ASK: Adaptive Sampling Kit

P. de Oliveira Castro, E. Petit, JC. Beyler, W. Jalby

Université de Versailles St-Quentin-en-Yvelines
–
Exascale Computing Research

2012/08/29

Outline

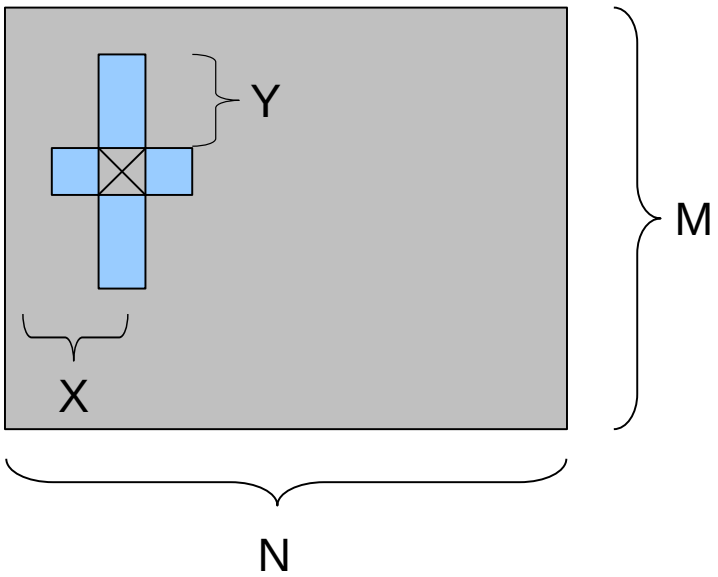
Motivation: Building Performance Models

- Building performance models is important to
  - Understand performance bottlenecks
  - Optimize applications
  - Find best architecture for a given application (co-design)

Motivation: Building Performance Models

- How to model performance ?
  - Using simulators or analytical models
    - ★ Architectures are complex and many factors interact (memory hierarchy, amount of parallelism, mapping, access patterns)
    - ★ Often models are too complex or costly
  - Black-box approach:
    - ★ Measure performance for different hardware or software configurations (the design space)
    - ★ Build an empirical model

Design Space example: Jacobi Stencil code



- T, number of OpenMP Threads, between 1 and 32
- N and M between 64 and 2048
- X,Y $\in \{1, 2, 4, 8, 16\}$
- Design space size around 31.10^8
- What is the performance for any combination of factors ?

Building empirical models

- Exhaustively measuring large design spaces is prohibitive.
- Build an accurate performance model with as few samples as possible

- Sampling method to select which points to measure
  - Samples must be chosen with care or the model will be biased.
- Regression model to estimate the missing samples
  - Linear, polynomial, SVM, Gaussian Process, Regression Trees, etc.

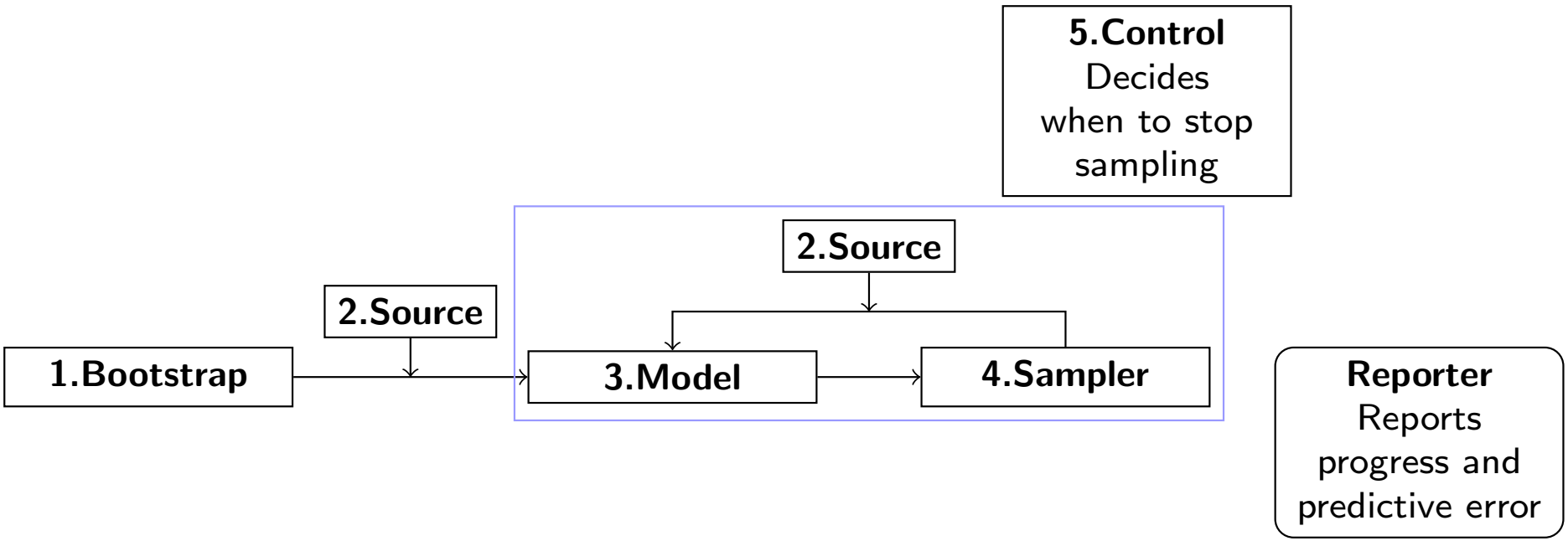
- No *one size fits all* strategy:
  - Depending on the design space response some models and sampling methods will work better than others
  - Important to try different strategies

Contributions

- The contributions of this work are:
  - ASK open-source toolkit to build empirical models
    - ★ Easy to try different sampling strategies
  - A novel sampling strategy HVS
    - ★ Evaluated on different performance characterization problems

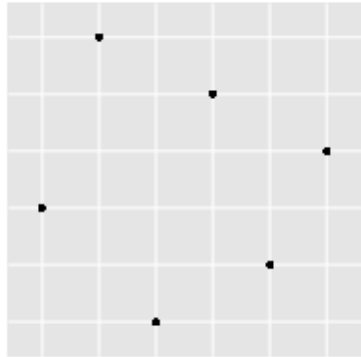
ASK: Adaptive Sampling Kit

- Adaptive Sampling Kit (ASK) is a toolkit for building empirical models
- Modular architecture for conducting experiments:
  - Easy to combine different sampling strategies and models
  - Gathers state-of-the art sampling methods
  - Provides visualization modules to supervise the sampling
  - Provides control modules to stop the sampling when its accurate enough

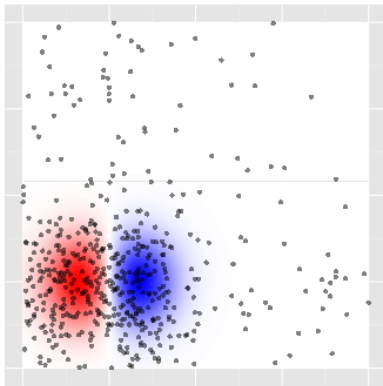


Sampling methods included in ASK

- Sampling methods fall in two main categories
- Static methods: Space Filling Designs
  - Select a set of samples covering the design space
  - All points are measured in a single batch
    - ★ Latin Hyper Cube
    - ★ Maximin
    - ★ Low discrepancy
    - ★ Random
- Adaptive methods:
  - Sampling iteratively adapts to the design space complexity
    - ★ AMART [Li09]: a Query-By-Comittee method
    - ★ TGP + ALC [Gramacy09]: an Error-reduction method
    - ★ HVS: a novel Error-reduction method that takes into account bias



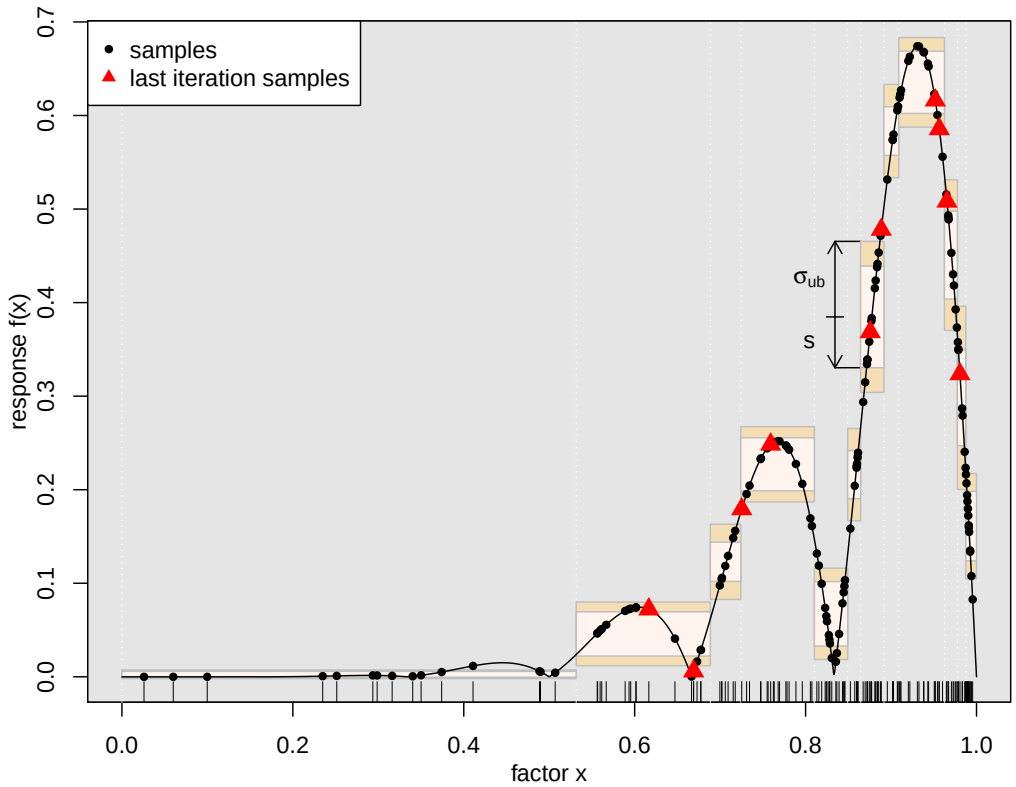
Latin Hyper Cube



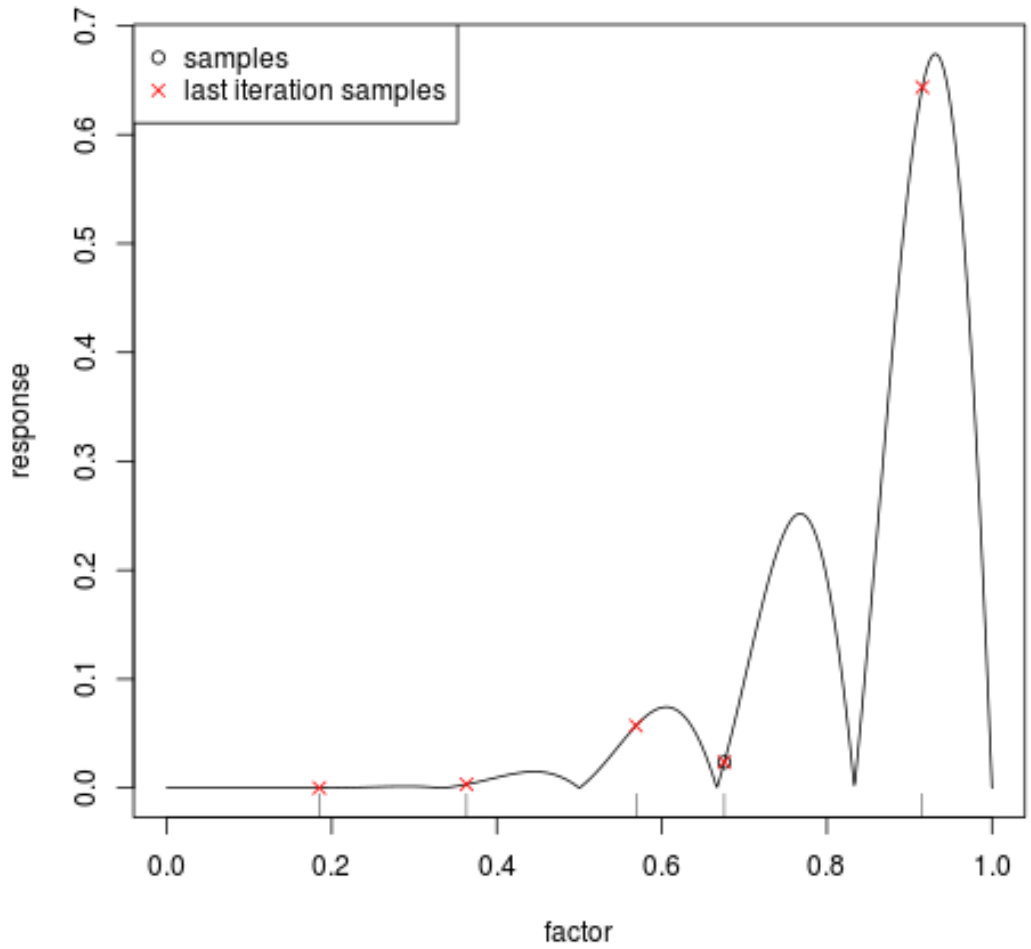
Adaptive Sampling

Hierarchical Variance Sampling (1/2)

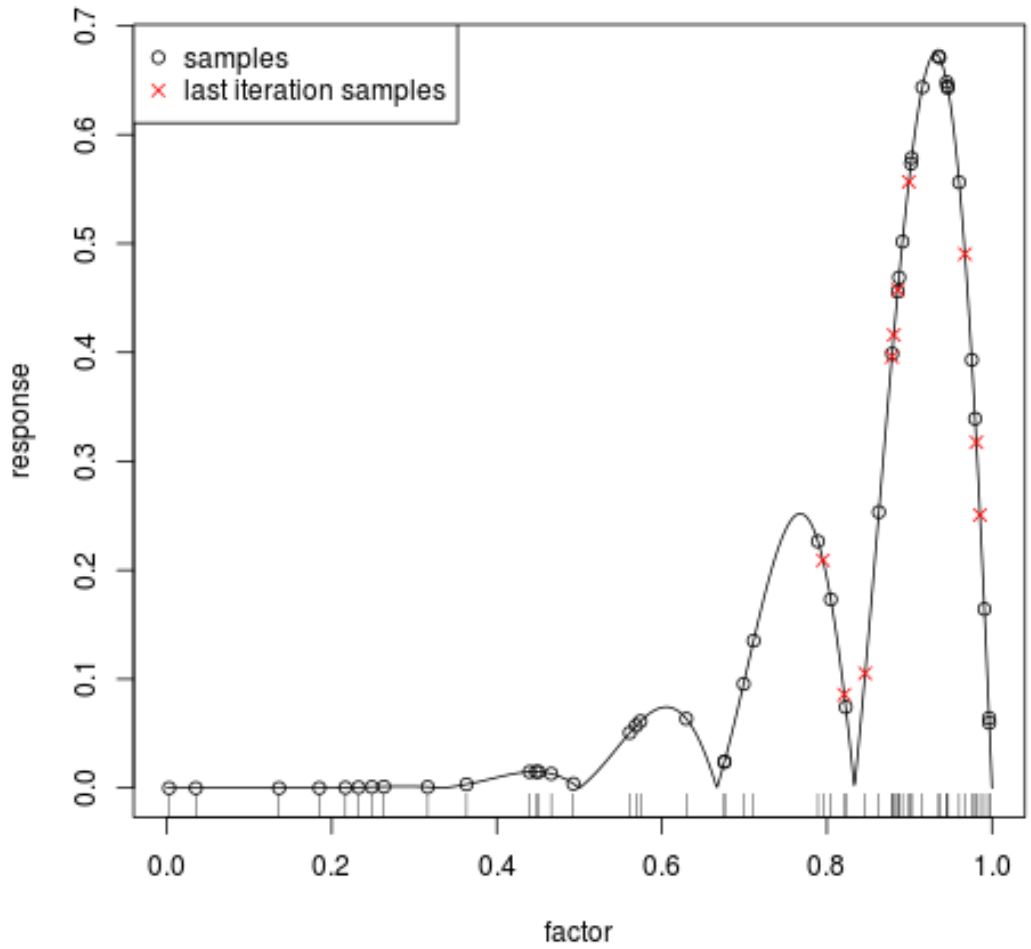
- Divide the space in regions using Regression Trees
- Compute the variance in each region
- Sample new points proportionally to: Variance upper bound $\times$ size of the region



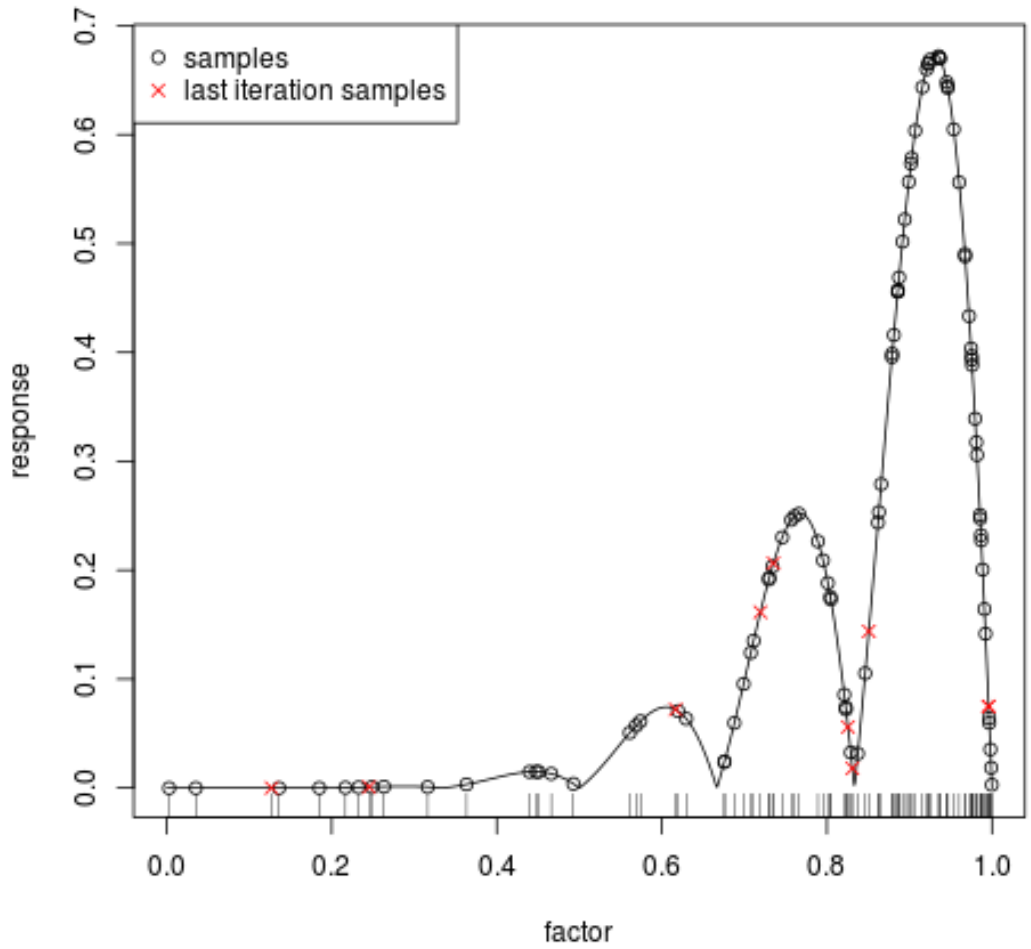
Hierarchical Variance Sampling (2/2)



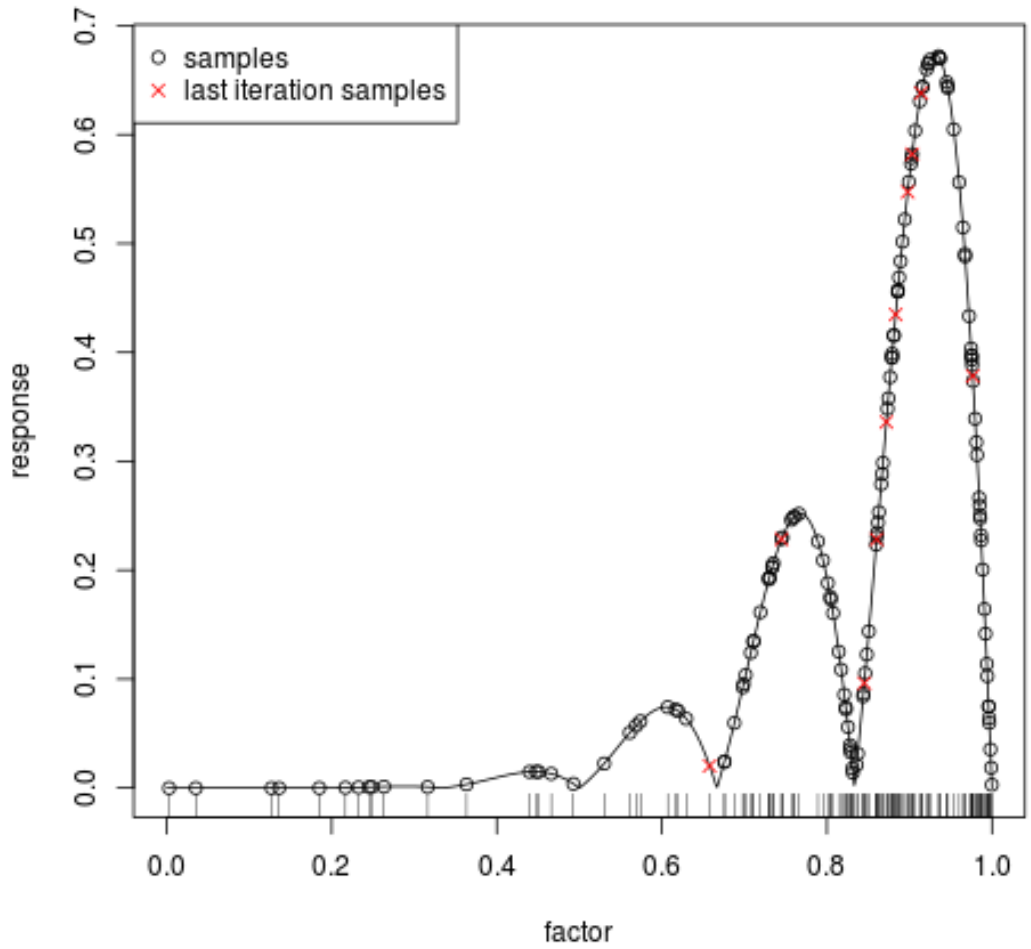
Hierarchical Variance Sampling (2/2)



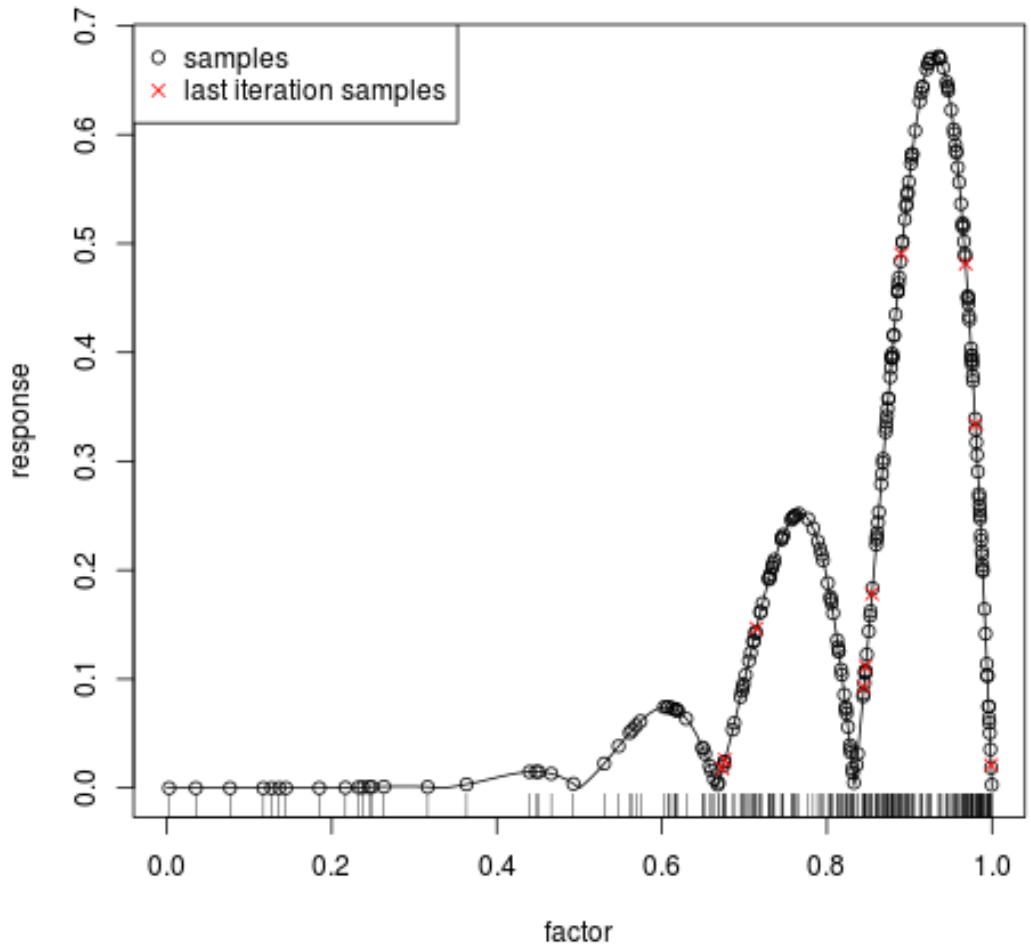
Hierarchical Variance Sampling (2/2)



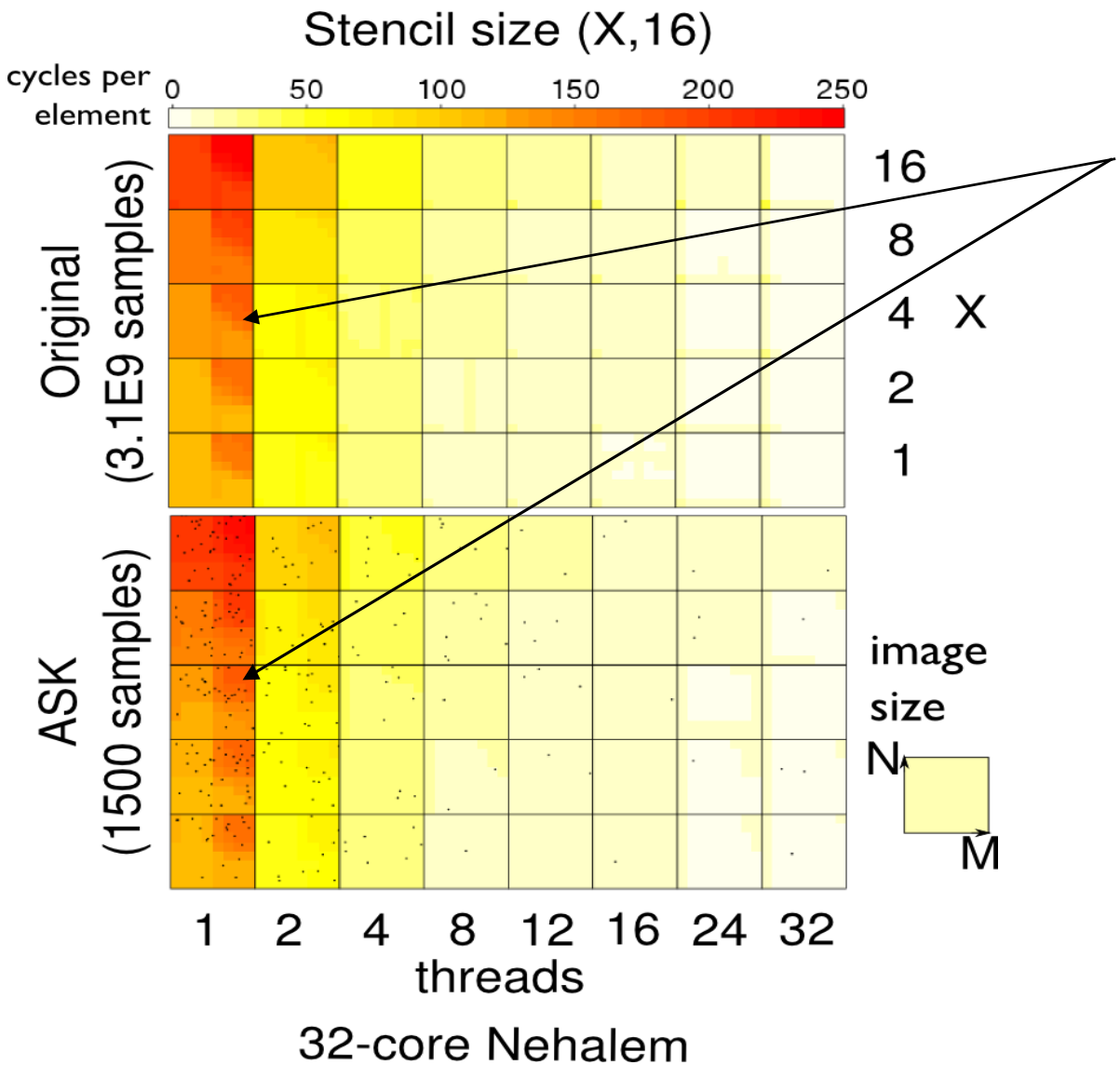
Hierarchical Variance Sampling (2/2)



Hierarchical Variance Sampling (2/2)



ASK: Stencil code evaluation



- Despite using only 1500 points, HVS+GBM captures the performance features of the application.
- (25600 samples used as original response test set)
- 32 cores Xeon X7550 2.00GHz

ASK: Evaluating estimation error

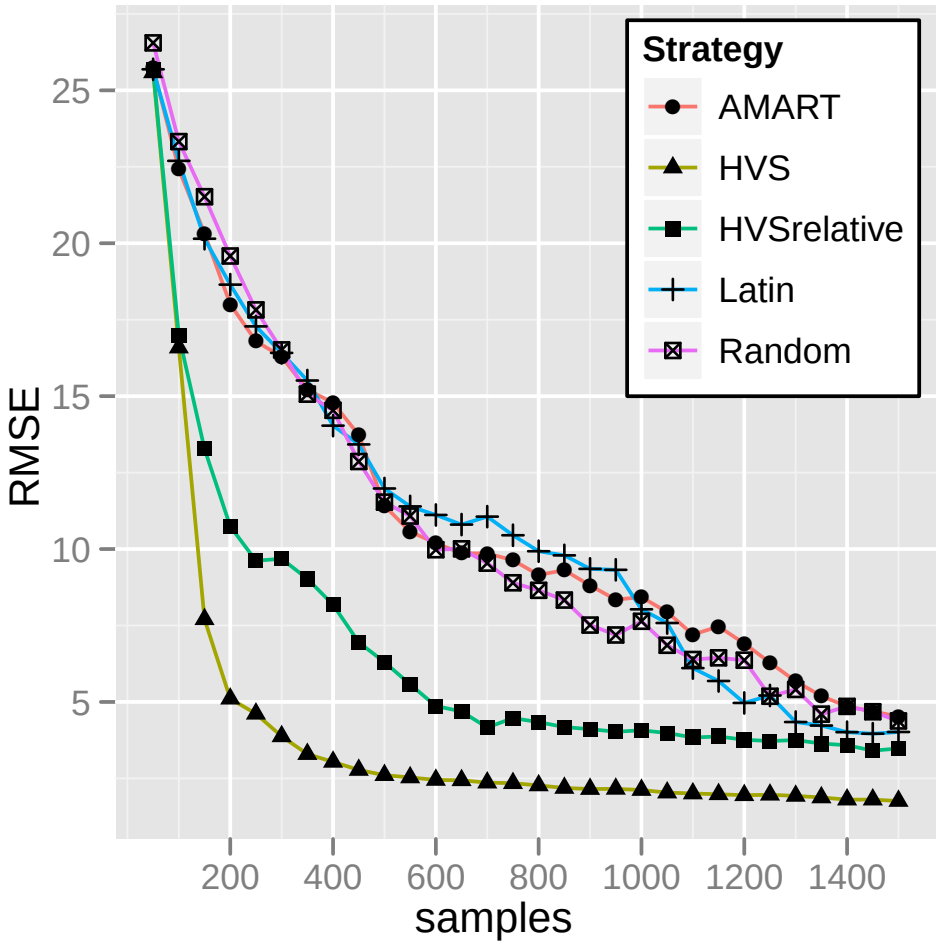


Figure: Stencils, Root Mean Square Error for different ASK sampling strategies

Using the Model for prediction

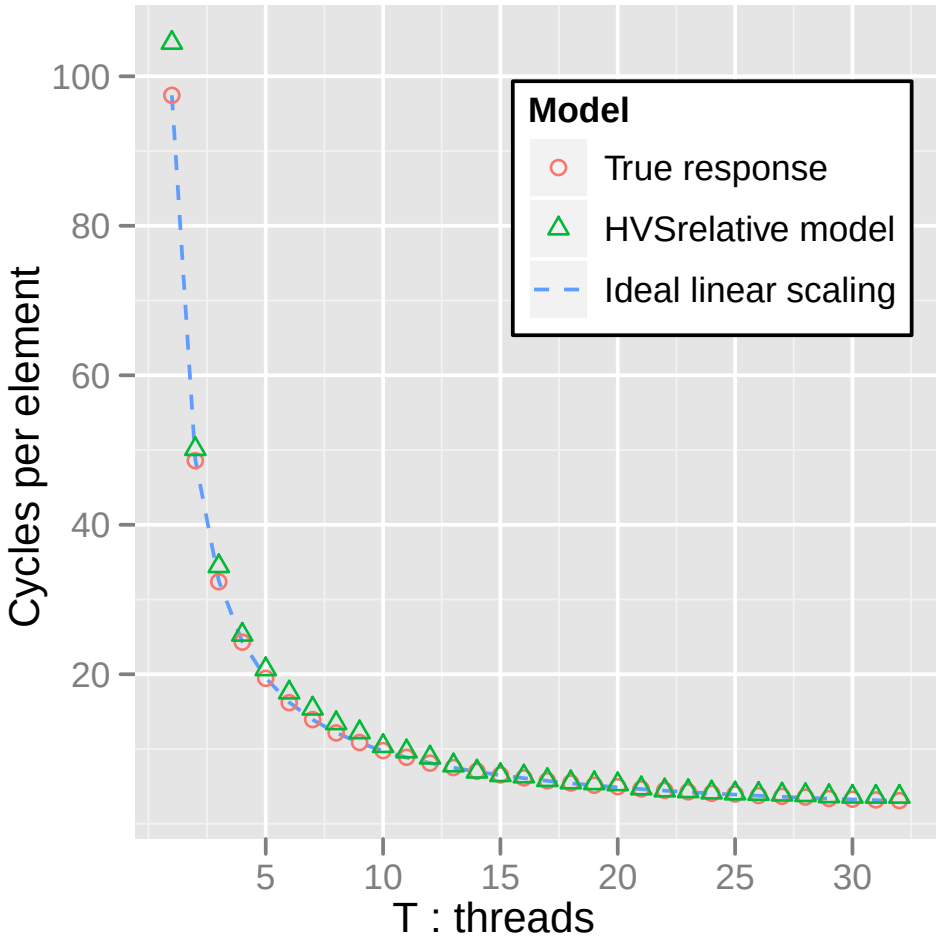
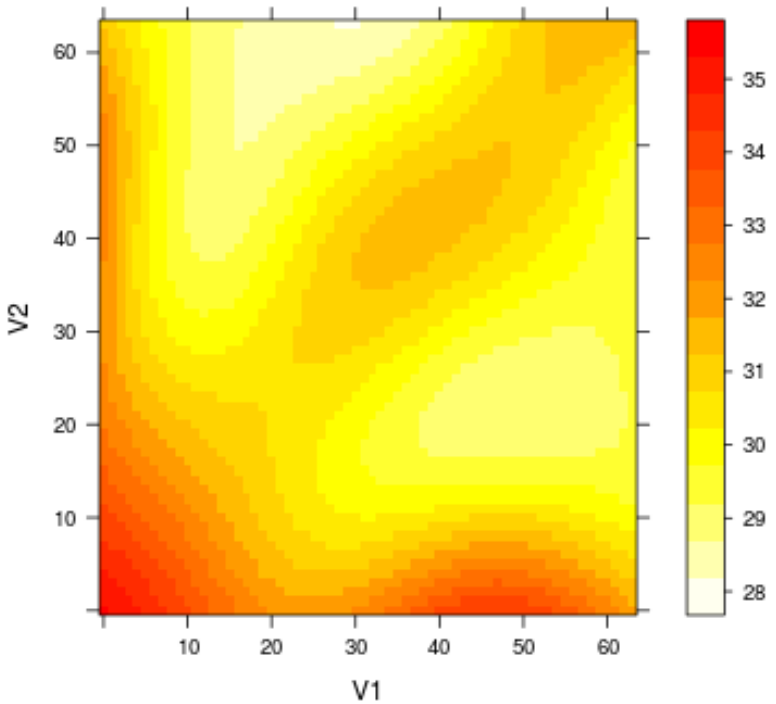


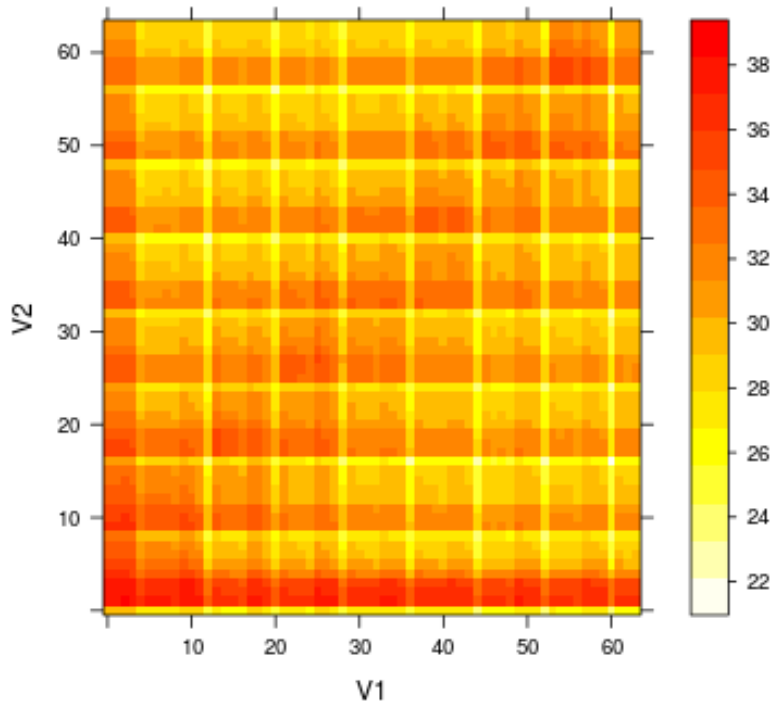
Figure: Scalability of the 8x8 stencil on a 1000x1000 matrix

Importance of selecting a good model

- Influence of alignment stream benchmark
  - Three store streams hitting memory
  - Memory offsets: $S(k)$, $S(V1+k)$, $S(V2+k)$
  - 4K aliasing
  - non aligned access overhead



SVM model



GBM model

Alternatives to ASK

- SUrrogate MOdeling Lab (SUMO) [Gorissen2010]
  - Mature toolbox
  - Includes many models and sampling methods
  - Automatic tuning of model parameters
  - Supports modeling of multiple responses
  - ASK specifically targets performance characterization
    - ⋆ AMART [Li09] and HVS methods have been evaluated on performance problems
  - Only supports real-valued inputs
  - Depends on Matlab and is not open-source (but freely available for academic use)
- Caret R package [Kuhn2012]
  - Includes many models
  - Automatic tuning of model parameters
  - Does not handle sampling

How to get ASK ?

- ASK is open-source and available at
  - `http://code.google.com/p/adaptive-sampling-kit/`

- The experimental data used in the paper is available at
  - `http://code.google.com/p/adaptive-sampling-kit/wiki/ExperimentalData`