Online Linear Regression of Sampling Data

Qiang Wu, Oskar Mencer

Carlos Tavares

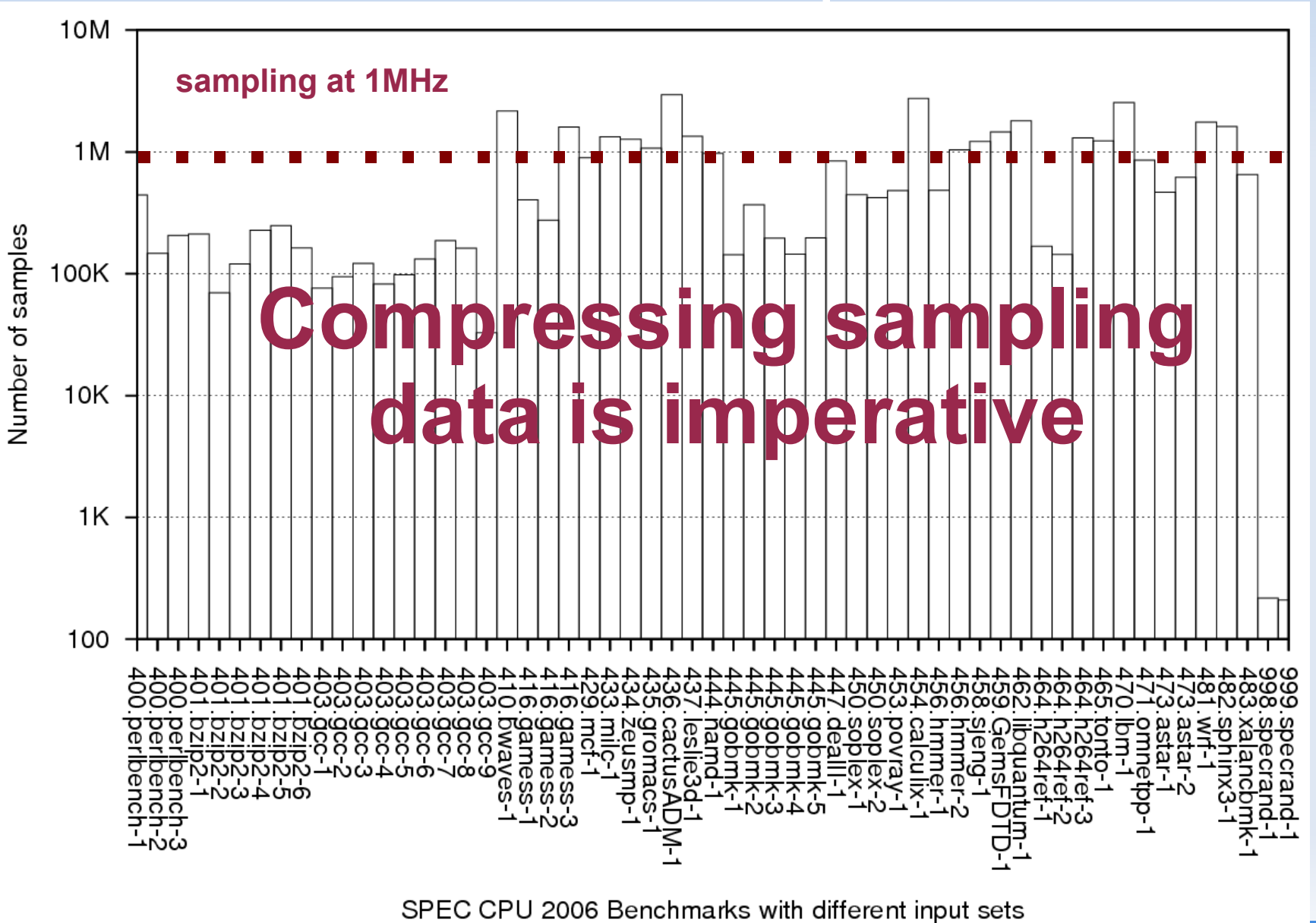
Imperial College London

Content

- Motivation
- Method
- Experiment
- Conclusion

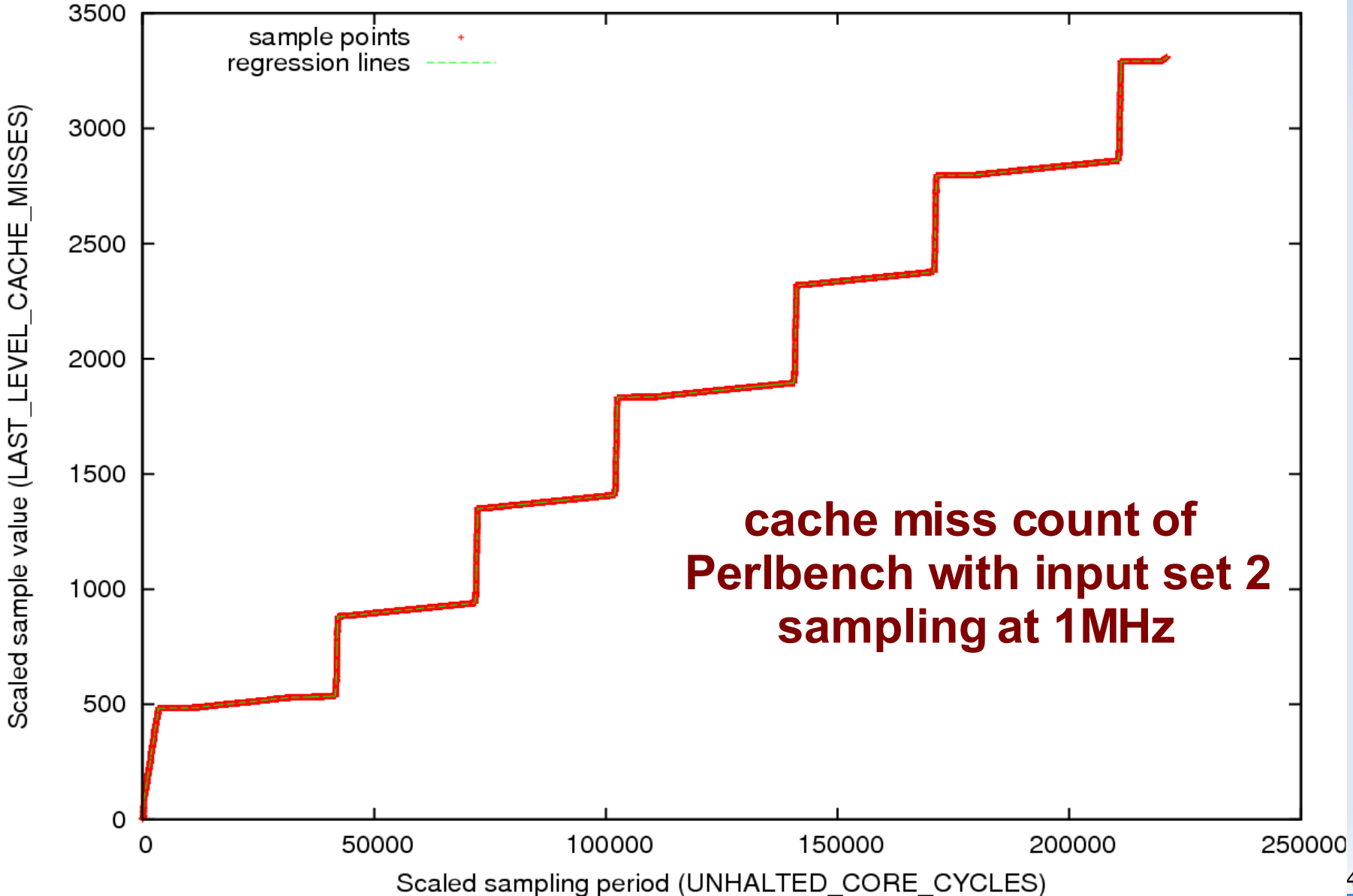
Motivation

Number of samples



Method

Capture the variation with lines?



Method
Online linear regression

- Online linear regression
  - *n* samples $(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)$
  - $y = k_n x + b_n$

$$k_n = \frac{nS_{xy(n)} - S_{x(n)}S_{y(n)}}{nS_{x^2(n)} - \left(S_{x(n)}\right)^2}$$

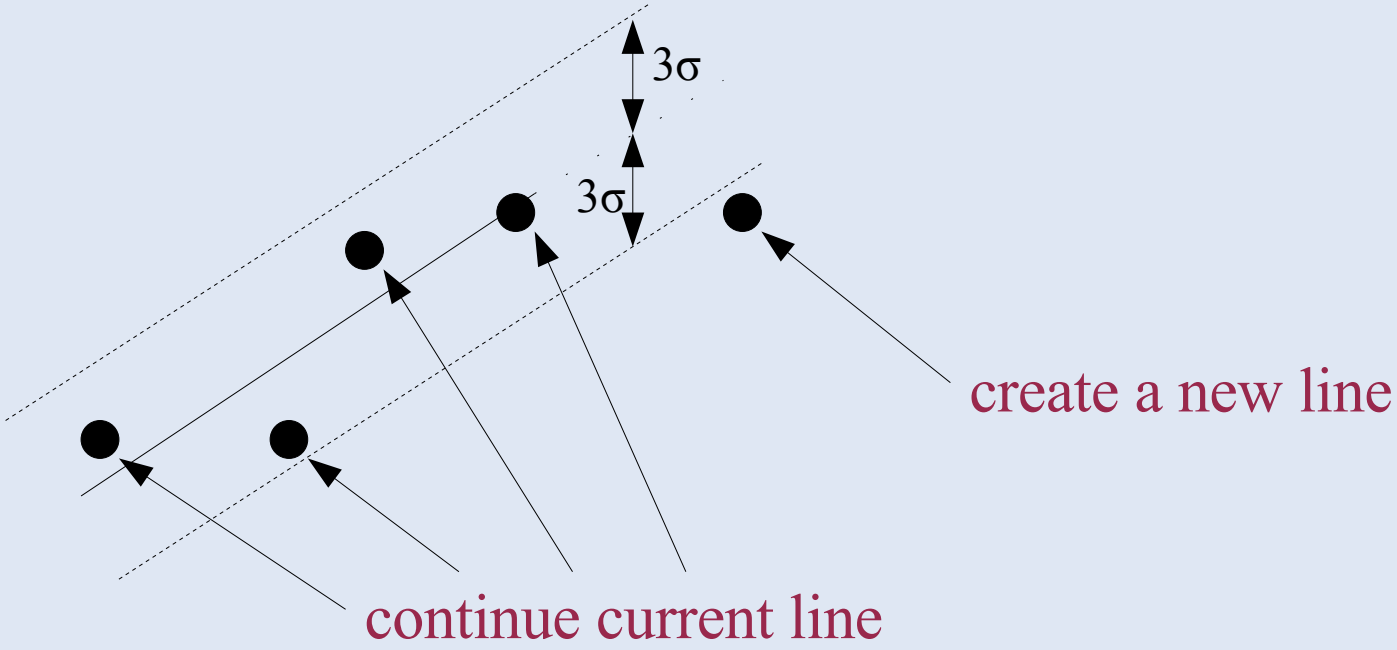
$$b_n = \frac{S_{y(n)}S_{x^2(n)} - S_{x(n)}S_{xy(n)}}{nS_{x^2(n)} - \left(S_{x(n)}\right)^2}$$

[1]

[1] W. Press, S. Teukuolsky, W. Vetterling, B. Flannery: Numerical Recipes in C, 2nd ed. Cambridge University Press, (1992)

Method
When to create a new line

- When to create a new line?
  - 3σ rule



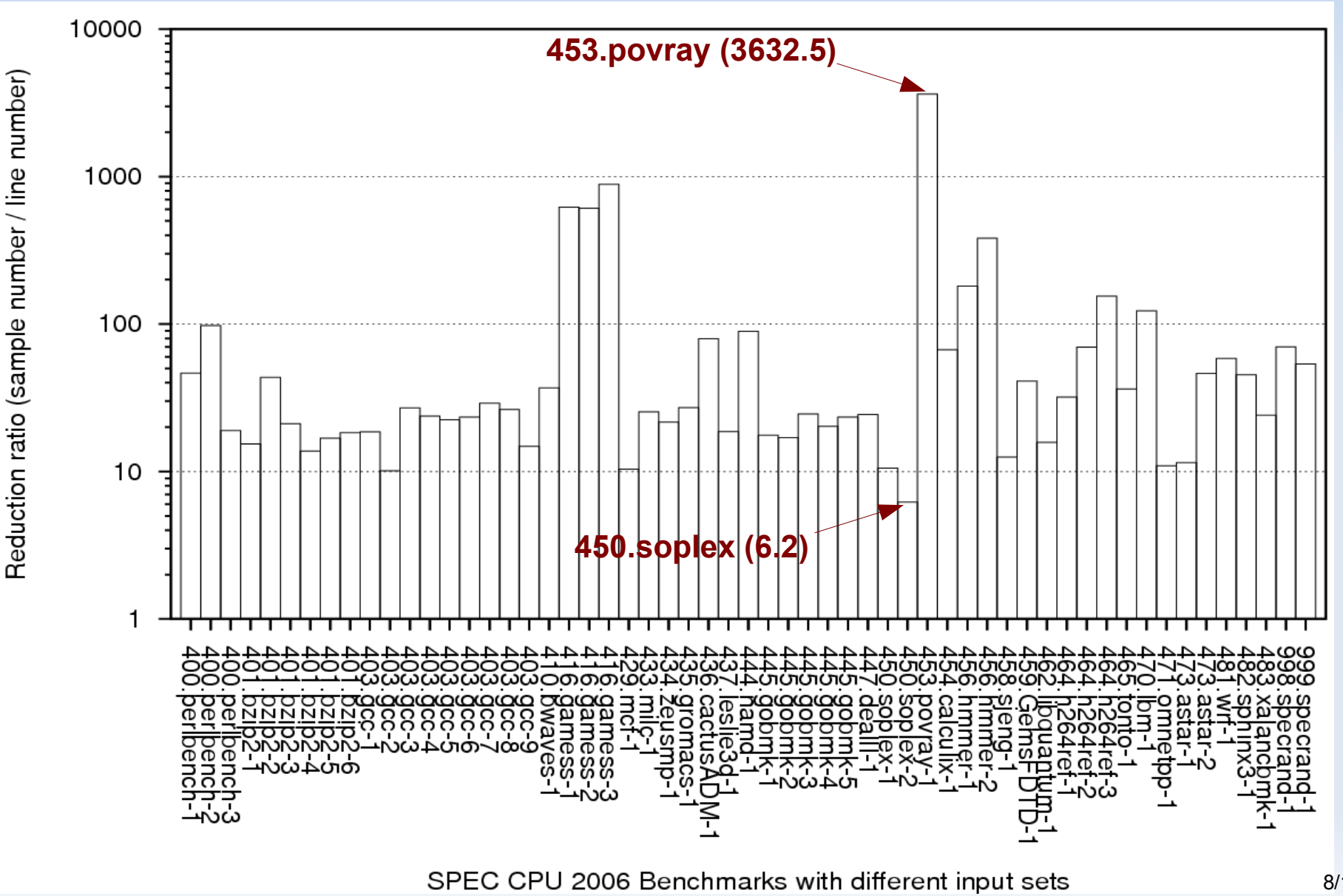
Experiment
Test platform

- Test machine
  - 2 quadcore Xeon @ 2GHz, 8 GB memory
  - Ubuntu Linux with *perfmon* [2]
- Test benchmarks
  - SPEC CPU 2006 [3] benchmarks
  - 57 test instances in total
- Test event
  - Last Level cache misses (Xeon L2 cache)

[2] Perfmon project. http://perfmon2.sourceforge.net/
[3] SPEC CPU 2006 benchmarks. http://www.spec.org/cpu2006/

Experiment

Data reduction ratio



Experiment

Regression error (MNESD)

- Maximal normalized estimated standard deviation (MNESD)

$$MNESD = \frac{\hat{\sigma}_{max}}{y_{max} - y_{min}}$$

Maximal estimated standard deviation in the sampling

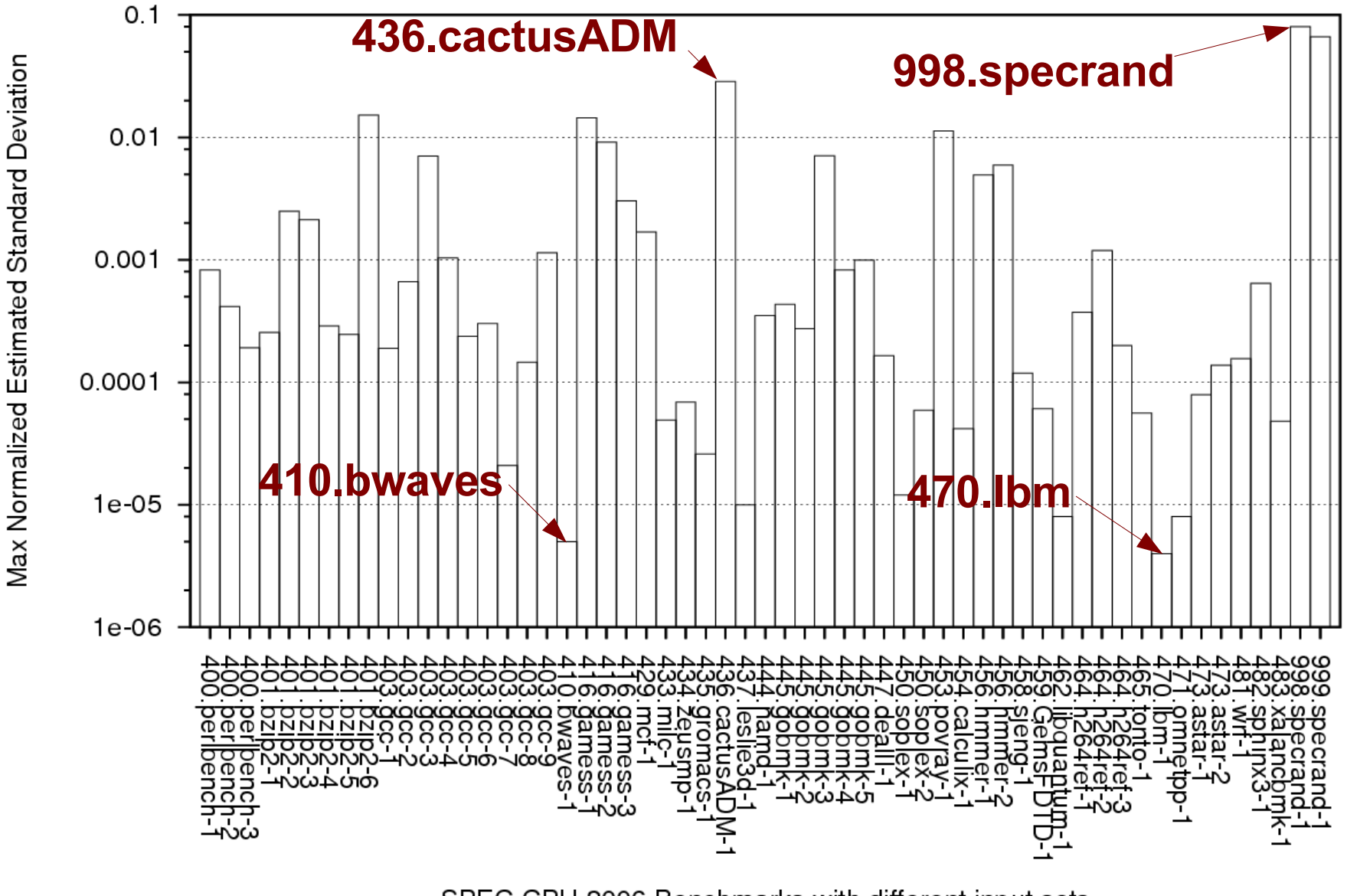
Maximal count value in the sampling

Minimal count value in the sampling

Normalization balances the differences among benchmarks

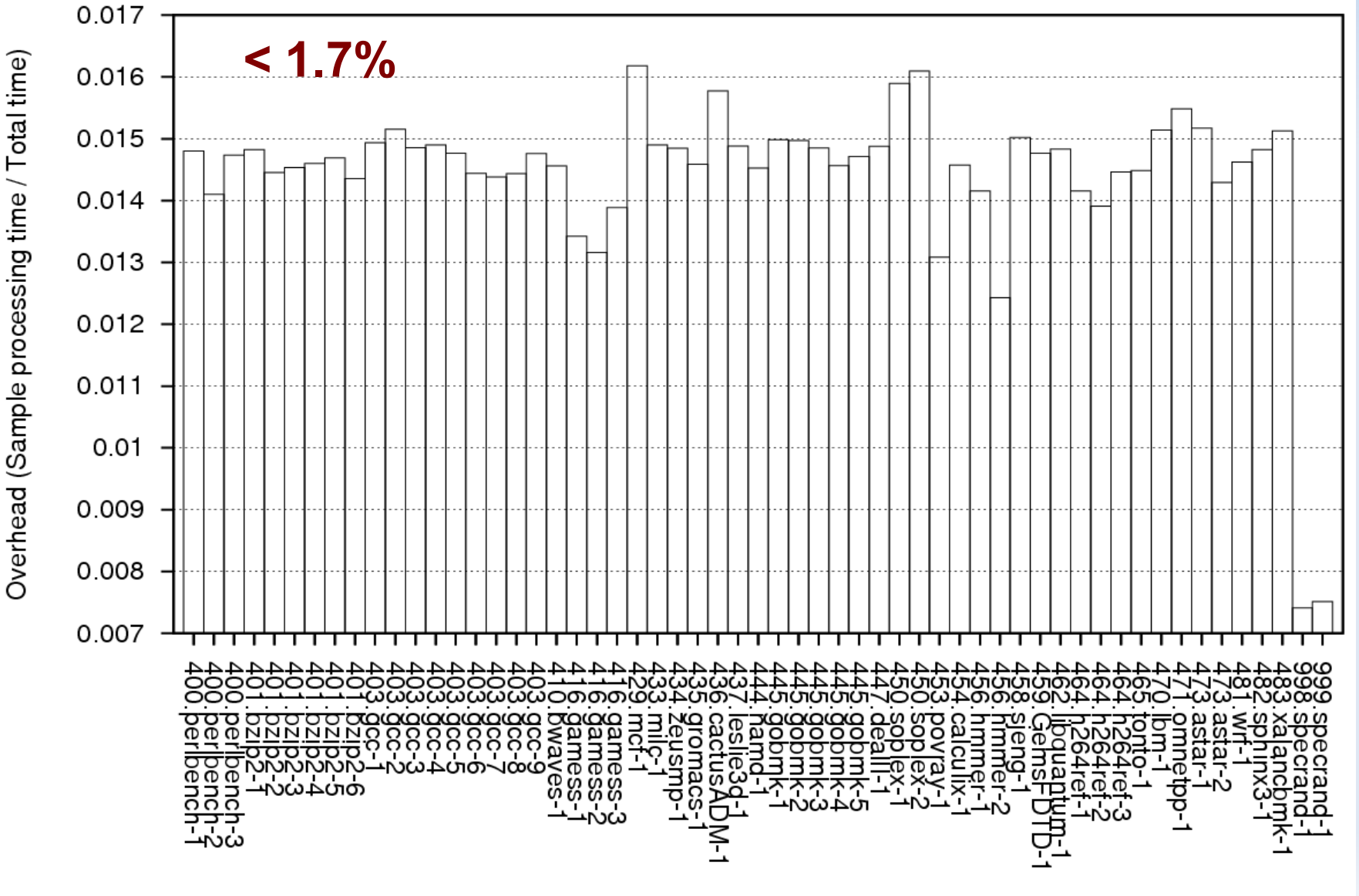
Experiment

Regression error (MNESD)



Experiment

Overhead



Conclusion

- Summary
  - Data reduction ratio > 10 typically
  - Regression error (MNESD) < 0.1
  - Overhead < 1.7%
- Tradeoff
  - More accuracy, less compression ratio
  - Picking the appropriate accuracy needed depends on the application