

# Performance Evaluation of Biological Applications that use FPGAs

Text

**Dr. Olaf O. Storaasli**  
**Future Technologies Group**  
**Computer Science & Mathematics Division**  
**Oak Ridge National Laboratory**  
**CUG 2007 Seattle**

# Research Team

**Olaf Storaasli**

**Weikuan Yu**



**Dave Strenski**



**Jim Maltby**



**mitrionics™**



# Acknowledgment

*This research was sponsored by the Laboratory Directed Research & Development Program of ORNL managed by UT-Battelle for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725. The U.S. Government retains a non-exclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes.*





# Contents

**Background:** FPGAs, Genome Sequencing

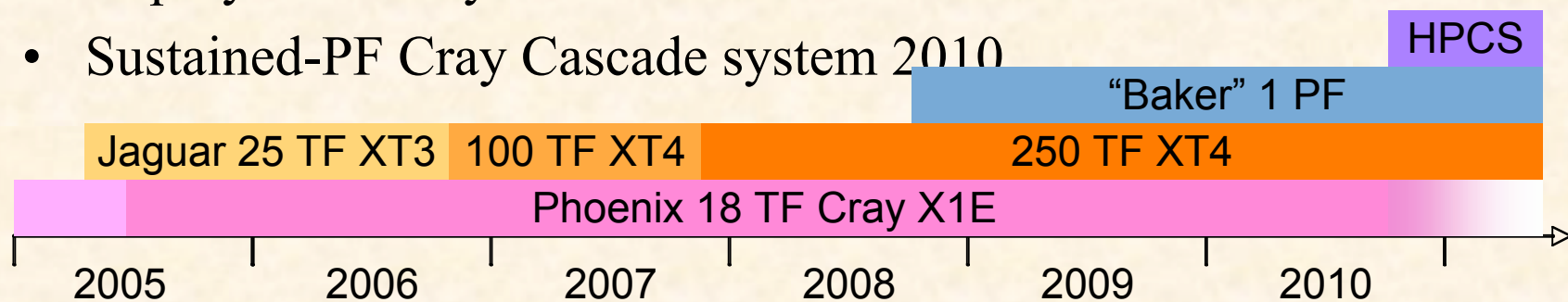
**Results:** FASTA for 3 [openfpga.org](http://openfpga.org) Cases

**Goal:** Speed Supercomputers with FPGAs

# ORNL Milestones: Deliver 1 PF system in 2008 Deliver 250 TF by 2007

## Roadmap

- Upgrade existing 25 TF XT3 to dual-core 100 TF system in 2006
- Upgrade 100 TF to 250 TF in late-2007
- Deploy 1 PF Cray “Baker” late 2008
- Sustained-PF Cray Cascade system 2010



18 TF Cray Phoenix and 25 TF Cray Jaguar currently in production

Storaasli - EUG07



# Future Supercomputer Technologies

**Commodity:** 2<sup>n</sup> core 2 GHz chips

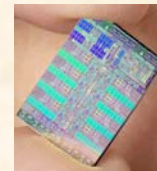
**Special:** El Dorado, Cyclops, PiM

## Accelerators to watch

- **FPGA:** DSP => HPEC => HPC



- **Cell:** IBM, Sony, Toshiba

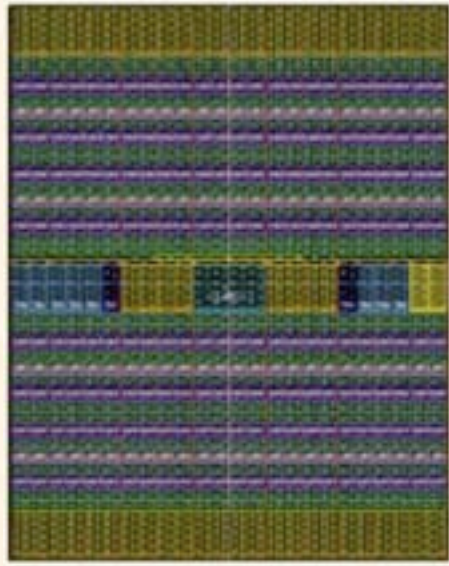


- **GPUs:** onboard  $\mu$ P?

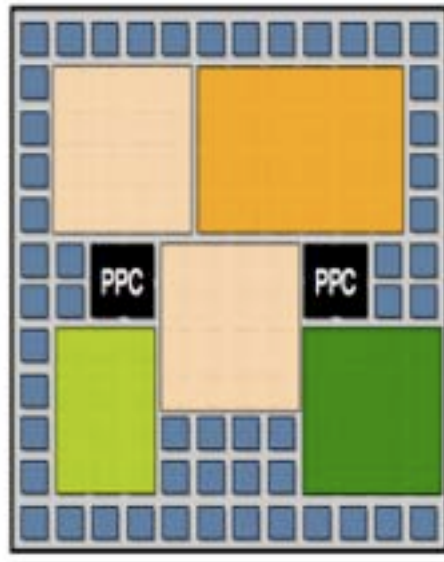


- **Array:** 

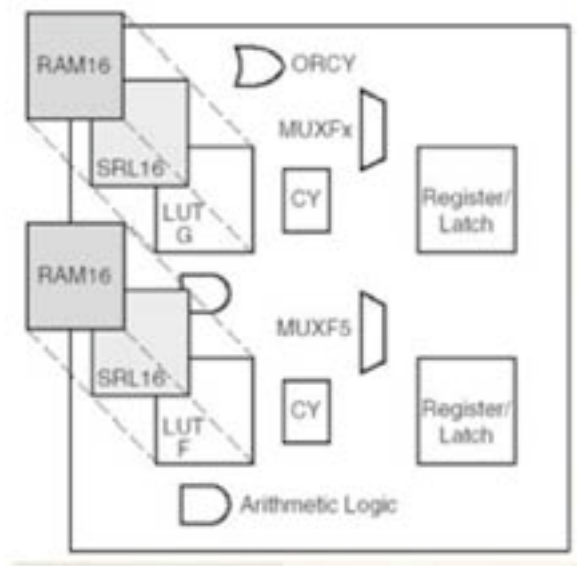
# Xilinx Virtex4 FPGA



**Logic array**



**PPC Processors**

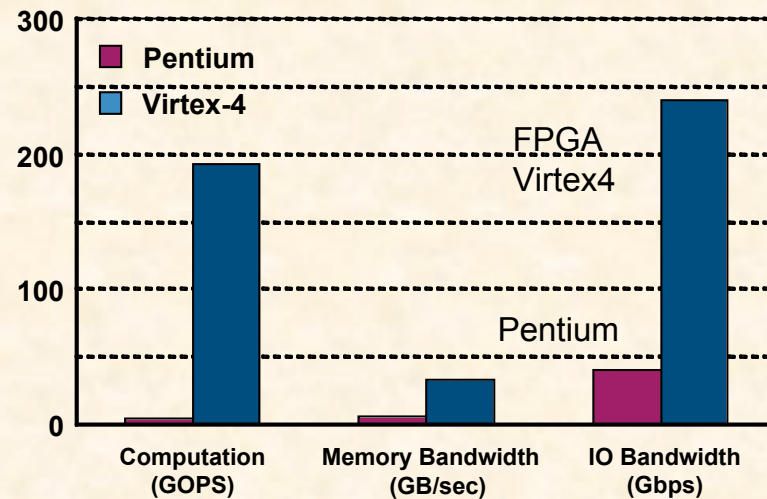
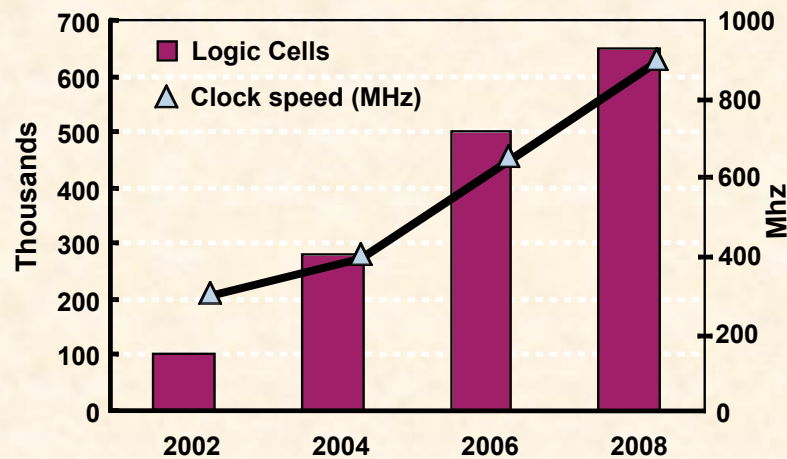


**67,584 slices\***

\*LX-160 (89,088 on LX-200)

# Why FPGAs?

- **Performance**—optimal silicon use (maximize parallel ops/cycle)
- **Rapid growth**—Cells, Speed, I/O
- **Power**—1/10th CPUs
- **Flexible**—tailor to application





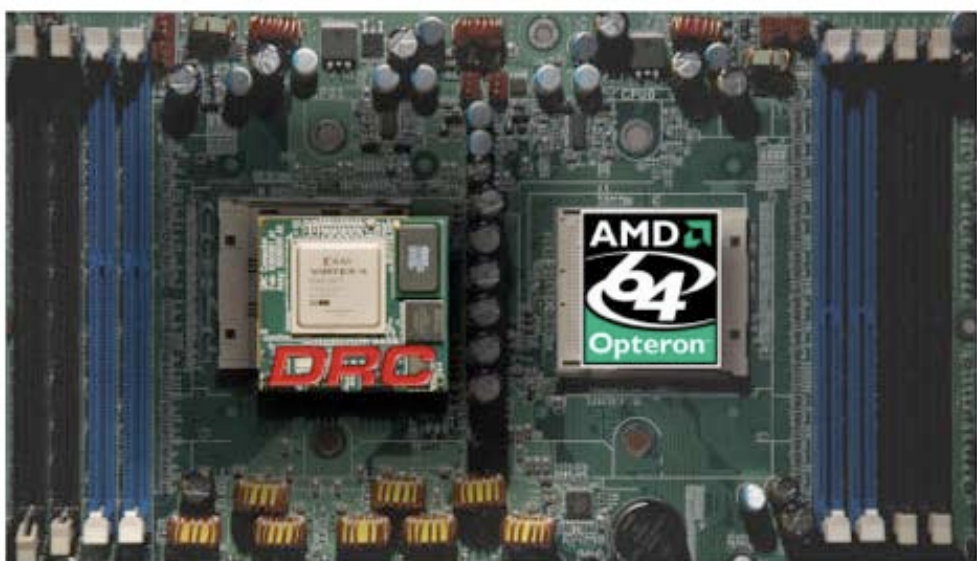
# Growing Industry Interest



“After exhaustive analysis, Cray concluded that, although *multi-core* commodity processors will deliver *some improvement*, exploiting parallelism through a variety of processor technologies using scalar, vector, multithreading & hardware accelerators (e.g., FPGAs or ClearSpeed co-processors) creates the greatest opportunity for application acceleration”

**HPC Future, Steve Scott, Cray CTO:**

*HPCWire* 24 March '06



**Cray Selects DRC FPGA Coprocessors for HPCS & future Supercomputers (*HPCWire*, 5 May '06)**

*Invited to test DRC - Jan Silverman, Cray VP Corp. Strategy*

 **Virtex4 FPGA blades to:**

*“Accelerate mission-critical applications by over 100x”*

 **FPGA accelerators**

 **FPGAs + FP chip**

**Potential: Petaflops/Exaflops at reduced power**



Storaasli - CUG07



# Smith-Waterman Algorithm Scoring

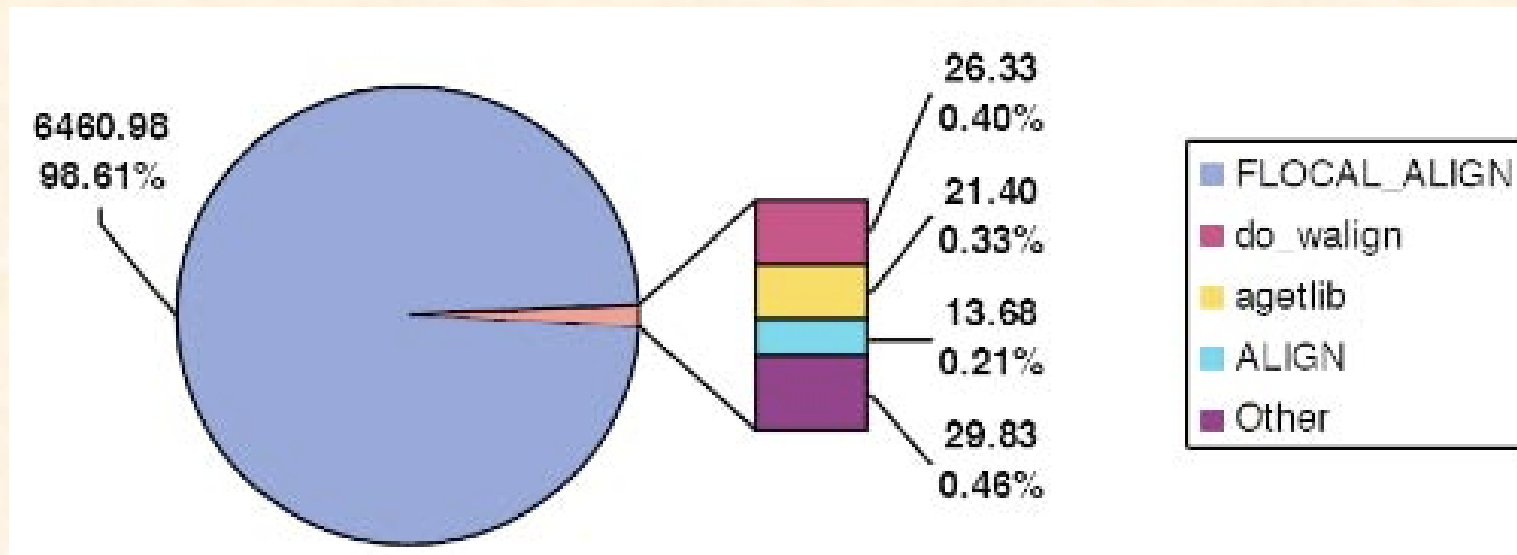
**Query Sequence**

**Database Sequence**

	0	A	C	G	T	...	C
0	0	0	0	0	0	0	0
A	0	2	0	0	0	2	0
C	0	0	4	2	1	0	2
G	0	0	2	6			
A	0						
A	0						
C	0						
...	0						
G	0						

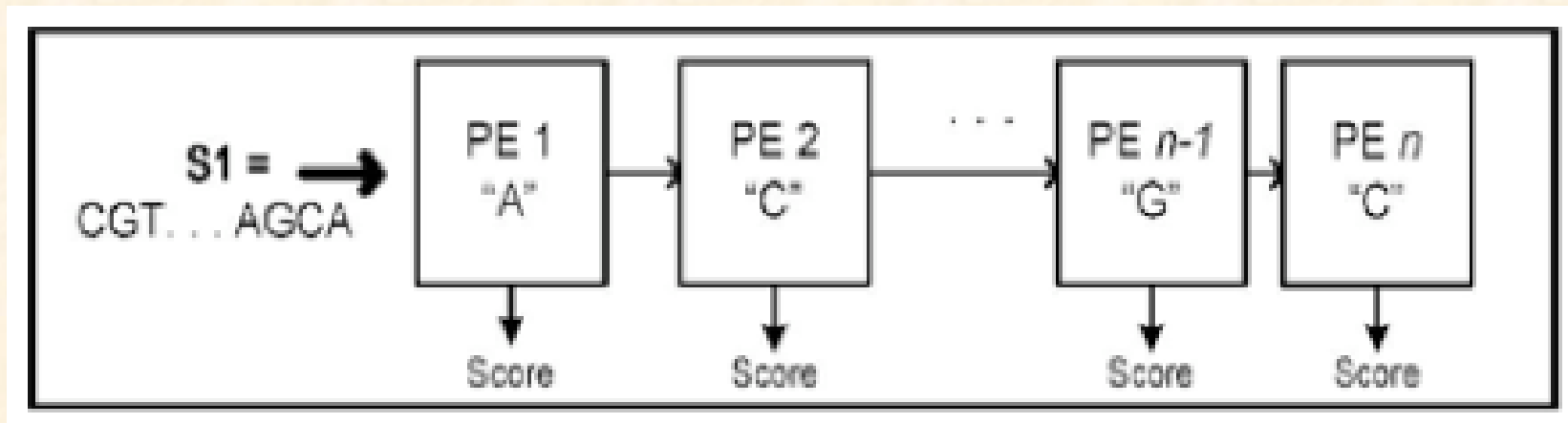
1. Initialize row & column 1 to 0
2. Score matches from upper left
3. Add to above-left score ( $2+4=6$ )

# Search34 Computation Profile



**98.61% is FLOCAL\_ALIGN**

# Smith-Waterman Pipeline



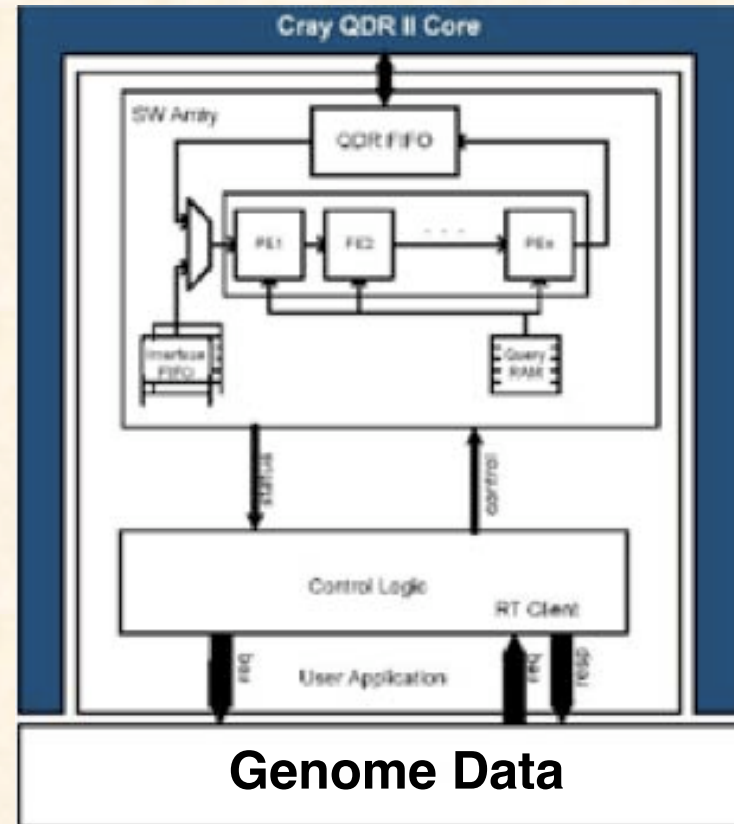
1. Query character preloaded into each PE
2. String  $S1$  shifted thru pipe to compare
3. Score generated

# Smith-Waterman

## Parallel Score Calculation

	0	A	C	G	T	...	C
0	0	0	0	0	0	0	0
C	0	0	0	2	1	2	1
G	0	0	2	1	0	3	2
T	0	0	1	4	3	2	
⋮	0	2	2	3	6		
T	0	1	1	2			
A	0	2	0				
A	0	0					
G	0						
C	0						
A	0						

## Overall Algorithm



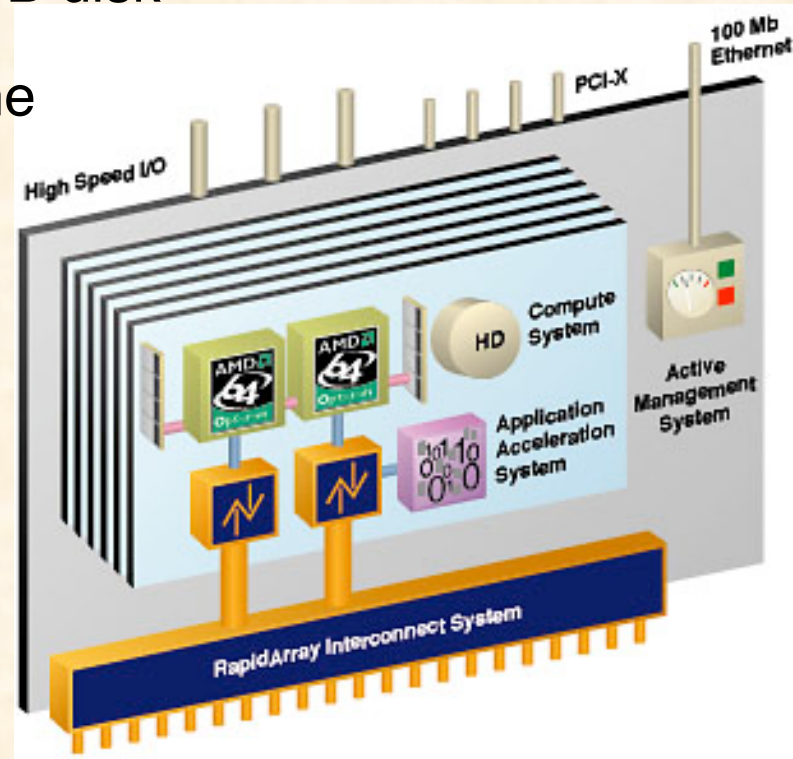


# Tiger

- 144 processors
- 2.2 Ghz AMD Opteron® processors
- 576 Gbytes of memory
- 18 TB Disk storage
- Fat Tree topology

# Cray XD1

- 144 processors => 633 GFLOPS peak  
(12 chassis 2.2GHz AMD Opterons)
- 6 Xilinx Virtex2Pro 5M gate FPGAs
- 576 GB Memory, 18TB disk
- Cray terabit backplane
- Porting applications



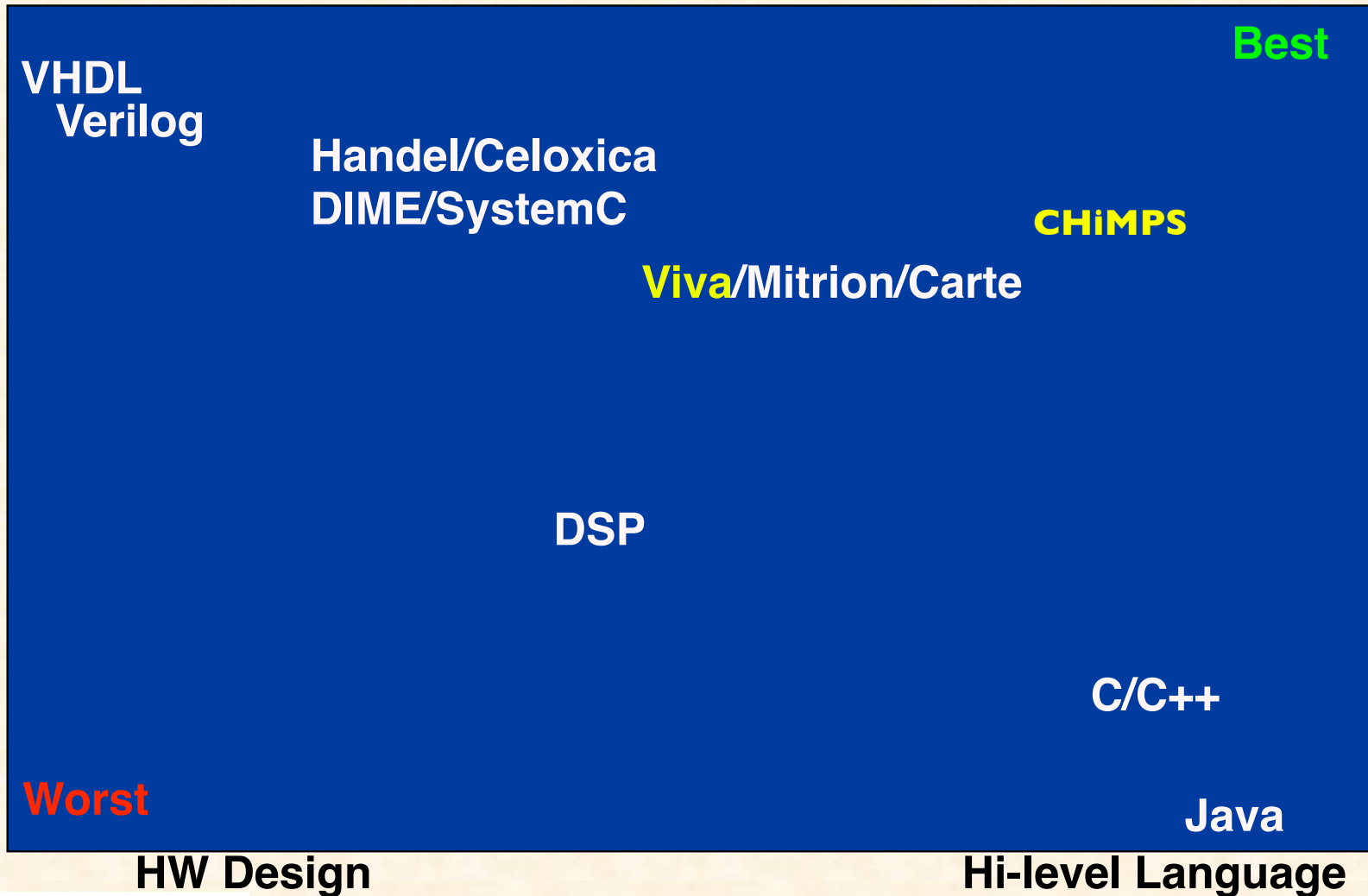
Storaasli - CUG07





# Tools: Performance vs. Coding Ease

Performance



Coding Ease



Storaasli - CUG07



# Openfpga.org Smith-Waterman Benchmark

- **FASTA** (University of Virginia) application  
<http://fasta.bioch.virginia.edu>
- Uses **search34** code & Cray **SWA** core
- Human Genome Data: 4GB compressed  
3685 searches (MPI on ORNL Cray XD1)



Alignment of ACGAADCCTTGC and ACGTATGC

	G	A	C	G	T	A	T	G	C
G	0	0	0	0	0	0	0	0	0
A	0	2	0	0	0	2	0	0	0
C	0	0	4	2	1	0	1	0	2
G	0	0	2	6	4	3	2	3	1
A	0	2	1	4	5	6	4	3	2
A	0	2	1	3	3	7	5	4	3
C	0	2	4	2	2	5	6	4	6
C	0	0	2	3	1	4	4	5	6
C	0	0	2	1	2	3	3	3	7
T	0	0	0	1	3	2	5	3	5
T	0	0	0	0	3	2	4	4	4
G	0	0	0	2	1	2	2	6	4
C	0	0	2	0	1	0	1	4	8

Final alignment

A	C	G	A	A	C	C	T	T	G	C	
A	C	G	T	A	-	-	-	-	T	G	C



Storaasli - CUG07



# Results

**Case 1: *Micro-RNA (DNA Short Reference Sequence)***

**Case 2: *Bacillus anthracis DNA comparison***

**Case 3: Amino Acid**

## Output Options (Speedup Impact)

**Detailed: -Q -H -f -l0 -g -3 -d 10 -b 10 -s OpenFPGA.mat -E 0.0001**

**Minimal: -Q -H -f -l0 -g -3 -d 0 -b 10 -s OpenFPGA.mat -E 0.0001**



# FPGA Performance

## ORNL XD1 (Virtex2): Initial Results

Case 1: *Micro-RNA*

### FPGA vs Opteron Time (hrs) for FASTA

	1	2	3	4	5
CPU 2.2GHz	75	-	-	-	-
FPGA(s) 0.2GHz	7.39	3.75	2.48	1.91	1.56
FPGA Speedup vs 1 CPU	10.15	20.0	30.2	39.3	48.1



Storaasli - CUG07



# Cray XD1 FPGA Speedup vs. 2.2 GHz Opteron

## Case 2: *Bacillus anthracis* DNA comparison

### Virtex2 Pro 50 Speedup

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	Avg	SD
8k	26	30	27	32	30	30	29	30	30	27	31	29	31	30	31	31	31	30	29.6	1.2
16k	22	25	26	31	30	30	28	31	28	27	30	29	29	29	32	31	32	29	28.7	2.5
8k	49	49	49	50	49	49	50	49	49	49	49	49	49	49	50	49	49	49	49.4	0.2
16k	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	49	50	50	49.9	0.3

### Virtex4 LX160 Speedup

	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	Avg	SD
8k	36	43	39	47	44	45	43	45	45	39	46	42	46	44	46	46	46	43	43.5	2.9
16k	29	33	37	45	44	43	39	47	41	37	46	41	43	41	47	46	48	43	41.5	4.9
8k	98	98	98	97	97	98	98	98	98	97	98	98	98	98	98	98	98	97	97.6	0.1
16k	100	101	101	100	100	100	101	101	101	101	101	101	101	101	100	100	101	100	100.7	0.4

24 => Sequence AE017024

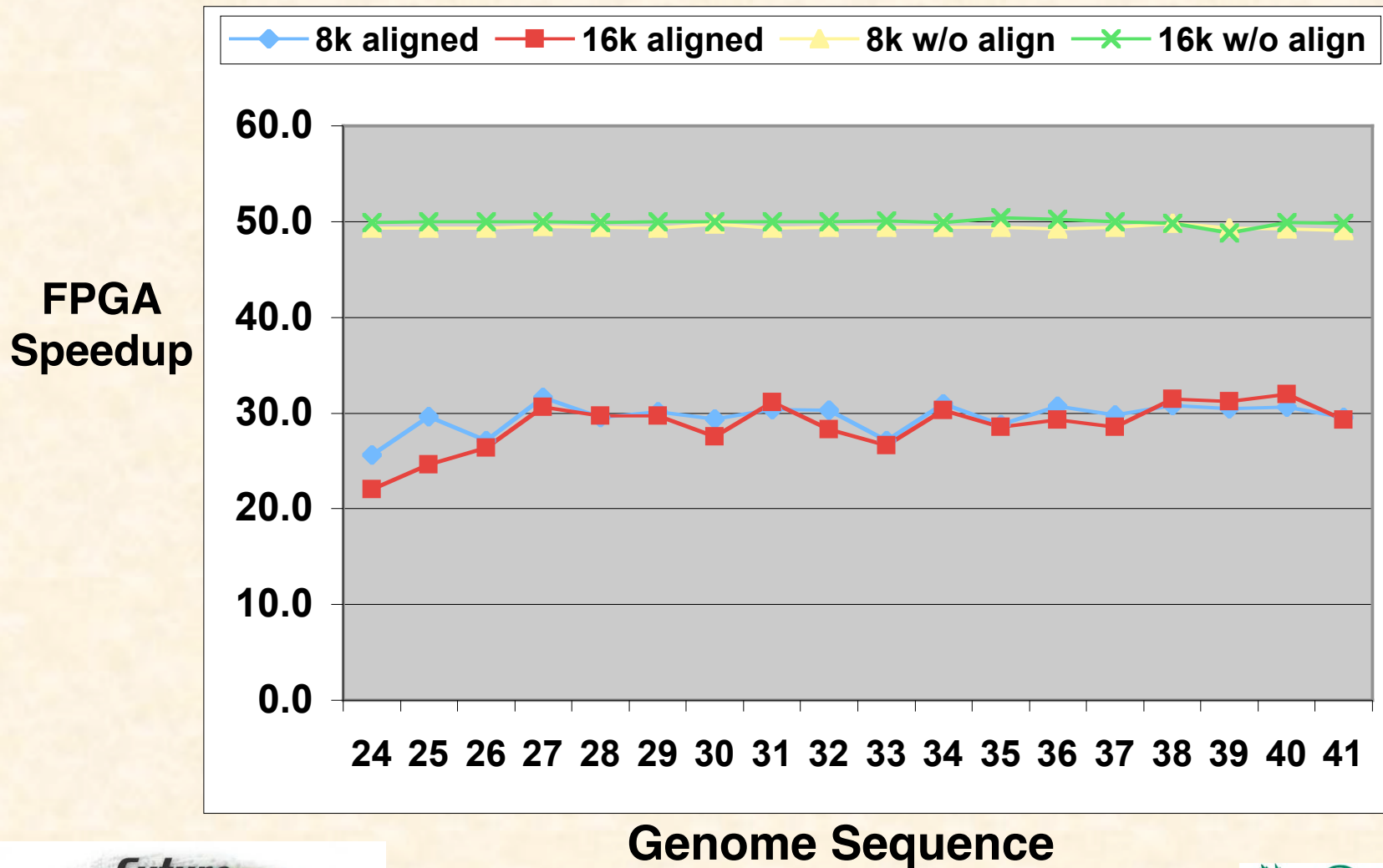


Storaasli - CUG07



# XD1 Virtex2 Speedup vs. 2.2 GHz Opteron

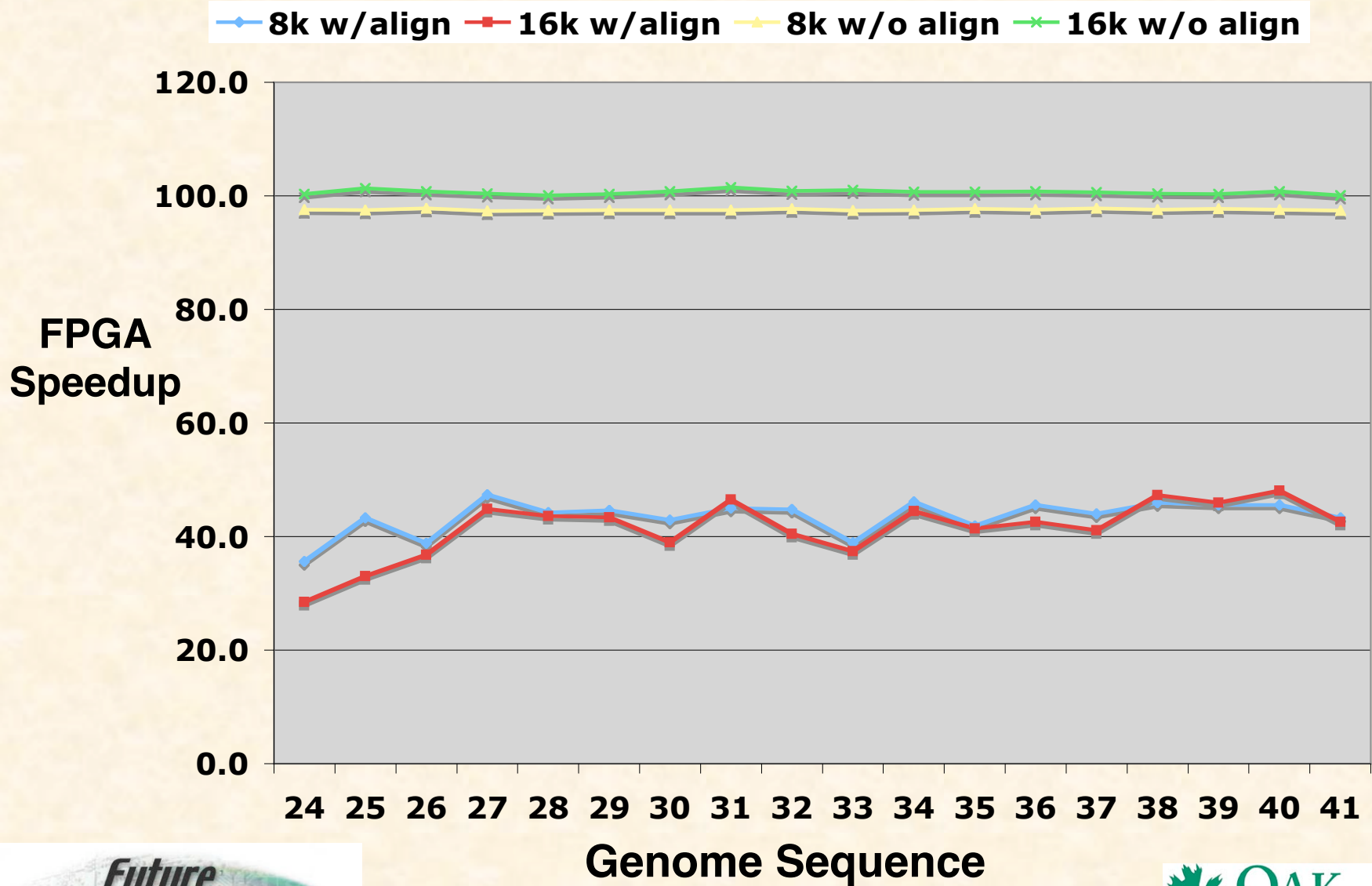
## Case 2: *Bacillus anthracis* DNA comparison



Storaasli - CUG07



# XD1 Virtex4 Speedup vs. 2.2 GHz Opteron



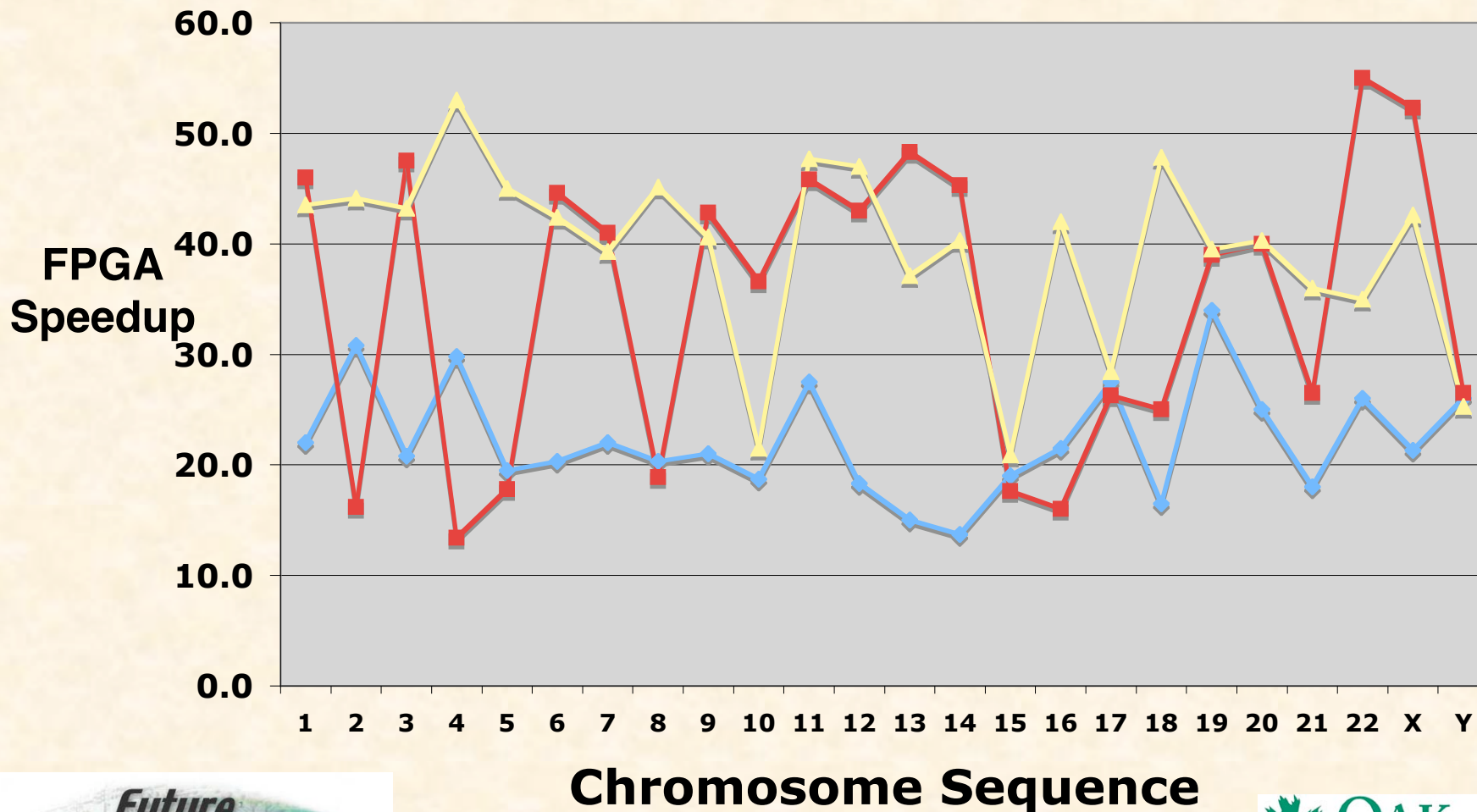
Storaasli - CUG07



# XD1 Virtex2 Speedup vs. 2.2 GHz Opteron

## Case 3: Amino Acid

◆ myc ■ ras ▲ src



Chromosome Sequence

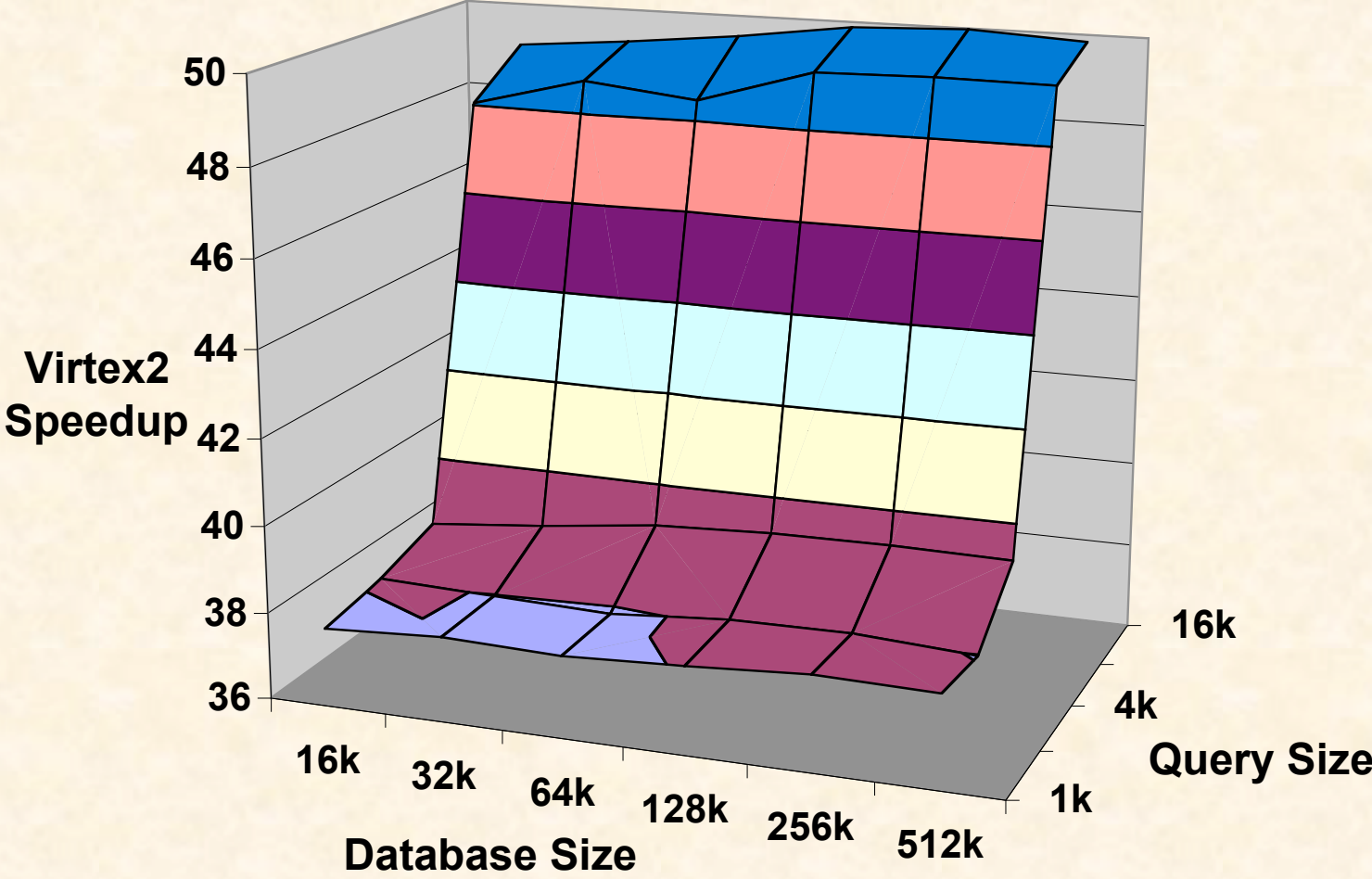


Storaasli - CUG07





# FPGA Speedup: Query and Database size



Storaasli - CUG07



# Future Opportunities

- Speedup XD1 code another 2X => **200X** (LX160)
- LX200 speedup (89k/68k = 1.3) => **262X** (LX200)
- DRC LX200 module => Cray XT supercomputers
- ORNL-Xilinx-Cray-DRC CHiMPS collaboration
  - widen range of applications (climate, MD, solvers,...)

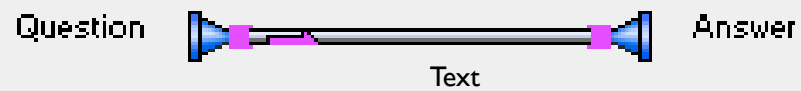


# Summary

- ***FPGA, Genome matching background***
  - *FASTA, search34, Smith-Waterman*
- **Results: 3 openfpga.org cases**  
**XD1 Speedups: 50X (Virtex2), 100X (Virtex4)**  
*(promise of 200X or more)*
- ***Future opportunities: DRC LX200 (Session 12C)***



# THANK YOU



**Google: Olaf ORNL**



Storaasli - CUG07

