Accelerating Genome Sequencing 100X with FPGAs

Olaf Storaasli - HPEC07, MIT Lincoln Labs

- Relevance
- Accomplishments
- Summary

Future Technologies Group ornl

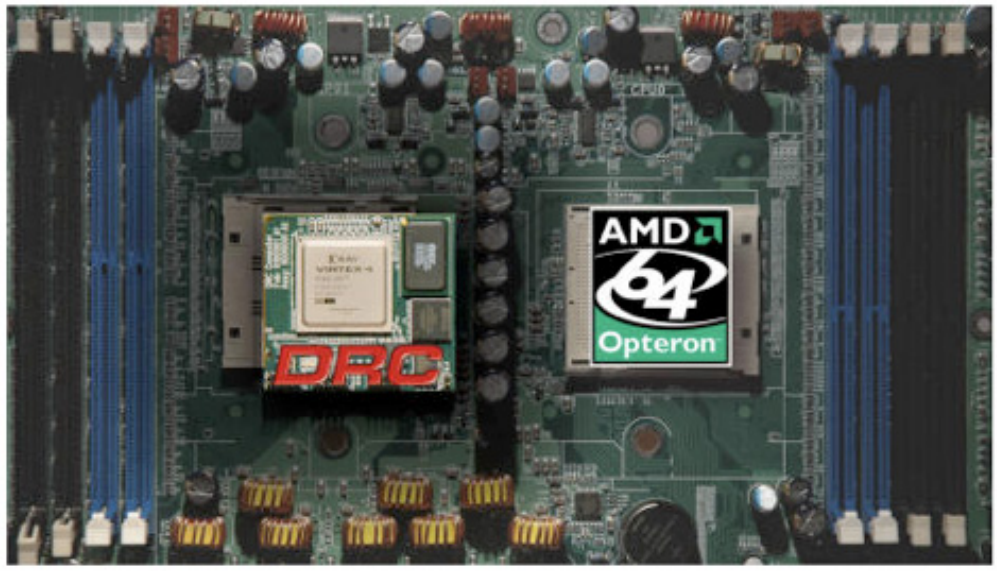
OAK RIDGE National Laboratory

Increasing FPGA Relevance to HPC

CRAY THE SUPERCOMPUTER COMPANY

*"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



DRC FPGA Coprocessors in HPCS & Cray XT

SGI Virtex4 FPGA blades to:

"Accelerate mission-critical applications by over 100x"

SRC FPGAs (Altera)

XtremeData, Inc. COMPUTING REDEFINED ***Opteron Socket***

Potential: Petaflops/Exaflops at reduced power

Future Technologies Group oml

OAK RIDGE National Laboratory

FPGA Coding: Graphical vs Text

Gauss matrix solver

Viva: Graphical Icons—3-dimensional

Compiler, simulator, and debugger

MitrionC: Text/flow—1-dimensional

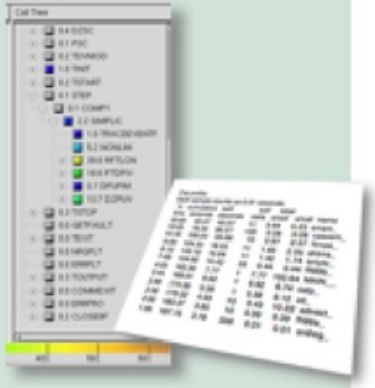
+ Carte/SRC, CHiMPS-VHDL/Xilinx , DSPlogic

Viva Algorithms Developed*

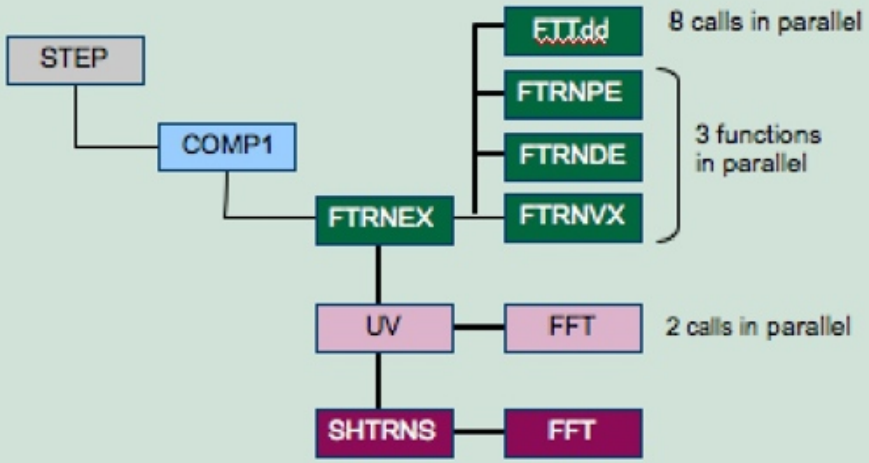
- **Matrix Algebra**: {V}, [M], $\{V\}^T\{V\}$, [M]x[M],GCD,…
- n! => Probability: Combinations/Permutations
- Cordic => Transcendentals: sin, log, exp, cosh…
- $\partial y/\partial x$ & $\int f(x)dx$ => Runge-Kutta: CFD, Newmark Beta: CSM
- **Matrix Equation Solvers**: [A]{x} = {b}, Gauss & Jacobi
- Dynamic Analysis: $[M]\{\ddot{u}\} + [C]\{u\} + [K]\{u\} + NL = \{P(t)\}$
- Nonlinear Analysis: reduces NL time
- **Structural Design/Optimization**
- **Unsolved App:** Traveling Salesman

Climate/Weather Code: CHiMPS FPGA Port

Profile

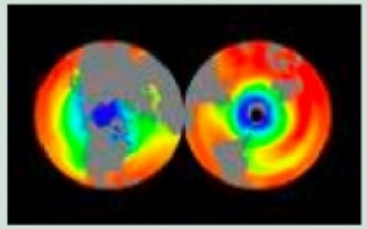


Find parallelism: 80% FFTs



Goal

More GF/$ GF/Watt



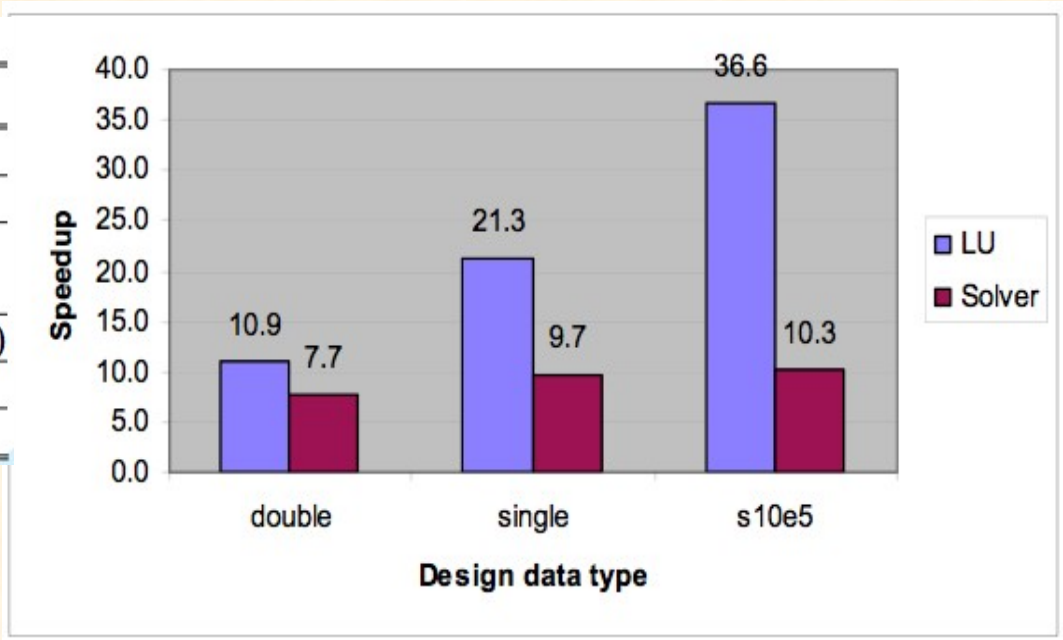
Model faster

*MD CHiMPS port took **20 min***
*+ **20 min** to optimize*

10X* FPGA Speedup for Matrix Equation Solver
36X for LU Decomposition

Table 6: LU implementation on XC2VP50-7

Design	Double FP	Single FP	S10e5
PE amount	8	16	32
Max size	128	256	256
Achievable Frequency	120MHz	150MHz	150MHz
Slices	27,005 (57%)	14792 (59%)	14730 (62%)
BRAMs	68 (29%)	129 (55%)	65 (28%)
MULT18X18	128 (55%)	64 (27%)	32 (13%)



Benefits:

High performance of LP arithmetic
High precision accuracy
Speedup increases with matrix size

First mixed-precision LU & solver for FPGAs

*vs 2.2 GHz Opteron

Smith-Waterman Benchmark

- *FASTA genome matching application*

  http://fasta.bioch.virginia.edu

- Uses search34 code & Cray SW core

- NCBI Human Genome Data: 4GB compressed
  Case 1: Micro-RNA, Case 2: DNA

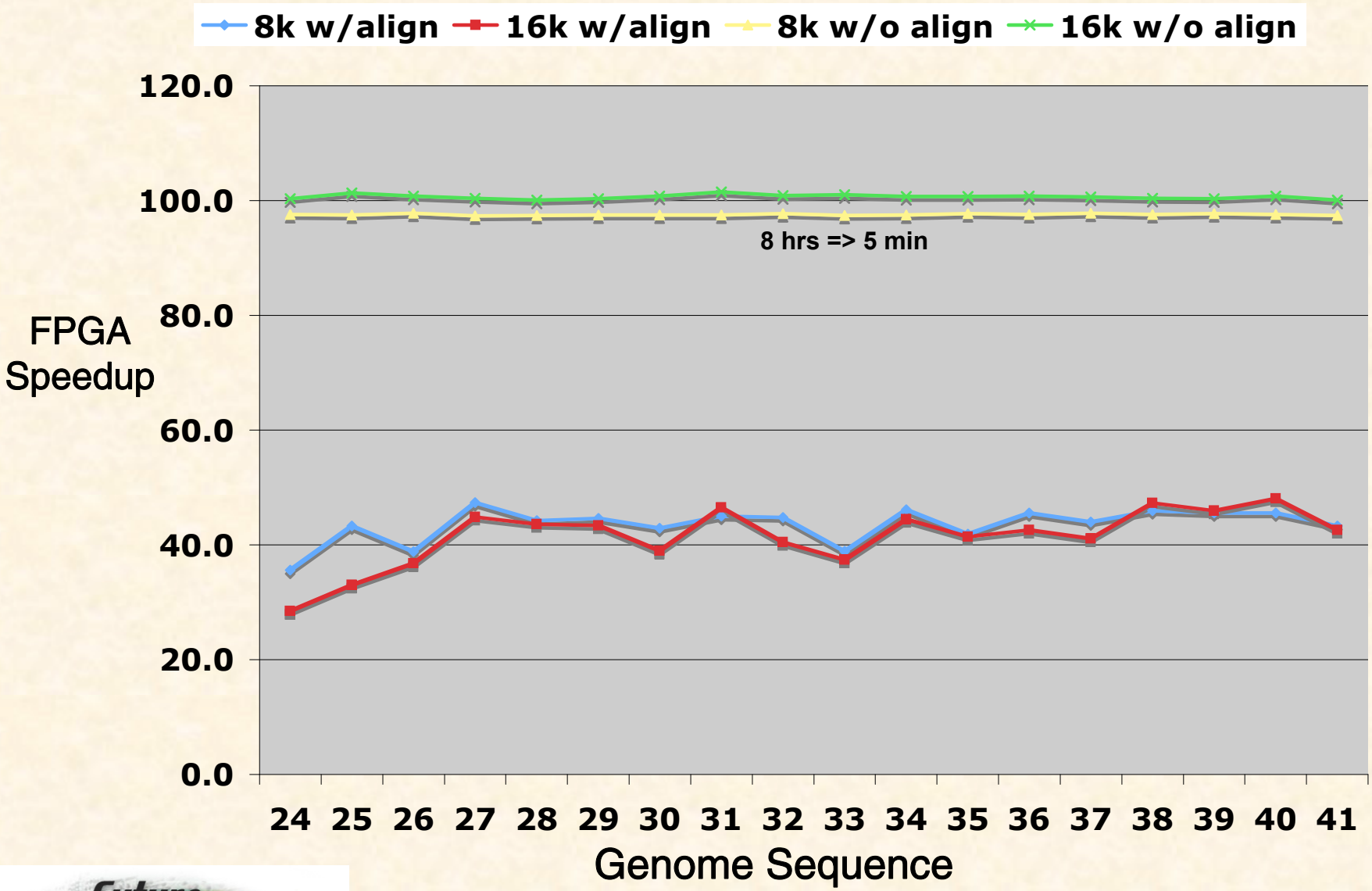
Alignment of ACGAACCCTTGC and ACGTATGC

	0	A	C	G	T	A	T	G	C
0	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

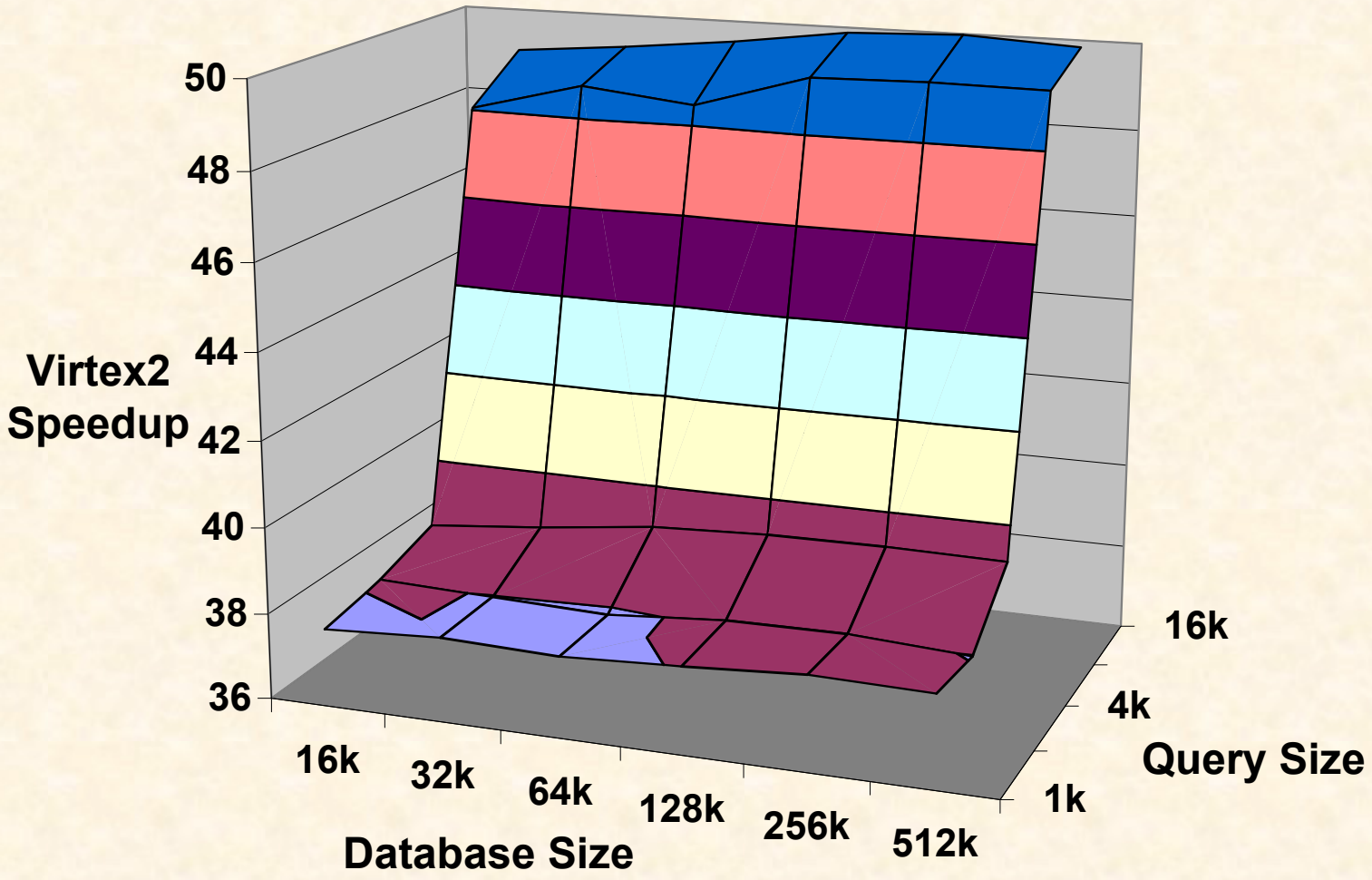
A	C	G	A	A	C	C	C	T	T	G	C
	|	|		|				|	|	|
A	C	G	T	A	-	-	-	-	T	G	C

100X* Cray XD1 Virtex-4 Speedup*



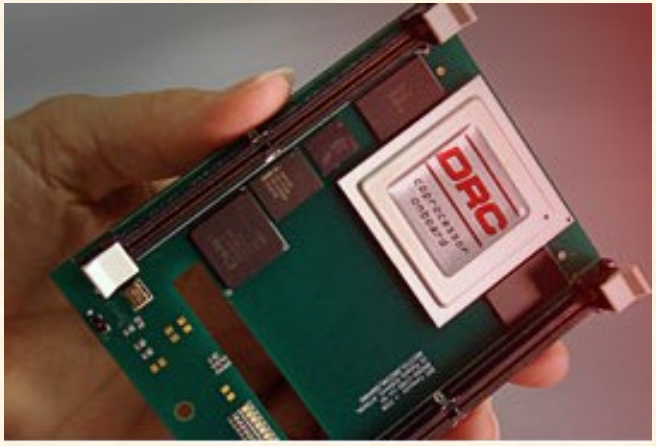
*vs 2.2 GHz Opteron

FPGA Speedup grows with Query Size



Next: Expected Results

- DRC LX200 (89k/68k = 1.3) => 130X => Cray XT4
- 144 Cray XD1 FPGAs => 144X50+ => 7,200X
- New CHiMPS (climate, MD, solver apps) => paper
- New SGI RASC* (Virtex4): BLAST 60X, FASTA,…



*Reconfigurable Application Specific Computing

Summary

- *Increased FPGA HPC relevance*: speed, power
- Growth: team, H/W, tools & Apps (+$0.0)
  - FPGAs: Cray, SGI, SRC, Nallatech, Digilent + Bee2, DRC
  - Tools: Mitrion-C, Carte, Viva, CHiMPS
  - Apps: STSWM, FASTA, MD, BLAST & Matrix Eqn. Solver
- Results: FASTA Speedups: 50X (V2), 100X (V4)
  Solver 10X (V2), BLAST 60X (V4) + MD 10X, STSWM NDA
- Next:
  - *144 FPGAs,* DRC LX200 => XT4 => *ORNL Path*
  - CHiMPS & SGI/RASC Application performance
  - Publish results