The price performance of performance models

Felix Wolf, Technical University of Darmstadt

ModSim 2021

Photo: Alex Becker / TU Darmstadt

Acknowledgement

TU Darmstadt

- Alexandru Calotoiu
- Alexander Geiß
- Alexander Graf
- Daniel Lorenz
- Benedikt Naumann
- Thorsten Reimann
- Marcus Ritter
- Sergei Shudler

ETH Zurich

- Alexandru Calotoiu
- Marcin Copik
- Tobias Grosser
- Torsten Hoefler
- Nicolas Wicki

LLNL

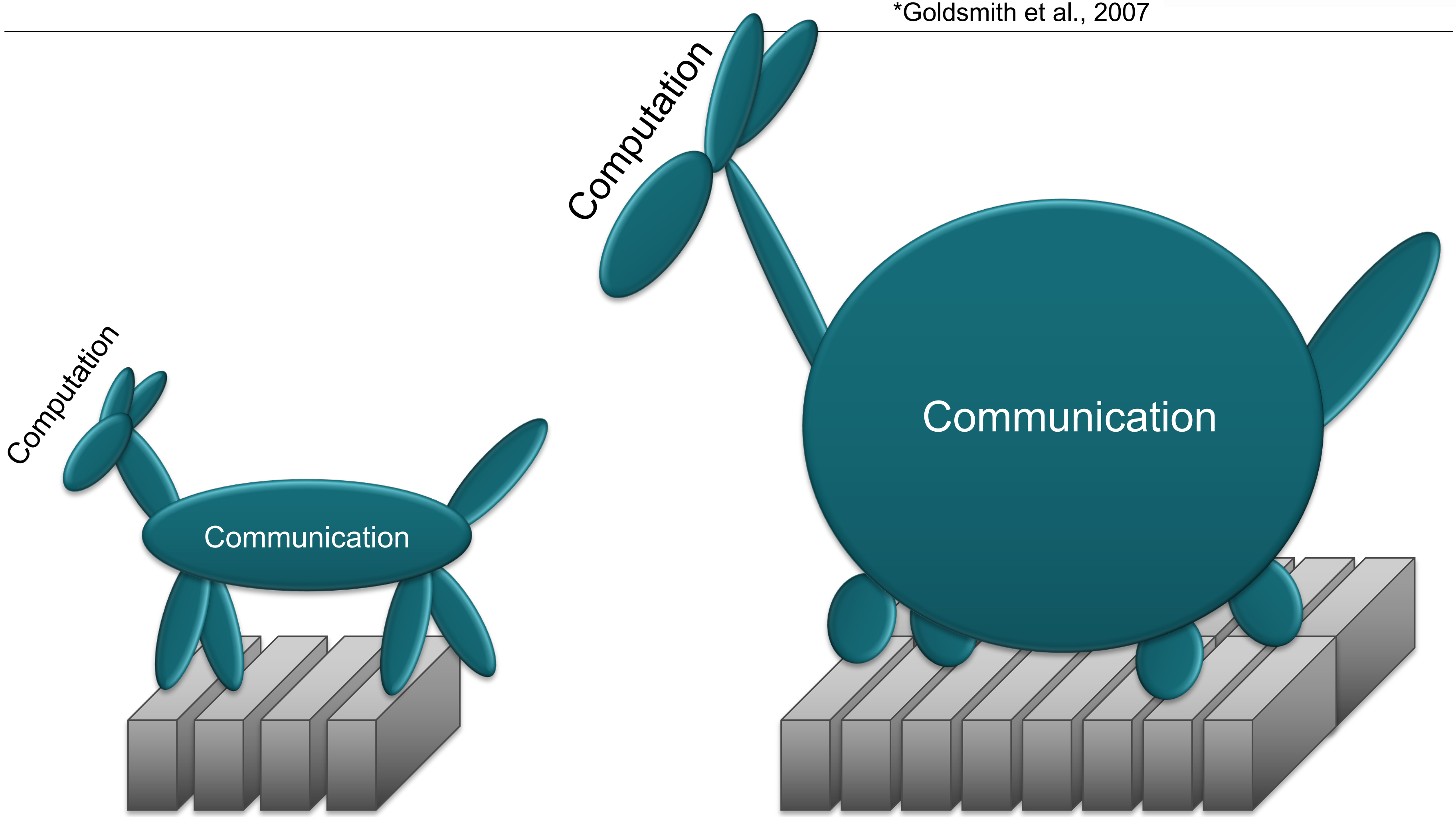
- David Beckingsale
- Christopher Earl
- Ian Karlin
- Martin Schulz

Bundesministerium für Bildung und Forschung

DFG FNSNF

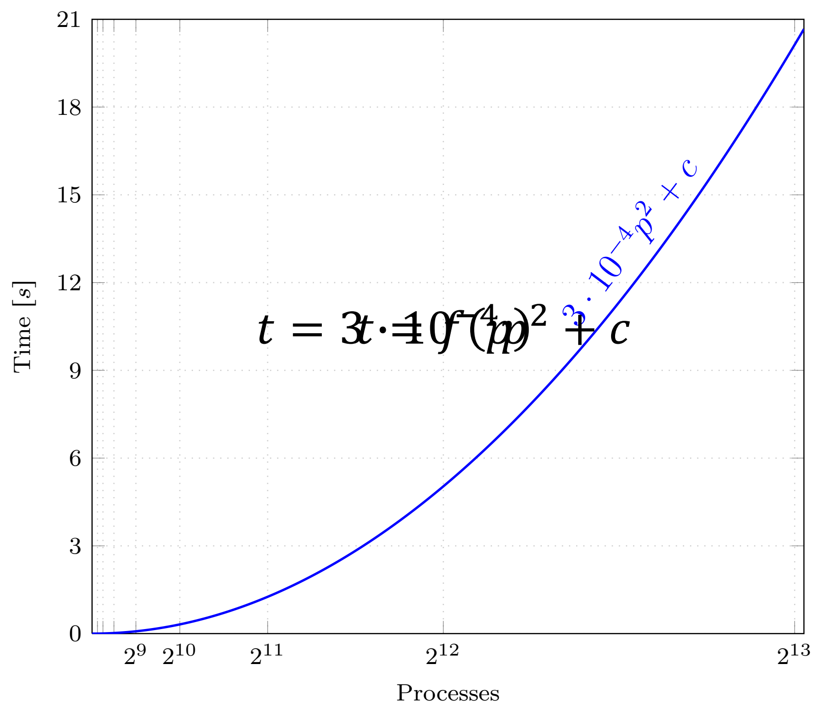
Scaling your code can harbor *performance surprises*[*]…

*Goldsmith et al., 2007



Performance model

Formula that expresses a relevant performance metric as a function of one or more execution parameters

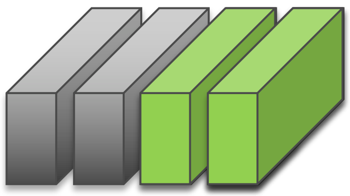


Analytical (i.e., manual) creation challenging for entire programs

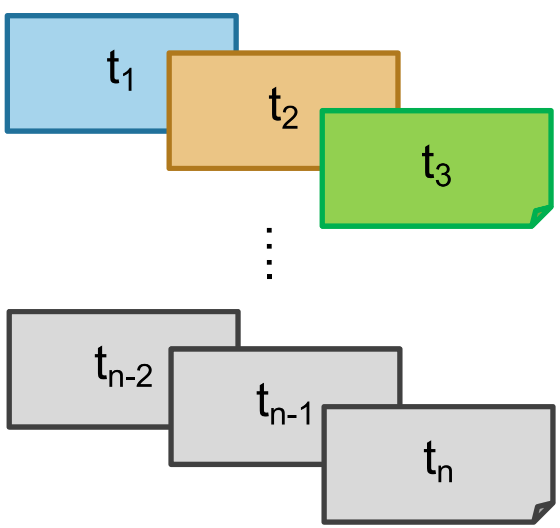
Identify kernels
- Incomplete coverage

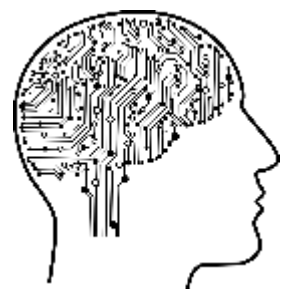
Create models
- Laborious, difficult

Empirical performance modeling



Performance measurements with different execution parameters $\mathbf{x_1,\ldots,x_n}$





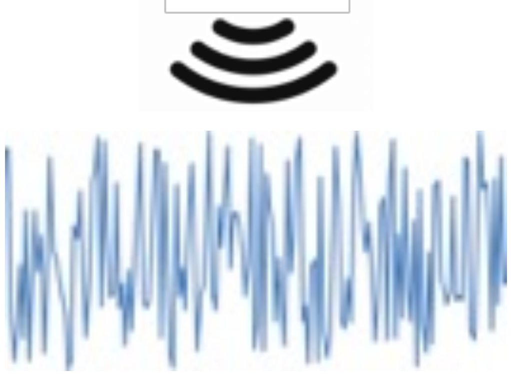
Machine learning

$$t = f(x_1, \ldots, x_n)$$

Alternative metrics: FLOPs, data volume…

Challenges

Applications



Run-to-run variation / noise

System

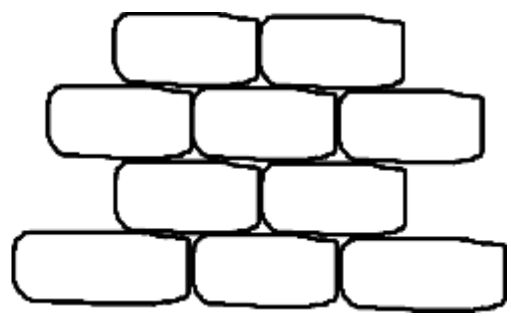
Cost of the required experiments

How to deal with noisy data

- Introduce **prior** into learning process
  - Assumption about the probability distribution generating the data

Time

~



Effort

- Computation
- Memory access
- Communication
- I/O

Typical algorithmic complexities in HPC

Computation	Communication
LU $t(p) \sim c$	LU $t(p) \sim c$
FFT $t(p) \sim \log_2(p)$	FFT $t(p) \sim c$
Naïve N-body $t(p) \sim p$	Naïve N-body $t(p) \sim p$
…	…
Samplesort $t(p) \sim p^2 \log_2^2(p)$	Samplesort $t(p) \sim p^2$

Performance model normal form (PMNF)

$$f(x) = \sum_{k=1}^{n} c_k \cdot p^{i_k} \cdot log_2^{j_k}(x)$$

Single parameter [Calotoiu et al., SC13]

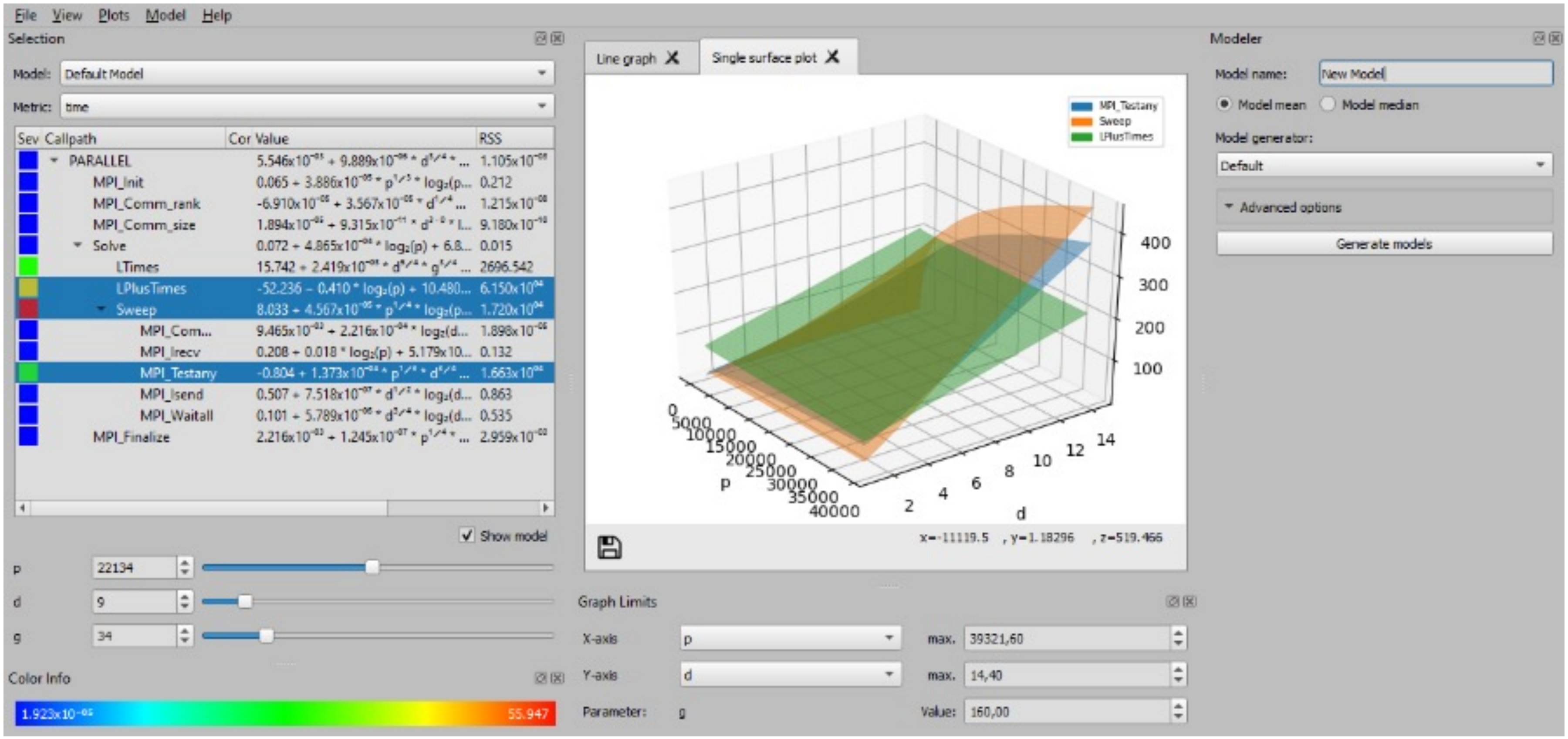
Parameter selection → Search space configuration → Linear regression + cross-validation → Quality assurance

$$f(x_1, \ldots, x_m) = \sum_{k=1}^{n} c_k \prod_{l=1}^{m} x_l^{i_{kl}} \cdot log_2^{j_{kl}}(x_l)$$

Multiple parameters [Calotoiu et al., Cluster’16]

Heuristics to reduce search space

Extra-P 4.0



Available at: https://github.com/extra-p/extrap

MPI implementations

[Shudler et al., IEEE TPDS 2019]

Platform	Juqueen	Juropa	Piz Daint
Allreduce [s]			**Expectation: *O* (log *p*)**
Model	$O(\log p)$	$O(p^{0.5})$	$O(p^{0.67} \log p)$
R^2	0.87	0.99	0.99
Match	✓	~	✗!
Comm_dup [B]			**Expectation: *O* (1)**
Model	2.2e5	256	$3770 + 18p$
R^2	1	1	0.99
Match	✓	✓	✗

Kripke - example w/ multiple parameters

TECHNISCHE UNIVERSITÄT DARMSTADT

SweepSolver

Main **computation** kernel

Expectation – Performance depends on **problem size**

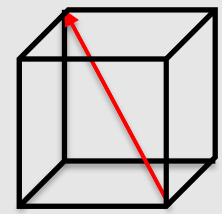
$$t \sim d \cdot g$$

Actual model:

$$t = 5 + d \cdot g + \underline{0.005 \cdot \sqrt[3]{p} \cdot d \cdot g}$$

MPI_Testany

Main **communication** kernel: 3D wave-front communication pattern



Expectation – Performance depends on **cubic root of process count**

$$t \sim \sqrt[3]{p}$$

Actual model:

$$t = 7 + \sqrt[3]{p} + \underline{0.005 \cdot \sqrt[3]{p} \cdot d \cdot g}$$

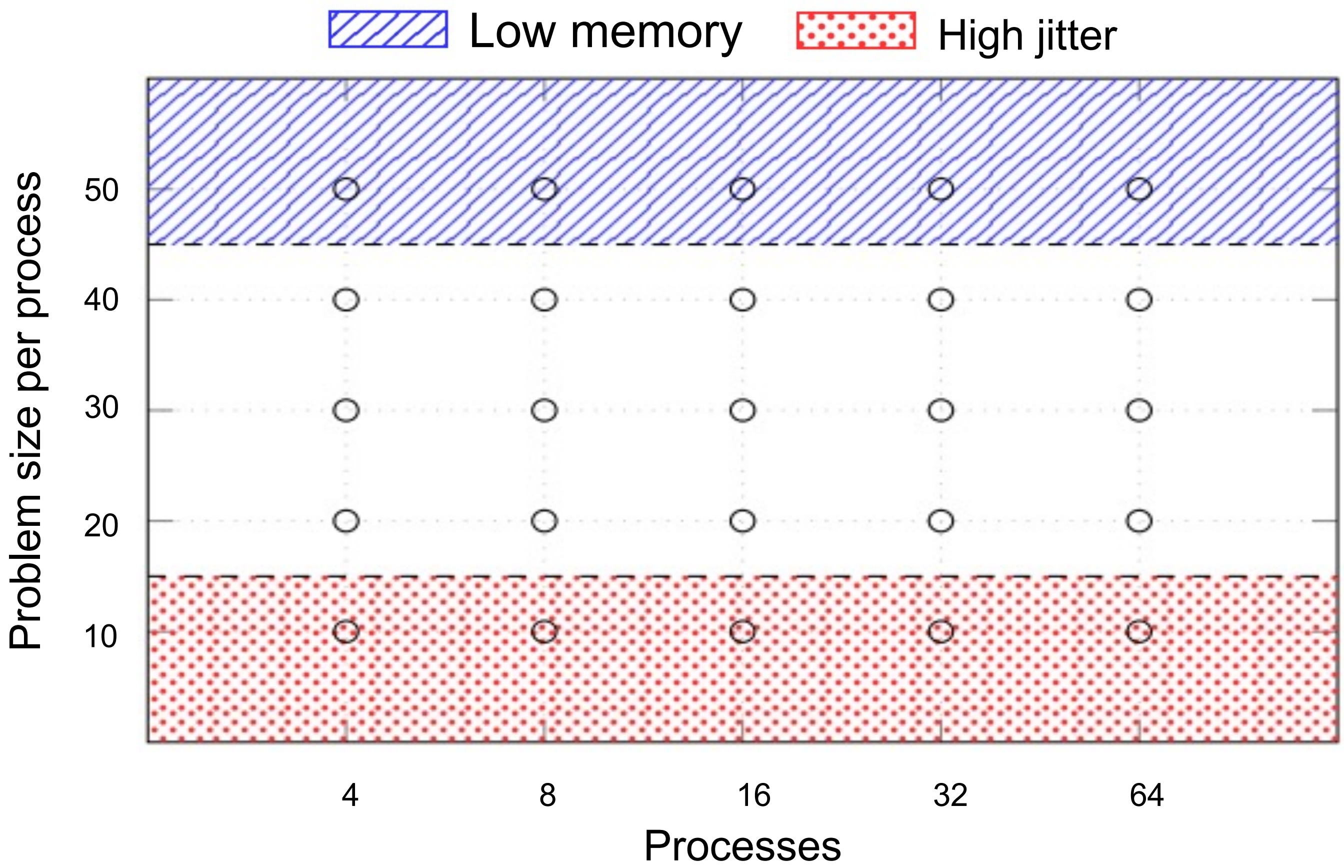
Kernels must wait on each other

Smaller compounded effect discovered

*Coefficients have been rounded for convenience

Experiments can be expensive

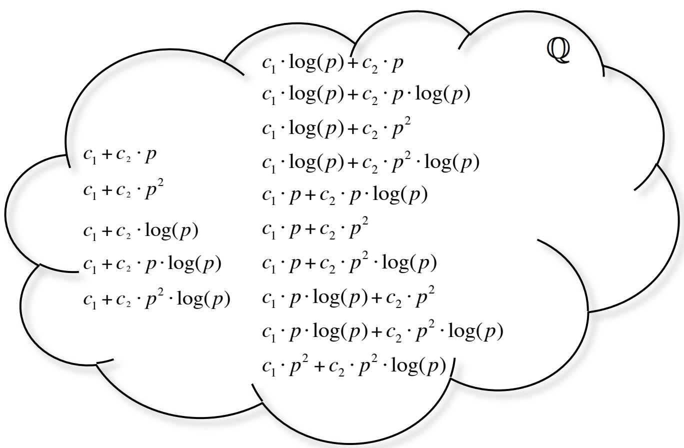
Need   experiments, = #parameters



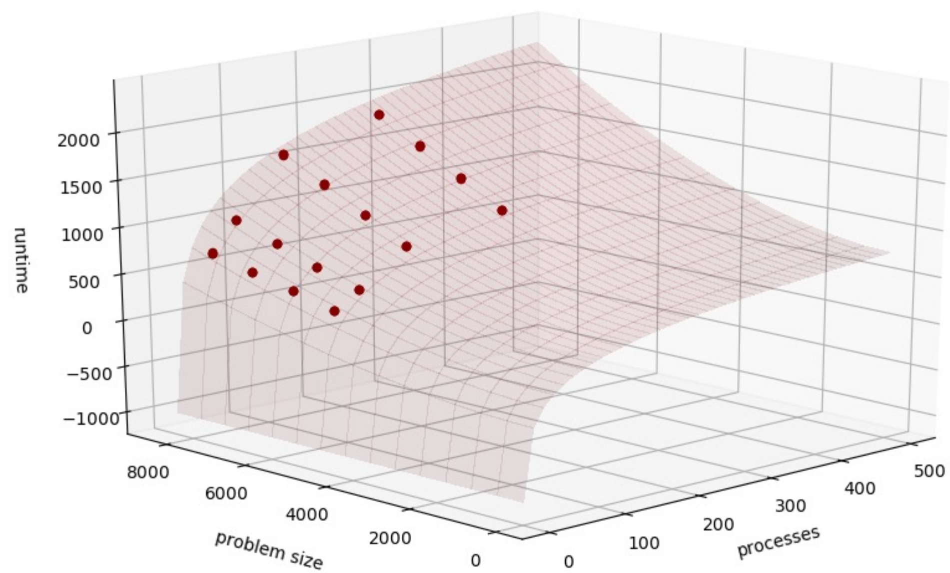
Multi-parameter modeling in Extra-P

Find best single-parameter model

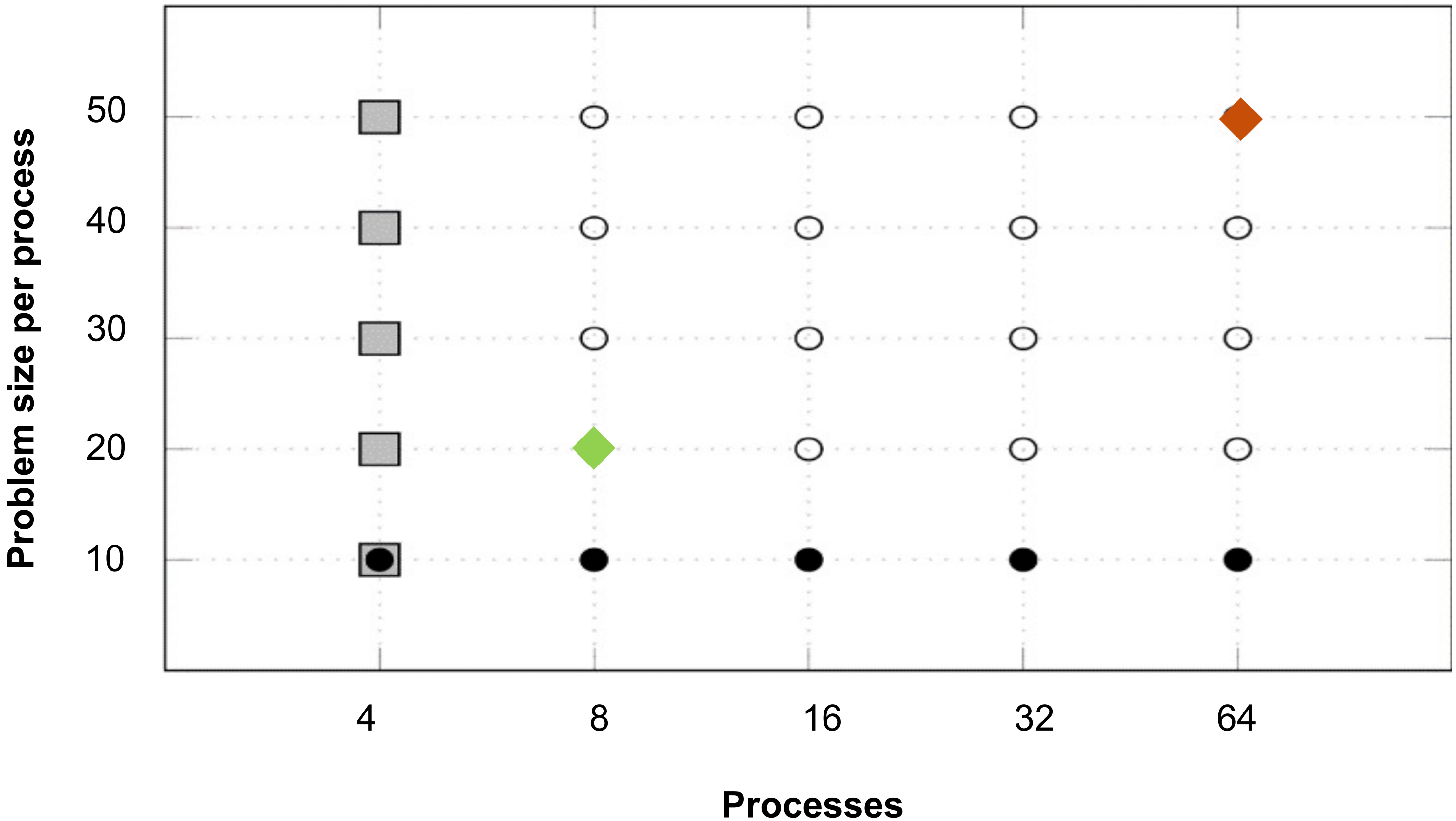
Combine them in the most plausible way (+, *, none)



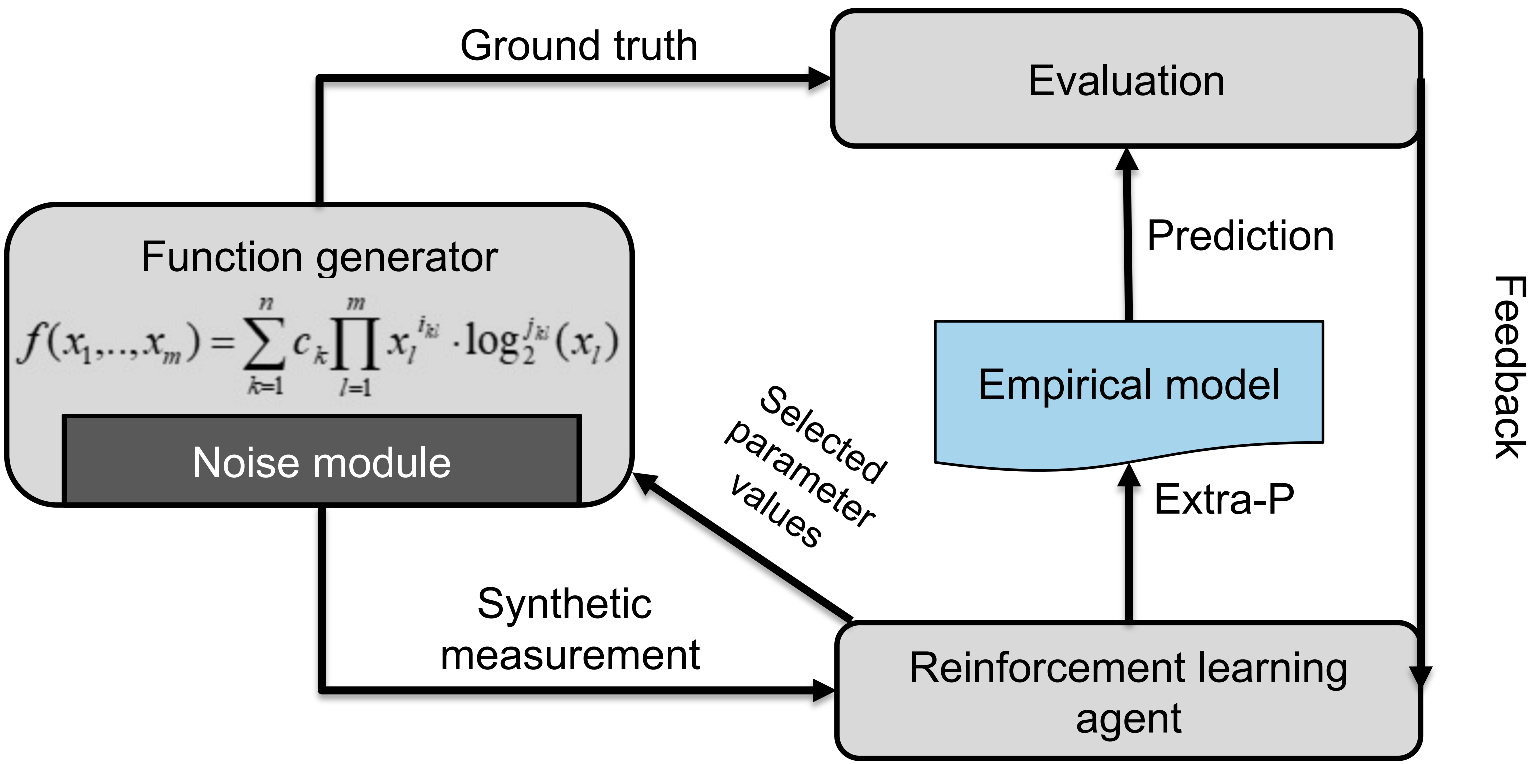
Generation of candidate models and selection of best fit



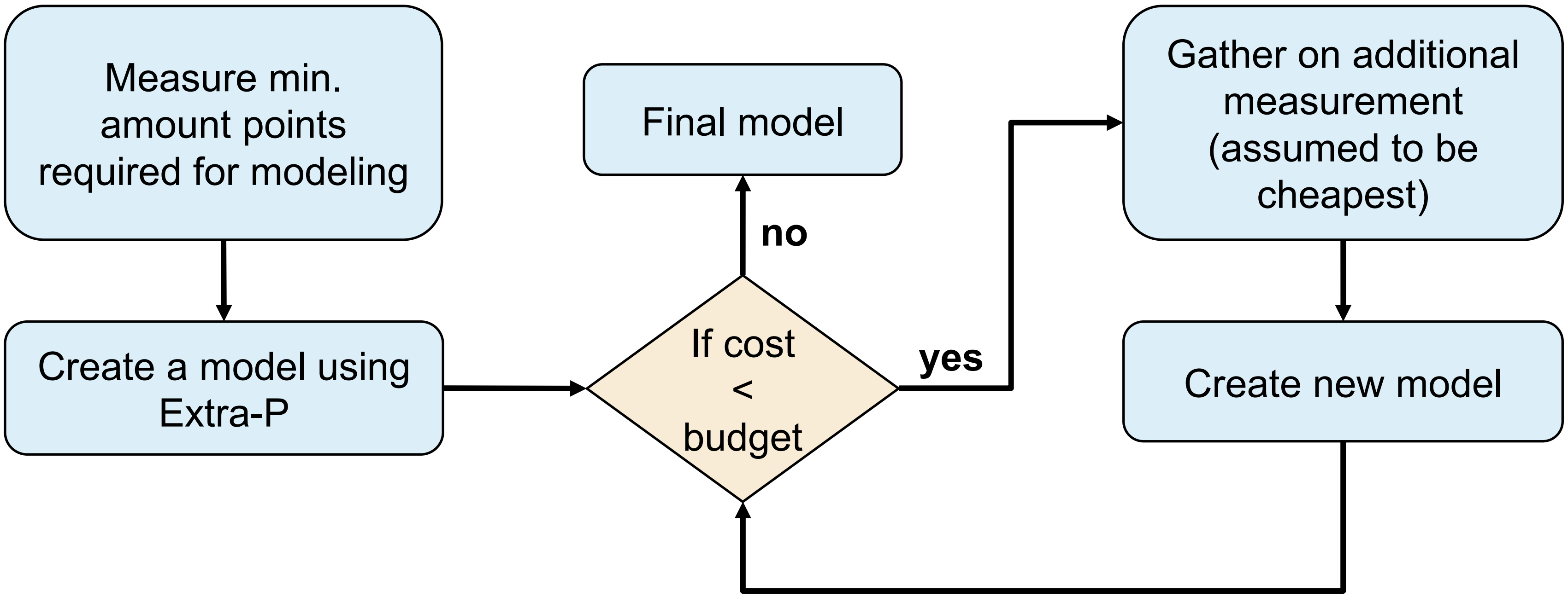
How many data points do we really need?



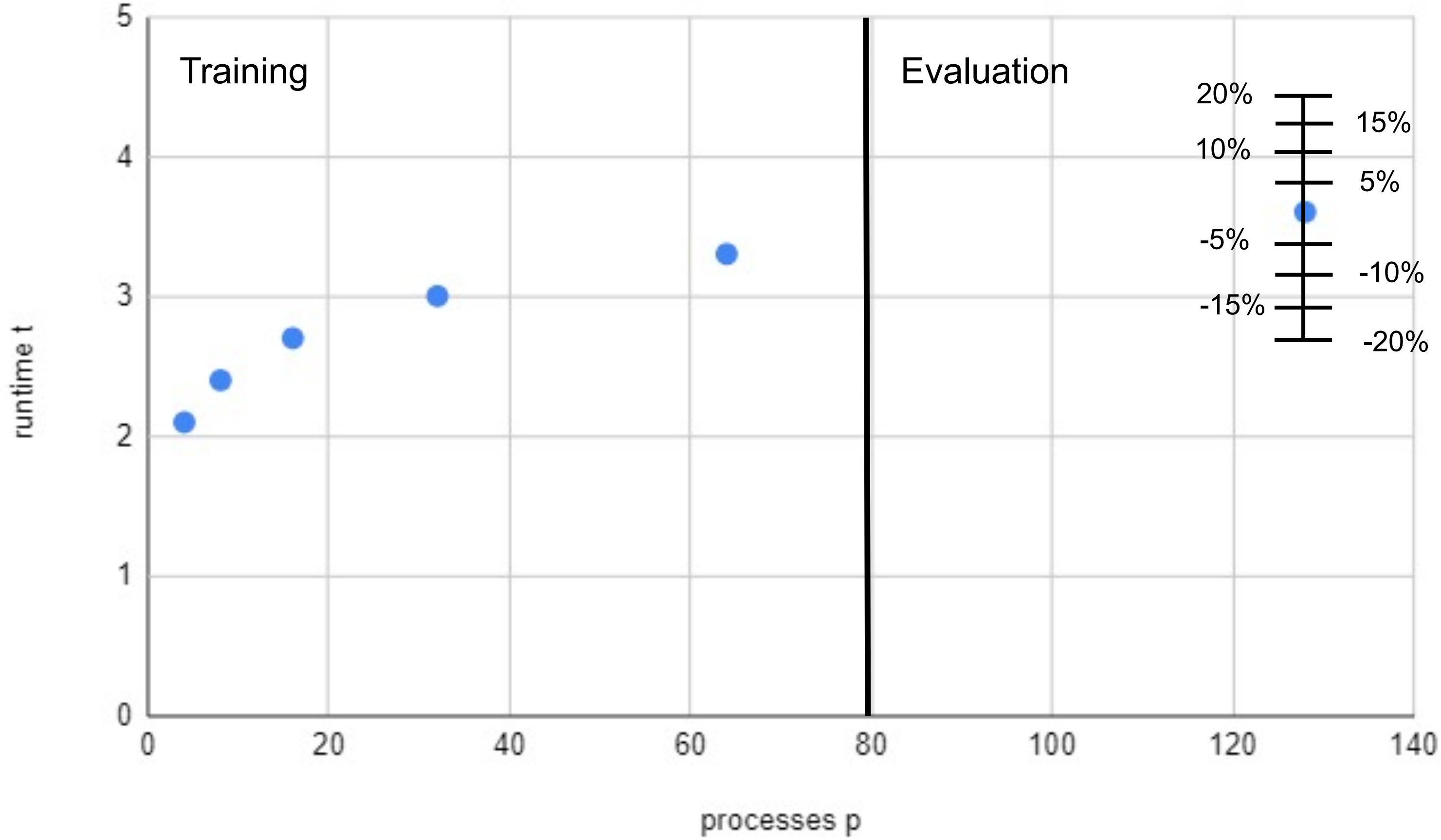
Learning cost-effective sampling strategies
[Ritter et al., IPDPS'20]



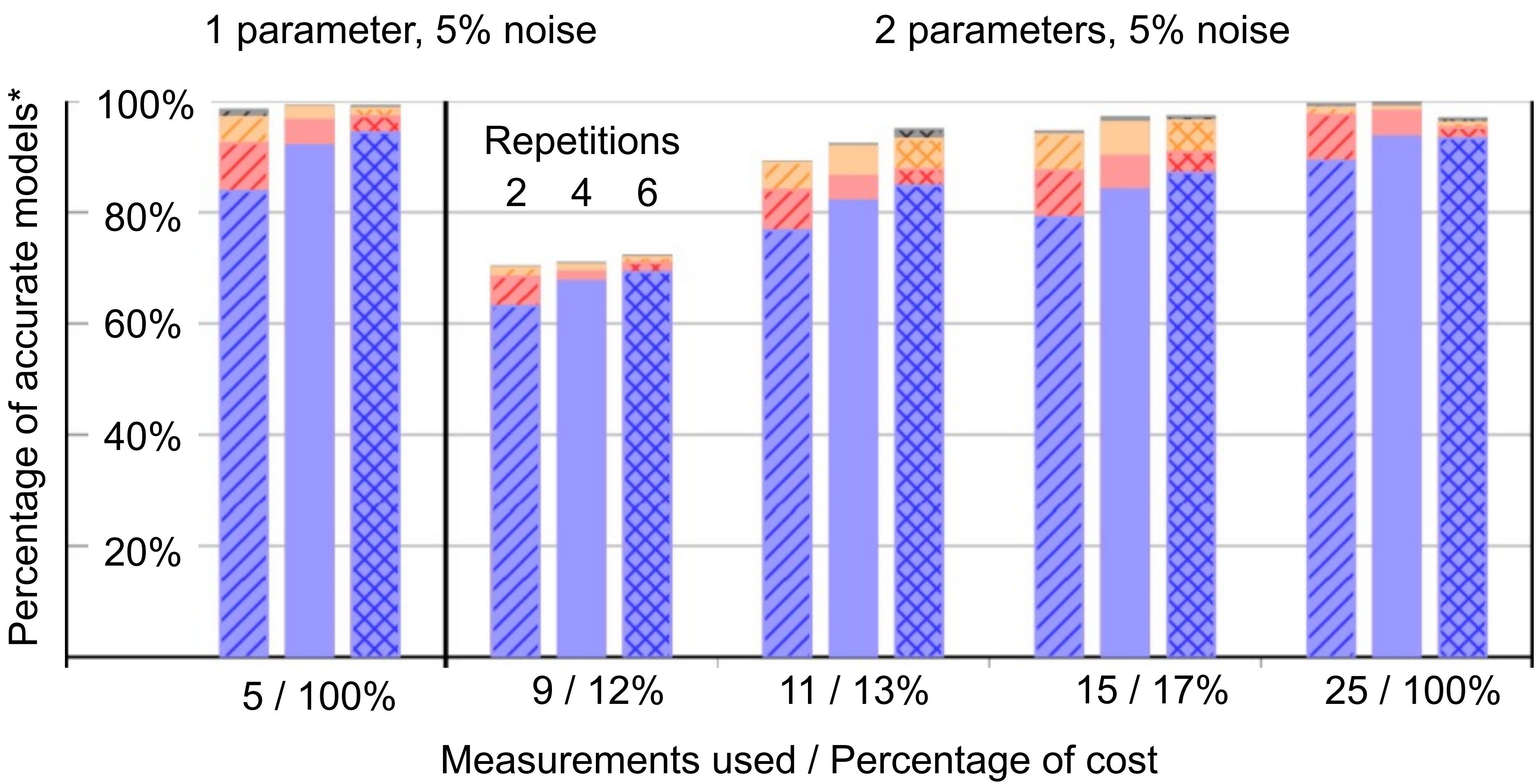
Heuristic parameter-value selection strategy



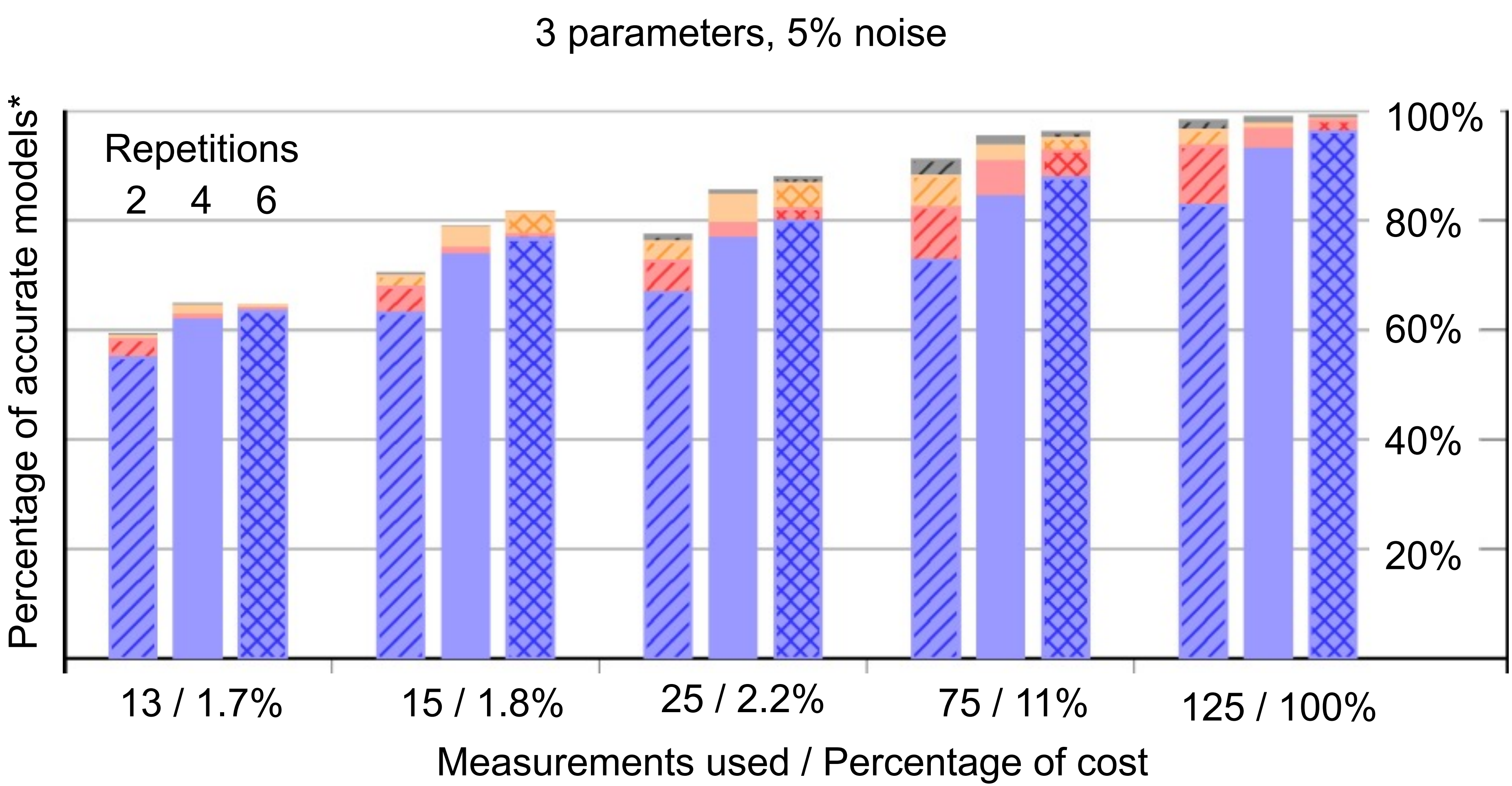
Synthetic data evaluation



Synthetic evaluation results

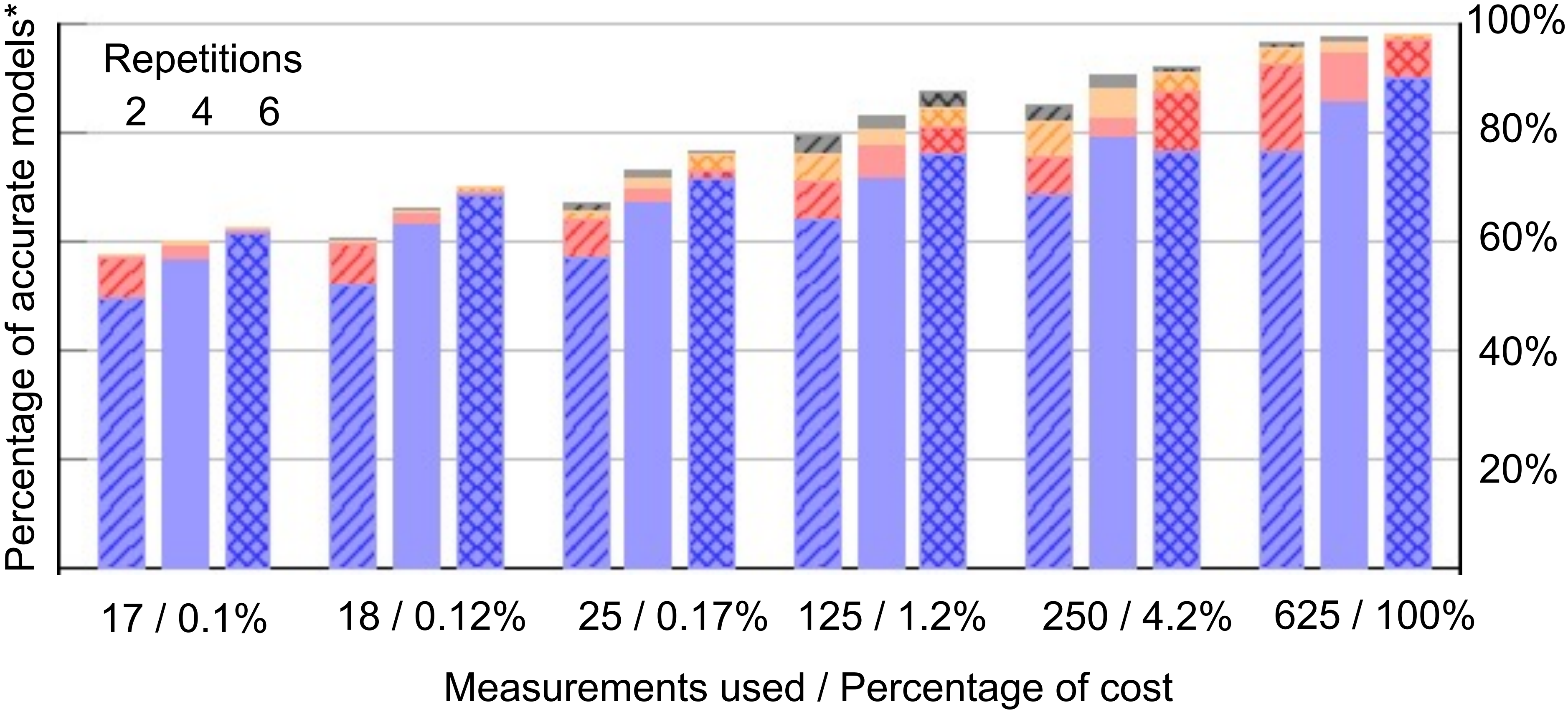


Synthetic evaluation results

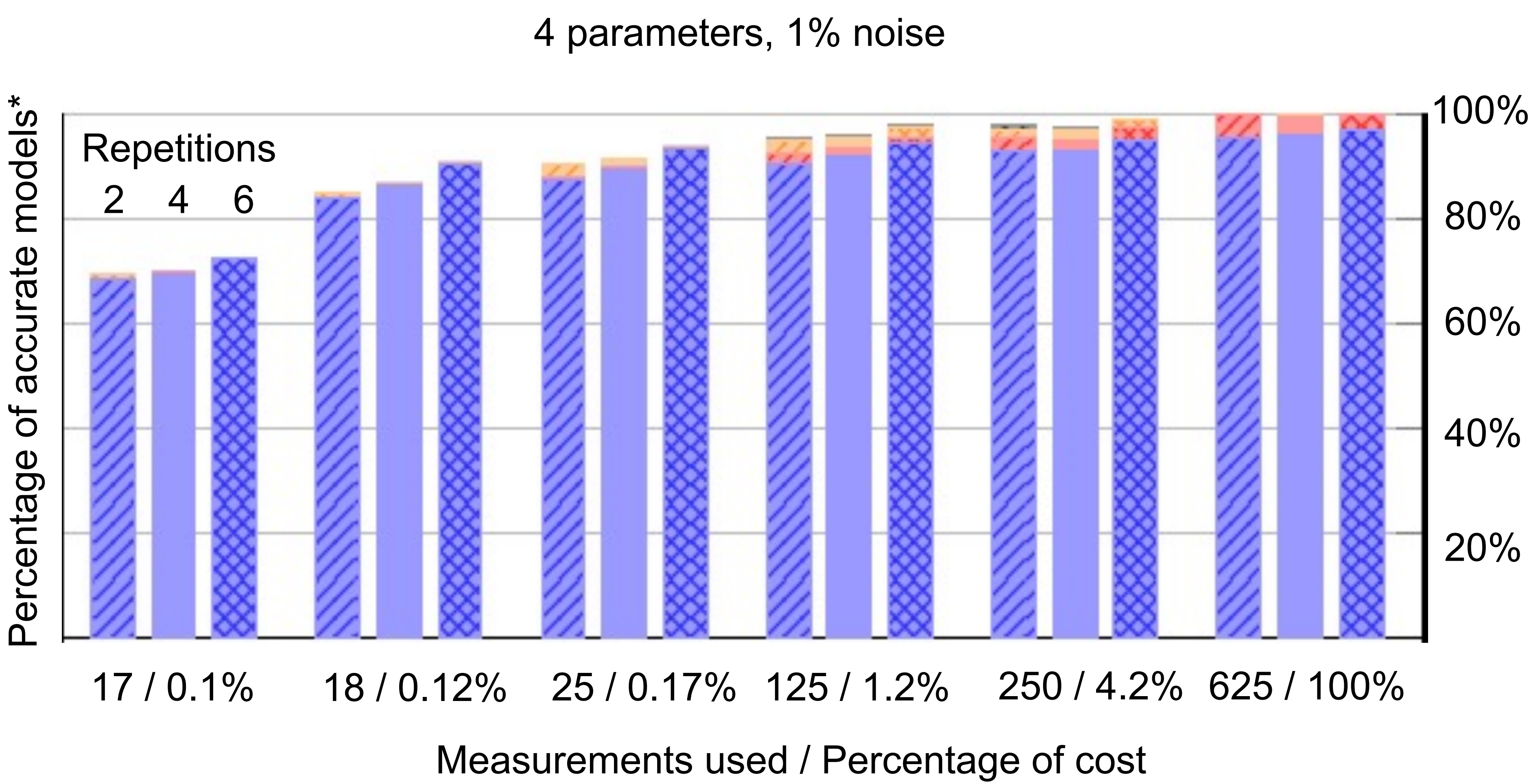


Synthetic evaluation results

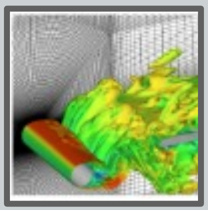
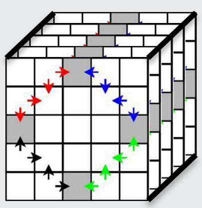
4 parameters, 5% noise



Synthetic evaluation results

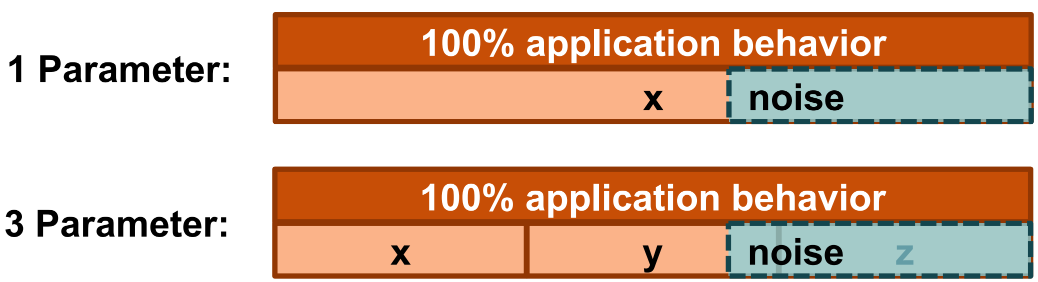


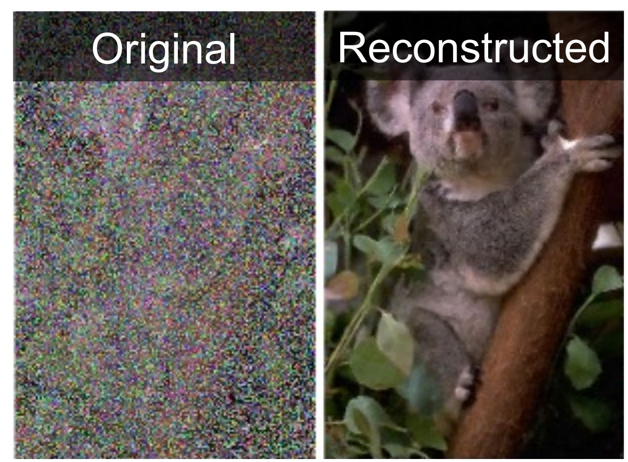
Case studies

Application	#Parameters	Extra points	Cost savings [%]	Prediction error [%]
FASTEST 	2	0	70	2
Kripke 	3	3	99	39
Relearn 	2	0	85	11

Noise-resilient performance modeling
[Ritter et al., IPDPS'21]

- Performance measurements frequently affected by noise
- Regression struggles with increased amounts of noise – especially w/ more parameters
- Neural networks are resilient to noise – **when noise is part of their training**

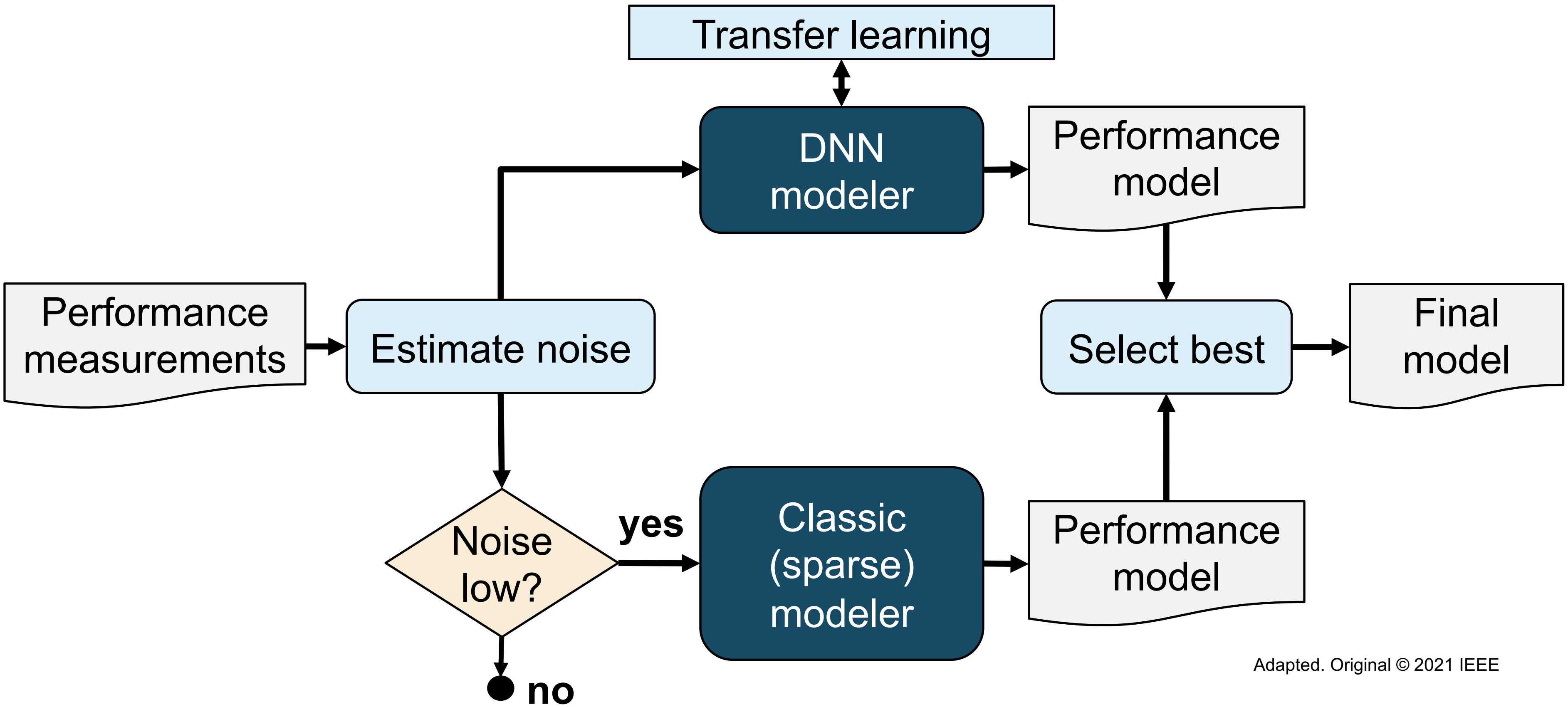




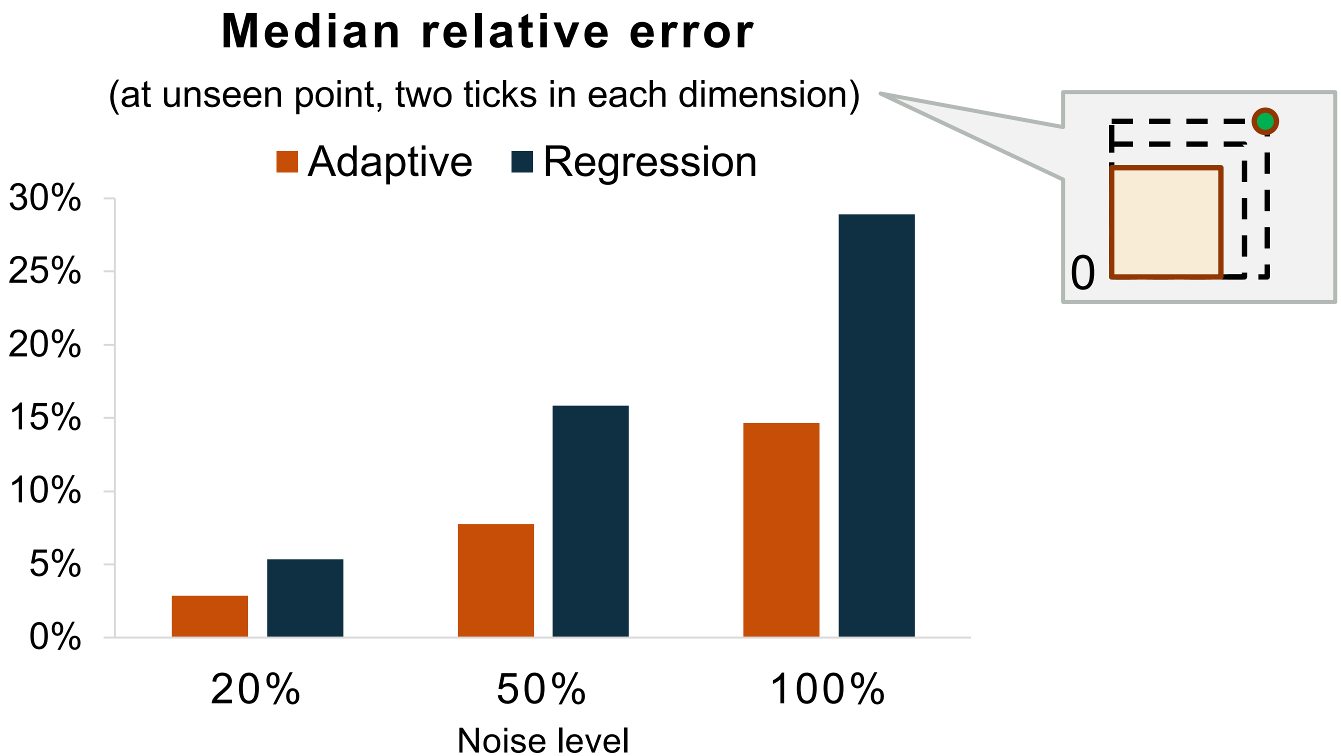
Adapted from: https://developer.nvidia.com/blog/ai-can-now-fix-your-grainy-photos-by-only-looking-at-grainy-photos/

Noise-resilient adaptive modeling

DNNs often better at guessing models in the presence of noise



Noise-resilient performance modeling
Synthetic evaluation

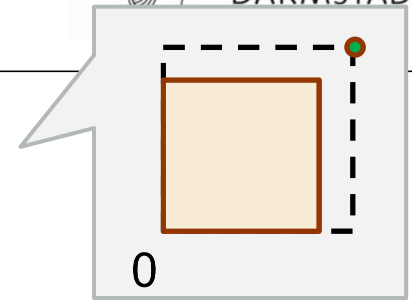


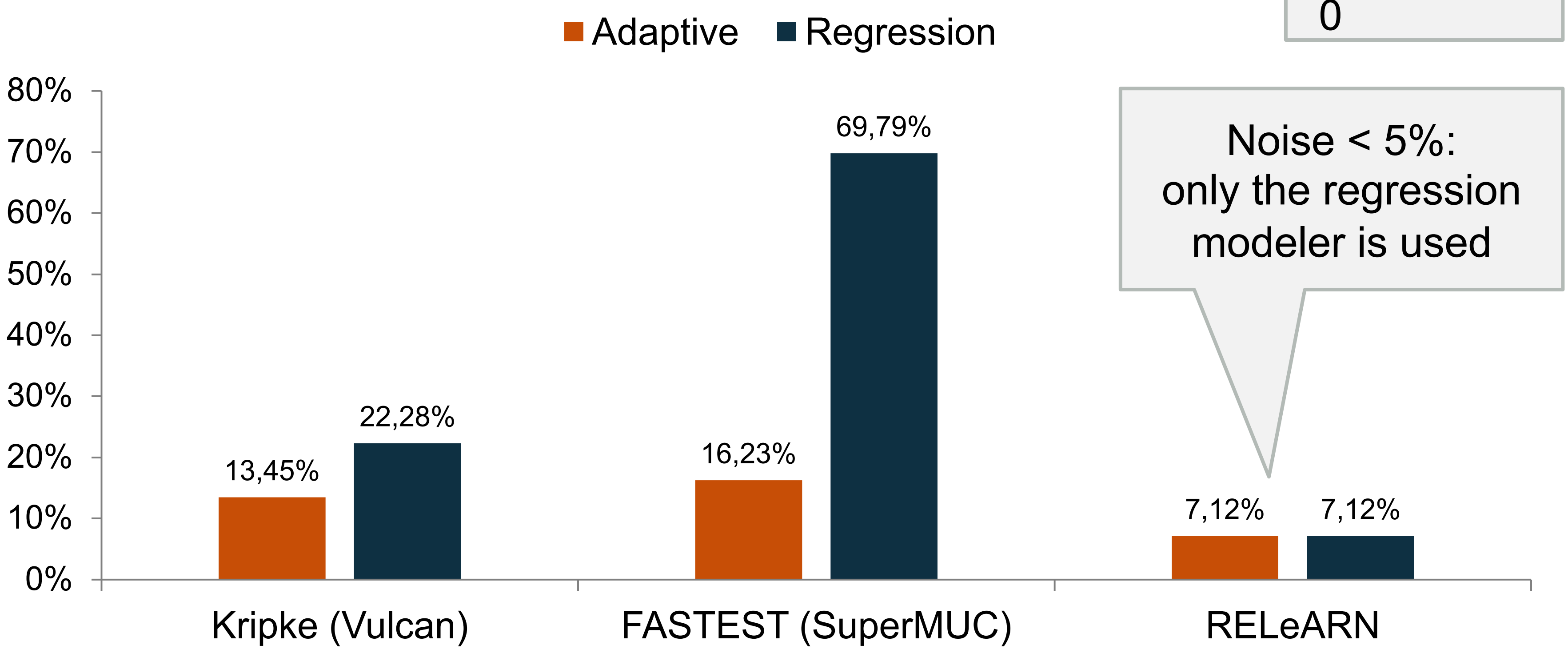
Adapted. Original © 2021 IEEE

2 parameters

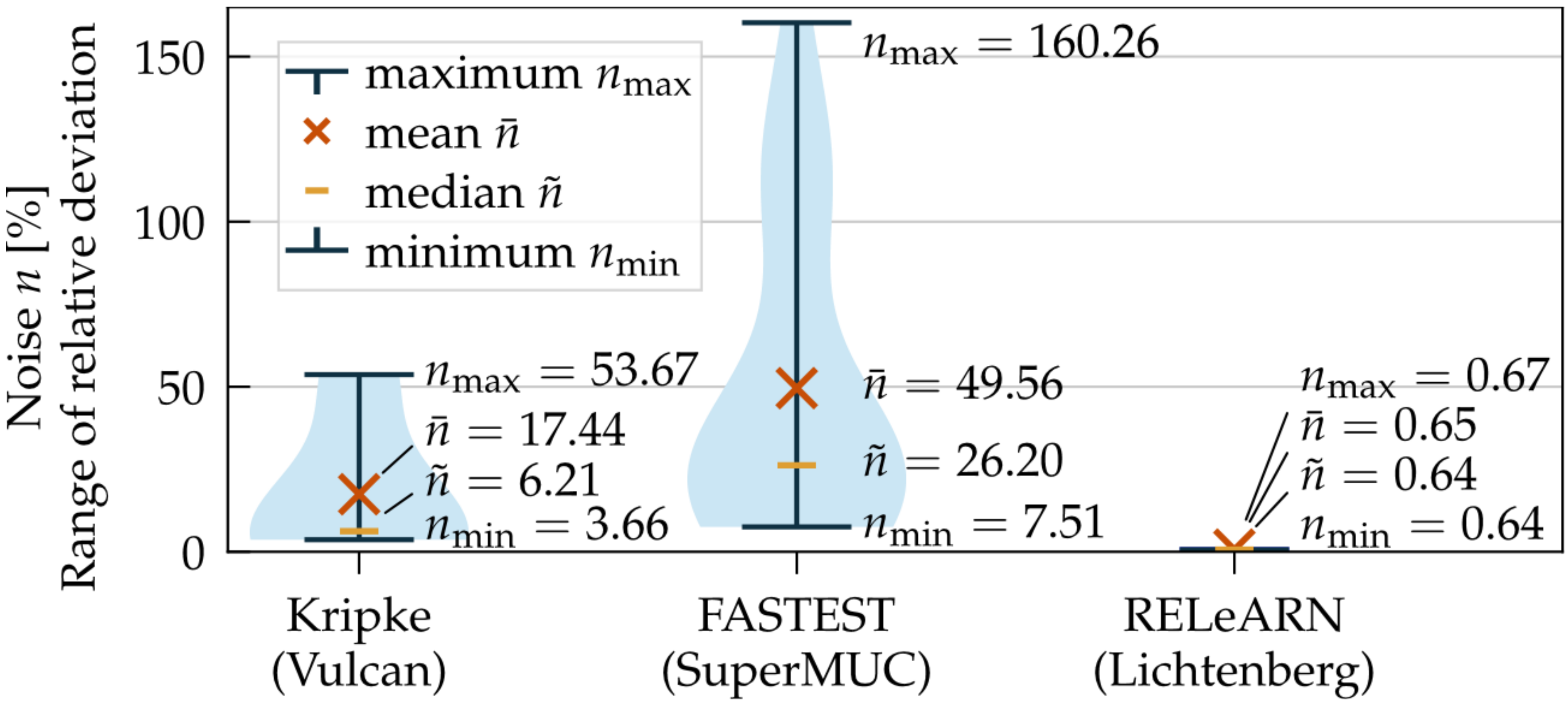
Noise-resilient performance modeling

Case studies – Results

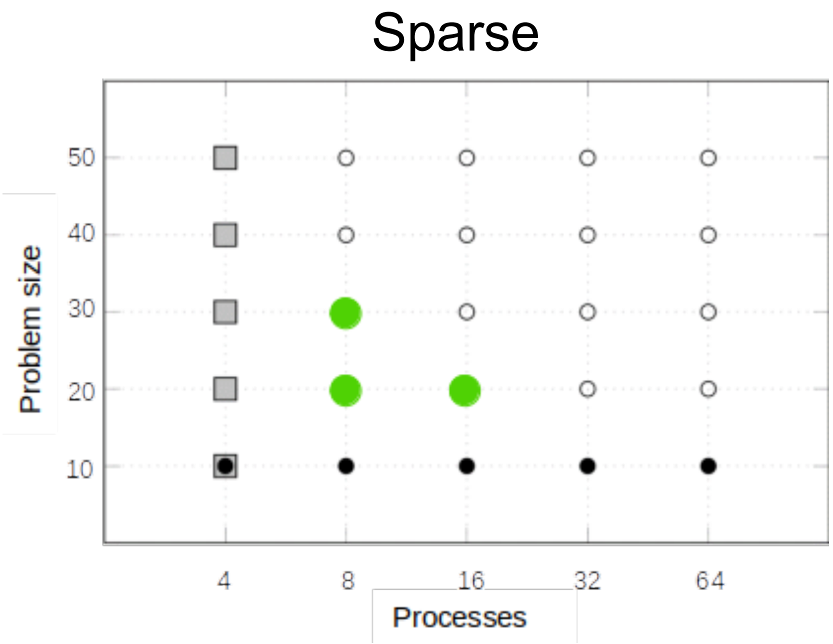


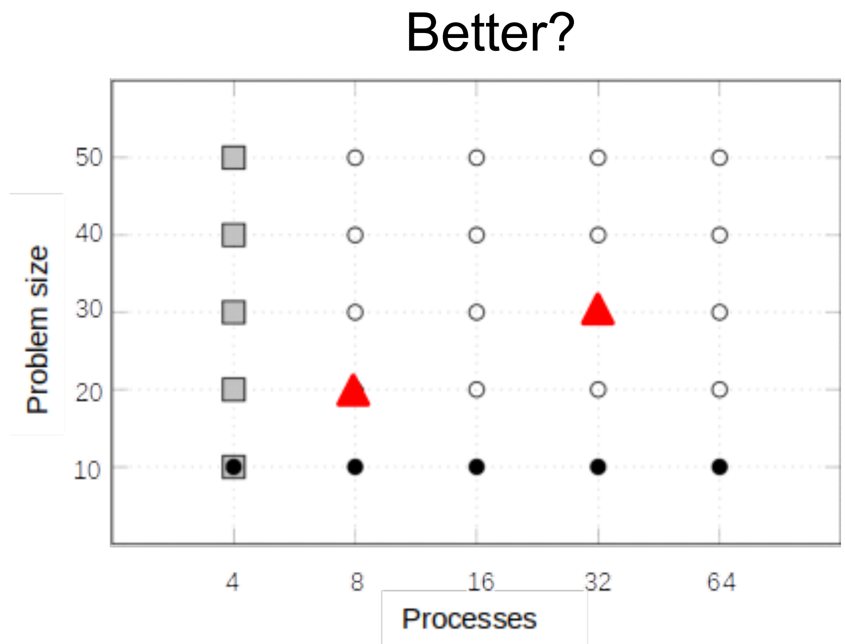


Noise-resilient performance modeling
Case studies – Noise

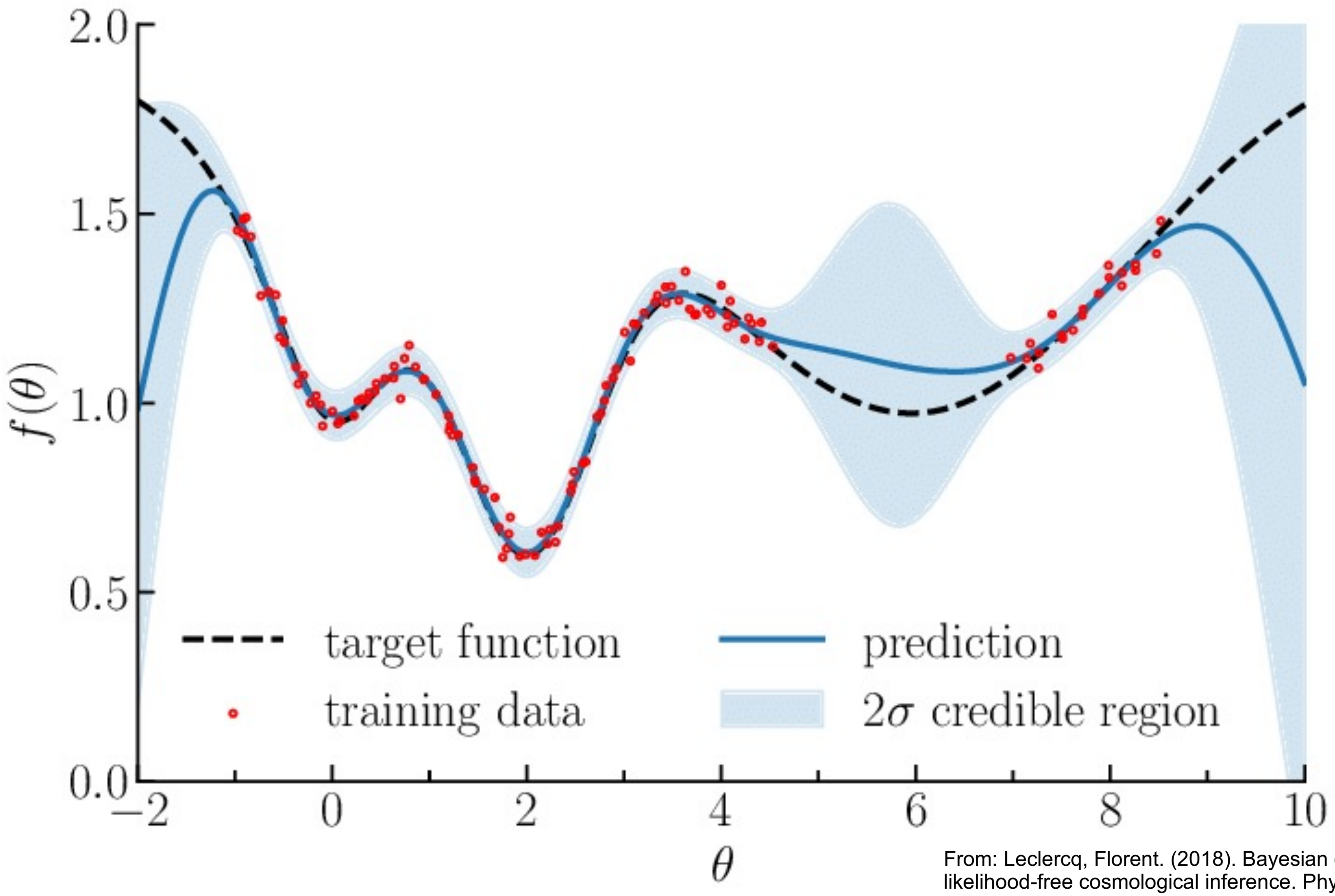


Optimized measurement point selection
[Naumann et al., in preparation]





Optimized measurement point selection
via Gaussian Process Regression (GPR) – Introduction



From: Leclercq, Florent. (2018). Bayesian optimization for likelihood-free cosmological inference. Physical Review D. 98. 10.1103/PhysRevD.98.063511. Used with permission.

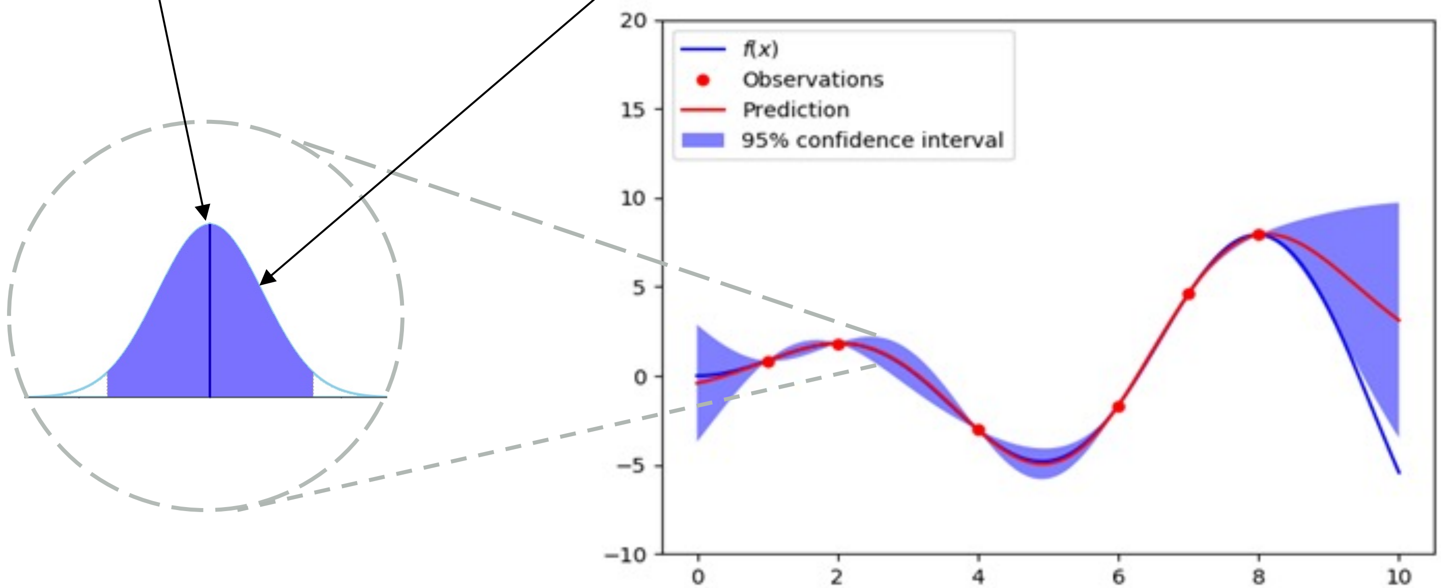
Optimized Measurement Point Selection
Gaussian Processes

Gaussian Process (GP): Series of normal distributions

$$GP(t) = N(\, m(t), c(t,t') \,)$$

Mean function

Covariance function

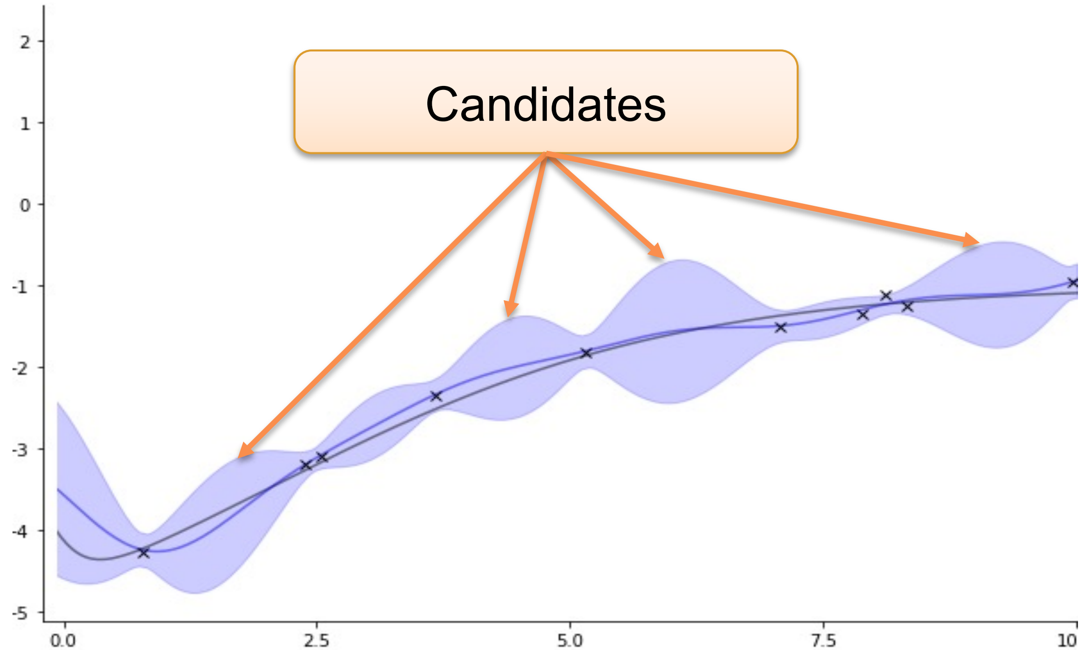


Optimized measurement point selection

Gaussian Processes

Idea: Use covariance function

High variance → High uncertainty → Promising measurement



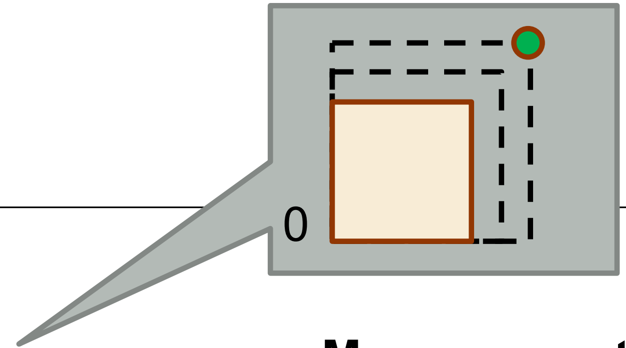
Measurement costs have to be considered!

↓

Cost-benefit calculation

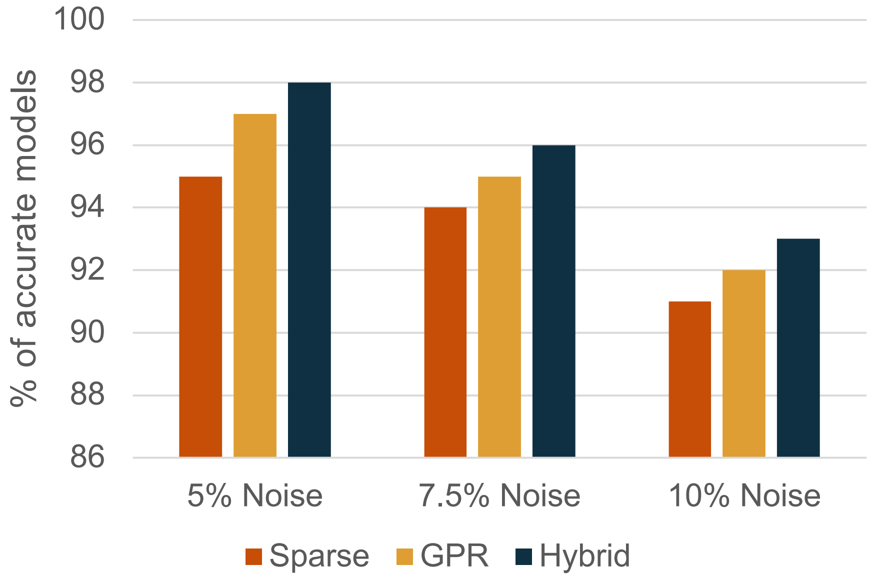
Results

Synthetic evaluation (3 parameters)



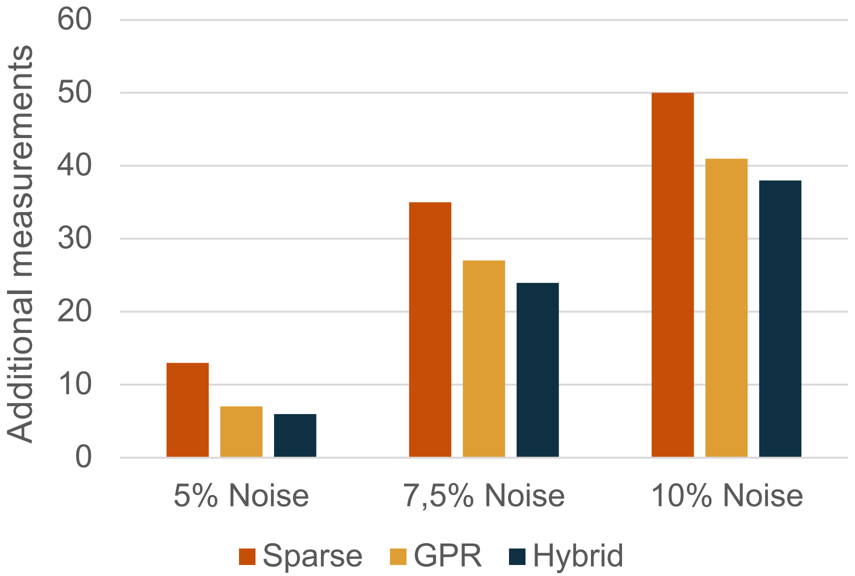
Relative error < 20%

(at unseen point, 2 ticks in each direction)

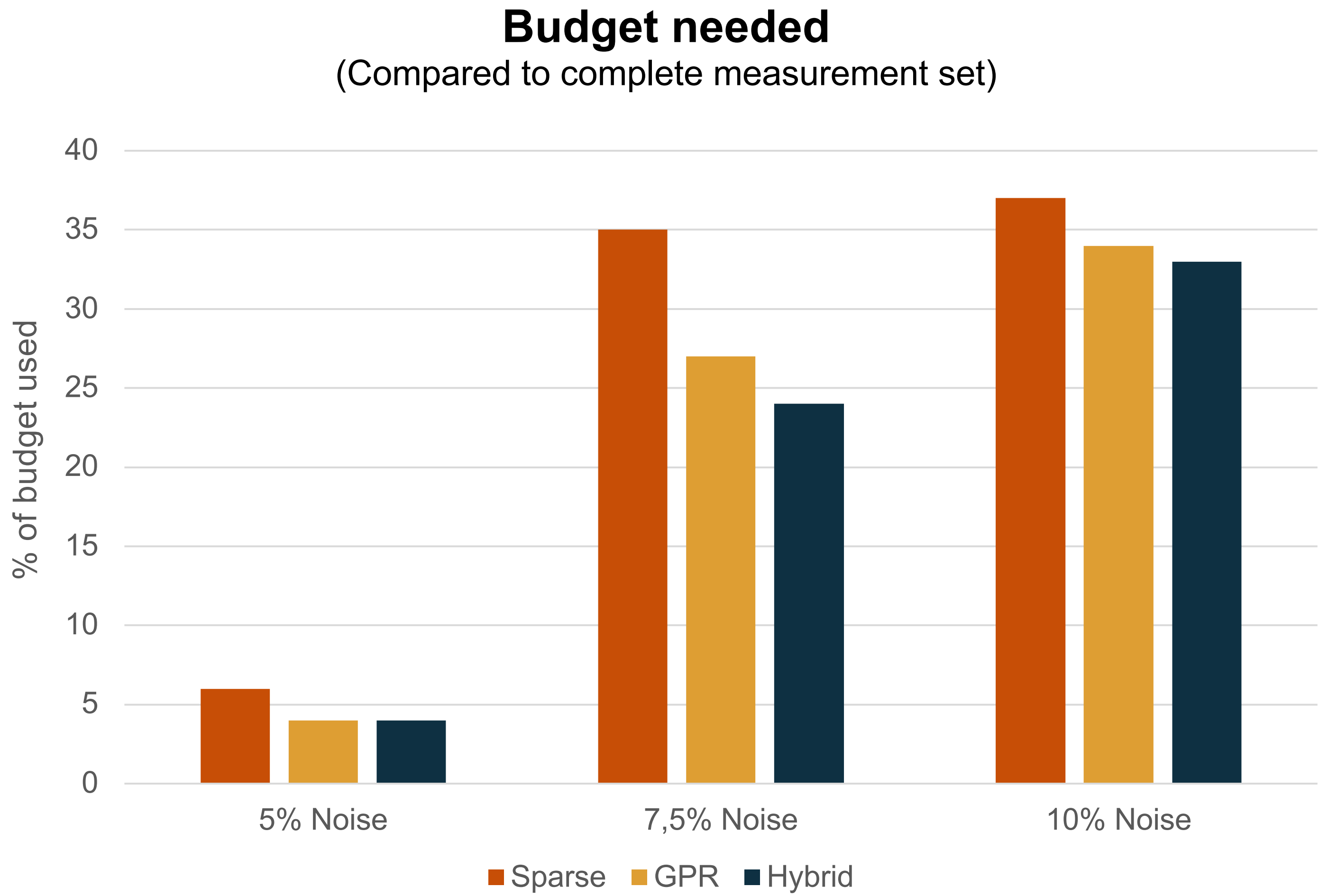


Measurements needed

(in addition to the minimal measurement set)



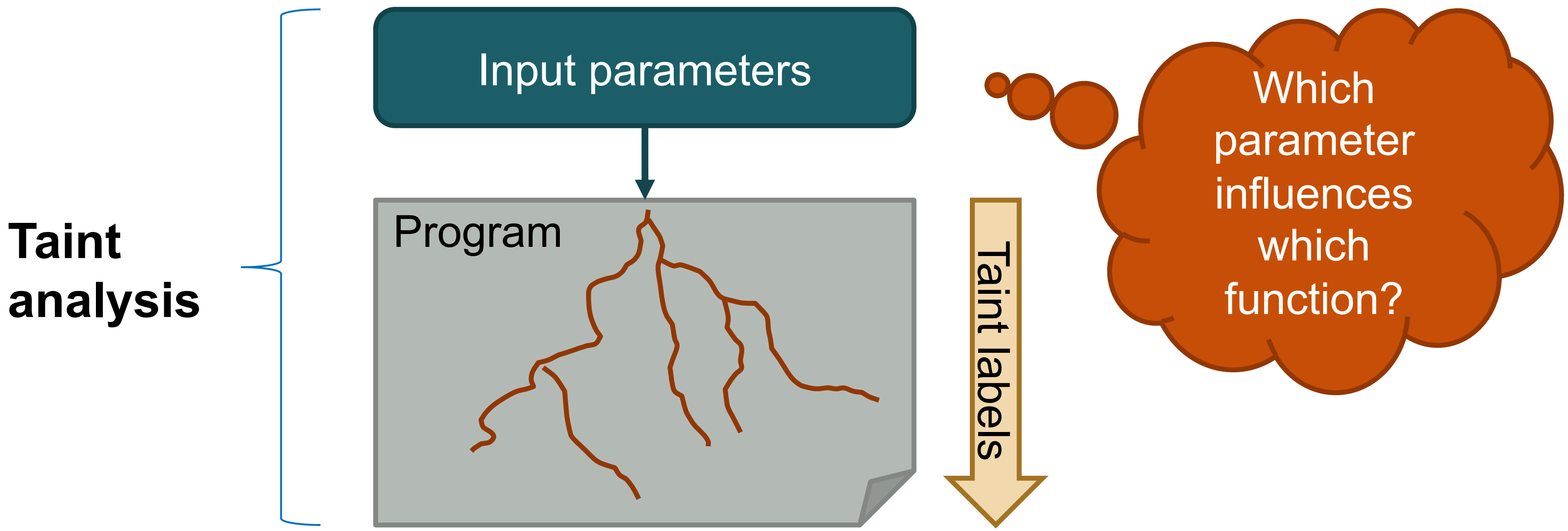
Results



Budget needed
(Compared to complete measurement set)
% of budget used
40
35
30
25
20
15
10
5
0
5% Noise
7,5% Noise
10% Noise
Sparse
GPR
Hybrid

Parameter selection
[Copik et al, PPoPP'21]

- The more paramters the more experiments
- Modeling parameters without performance impact is harmful



Selected papers

Topic	Bibliography
Foundation (single model paramter)	Alexandru Calotoiu, Torsten Hoefler, Marius Poke, Felix Wolf: Using Automated Performance Modeling to Find Scalability Bugs in Complex Codes. **SC13**.
MPI case study	Sergei Shudler, Yannick Berens, Alexandru Calotoiu, Torsten Hoefler, Alexandre Strube, Felix Wolf: Engineering Algorithms for Scalability through Continuous Validation of Performance Expectations. **IEEE TPDS**, 30(8):1768–1785, 2019.
Multiple model parameters	Alexandru Calotoiu, David Beckingsale, Christopher W. Earl, Torsten Hoefler, Ian Karlin, Martin Schulz, Felix Wolf: Fast Multi-Parameter Performance Modeling. **IEEE Cluster 2016**.
Cost-effective sampling strategies	Marcus Ritter, Alexandru Calotoiu, Sebastian Rinke, Thorsten Reimann, Torsten Hoefler, Felix Wolf: Learning Cost-Effective Sampling Strategies for Empirical Performance Modeling. **IPDPS 2020**.
Noise resilience	Marcus Ritter, Alexander Geiß, Johannes Wehrstein, Alexandru Calotoiu, Thorsten Reimann, Torsten Hoefler, Felix Wolf: Noise-Resilient Empirical Performance Modeling with Deep Neural Networks. **IPDPS 2021**.
Taint-based performance modeling	Marcin Copik, Alexandru Calotoiu, Tobias Grosser, Nicolas Wicki, Felix Wolf, Torsten Hoefler: Extracting Clean Performance Models from Tainted Programs. **PPoPP 2021**.

Thank you!

Q&A