

**Technical University of Crete**  
**Electronic and Computer Engineering Department**  
**Microprocessor and Hardware Laboratory**

*A General Reconfigurable Architecture for the  
BLAST algorithm*

Euripides Sotiriades  
Apostolos Dollas

**HBMIM 4-5 October, 2007**

# Outline

- **Motivation**
- **Algorithms**
- **Implementations up to Date**
- **TUC Architecture**
- **Performance and Comparison**
- **Future work**

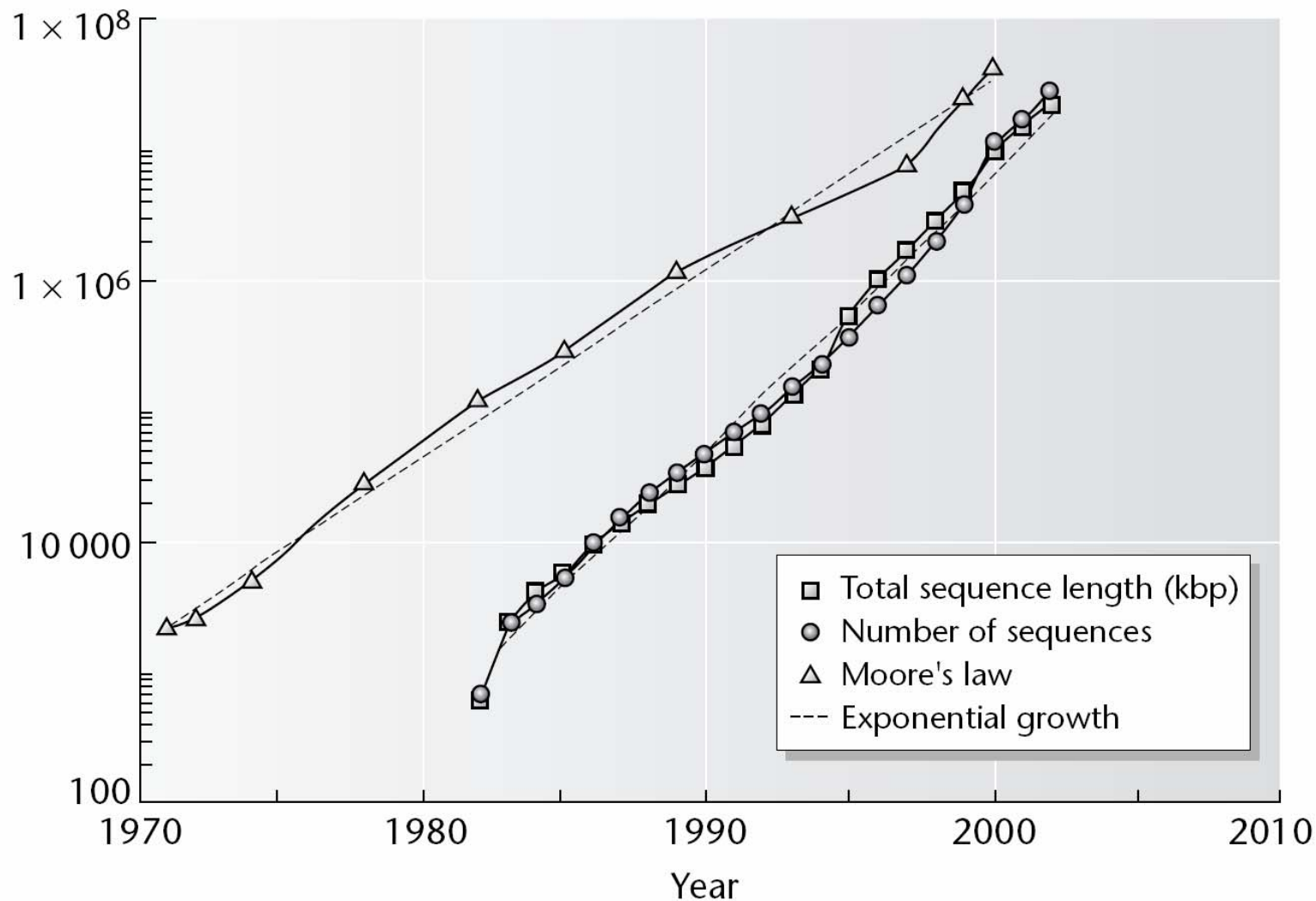
# Problem Description

Major Data Banks are created with huge scientific effort producing data at rate faster than Moore's law

Bioinformatics consist of several Computational Intensive Problems

The major operation in computational biology is sequence comparison.

# Data Size



# Implementations to Date

SW implementation of NCBI is dominating the area

NCBI software runs in Parallel systems as benchmark example for IBM, DELL, Apple

Smith Waterman HW implementation

- “Searching Genetic Databases on Splash 2” D. Hoang et al.
- “Gene Matching Using JBits “ S. Guccione *et. al.*
- “A Run-Time Reconfigurable System for Gene-Sequence Searching” @ Virginia Tech

# Implementations to Date (2)

## BLAST

- “RC-BLAST”, Clemson University-Sandia National Labs-University of Kansas
- “Mercury BLAST”, Washington University, St. Louis
- “TreeBLAST”, Boston University
- “FPGA/FLASH Accelerator”, IRISA / CNRS - Institute of Computing Technology Beijing

## Commercial

- DeCypher

# Reconfigurable Computing for DNA

Dynamic Programming Algorithm  
Implementation due to FPGAs Structure

Implementation of BLAST Algorithm

RC BLAST

TreeBLAST

MercuryBLAST

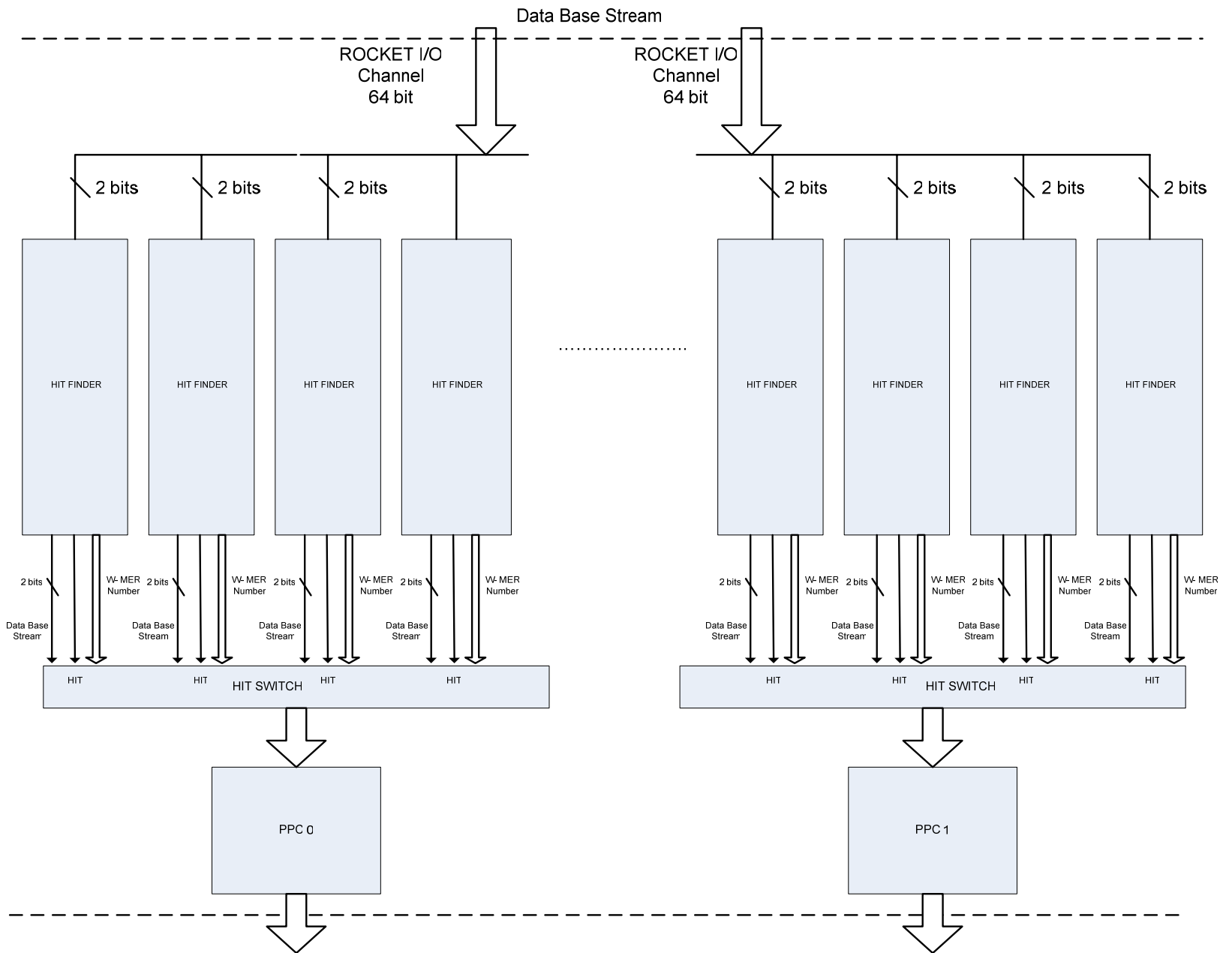
DeCypher

# Recent Technology FPGA Features for BLAST Implementation

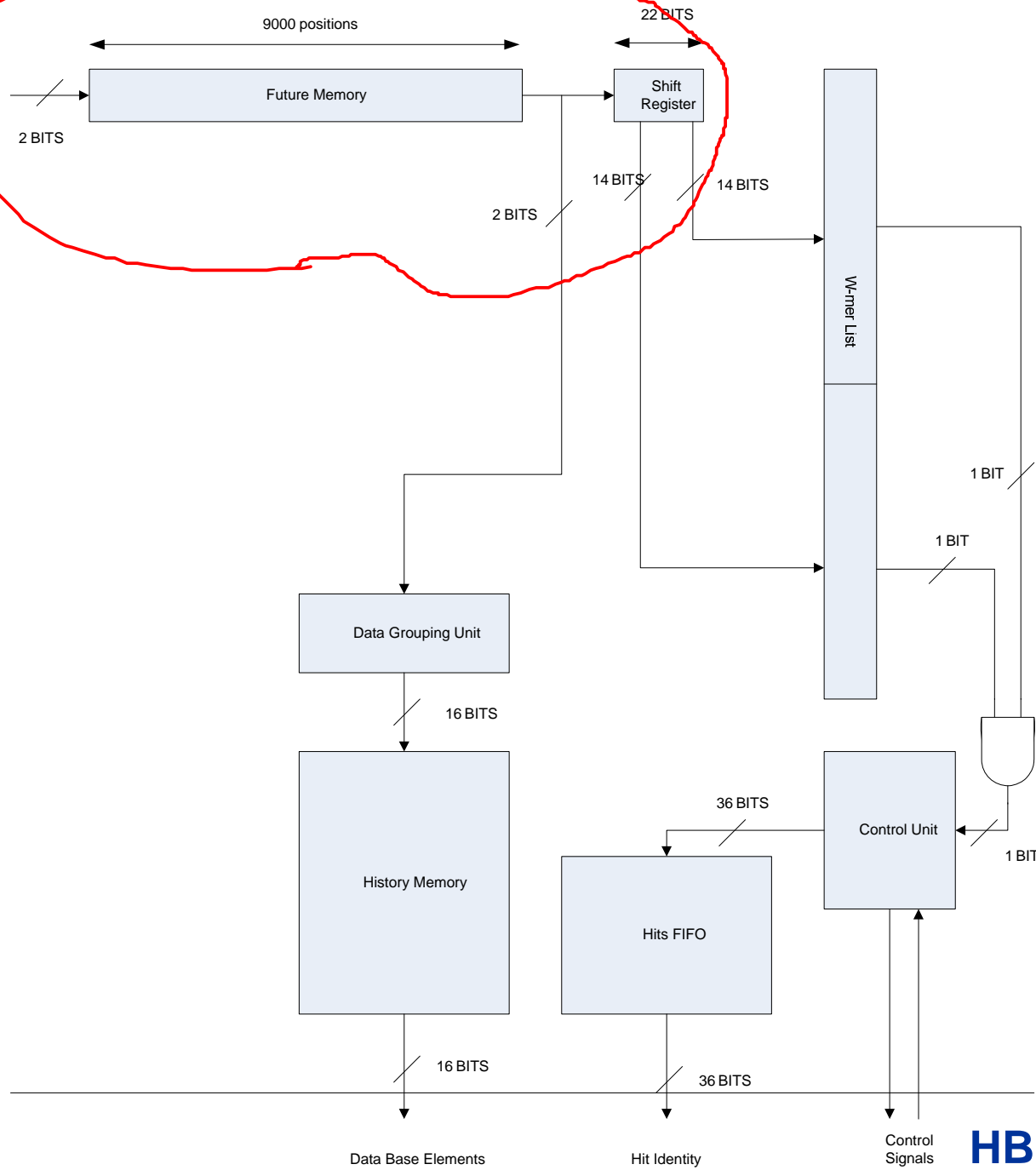
- High I/O Throughput
  - Up to 24 Transceivers
  - Up to 3.75 GBs Baud Rate (90 aggregate).
  - Many different Protocol Implementation
- Large Amount of RAM
- Existence of Powerful PowerPC processors

# TUC Architecture

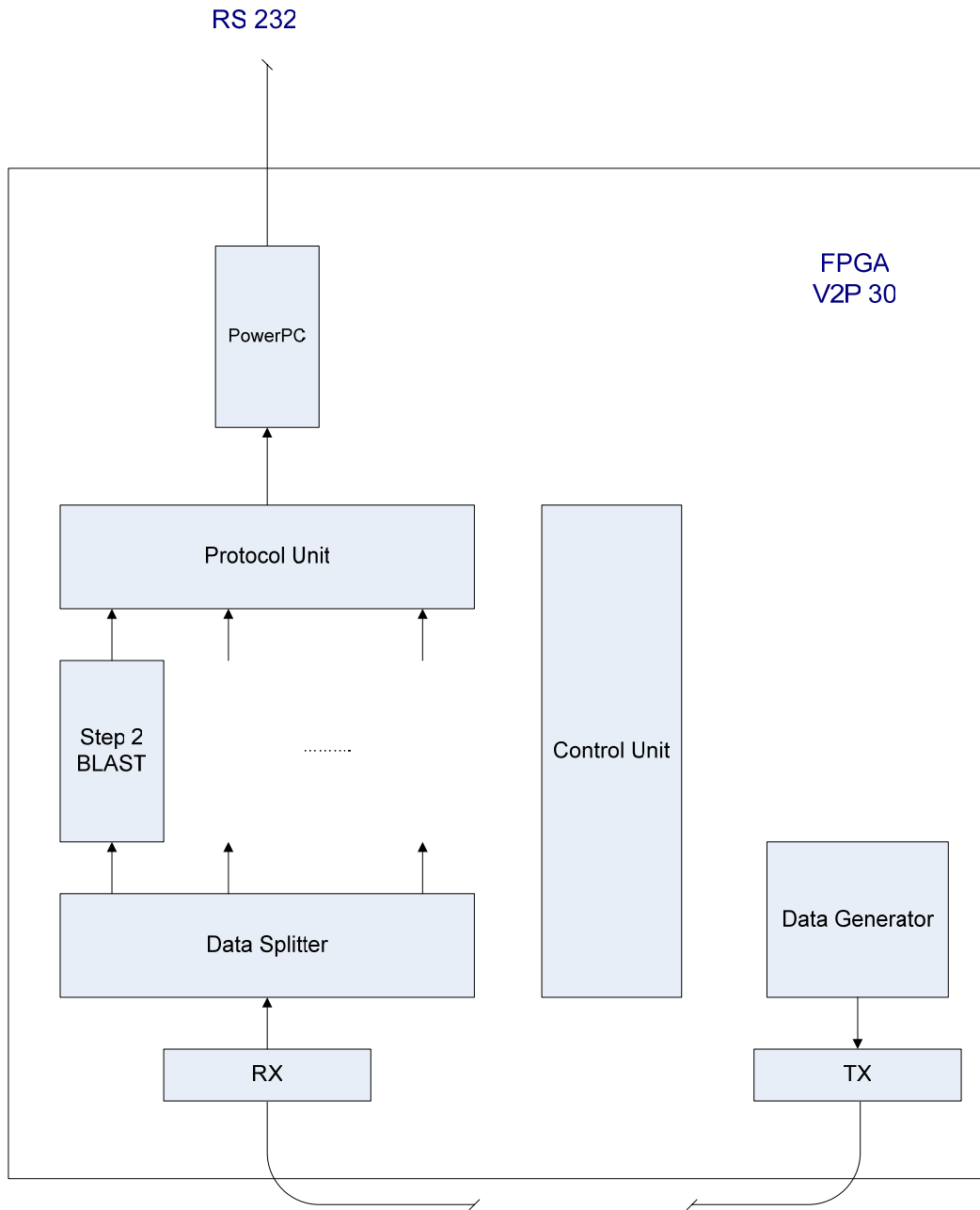
- Development of 2 TUC generations
- FPGA based Architecture
- Parallel Processing Elements
- Implementation of all BLAST variations
- Any size of database
- 1<sup>st</sup> Step is precalculated initializes the system
- Queries of any size up to 200.000



# Datapath



# Prototyping



Infiniband 2Gbps

# Performance

<i>TUC system Throughput (char.10<sup>6</sup>/sec)</i>		
<i>BLAST<sub>n</sub></i>	<i>BLAST<sub>p</sub></i>	<i>BLAST<sub>x</sub>/ TBLAST<sub>n</sub> TBLAST<sub>x</sub></i>
8192	8192	1365
8192	8192	1365
8192	4096	683
4096	2688	341
2048	896	128
1344	512	75
640	256	32
448	128	21
320	128	11

# Performance Comparison

*Software Implementations*

*Hardware Implementations*

	<i>IBM System</i>			<i>Intel Pentium 4 @ 3.00 GHz</i>			<i>Mercury BLAST</i>	<i>TreeBlast</i>
<i>Query Length (Char.)</i>	<i>BLASTn</i>	<i>BLASTp</i>	<i>BLASTx/ TBLASTn TBLASTx</i>	<i>BLASTn</i>	<i>BLASTp</i>	<i>BLASTx/ TBLASTn TBLASTx</i>	<i>BLASTn</i>	<i>BLASTp</i>
<b>600</b>	-	-	-	-	-	-	-	<b>74</b>
<b>1000</b>	<b>6,82</b>	<b>169,15</b>	<b>72,15</b>	<b>13,91</b>	<b>413,74</b>	<b>65,09</b>	-	-
<b>2000</b>	<b>6,82</b>	<b>169,15</b>	<b>72,15</b>	<b>20,02</b>	<b>644,03</b>	<b>104,04</b>	-	-
<b>5000</b>	<b>51,41</b>	<b>413,74</b>	<b>497,09</b>	<b>37,53</b>	<b>803,14</b>	<b>137,42</b>	-	-
<b>10000</b>	<b>25,70</b>	<b>271,52</b>	<b>248,18</b>	<b>31,67</b>	<b>890,07</b>	<b>124,00</b>	-	-
<b>25000</b>	-	-	-	-	-	-	<b>1400</b>	-
<b>30000</b>	<b>12,85</b>	<b>90,51</b>	<b>93,16</b>	<b>49,64</b>	<b>869,90</b>	<b>124,27</b>	-	-
<b>50000</b>	<b>25,29</b>	<b>343,62</b>	<b>396,83</b>	<b>37,20</b>	<b>867,80</b>	<b>136,36</b>	-	-
<b>100000</b>	<b>12,04</b>	<b>171,81</b>	<b>169,31</b>	<b>34,17</b>	<b>984,62</b>	<b>152,38</b>	-	-
<b>150000</b>	<b>8,43</b>	<b>85,91</b>	<b>111,11</b>	<b>34,70</b>	<b>853,33</b>	<b>190,91</b>	-	-
<b>200000</b>	<b>6,02</b>	<b>85,91</b>	<b>58,20</b>	<b>31,59</b>	<b>1422,22</b>	<b>157,14</b>	-	-

# Conclusions - Future Work

- BLAST Implementation and Bioinformatics in general are Interesting Problems area for reconfigurable logic
- Reconfigurable logic can offer a significant performance boosting providing a system level solution.
- Gapped BLAST
- Systematic Approach of Bioinformatics Problems and Algorithms from Reconfigurable HW
- Build a Reconfigurable based platform for Bioinformatics computations
- Create IP cores for several computationally intensive problems (as phylogenetics, prediction of RNA secondary structure, protein folding problem, etc)



Thank you for your attention