choice
  - Upgrades

- On Line Web Documentation
  - Software installation Guides
  - Users Guides
  - Configuration Guides

- CD ROM Available on request