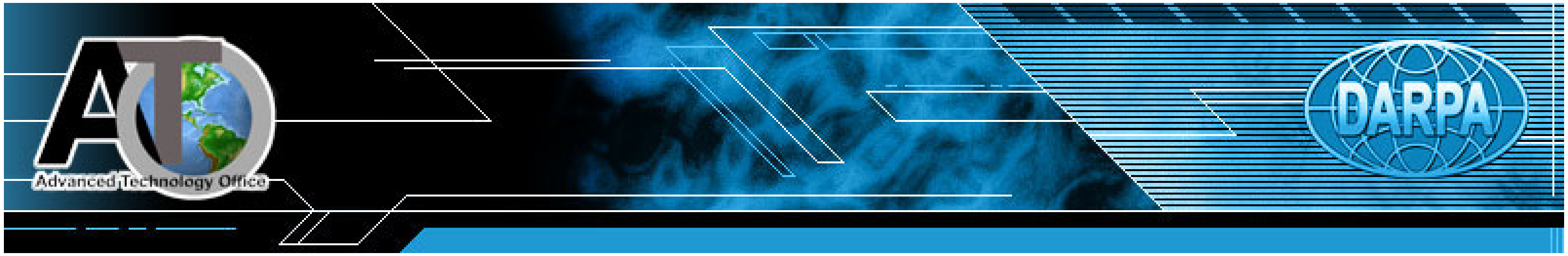
USING FIELD PROGRAMMABLE GATE ARRAYS IN A BEOWULF CLUSTER

Matthew J. Krzych

Naval Undersea Warfare Center

Approved for Public Release, Distribution Unlimited.

Sponsor



- DARPA - Advanced Technology Office
  - Robust Passive Sonar Program
  - Program Manager – Ms. Khine Latt

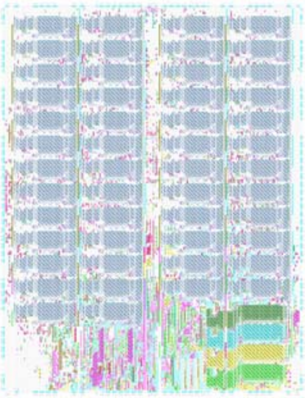
Problem Description

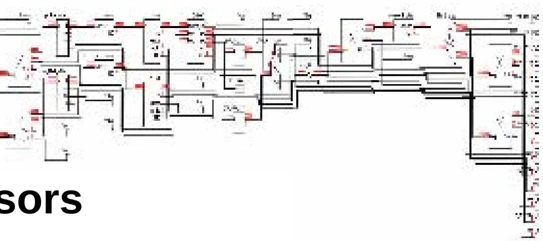
- Building an embedded tera-flop machine
  - Low Cost
  - Small footprint
  - Low power
  - High performance
- Utilize commercially available hardware & software
- Application:
  Beamform a volume of the ocean
  - Increase the number of beams from 100 to 10,000,000

On February 9, 2000 IBM formally dedicated Blue Horizon, the teraflops computer. Blue Horizon has 42 towers holding 1,152 compute processors, and occupying about 1,500 square feet. Blue Horizon entered full production on April 1, 2000.

Approach

- Compile matched field “beamformer” onto a chip





  - Specialized circuitry
    - 10x over Digital Signal Processors
    - 100x over General Purpose Processors

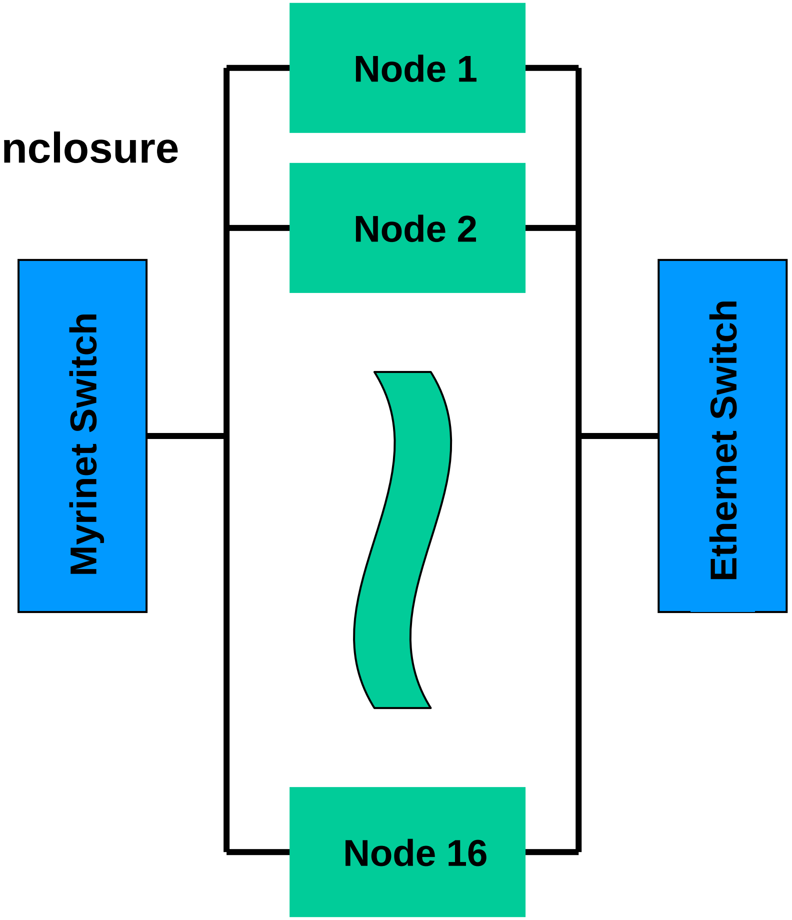
- DARPA Embedded High Performance Computing Technology
  - » Adaptive Computing FPGAs
  - » Message Passing Interface (MPI)
  - » Myrinet – High Speed Interconnect

Beowulf Cluster

Sustained 65 Gflops with FPGA’s

System Hardware

- 16 Node Cluster
  - AMD 1.6 GHz and Intel Pentium 2.2 GHz
  - 1 to 4 GBytes memory per node
  - 2U & 4U Enclosures w/ 1 processor per enclosure
  - $2,500 per enclosure [1]
- 8 Embedded Osiris FPGA Boards
  - Xilinx XC2V6000
  - $15,000 per board [1]
- Myrinet High Speed Interconnect
  - Data transfer: ~250 MBytes/sec
  - Supports MPI
  - $1,200 per node [1]
  - $10,500 per switch [1]
- 100 BASE-T Ethernet
  - System control
  - File sharing



Total Hardware Cost[1]: $190K

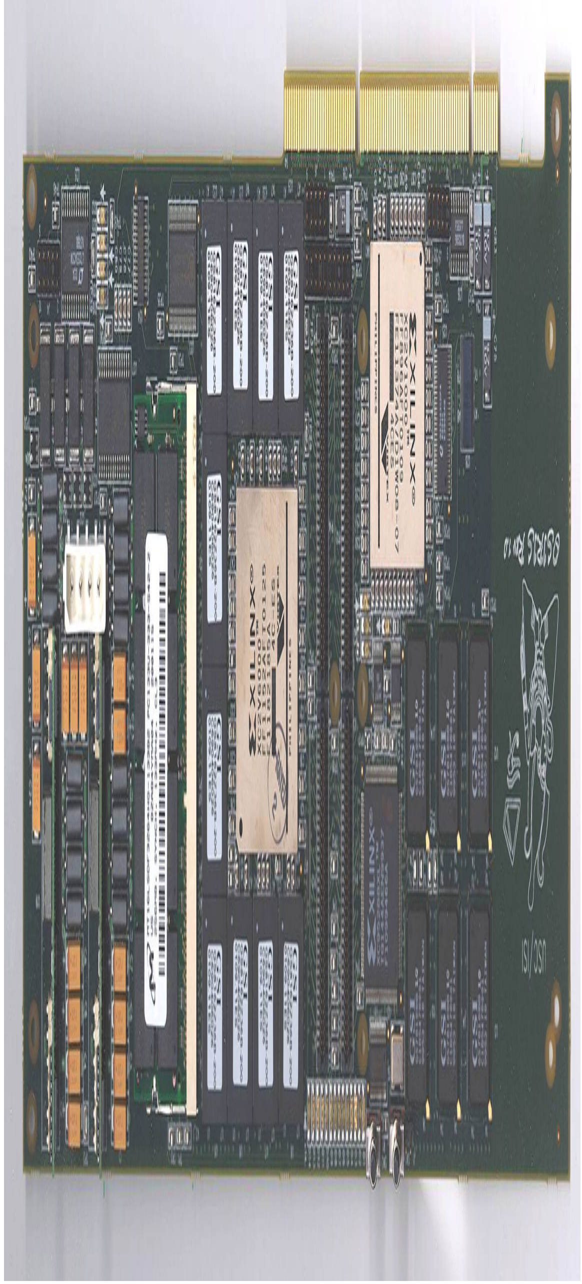
[1] Cost based on 2001 dollars. Moore's Law asserts processor speed doubles every 18 months. 2004 dollars will provide more computation or equivalent computation for fewer dollars.

Hardware Accelerator

- Osiris FPGA board
  - Developed by ISI / USC
  - Sponsored by DARPA ITO
    Adaptive Computing Systems Program
  - 256 Mbyte SDRAM

- Xilinix XC2V6000 chip
  - ~ 6,000,000 gates
  - 2.6 Mbits on chip memory
  - 144 18 by 18 bit multipliers

- PCI bus 64 bit / 66MHz Interface
- Sustained 65 Gflops
- Numerous commercial vendors



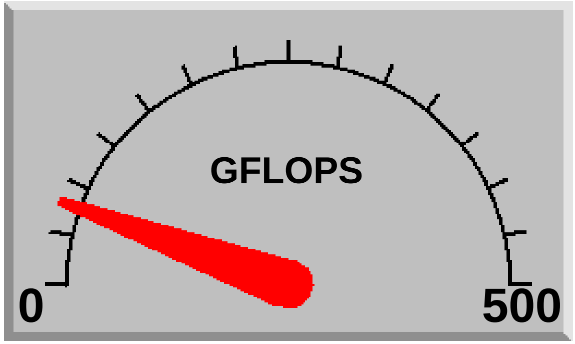
System Software

- Multiple programming languages used:
  - C, C++, Fortran77, Fortran90, Matlab MEX, VHDL
- Message Passing Interface (MPI)
- Red Hat Linux v7.3
- Matlab
  - System displays
    - Interface to MPI via shared memory
  - Post processing analysis
- Run-time cluster configuration
  - Supports run-time cluster configuration (hardware & software)

Computational Performance

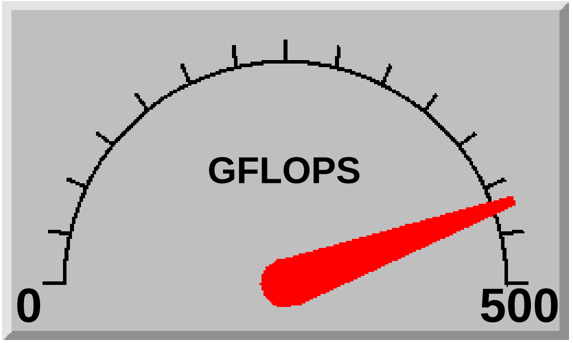
- WITHOUT hardware accelerator
  - 16 nodes (2.2 GHz)
  - 5 GFLOPS sustained
    - Single precision

- WITH hardware accelerator
  - 8 FPGA boards
  - 500 GFLOPS
    - Fixed point
    - Pipelining
    - Parallelism



Hardware

Accelerator



Run-time Cluster Configuration

- Developed in-house
  - Exploits MPI communication constructs
  - Uses Linux shell scripts & remote shell command ‘rsh’

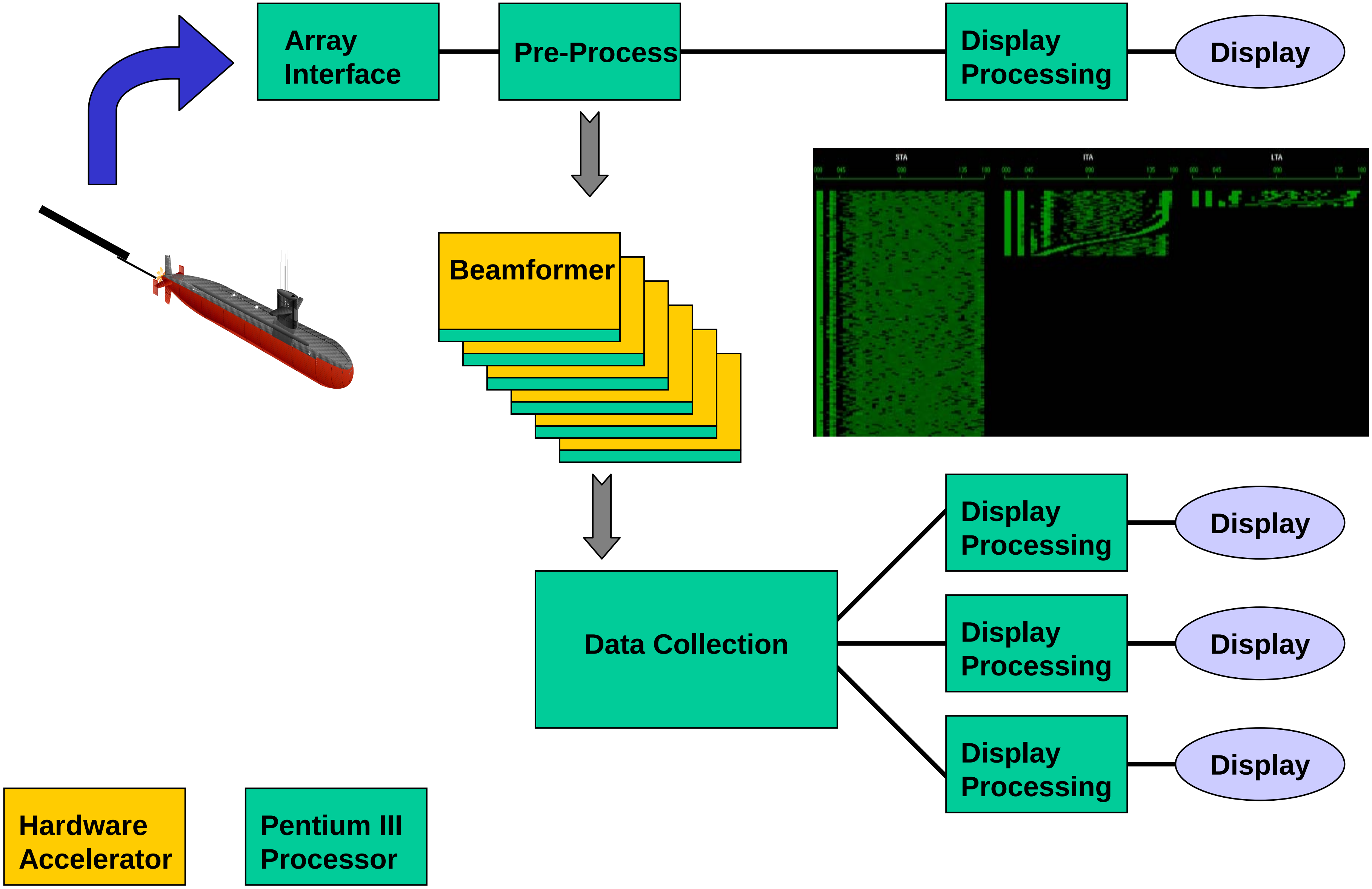
- Based on user specified configuration
  - Configuration defined in text file

- Allocates system resources at start-up
  - Identify hardware availability
  - Identify which functionality to execute

- Map functionality to specific nodes at run-time

```
Functional Description File
==========================================
FUNCTION      NUMBER     VALID HOSTS
***
array_if23      1            x0
frontend        1            x0
disp_conv       0            xb
mfp             3            x3, x1, x2, xa
collector       1            xa
disp_mbtr       1            xc, xb
disp_mrtr       1            xb, xc
```

Sonar Application



Array Interface

Pre-Process

Display Processing

Display

Beamformer

Data Collection

Display Processing

Display

Display Processing

Display

Display Processing

Display

Hardware Accelerator

Pentium III Processor

Benefits

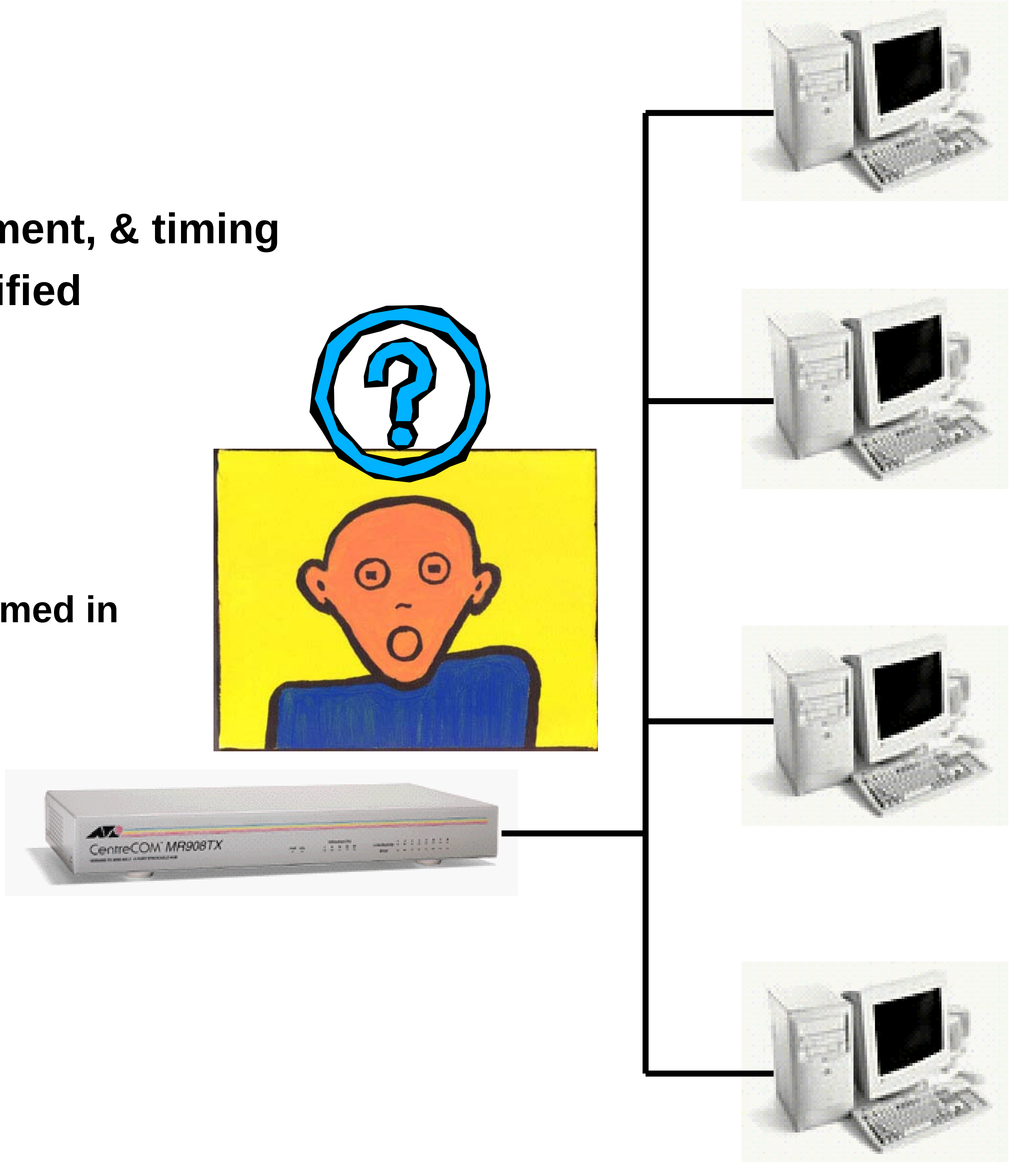
- High performance (500 GFLOPS), low cost solution (<200K)
- FPGAs
  - Performance (100x increase)
  - Small footprint (PCI board)
  - Power
- Beowulf Cluster
  - Flexibility /robustness
    - Supports heterogeneous hardware
    - Run-time selection of processors
    - Run-time selection of functions to instantiate
    - Run-time selection of system parameters
  - Scalability
    - Add / remove hardware assets
    - Add / remove functionality
- MPI
  - Facilitates flexibility & scalability
  - Runs on multiple hardware platforms & operating systems
  - Supports multiple communication schemes

    (point-to-point, broadcast, etc.)

Issues

- ❑ FPGAs
  - ❑ Lengthy development time
  - ❑ Difficult to debug
  - ❑ Bit file tuning: sizing, placement, & timing
  - ❑ Bit files are NOT easily modified
  - ❑ Bit files are NOT portable

- ❑ Beowulf Cluster
  - ❑ Functional mapping
    - Flexibility must be programmed in
  - ❑ Performance optimization
    - Identifying bottlenecks
    - Load balancing
  - ❑ Configuration Control
    - System maintenance
    - Keeping track of assets
    - Asset compatibility
  - ❑ Tool availability



Summary

- Computationally demanding sonar application successfully implemented
  - Could NOT have been implemented using traditional methods
- 16 node Beowulf cluster developed using 8 embedded FPGAs
  - Fits in 1 ½ standard 19” racks
  - Hardware costs < $200k
  - FPGA software tools < $40k
- 500 GFLOPS sustained processing achieved