Statistical Error Bounds for Data Parallel Applications

Parker Hill, Michael Laurenzano, Babak Zamirai, Mehrzad Samadi, Scott Mahlke, Jason Mars, Lingjia Tang
Using Approximation

Exact
Binarize
Using Approximation

Exact Binarize
Using Approximation
Using Approximation

Exact Binarize

Approximation
Using Approximation

Approximate Binarize
Using Approximation

Approximate Binarize
Using Approximation (is hard)
Previous Work
Previous Work

- Calibration, profiling — coarse-grain, quality is loose
Previous Work

- Calibration, profiling — coarse-grain, quality is loose
- Formal reasoning — difficult for programmer
Previous Work

- Calibration, profiling — coarse-grain, quality is loose
- Formal reasoning — difficult for programmer
- Application specific solutions — limited scope
Dynamic Selection

• Determine approximation for each input, but:
  • Must be quick
  • Must be correct
Dynamic Selection

- Determine approximation for each input, but:
  - Must be quick
  - Must be correct
Dynamic Selection

• Determine approximation for each input, but:
  • Must be quick
  • Must be correct
Dynamic Selection

• Determine approximation for each input, but:
  • Must be quick
  • Must be correct
Dynamic Selection

• Determine approximation for each input, but:
  • Must be quick
  • Must be correct
Overview

• Statistical error bounds for data parallel applications:
  1. Randomly sample approximation error
  2. Build final error model from the error samples
  3. Build an error bound from the model
Error Samples

• Data parallel model
• Sample the output space → error samples
Error Samples

- Data parallel model
- Sample the output space $\rightarrow$ error samples
Error Samples

- Data parallel model
- Sample the output space → error samples
Error Samples

- Data parallel model
- Sample the output space → error samples

![Diagram of error samples](image)
Error Samples

- Data parallel model
- Sample the output space $\rightarrow$ error samples

![Diagram showing input, exact output, approximate output, and error samples with 4 correct and 2 incorrect samples.](image)
Error Model

- Create a statistical model from the error samples
- Represent final error in terms of component errors
- Bayesian inference → refine statistical model

![Diagram showing normalized PDFs for different sample sizes]
Error Bound

- Find the error bound from the statistical error model
- 90th percentile $\rightarrow$ 90% confidence (error $<$ bound)
Evaluation of Accuracy

- How accurate is this error bound in practice?
- Try multiple confidence levels, 800 images
- Tiling approximation*
- 1% of error space was sampled

*Samadi et al. ASPLOS 2014
Potential Speedup

• Assuming:
  • 8 of 64 approximations checked to find ideal
  • X% sampled overhead = X% of exact computation
  • Error target set to ≤10%
Conclusion

• Error can be statistically modeled:
  • Given the ability to sample the error space
  • Given some knowledge about the error space

• Can use statistical model to bound error

• Expected low enough overhead to compute per input