The Merrimac streaming supercomputer project aims to develop a scientific computer that offers an order of magnitude or more improvement in performance per unit cost compared to cluster-based scientific computers built from the same underlying semiconductor and packaging technology. We expect this efficiency to arise from two innovations: stream architecture and advanced interconnection networks. Organizing the computation into streams and exploiting the resulting locality using a register hierarchy enables a stream architecture to reduce the memory bandwidth required by representative computations by an order of magnitude or more.

### Merrimac Node

- 16 data-parallel compute clusters
- Integrated scalar processor
- Indexable SRF (poster)
- Capable memory system
  - 16 GBytes/s (random access)
  - Stream cache
  - ScatterOp (poster)

### Merrimac System

- Flattened memory bandwidth within a 16-node board
- 4:1 Concentration within a 32-node backplane, 8:1 across a 32 backplane system
- Routers with bandwidth 8+460Gb/s route messages with length L=128b
- Requires high radix to exploit

### Reliability, Availability, and Serviceability

- Detect data errors using parity
  - Parity protect all arrays
    - DRAM, SRF, Cache, on-board store, Reorder buffers, Reorder buffers

- Checkpoint - rollback
- MTBF of 166 hours per board
- Replace 1 board per month
- MTBF of 16 hours per component
- Checkpoint duration ~5 minutes
- Recovery time ~10 minutes
- Checkpoint every 5 hours
- Slowdown of less than 3%

- Redundant power and cooling
  - Diode voting on boards
- Hot extra board per cabinet
- Network gracefully degrades

### Stream Applications

- Exploits locality
  - producer-consumer within kernel in the LRF
  - producer-consumer across kernels in the SRF
- Reduces the distance data travels
- Supports large number of ALUs

### Future Research

- Interface between stream and scalar processor
  - Short-stream effects
    - Amount of software control
- Mechanisms for variable-rate streams
- Stream-cache studies
- Multi-node execution and simulation
- Work partitioning between static and run-time
  - Stream scheduling
  - Non-contiguous SRF

### Bandwidth/Register Hierarchy

- Mechanisms for variable-rate streams
- High-performance mode
- Redundant power and cooling
- Diode voting on boards
- Hot extra board per cabinet
- Network gracefully degrades

### Merrimac Floorplan

- Inter-cluster switch between clusters
- 127.5 mm² (small ~2x10)
  - Stanford Imagine is 16mm x 16mm
  - MIT Raw is 18mm x 18mm
- 32 Watts (P4 = 75 W)

### Input/Output

- Interfaces to industry standard I/O devices
  - 10 Gbit Ethernet (XAU) interfaces/router
  - 2.5Gb/s I/O bandwidth per node

### Board and Network

- Flat memory bandwidth within a 16-node board
- 4:1 Concentration within a 32-node backplane, 8:1 across a 32 backplane system
- Routers with bandwidth 8+460Gb/s route messages with length L=128b
- Requires high radix to exploit

### Table

<table>
<thead>
<tr>
<th>Application</th>
<th>Sustained GFLOPS</th>
<th>FP Ops / Mem Ref</th>
<th>LRF Ref.</th>
<th>SRF Ref.</th>
<th>Mem Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>StreamFEM 2D (Euler-quotientic)</td>
<td>32.2</td>
<td>25.5</td>
<td>105.50, 56, 486 (93.6%)</td>
<td>1,100, 777 (5.7%)</td>
<td>1,250, 486 (9.7%)</td>
</tr>
<tr>
<td>StreamFEM 2D (Navier-Cubic)</td>
<td>33.5</td>
<td>50.6</td>
<td>733, 294, 6, 0 (94.0%)</td>
<td>8,512, 751 (5.6%)</td>
<td>9,150, 108 (0.4%)</td>
</tr>
<tr>
<td>StreamMD 3D</td>
<td>21.3</td>
<td>14.3</td>
<td>427, 742, 24, 0 (98.3%)</td>
<td>9,057, 009 (2.3%)</td>
<td>5,978, 846 (1.4%)</td>
</tr>
<tr>
<td>StreamFLO (streamed-estimation)</td>
<td>50</td>
<td>98%</td>
<td>12%</td>
<td>12%</td>
<td></td>
</tr>
</tbody>
</table>