

# Workload Characterization of Bioinformatics Applications on Pentium 4 Architecture

Yue Li, Tao Li, Tamer Kahveci, José Fortes  
University of Florida

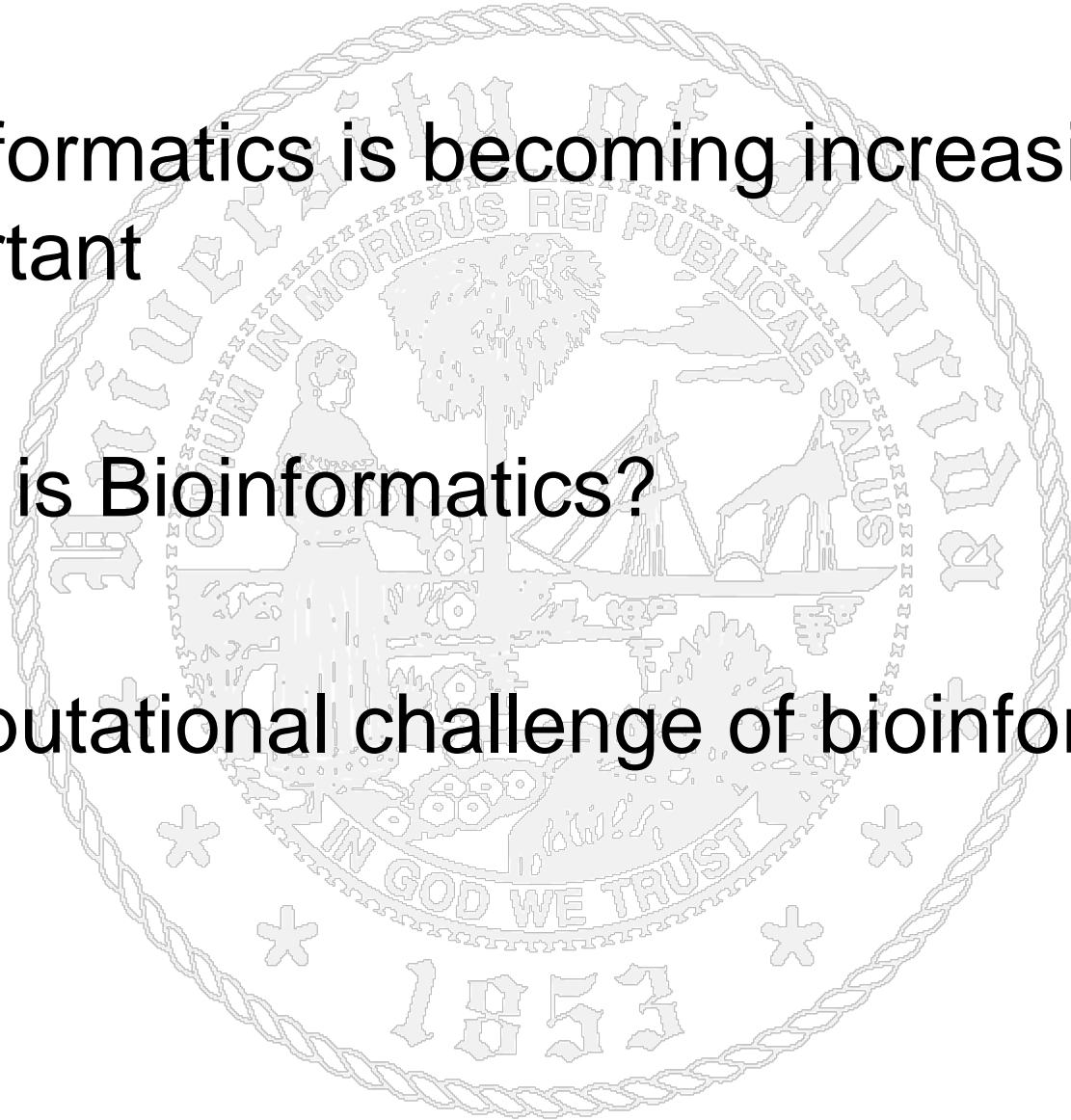
Sept. 27, 2005

1853

# Motivation



- Bioinformatics is becoming increasingly important
- What is Bioinformatics?
- Computational challenge of bioinformatics



# Previous Work

---



- Analysis of single program
- Focus on parallel programming issues at application level
- Largely ignore architecture design issues

# Contributions of this Work



- Characterize 17 representative bioinformatics applications on Intel Pentium 4 Architecture
  - Focus on various microarchitecture characteristics
- Propose a benchmark suite - BioInfoMark

# Outline



- Background
- Experimental Methodology
- A Characterization of Bioinformatics Benchmarks
- Conclusions

# The Area of Bioinformatics



- Sequence Analysis
- Phylogeny Analysis
- Protein Structure Analysis
- Molecular Dynamics

# Selected Benchmarks



Sequence Analysis	Blast, Fasta, ClustalW, Hmmer, Glimmer, Emboss
Phylogeny Analysis	Phylip
Protein Structure Analysis	Dali, CE, Predator, Threader
Molecular Dynamics	Gamess, Amber

# Outline



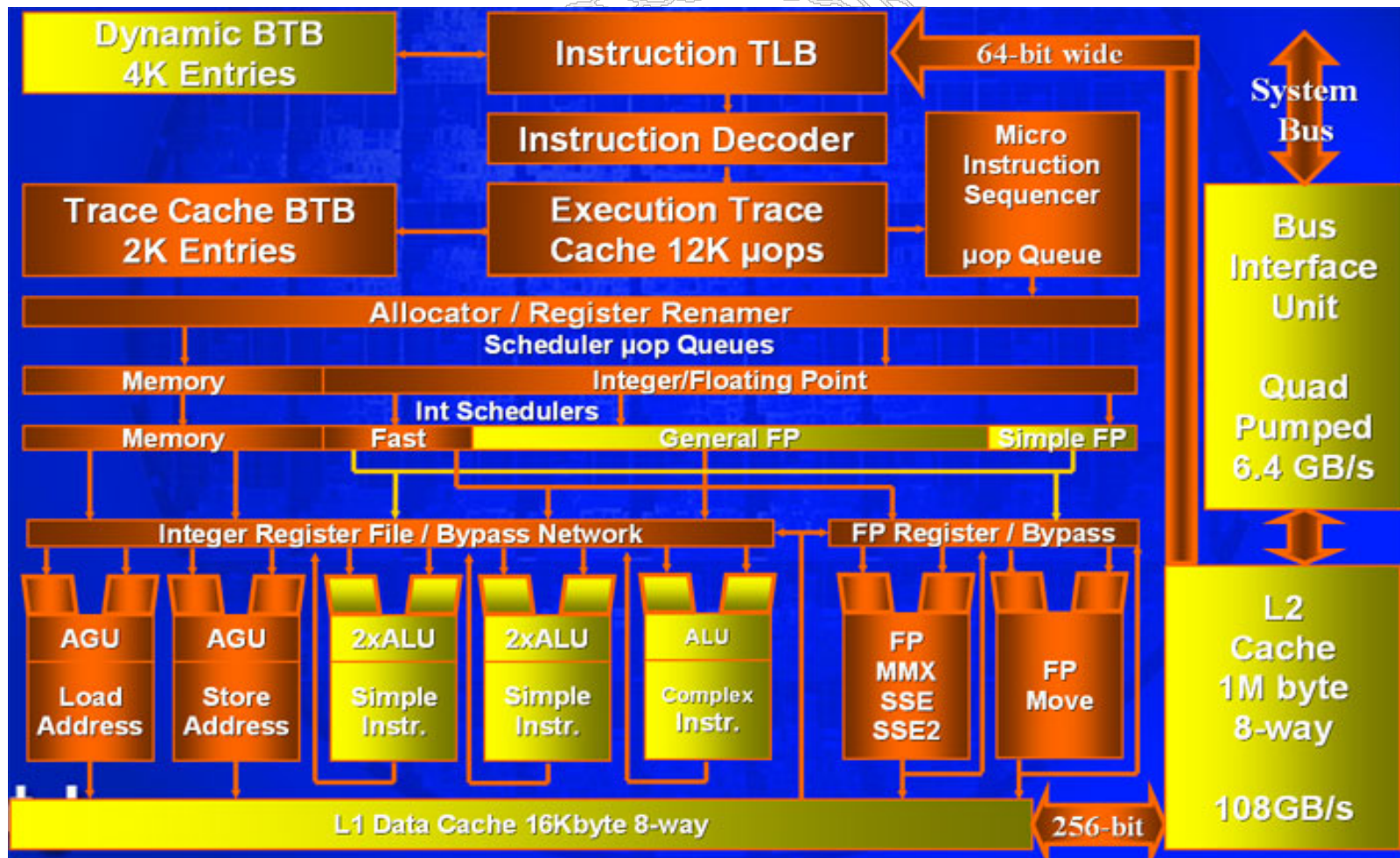
- Background
- **Experimental Methodology**
- A Characterization of Bioinformatics Benchmarks
- Conclusions

# Experimental Configuration



<p>Processor</p>	<p>2.4GHz Intel Pentium 4 Processor</p>
<p>Performance Counter</p>	<p>Brink and Abyss Performance Counter Tools</p>

# Pentium 4 Microarchitecture

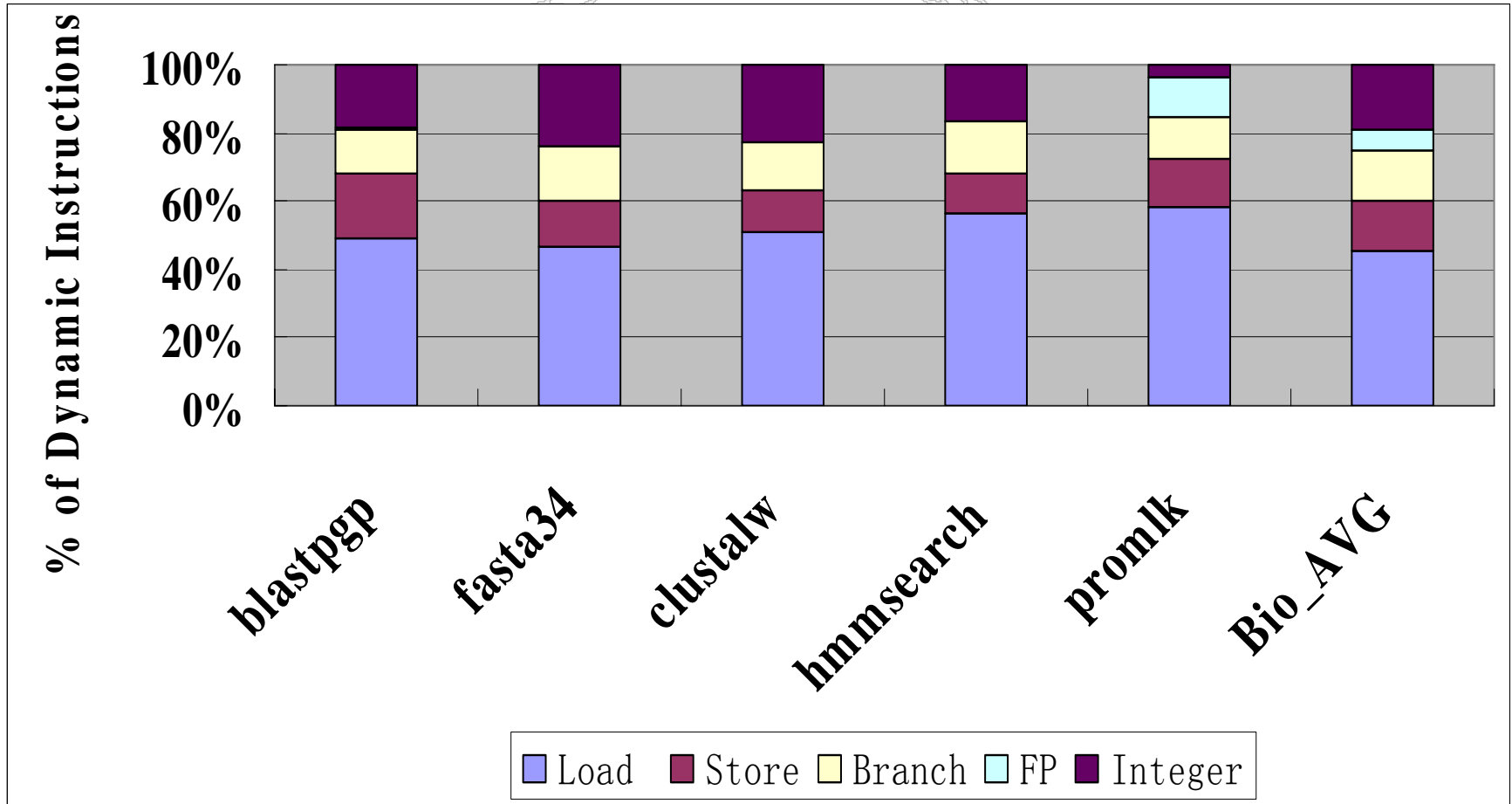


# Outline

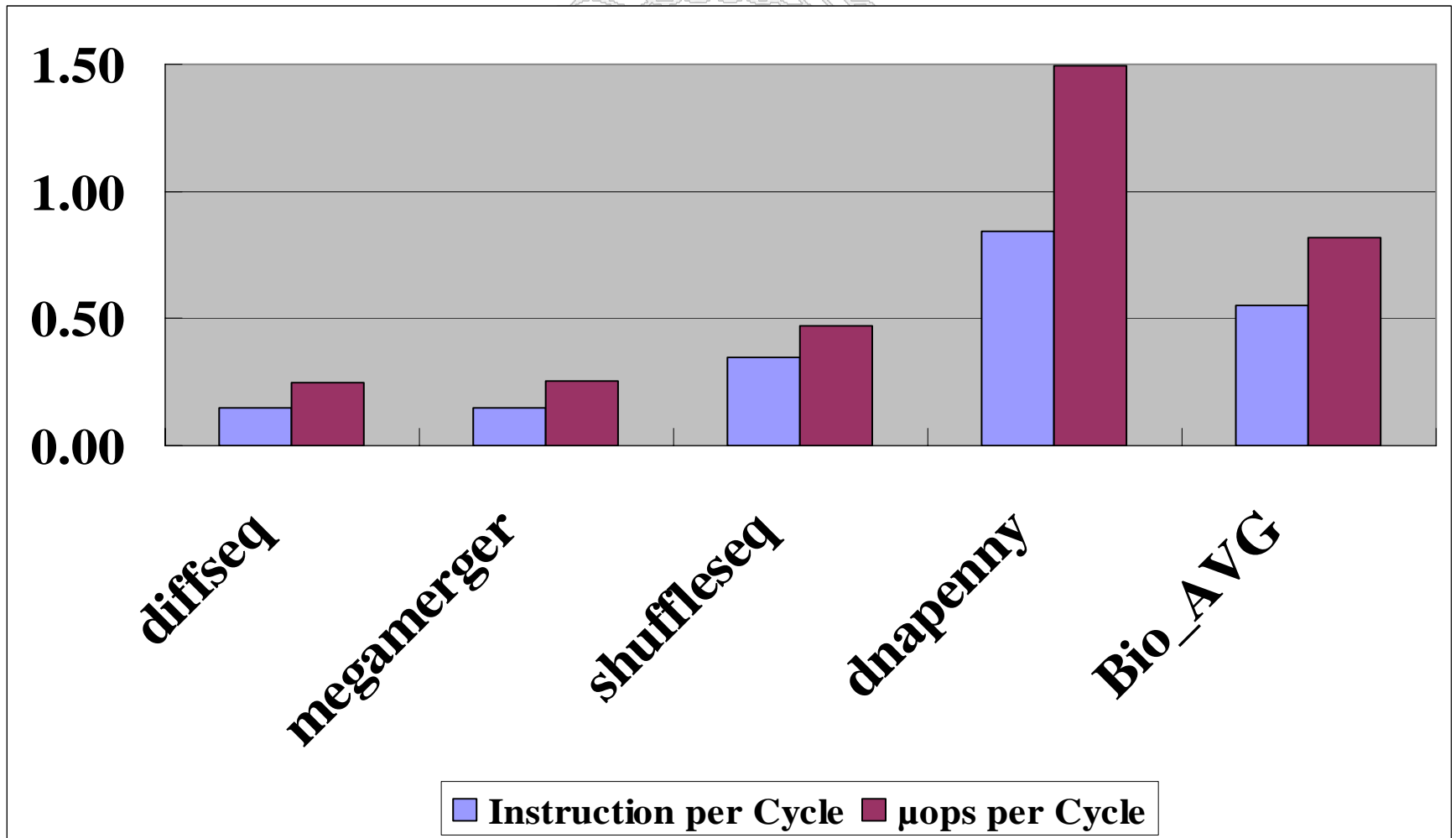


- Background
- Experimental Methodology
- A Characterization of Bioinformatics Benchmarks
- Conclusions

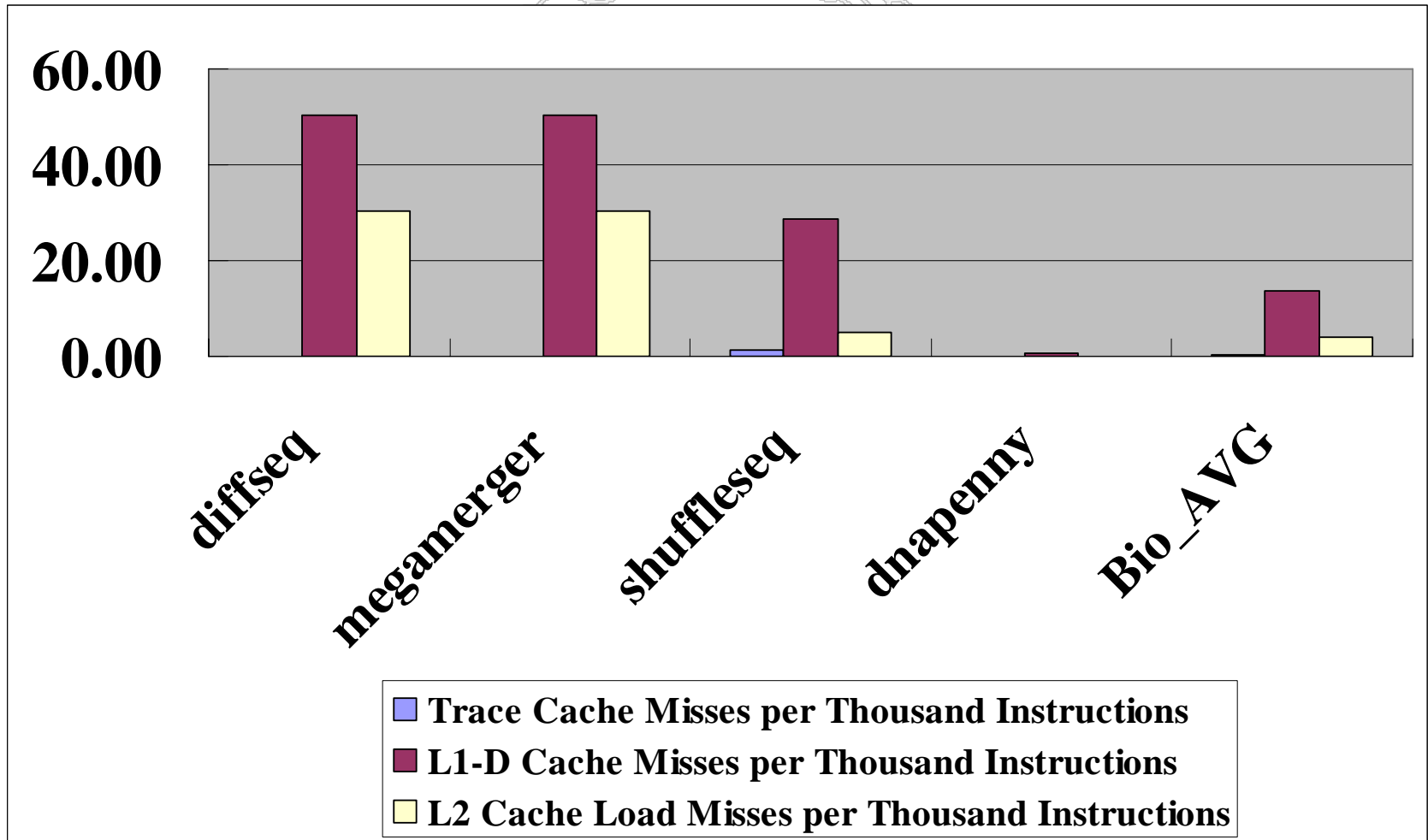
# Dynamic Instruction Profile



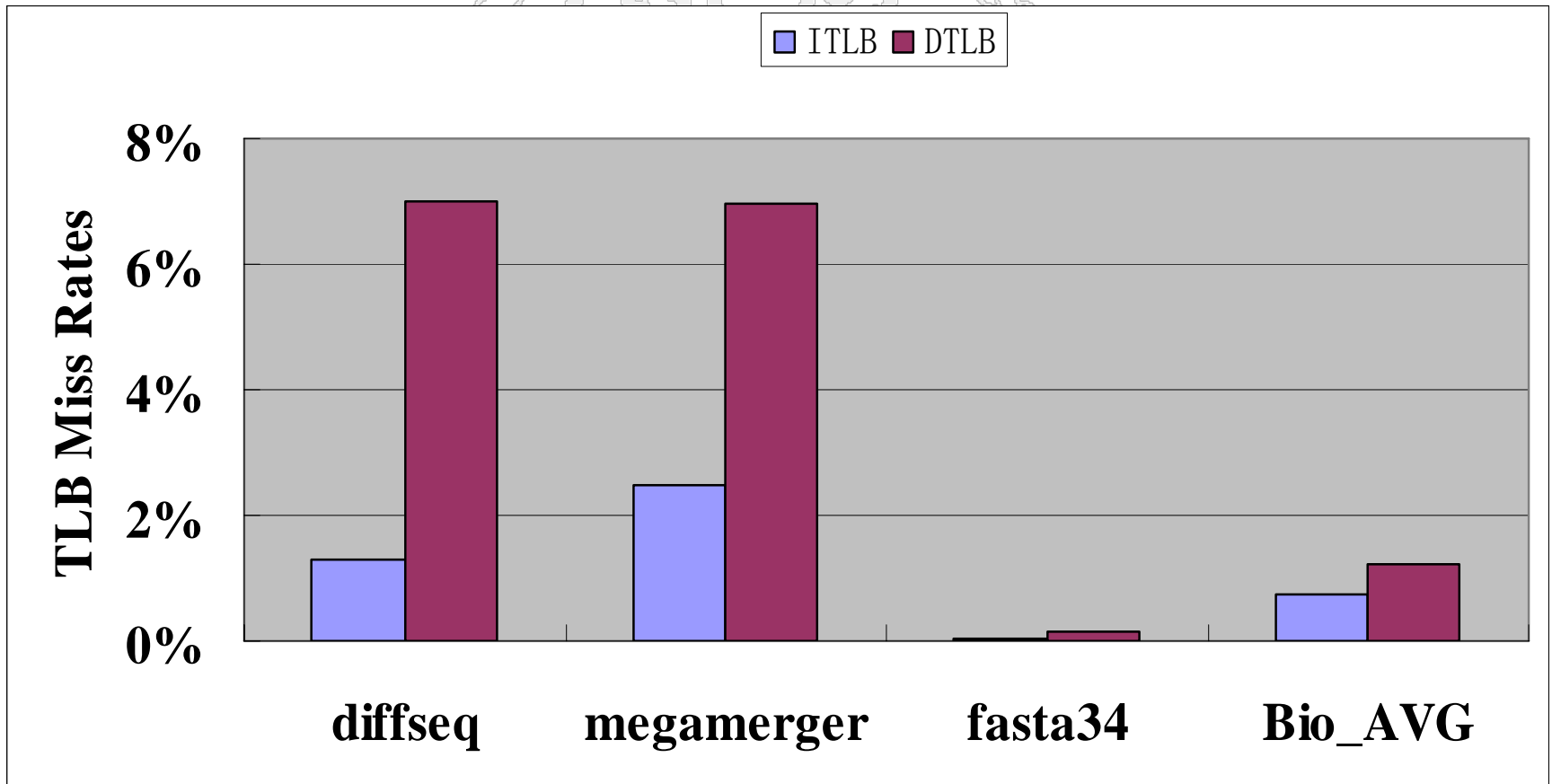
# IPC & $\mu$ PC



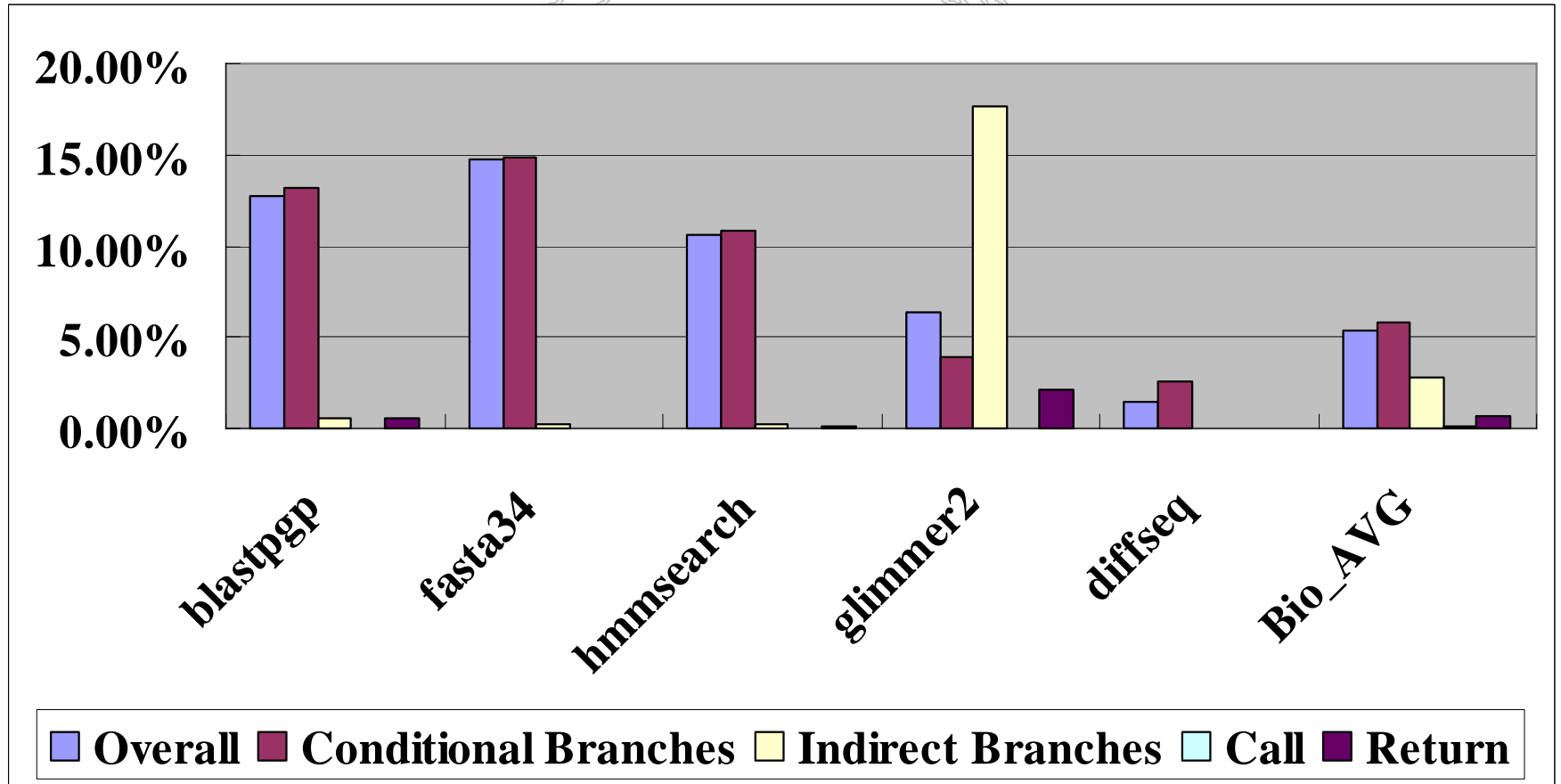
# Cache Misses per 1000 inst.



# TLB Miss Rates



# Branch Misprediction Rate



# A Comparison with SPEC Int 2K



	Bio_AVG	SPEC_AVG
Memory Reference	60.25%	59.78%
IPC	0.55	0.52
$\mu$ PC	0.82	0.79
L1-D Cache Misses	13.82	41.53
L2 Cache Load Misses	3.93	7.72
DTLB Misses*	1.23%*	6.60%
Overall Branch Mispred.	5.41%	4.64%

# Outline



- Background
- Experimental Methodology
- A Characterization of Bioinformatics Benchmarks
- Conclusions

# Conclusions



- Bioinformatics benchmarks are memory intensive workloads
- Prefetching and large data cache can satisfy the benchmarks efficiently
- Branch misprediction is another bottleneck

# The Website

---



The proposed BioInfoMark benchmark suite can be downloaded from:

<http://www.ideal.ece.ufl.edu/BioInfoMark>

