

# Repeatable, Reproducible, or Useful?

Amer Diwan and Robert Hundt  
Google

# Repeatable

- I conduct the experiment twice using the same setup and get the same results
- Why should we care?
  - If even I don't get consistent results from my experiment, then my experiment is doomed!
- Challenge: inter-run variation
  - Page mappings, interference with other jobs, ...

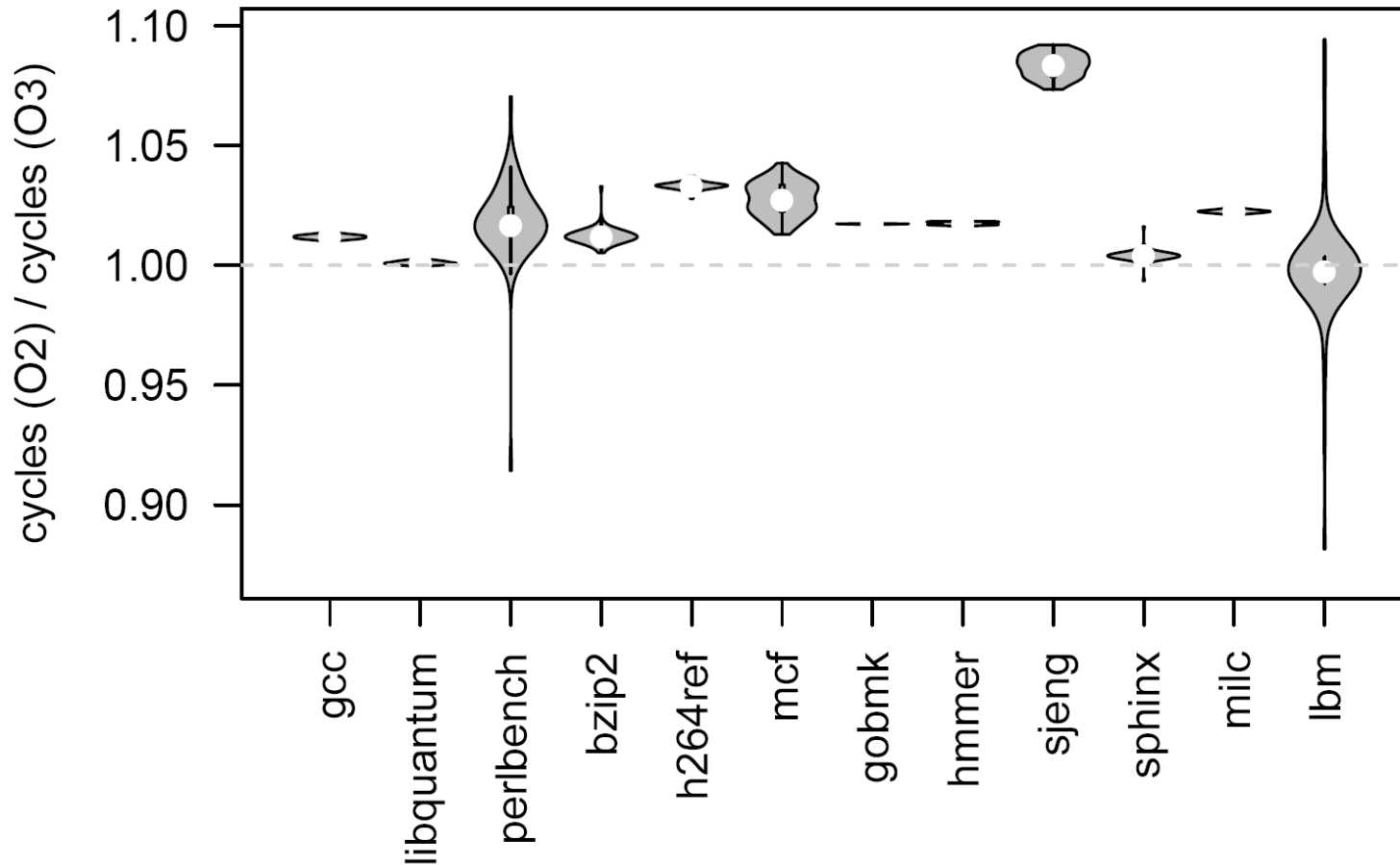
# What can we do?

- Repeat experiments as many times as needed to obtain tight confidence intervals
  - T-test, ...
- Report/record results with confidence intervals

# Reproducible

- My friend and I conduct the same experiment using the “same” setup and get the same results
- Why should we care?
  - If others cannot reproduce our experiments then are they actually correct?
- Challenge: bias

# Biases hiding under every rock...



The setting of irrelevant environment variables can lead to contradictory conclusions

# What can we do

- Account and control for all sources of bias
  - ... yeah, right!
- Account and control for all known sources of bias
  - Try to interactively discover sources of bias by repeatedly submitting to the archive

# Sources of bias

- Anything that affects memory layout
  - Environment variables, link order, heap size (Java), ...
- Benchmarks
  - What exactly does the benchmark test?
- Software and hardware components (e.g., microprocessors)
- etc.
- If we control for all sources of bias, we should get reproducible results

# Useful

- Real users should get results consistent with our experiments
- Why should we care?
  - If our results only apply to lab settings, then they are irrelevant!
- Challenge: “Controlling” bias is not a solution

# The problem with controlling bias

- Repeating an experiment with the “same” bias gives reproducible but not useful results
  - e.g., Every time anyone ask my wife she predicts the same winner for the election—this is repeatable but always has the same bias!
- Need randomized trials

# Randomized trials

- Randomly pick values for variables that cause bias
- Run an experiment
- Repeat

Use statistical methods to summarize the trials

# The vision for an archival system

Self-contained script for running experiment



Repeat every experiment multiple times and use t-test

Repeatable



Sources of bias



Control for known sources of bias  
(benchmarks, environment variables...)

Reproducible



Randomized trials for known sources of bias

Useful

