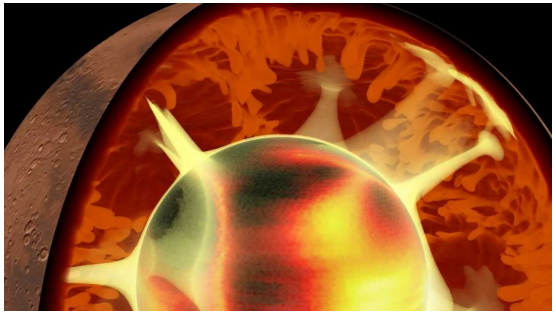
Finding Performance-Optimal Configurations for High-Performance Computing

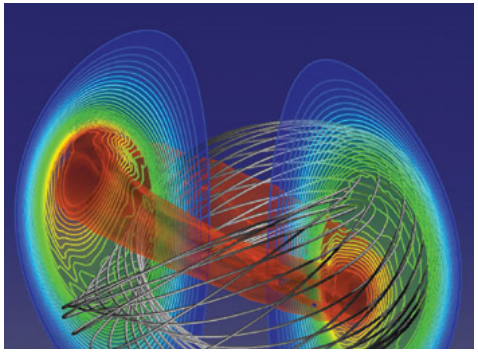
Alexander Grebhahn, Norbert Siegmund, Sven Apel

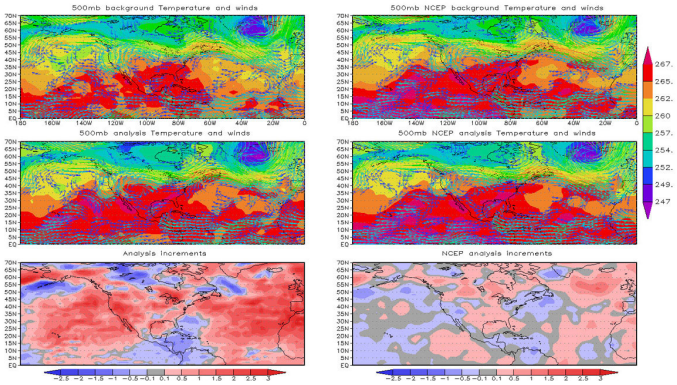
University of Passau

FOSD Meeting 2014, Dagstuhl

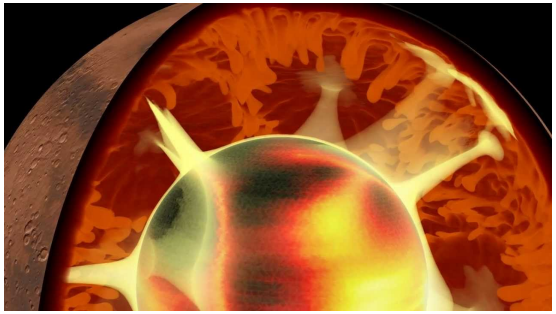
High-Performance Computing and ExaStencils

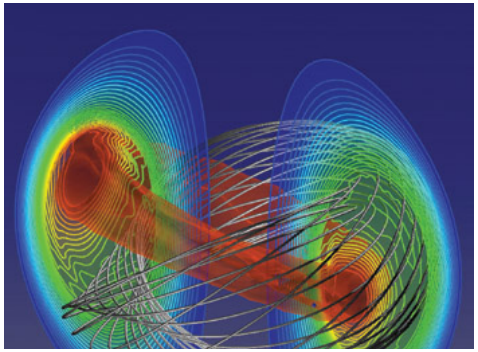


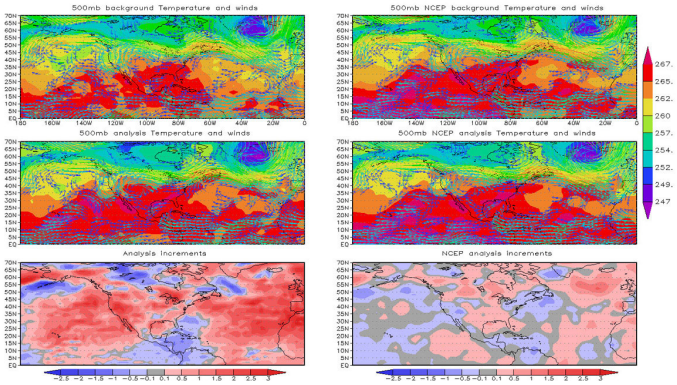


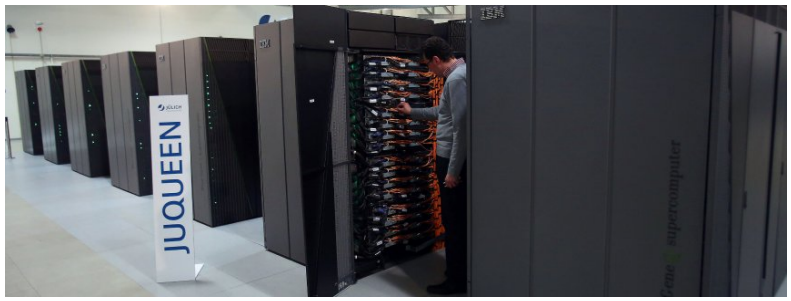


High-Performance Computing and ExaStencils

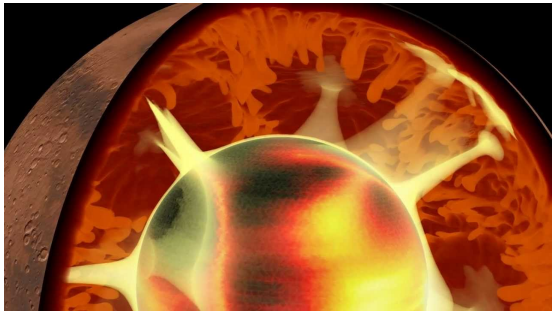


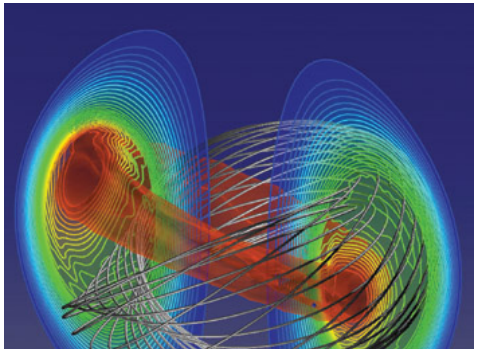


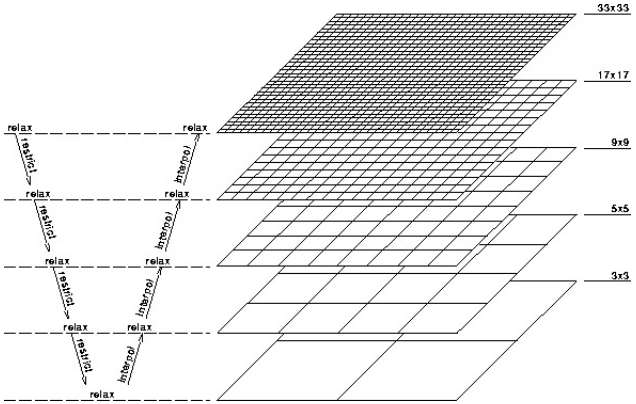


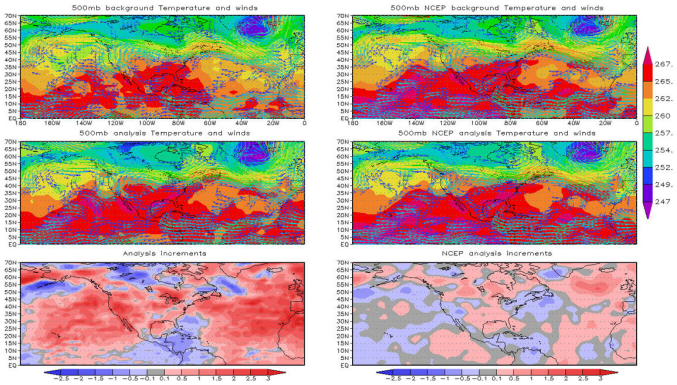


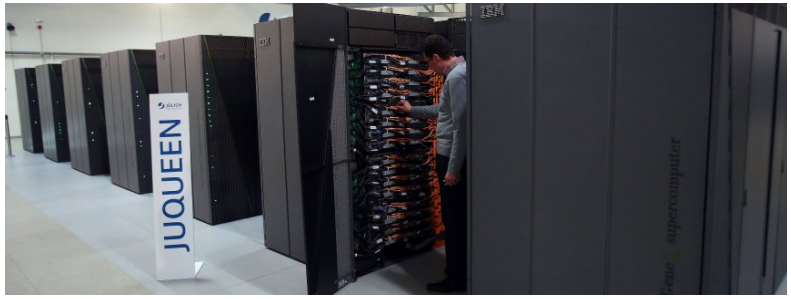
High-Performance Computing and ExaStencils



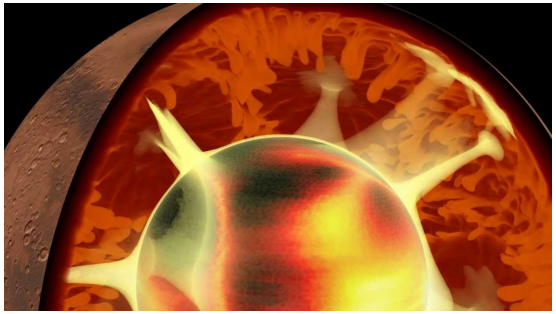


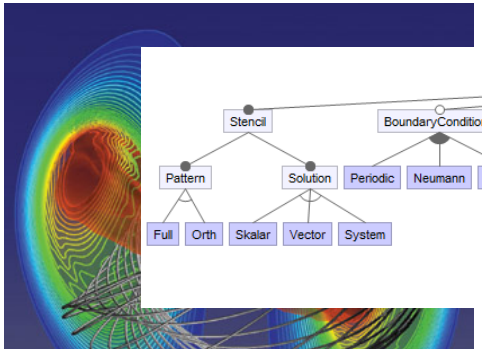


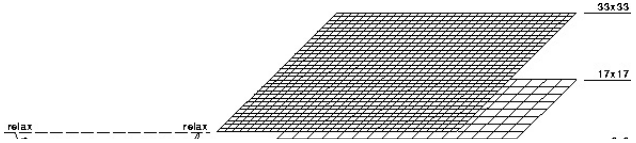


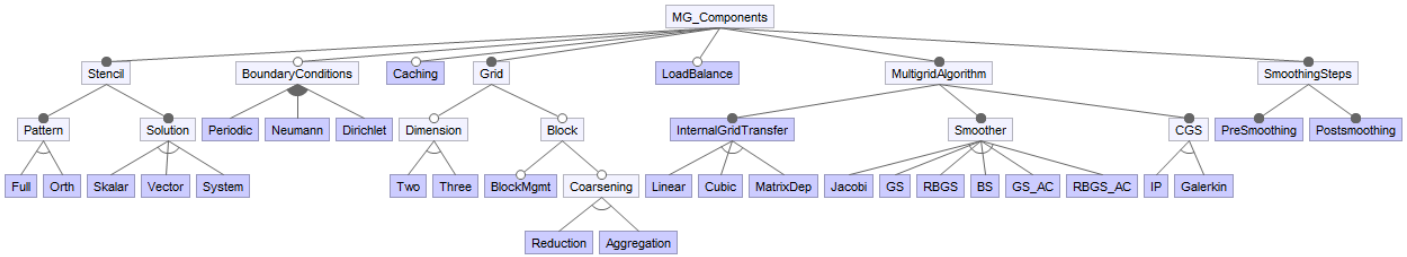


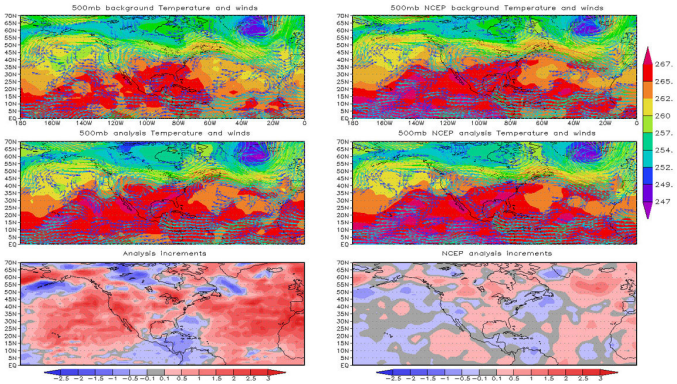
High-Performance Computing and ExaStencils

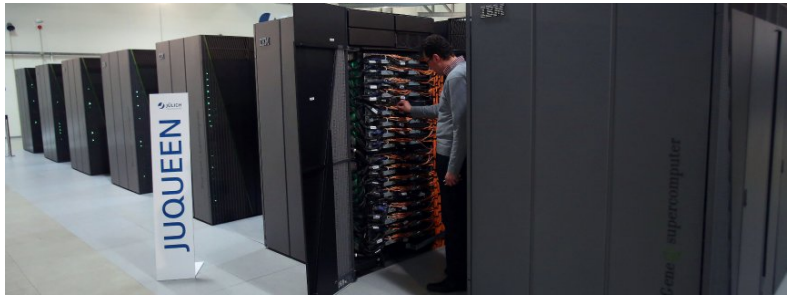




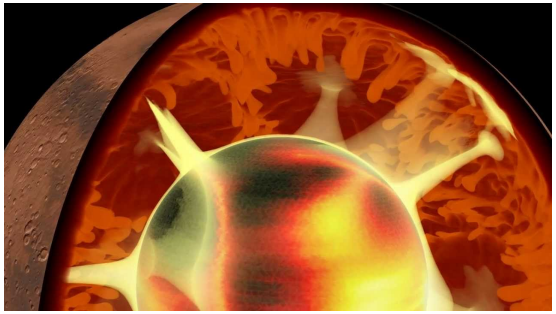


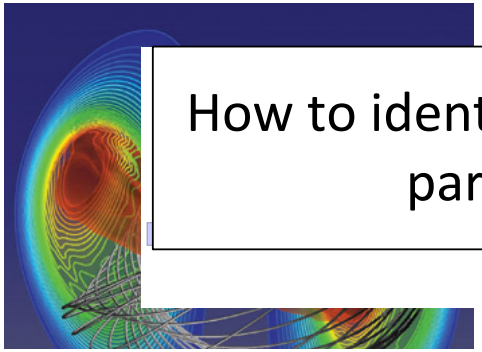


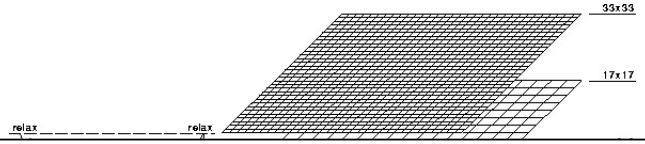




High-Performance Computing and ExaStencils

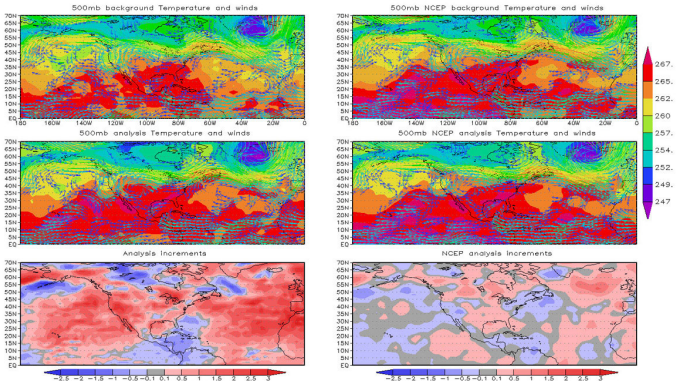


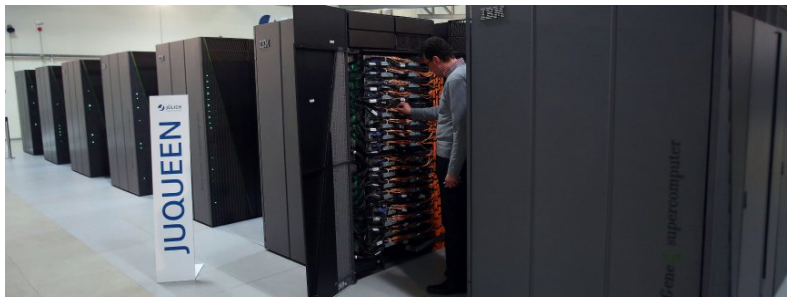




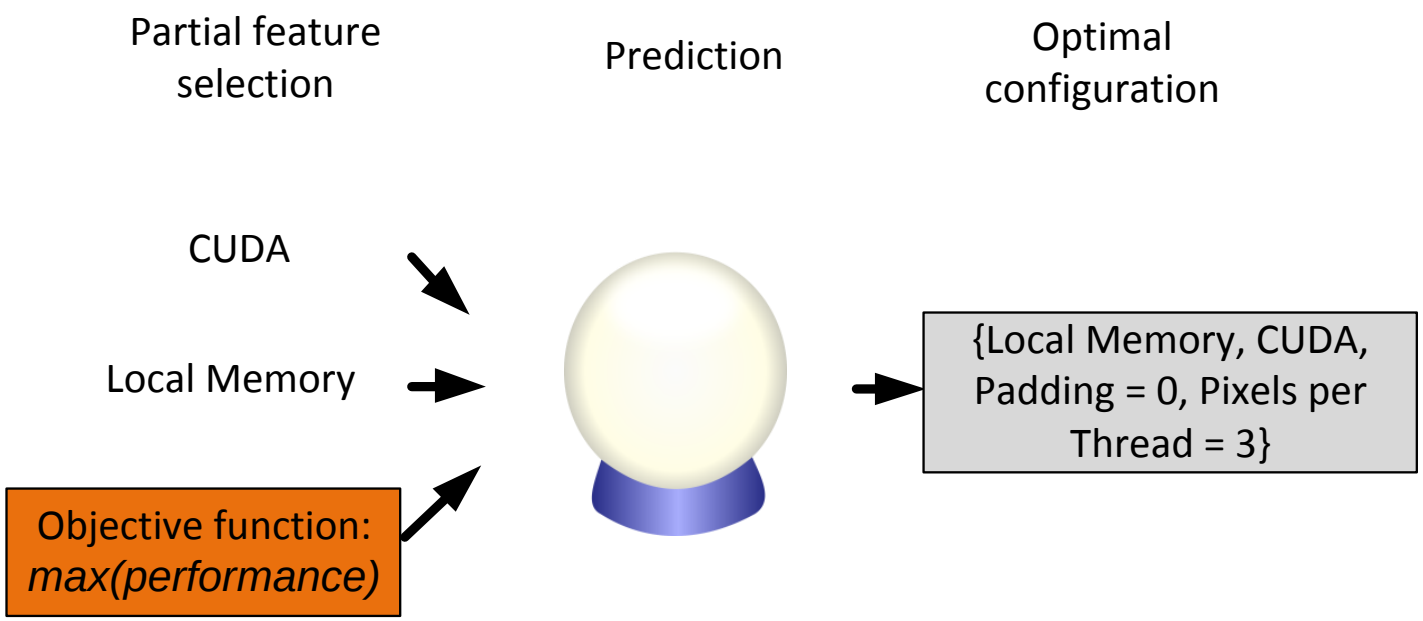
How to identify performance-optimal components and parameters for a specific hardware?

Reduction Aggregation





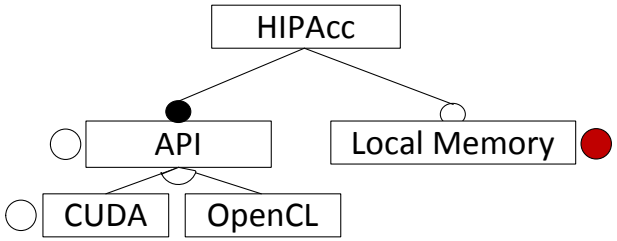
SPL Conqueror [Siegmund et al., 2012]



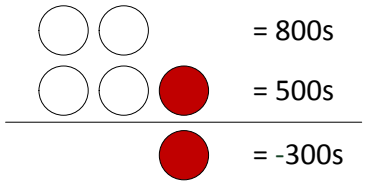
Advantages:

- Detection of feature interactions
- Transparent (i.e., influences of individual features and feature interactions explicitly modeled and quantified)

Influence of Individual Features

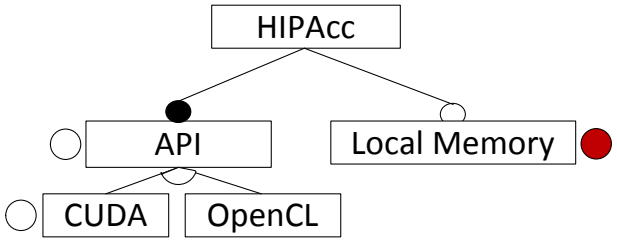


Identification:

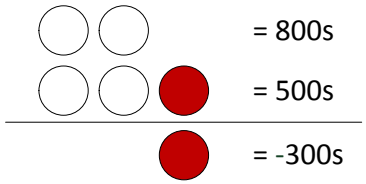


Performance difference is interpreted as contribution of the feature in question

Influence of Individual Features



Identification:

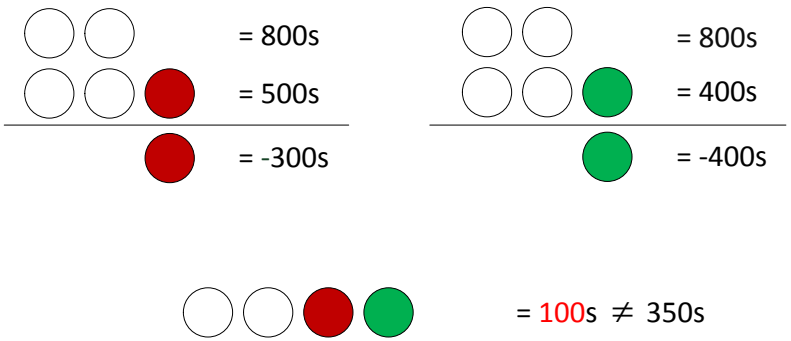


Performance difference is interpreted as contribution of the feature in question

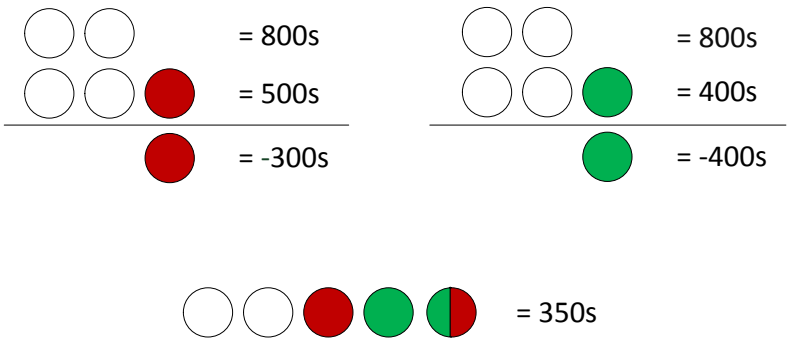
Heuristics:

Feature-wise (FW) heuristic: Quantifies the influence of individual features on performance

Interactions Between Features



Interactions Between Features



Heuristics:

- Pair-wise (PW) heuristic: interactions between two features
- Higher-order (HO) heuristic: interactions between three or more features
- Hot-spot (HS) heuristic: interactions of "hot-spot" features

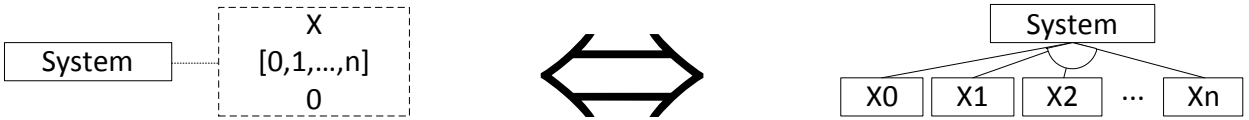
Numerical Parameters (Non-Boolean Features)

Existing heuristics work for boolean features only!

Numerical Parameters (Non-Boolean Features)

Existing heuristics work for boolean features only!

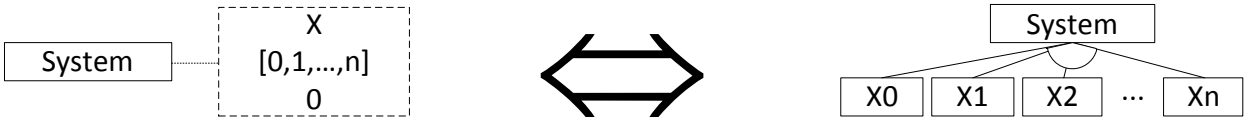
Discretization:



Numerical Parameters (Non-Boolean Features)

Existing heuristics work for boolean features only!

Discretization:

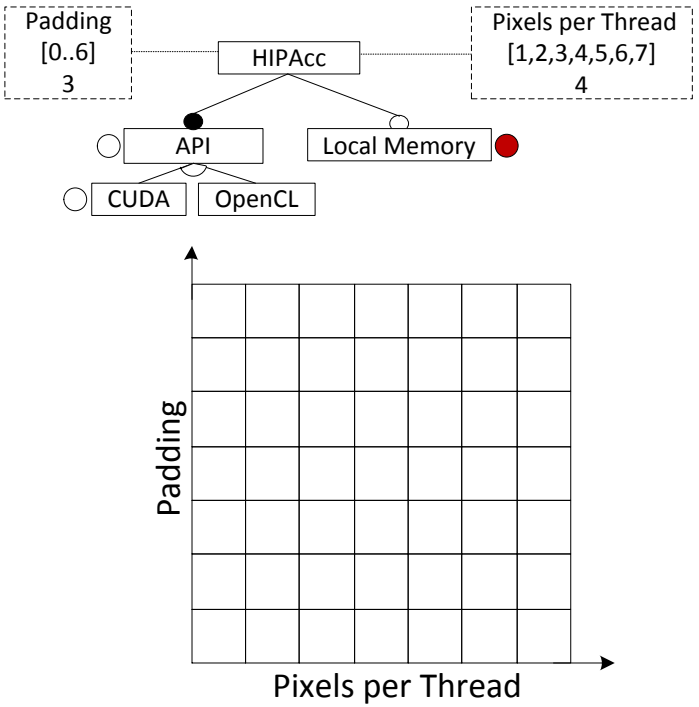


Disadvantages:

- Increasing number of features
- Loss of connection between parameter values

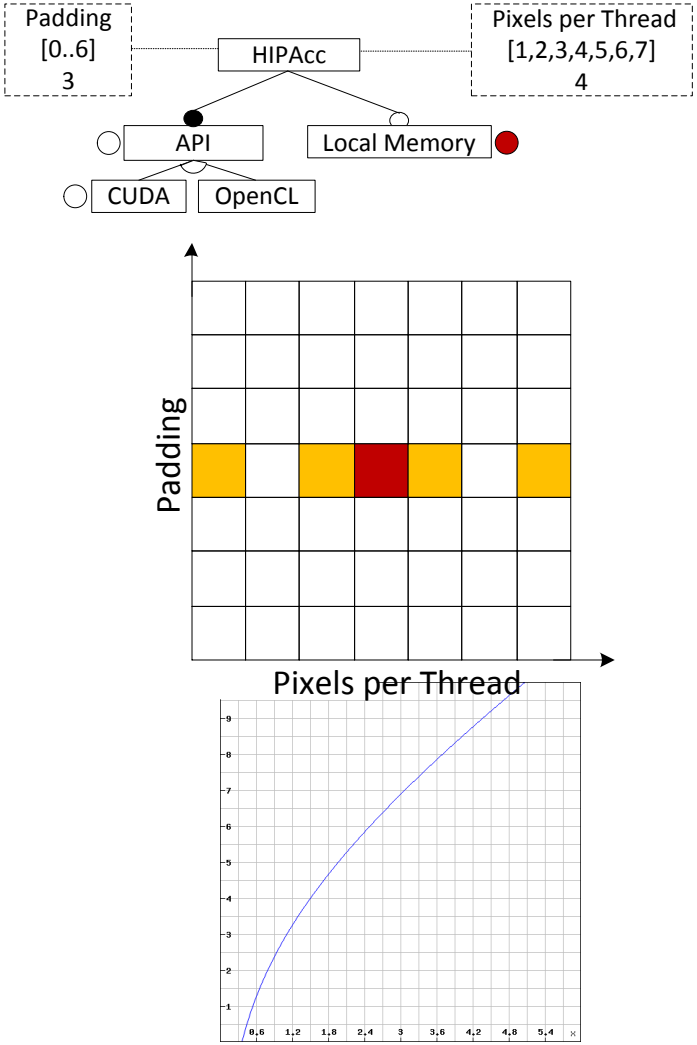
Influence of Parameters

- Determine influence of parameter values on performance
- Learn function for each pair of parameter and feature
- Independent sampling of parameters



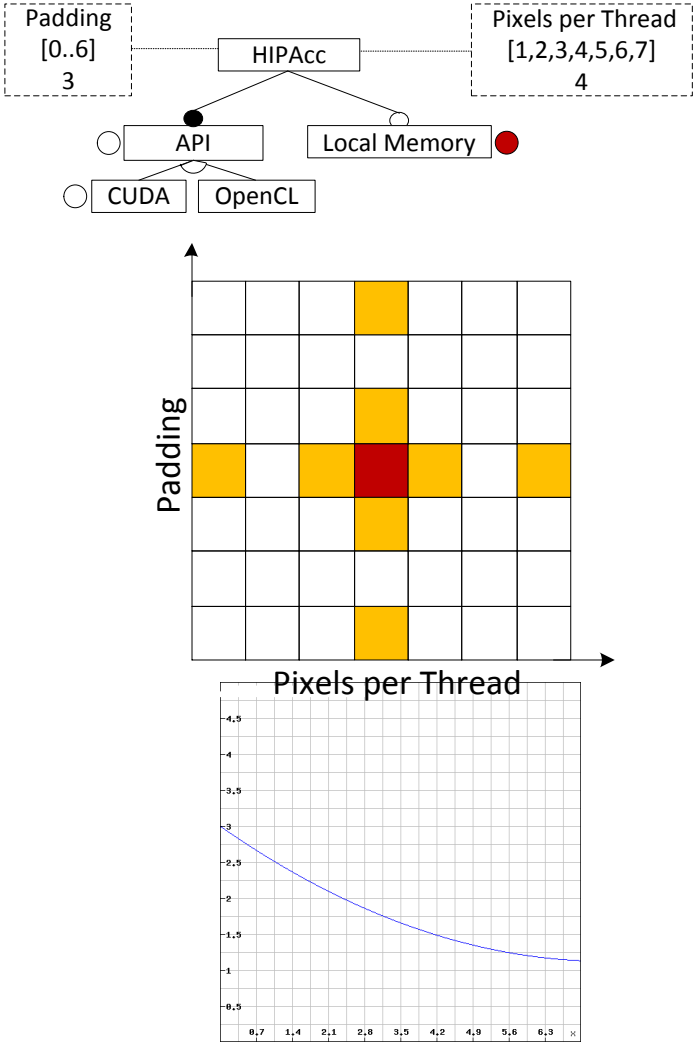
Influence of Parameters

- Determine influence of parameter values on performance
- Learn function for each pair of parameter and feature
- Independent sampling of parameters



Influence of Parameters

- Determine influence of parameter values on performance
- Learn function for each pair of parameter and feature
- Independent sampling of parameters



Heuristics:

- Function learning (FL) heuristic

First Results [Grebhahn et al., 2014]

Research questions:

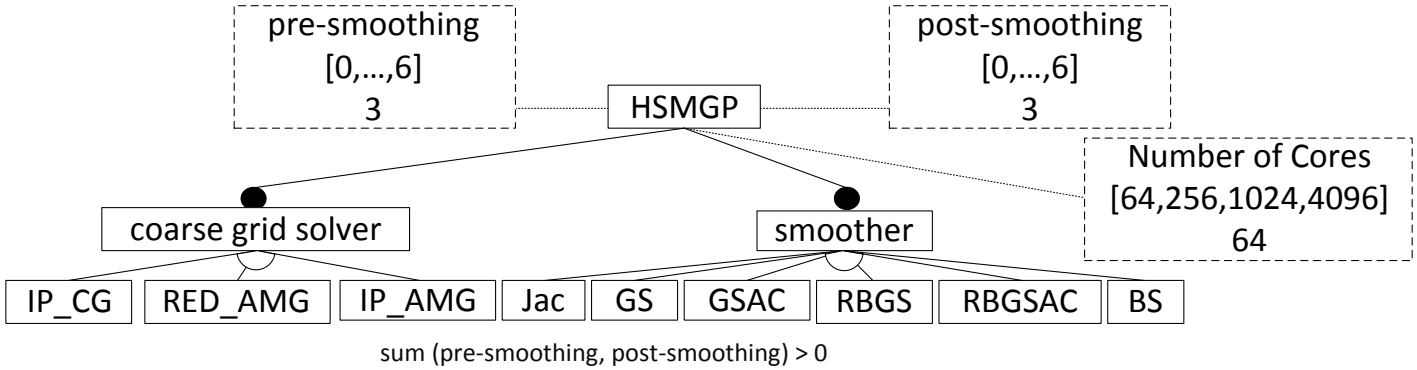
- What is the prediction accuracy of the different heuristics?
- Can we predict the performance-optimal configuration?

Customizable programs:

HIPACC Highly Scalable Multi-Grid Solver (HSMGS)

Multi-Grid Solver using DUNE (DUNE MGS)

HSMGS


pre-smoothing
[0,...,6]
3
HSMGP
post-smoothing
[0,...,6]
3
Number of Cores
[64,256,1024,4096]
64
coarse grid solver
smoother
IP_CG
RED_AMG
IP_AMG
Jac
GS
GSAC
RBGS
RBGSAC
BS
sum (pre-smoothing, post-smoothing) > 0

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF					
FW					
PW					
HO					
HS					
FL					

Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	3 456 (100)	0	0	0	1
FW					
PW					
HO					
HS					
FL					

Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	3 456 (100)	0	0	0	1
FW	26 (0.8)	23.4 ± 18.7	19.0	3.8	40
PW					
HO					
HS					
FL					

Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	3 456 (100)	0	0	0	1
FW	26 (0.8)	23.4 ± 18.7	19.0	3.8	40
PW	274 (7.9)	4.8 ± 8.6	1.8	31.4	77
HO					
HS					
FL					

Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	3 456 (100)	0	0	0	1
FW	26 (0.8)	23.4 ± 18.7	19.0	3.8	40
PW	274 (7.9)	4.8 ± 8.6	1.8	31.4	77
HO	1 331 (38.5)	60.7 ± 67.2	41.5	270.0	312
HS					
FL					

Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	3 456 (100)	0	0	0	1
FW	26 (0.8)	23.4 ± 18.7	19.0	3.8	40
PW	274 (7.9)	4.8 ± 8.6	1.8	31.4	77
HO	1 331 (38.5)	60.7 ± 67.2	41.5	270.0	312
HS	2 902 (84.0)	8.0 ± 33.9	0	270.0	55
FL					

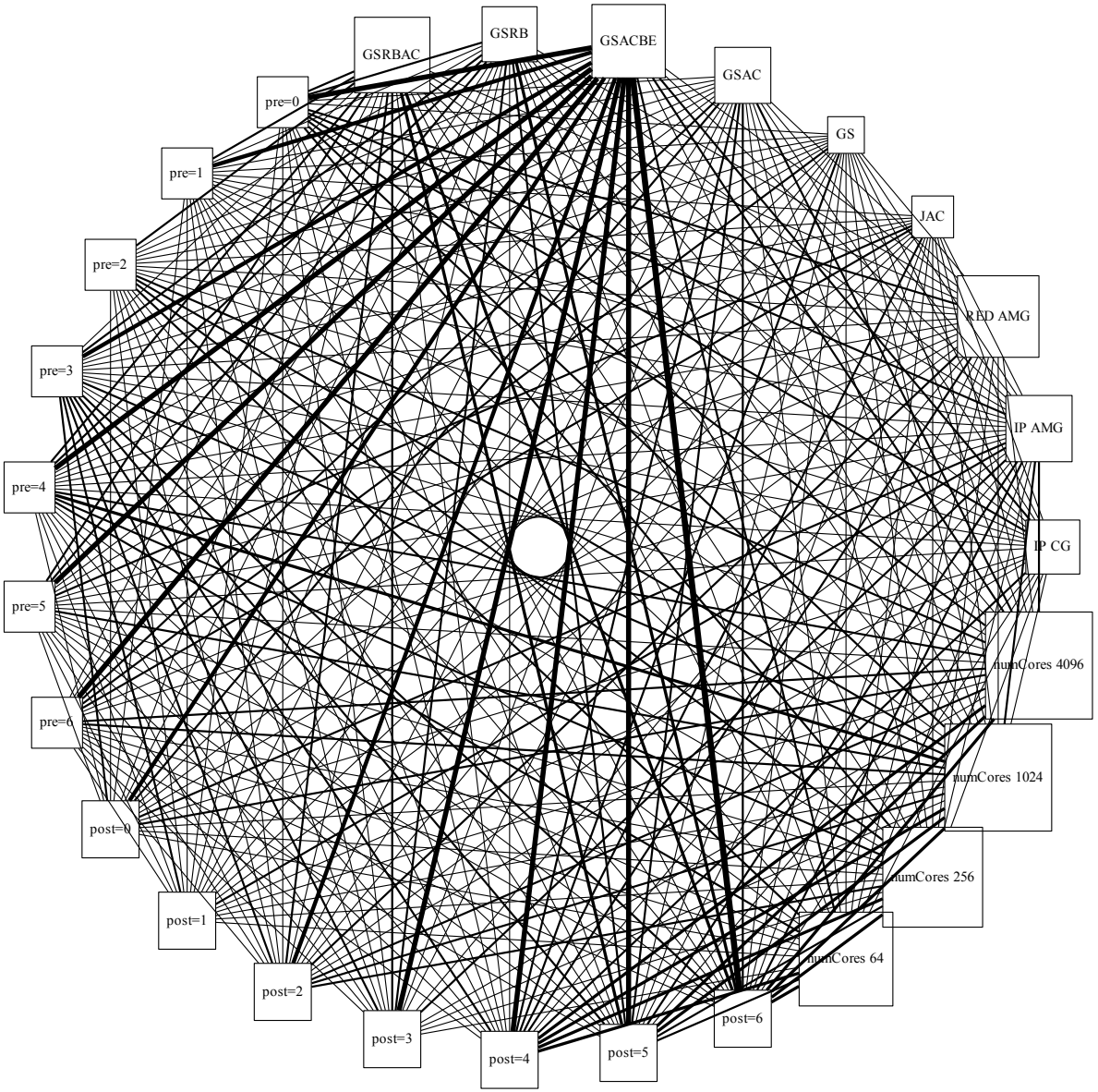
Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	3 456 (100)	0	0	0	1
FW	26 (0.8)	23.4 ± 18.7	19.0	3.8	40
PW	274 (7.9)	4.8 ± 8.6	1.8	31.4	77
HO	1 331 (38.5)	60.7 ± 67.2	41.5	270.0	312
HS	2 902 (84.0)	8.0 ± 33.9	0	270.0	55
FL	112 (3.2)	2.5 ± 3.1	1.8	0	1

Table: BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HSMGS – Feature Interactions (Pair-Wise)



HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF					
	HO					
	HS					
	FL					
DUNE MGS	BF					
	HO					
	HS					
	FL					

HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO					
	HS					
	FL					
DUNE MGS	BF					
	HO					
	HS					
	FL					

HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS					
	FL					
DUNE MGS	BF					
	HO					
	HS					
	FL					

HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
	FL					
DUNE MGS	BF					
	HO					
	HS					
	FL					

HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
	FL	216 (1.6)	32.9 ± 31.1	23.5	0.75	3544
DUNE MGS	BF					
	HO					
	HS					
	FL					

HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
	FL	216 (1.6)	32.9 ± 31.1	23.5	0.75	3544
DUNE MGS	BF	2 304 (100)	0	0	0	1
	HO					
	HS					
	FL					

HIPA^{cc}, DUNE MGS – Results

	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
	FL	216 (1.6)	32.9 ± 31.1	23.5	0.75	3544
DUNE MGS	BF	2 304 (100)	0	0	0	1
	HO	749 (32.6)	36.3 ± 51.7	18.2	226.9	133
	HS					
	FL					

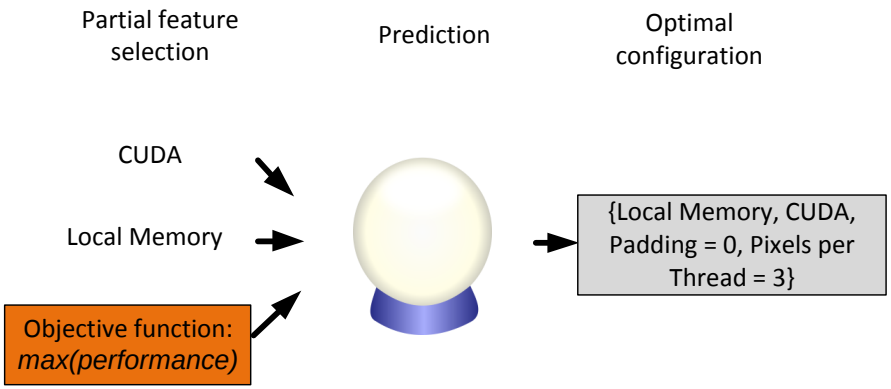
HIPA^{cc}, DUNE MGS – Results

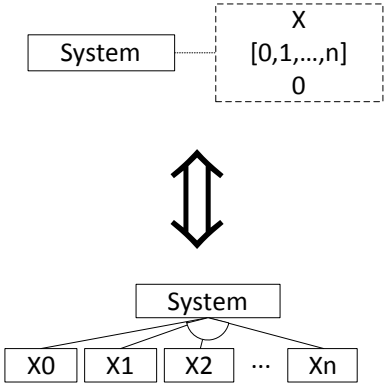
	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
	FL	216 (1.6)	32.9 ± 31.1	23.5	0.75	3544
DUNE MGS	BF	2 304 (100)	0	0	0	1
	HO	749 (32.6)	36.3 ± 51.7	18.2	226.9	133
	HS	1 643 (71.4)	49 ± 164.7	0	161.4	215
	FL					

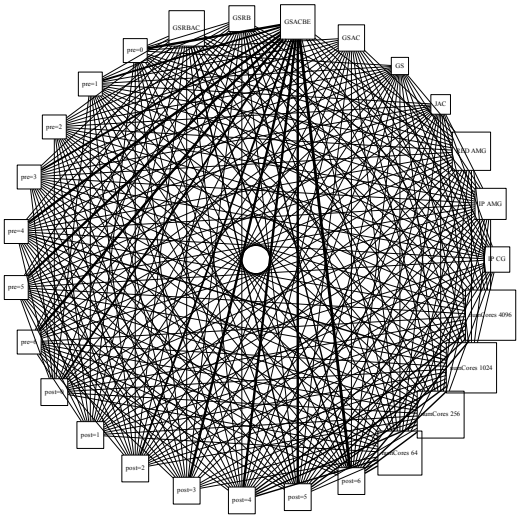
HIPA^{cc}, DUNE MGS – Results

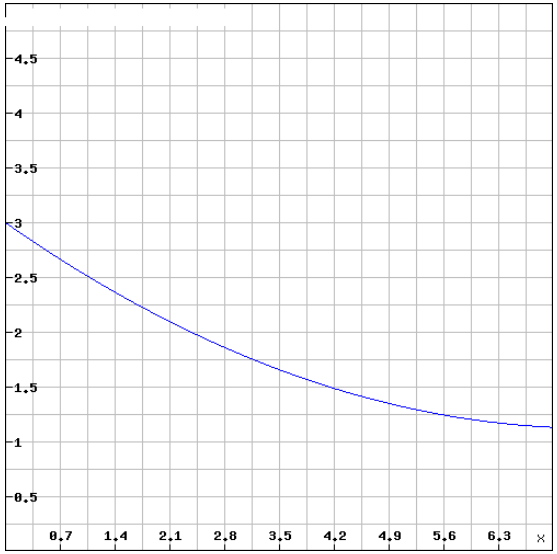
	Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
HIPA^{cc}	BF	13 485 (100)	0	0	0	1
	HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
	HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
	FL	216 (1.6)	32.9 ± 31.1	23.5	0.75	3544
DUNE MGS	BF	2 304 (100)	0	0	0	1
	HO	749 (32.6)	36.3 ± 51.7	18.2	226.9	133
	HS	1 643 (71.4)	49 ± 164.7	0	161.4	215
	FL	75 (3.3)	13.7 ± 12.6	10.2	48.5	10

Conclusion


Partial feature selection
Prediction
Optimal configuration
CUDA
Local Memory
Objective function: *max(performance)*
{Local Memory, CUDA, Padding = 0, Pixels per Thread = 3}


System
X
[0,1,...,n]
0
System
X0
X1
X2
...
Xn




4.5
4
3.5
3
2.5
2
1.5
1
0.5
0.7
1.4
2.1
2.8
3.5
4.2
4.9
5.6
6.3
x

Future Work

- Interactions between parameters
- Use domain knowledge
  - Degree of the function
  - Known interactions and dependencies between features and parameters
- Exploit Multi-Grid characteristics
  - Different configuration options are used in different computation parts

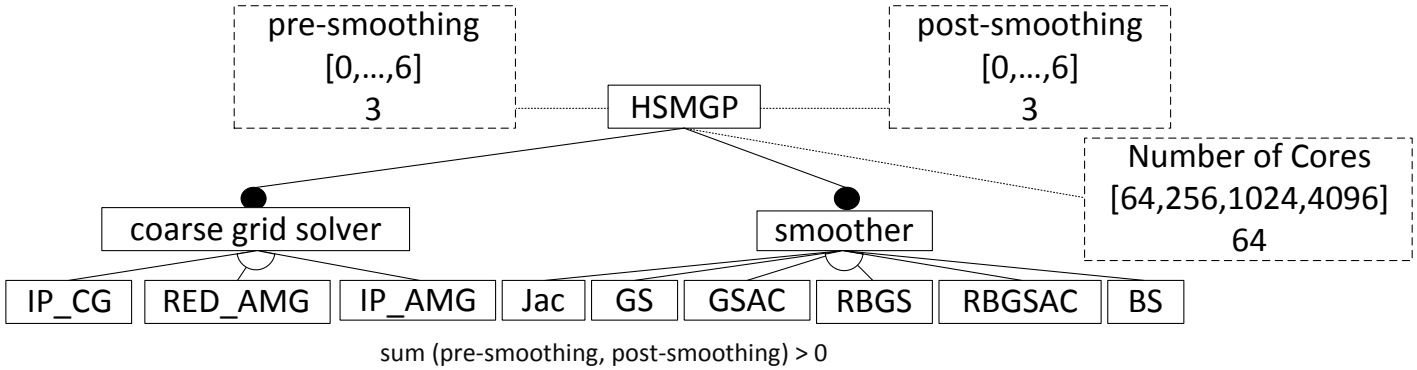
Questions

grebhahn@fim.uni-passau.de

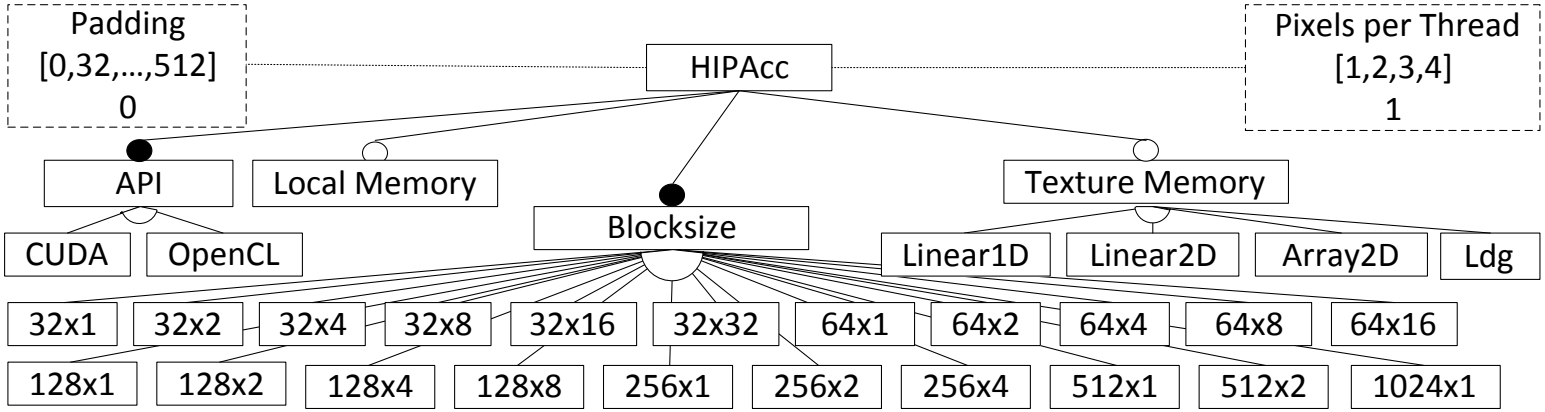
References

- Grebhahn, A., Kuckuk, S., Schmitt, C., Köstler, H., Siegmund, N., Apel, S., Hannig, F., and Teich, J. (2014).
  Experiments on Optimizing the Performance of Stencil Codes with SPL Conqueror.
  submitted to Parallel Processing Letters.
- Siegmund, N., Kolesnikov, S. S., Kästner, C., Apel, S., Batory, D., Rosenmüller, M., and Saake, G. (2012).
  Predicting performance via automated feature-interaction detection.
  In *Proc. ICSE*, pages 167–177. IEEE.

HSMGS


pre-smoothing
[0,...,6]
3
HSMGP
post-smoothing
[0,...,6]
3
Number of Cores
[64,256,1024,4096]
64
coarse grid solver
smoother
IP_CG
RED_AMG
IP_AMG
Jac
GS
GSAC
RBGS
RBGSAC
BS
sum (pre-smoothing, post-smoothing) > 0

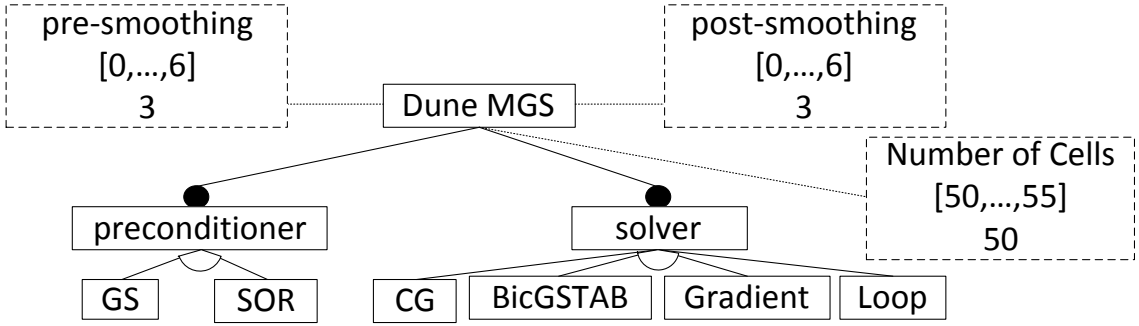
HIPA^{cc}



¬(Local Memory ^ 1024x1 ^ Pixel Per Thread = 2)
¬(Local Memory ^ 32x32 ^ Pixel Per Thread = 3)
¬(Local Memory ^ 64x16 ^ Pixel Per Thread = 3)
¬(Local Memory ^ 1024x1 ^ Pixel Per Thread = 3)
¬(Local Memory ^ 32x32 ^ Pixel Per Thread = 4)
¬(Local Memory ^ 64x16 ^ Pixel Per Thread = 4)

¬(Local Memory ^ 128x8 ^ Pixel Per Thread = 3)
¬(Local Memory ^ 256x4 ^ Pixel Per Thread = 3)
¬(Local Memory ^ 512x2 ^ Pixel Per Thread = 3)
¬(Local Memory ^ 128x8 ^ Pixel Per Thread = 4)
¬(Local Memory ^ 256x4 ^ Pixel Per Thread = 4)
¬(Local Memory ^ 512x2 ^ Pixel Per Thread = 4)
¬(Local Memory ^ 1024x1 ^ Pixel Per Thread = 4)
(Array2D $\Longrightarrow$ Padding = 0)

DUNE MGS



sum (pre-smoothing, post-smoothing) > 0

HIPA^{cc} – Results

Heu.
BF
FW
PW
HO
HS
FL

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HIPA^{cc} – Results

Heu.	# M (in %)
BF	13 485 (100)
FW	47 (0.3)
PW	702 (5.2)
HO	1 516 (11.2)
HS	2 881 (21.4)
FL	216 (1.6)

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HIPA^{cc} – Results

Heu.	# M (in %)	$\bar{e} \pm s$
BF	13 485 (100)	0
FW	47 (0.3)	80.8 ± 56.3
PW	702 (5.2)	17.2 ± 16.0
HO	1 516 (11.2)	7.8 ± 10.1
HS	2 881 (21.4)	3.8 ± 4.8
FL	216 (1.6)	32.9 ± 31.1

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HIPA^{cc} – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$
BF	13 485 (100)	0	0
FW	47 (0.3)	80.8 ± 56.3	73.6
PW	702 (5.2)	17.2 ± 16.0	13.4
HO	1 516 (11.2)	7.8 ± 10.1	4.4
HS	2 881 (21.4)	3.8 ± 4.8	3.3
FL	216 (1.6)	32.9 ± 31.1	23.5

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HIPA^{cc} – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]
BF	13 485 (100)	0	0	0
FW	47 (0.3)	80.8 ± 56.3	73.6	1.50
PW	702 (5.2)	17.2 ± 16.0	13.4	14.60
HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38
HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22
FL	216 (1.6)	32.9 ± 31.1	23.5	0.75

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

HIPA^{cc} – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	13 485 (100)	0	0	0	1
FW	47 (0.3)	80.8 ± 56.3	73.6	1.50	2180
PW	702 (5.2)	17.2 ± 16.0	13.4	14.60	428
HO	1 516 (11.2)	7.8 ± 10.1	4.4	9.38	1735
HS	2 881 (21.4)	3.8 ± 4.8	3.3	18.22	955
FL	216 (1.6)	32.9 ± 31.1	23.5	0.75	3544

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

DUNE MGS – Results

Heu.
BF
FW
PW
HO
HS
FL

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

DUNE MGS – Results

Heu.	# M (in %)
BF	2 304 (100)
FW	25 (1.1)
PW	191 (8.3)
HO	749 (32.6)
HS	1 643 (71.4)
FL	75 (3.3)

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

DUNE MGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$
BF	2 304 (100)	0
FW	25 (1.1)	32.2 ± 51
PW	191 (8.3)	42.2 ± 38.6
HO	749 (32.6)	36.3 ± 51.7
HS	1 643 (71.4)	49 ± 164.7
FL	75 (3.3)	13.7 ± 12.6

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

DUNE MGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$
BF	2 304 (100)	0	0
FW	25 (1.1)	32.2 ± 51	12.4
PW	191 (8.3)	42.2 ± 38.6	32.8
HO	749 (32.6)	36.3 ± 51.7	18.2
HS	1 643 (71.4)	49 ± 164.7	0
FL	75 (3.3)	13.7 ± 12.6	10.2

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

DUNE MGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]
BF	2 304 (100)	0	0	0
FW	25 (1.1)	32.2 ± 51	12.4	55.3
PW	191 (8.3)	42.2 ± 38.6	32.8	97.3
HO	749 (32.6)	36.3 ± 51.7	18.2	226.9
HS	1 643 (71.4)	49 ± 164.7	0	161.4
FL	75 (3.3)	13.7 ± 12.6	10.2	48.5

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning

DUNE MGS – Results

Heu.	# M (in %)	$\bar{e} \pm s$	$\tilde{x}$	δ [%]	rank
BF	2 304 (100)	0	0	0	1
FW	25 (1.1)	32.2 ± 51	12.4	55.3	498
PW	191 (8.3)	42.2 ± 38.6	32.8	97.3	273
HO	749 (32.6)	36.3 ± 51.7	18.2	226.9	133
HS	1 643 (71.4)	49 ± 164.7	0	161.4	215
FL	75 (3.3)	13.7 ± 12.6	10.2	48.5	10

Table : BF: brute force, FW: feature-wise, PW: pair-wise, HO: higher-order, HS: hot-spot, FL: function learning