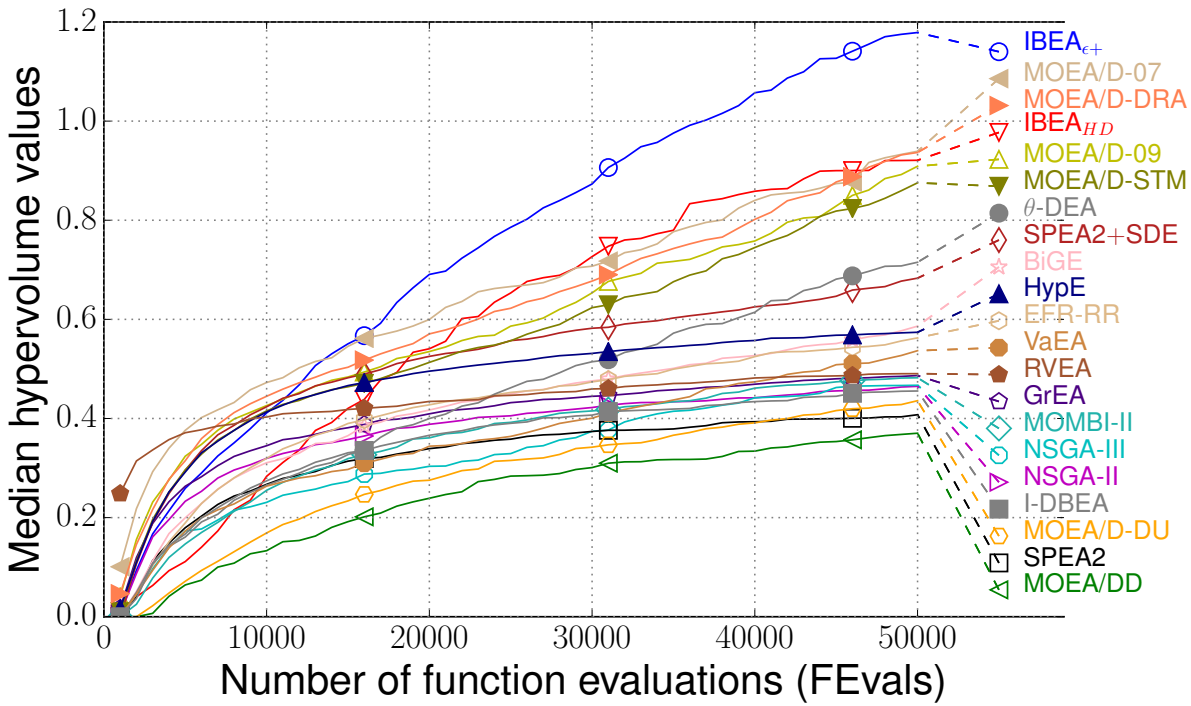
Benchmarking MOEAs for Multi- and Many-objective Optimization Using an Unbounded External Archive

Ryoji Tanabe (SUSTech) and Akira Oyama (JAXA)



A large number of MOEAs for MaOPs have been proposed

MaOPs: Many-objective Optimization Problems

- Multi-objective problems (MOPs) with more than 4-objectives
- MaOPs frequently appear in engineering applications
  - E.g., Radar waveform design problems [Hughes EMO'07], ...

Classical MOEAs do not perform well on MaOPs [Wagner EMO'07]

- E.g., NSGA-II [Deb TEVC'02], SPEA2 [Zitzler 01], ...

A large number of MOEAs for MaOPs have been proposed

- GrEA [Yang TEVC'13], NSGA-III [Deb TEVC'14], θ-DEA [Yuan TEVC'16], ...
- MOEA/DD [Li TEVC'15], I-DBEA [Asafuddoula TEVC'15], ...
- MOMBI-II [Gómez GECCO'15], BiGE [Li AIJ'15], ...

- A novel state-of-the-art MOEA is proposed for every week

Which MOEA is best?

This presentation: Benchmarking study of 21 MOEAs

An exhaustive benchmarking study has never been performed

- The performance of MOEAs has not been well understood
- In many cases, MOEAs were evaluated under a final population scenario, which has several critical issues

The anytime performance of 21 MOEAs is investigated under an Unbounded External Archive (UEA) scenario

- The 21 MOEAs = {NSGA-II, NSGA-III, ..., VaEA}
- The UEA stores all nondominated solutions found
  - The UEA can be introduced into all MOEAs with no changes in the original algorithmic framework

Our results will be helpful to both users and algorithm designers

- Users want to apply the best MOEA to real-world problems
- Designers develop novel MOEAs by improving good MOEAs
  - All of the experimental data are available on:
  - https://sites.google.com/site/benchmarkingmoeas/

Three optimization scenarios to evaluate the performance of MOEAs

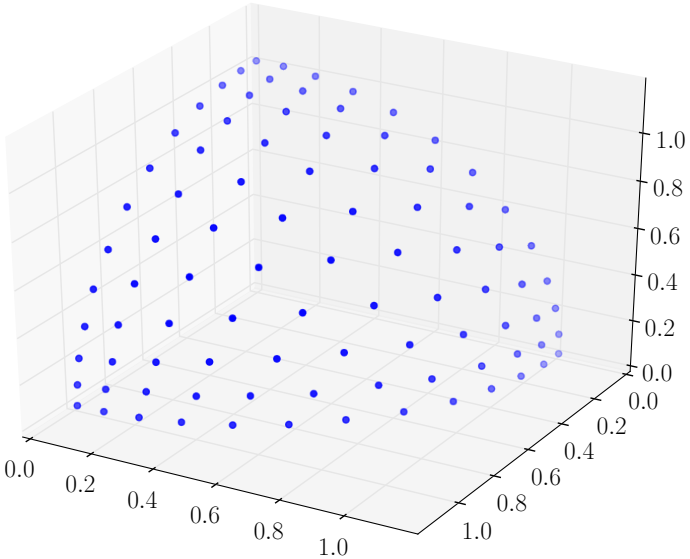
1. Final population scenario
   - All nondominated solutions in the final population are used
   - Traditional frequently-used performance evaluation scenario

2. UEA scenario (the UEA stores all nondominated solutions)
   - All nondominated solutions in the UEA are used
   - In some applications, it is important to know the exact shape of the entire PF in detail
   - The size of the UEA obtained by each MOEA is different

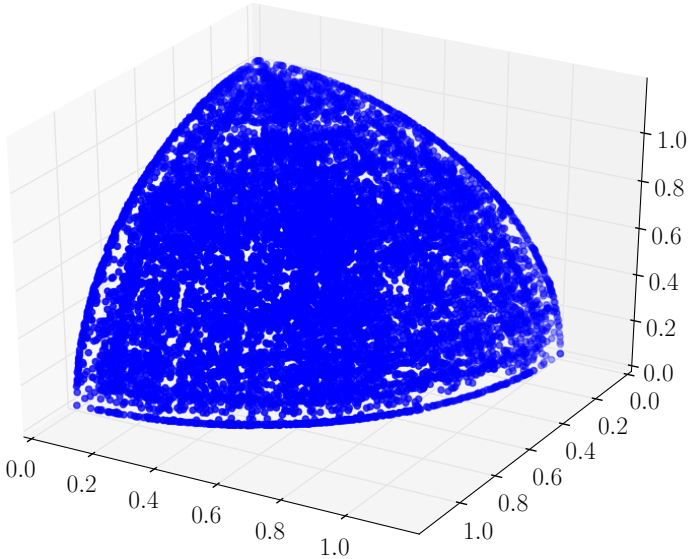
3. Reduced UEA scenario
   - A pre-specified number of selected nondominated solutions from the UEA are used for the performance assessment
   - Users often want to know only a small number of well-distributed solutions
   - This scenario is to compare MOEAs using the obtained solution sets of the same size, in contrast to the UEA scenario

Distribution of nondominated solutions under each scenario (NSGA-III)

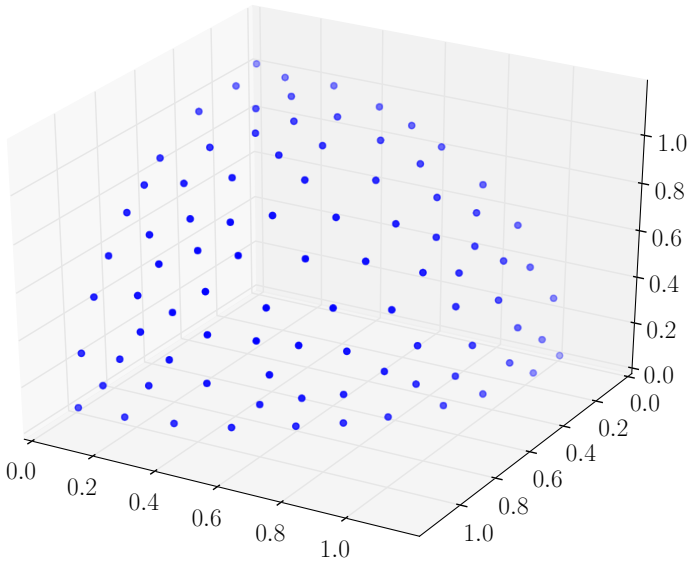
1. Final population scenario



2. UEA scenario



3. Reduced UEA scenario



Two problems of the traditional, final population scenario

Note: many previous work have already pointed out the problems

- E.g., [Nebro PPSN'08], [Bringmann PPSN'14], [Brockhoff GECCO'15]

P1. The performance rank strongly depends on the setting of the maximum number of function evaluations ($\text{FEvals}^{\max}$)

- An experiment with $\text{FEvals}^{\max} = 5,000$ shows:
  - "MOEA1 performs better than MOEA2"
- This conclusion cannot be applied when $\text{FEvals}^{\max} = 2,500$
- Solution1: Measure the anytime performance

P2. Only a set of nondominated solutions with the population size can be stored in the population

- When the number of nondominated solutions found exceeds the population size, they are removed from the population
- A good potential solution is discarded [Bringmann PPSN'14]
- Solution2: Use the unbounded external archive (UEA)

The UEA and reduced UEA scenarios are more reasonable?

Our benchmarking study vs. previous benchmarking studies

The first study that evaluates the anytime performance of a number of MOEAs on MaOPs with up to 6 objectives

Articles	Anytime performance?	UEA?	# of objectives	# of MOEAs
Wagner EMO'07			$3, 4, 5, 6$	8
Nebro PPSN'08	✓		2	6
Hadka ECJ'12	✓		$2, 4, 6, 8$	9
Li EMO'13			$5, 10$	8
Bringmann PPSN'14	✓	✓	2	4
Brockhoff GECCO'15	✓	✓	2	3
Ishibuchi CEC'16		✓	$3, 5$	4
Ma CEC'16			$3, 4, 6, 8$	7
Maltese TEVC'17			$3, 5, 8, 10, 15$	9
GECCO BBOB WS	✓	✓	2	∞
Our study	✓	✓	$2, 3, 4, 5, 6$	21

Experimental settings

Benchmark problems: Nine *normalized* WFG functions [Huband 06]

- To remove the effect of the normalization strategies
- The number of objectives $M \in \{2, 3, 4, 5, 6\}$
- The number of runs $= 21$
- The maximum number of function evaluations $= 50,000$

Performance indicator: The hypervolume (HV) indicator

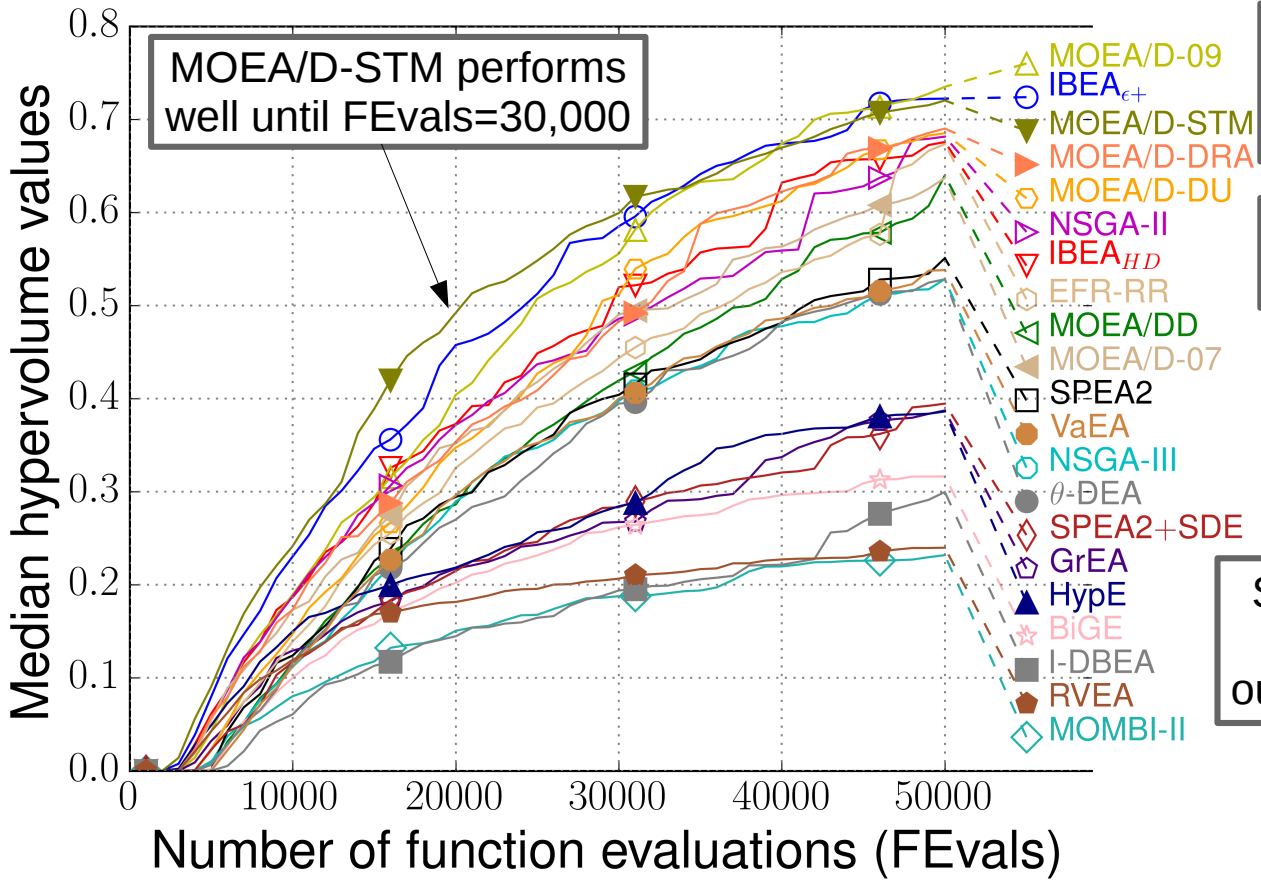
- The WFG algorithm [While TEVC'12] was used for the calculation

Settings for the 21 MOEAs

- SBX and polynomial mutation were used for all 21 MOEAs
  - MOEA/D-09, -DRA, -STM originally use DE operators
  - Their DE operators were replaced by the GA operators
- Source codes downloaded from author's websites were used for most of the 21 MOEAs
- The pop. size (M): $100(2)$, $91(3)$, $220(4)$, $210(5)$, $182(6)$

Anytime performance of the MOEAs on the 2-objective WFG1 function

- MOEA/D-STM, MOEA/D-09, and IBEA$_{\epsilon+}$ perform well



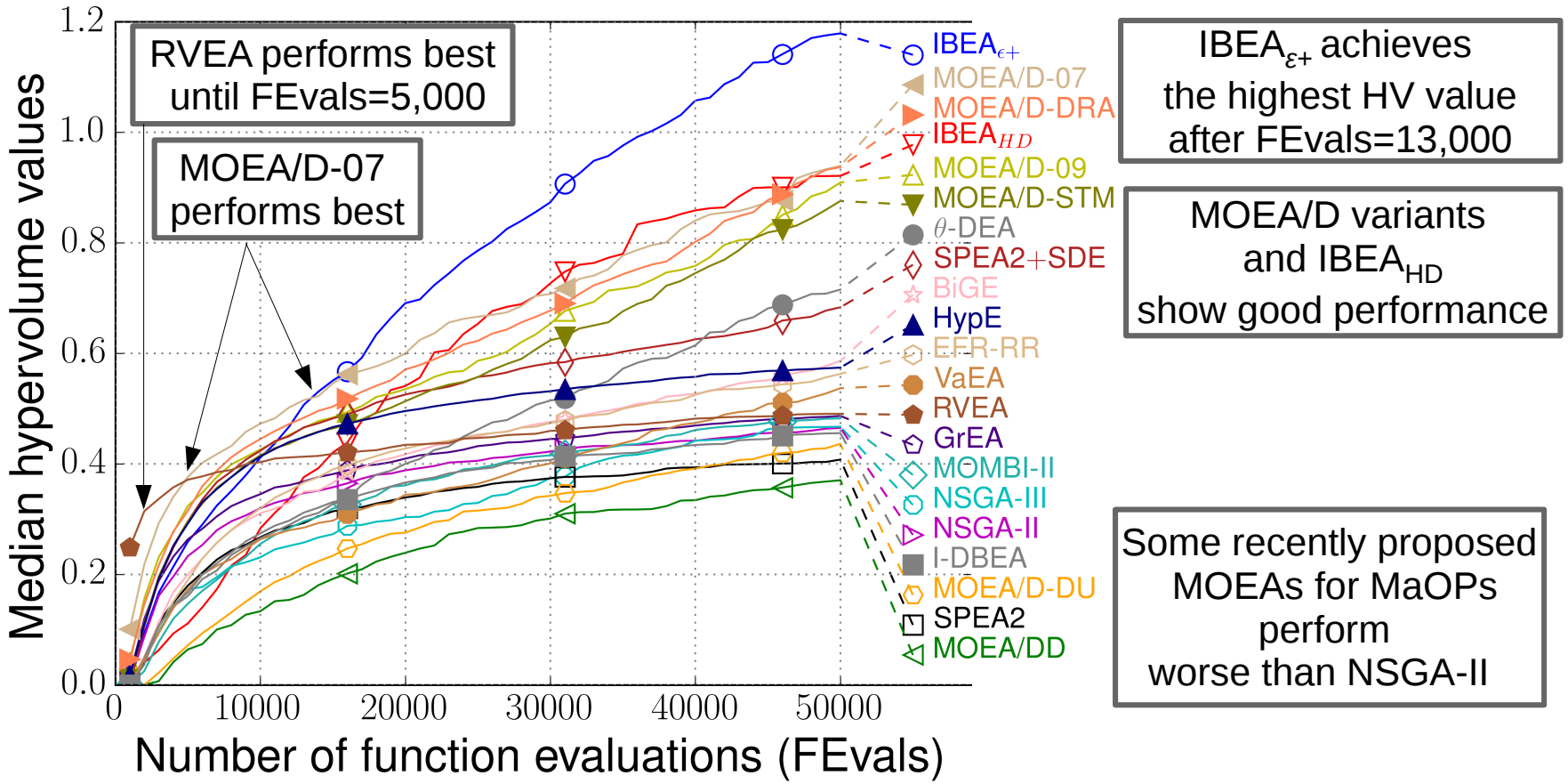
MOEA/D-09 and IBEA$_{\epsilon+}$ perform well after FEvals=40,000

NSGA-II shows good performance

Some recently proposed MOEAs are outperformed by NSGA-II

Anytime performance of the MOEAs on the 6-objective WFG1 function

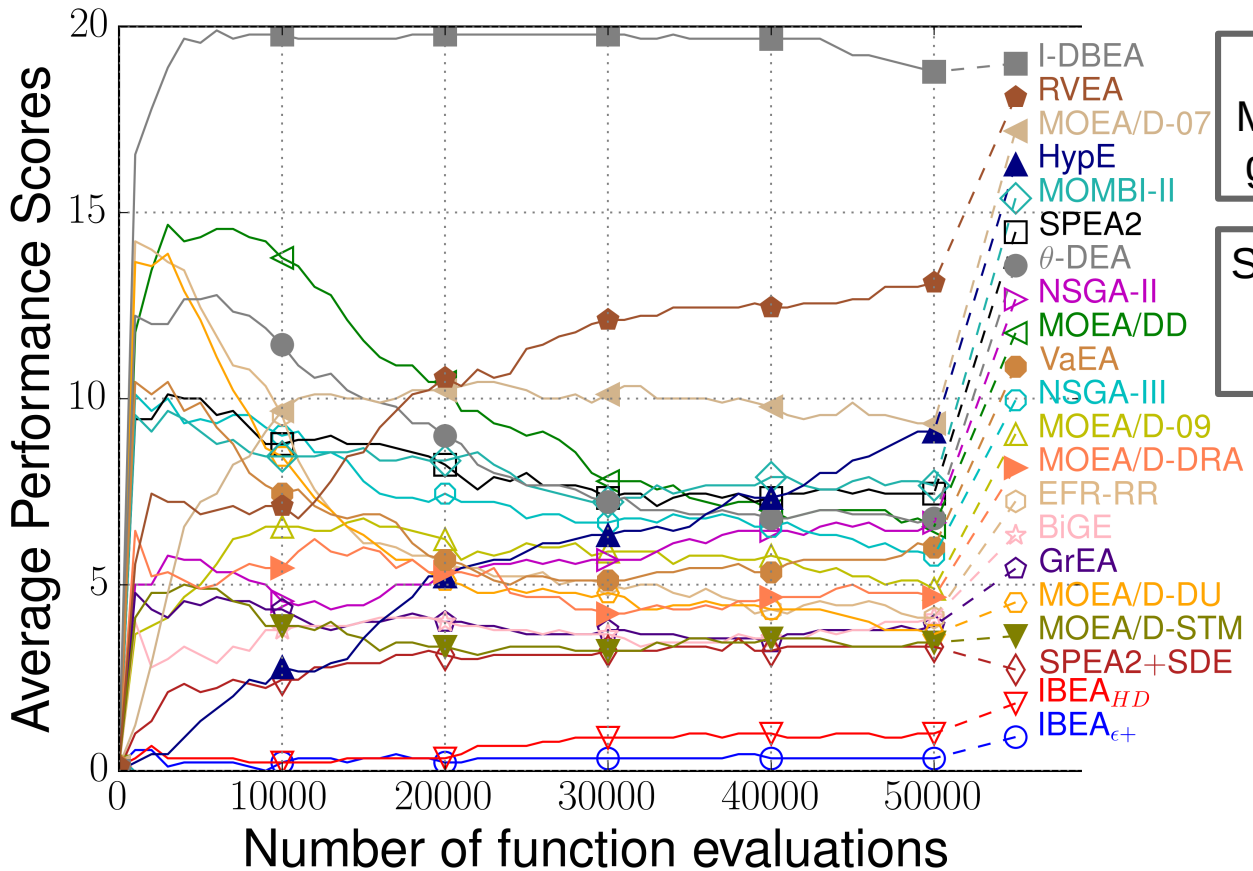
- RVEA, MOEA/D-07, and IBEA$_{\epsilon+}$ perform well



Overall performance of the 21 MOEAs on the 2-objective 9 WFGs

Average Performance Score (APS) [Bader ECJ'11]

- The APS value of one MOEA is determined by comparing it with the remaining MOEAs (with the Wilcoxon rank-sum test)
- A small APS value indicates that its performance is good



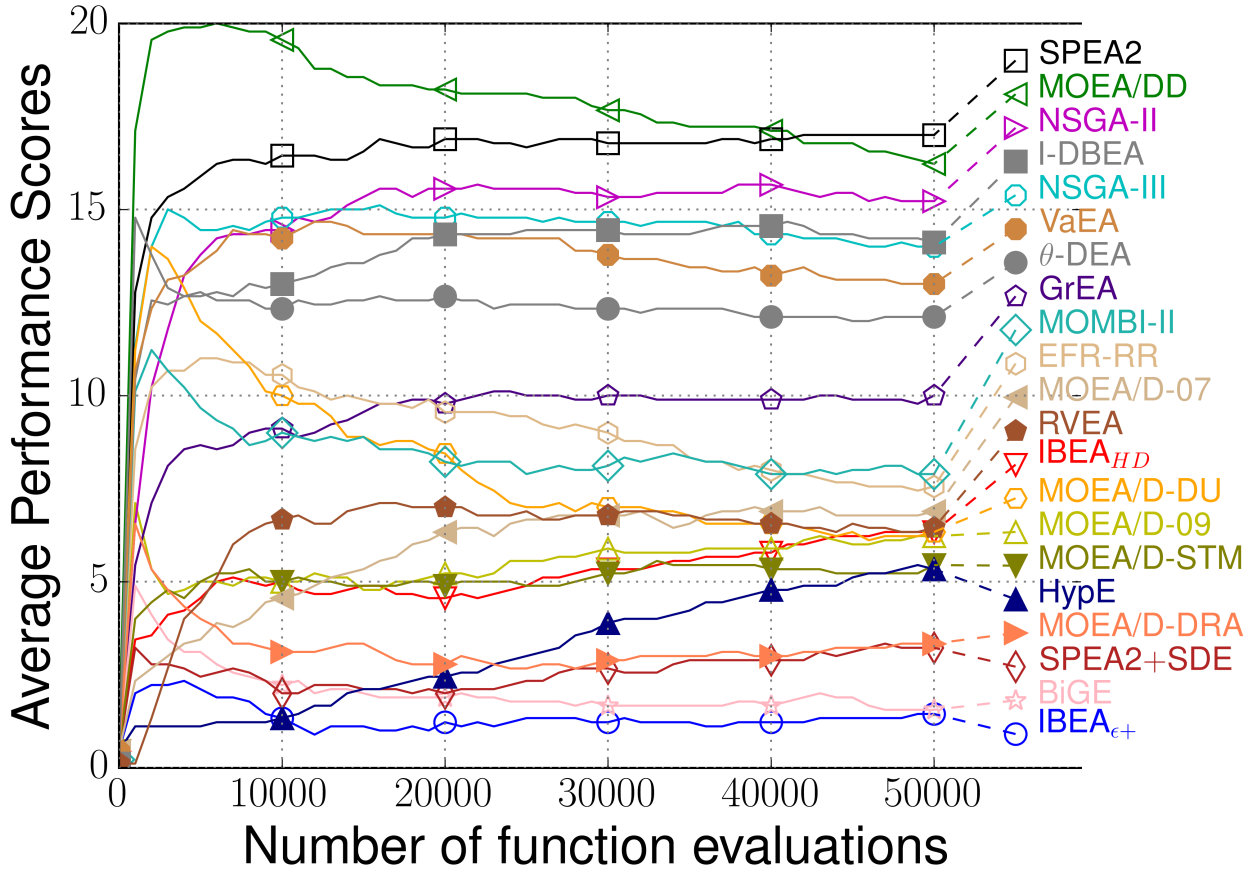
The APS value of MOEA/D-07 and HypE gradually deteriorates

Some recently proposed MOEAs perform worse than NSGA-II

SPEA2+SDE and MOEA/D-STM perform well

The two IBEAs are the best MOEA

Overall performance of the 21 MOEAs on the 6-objective nine WFGs



RVEA performs best Until FEvals = 2,000

HypE performs best Until FEvals = 10,000

IBEA$_{\epsilon+}$ and BiGE performs well

Why do some latest MOEAs perform poorly in our study?

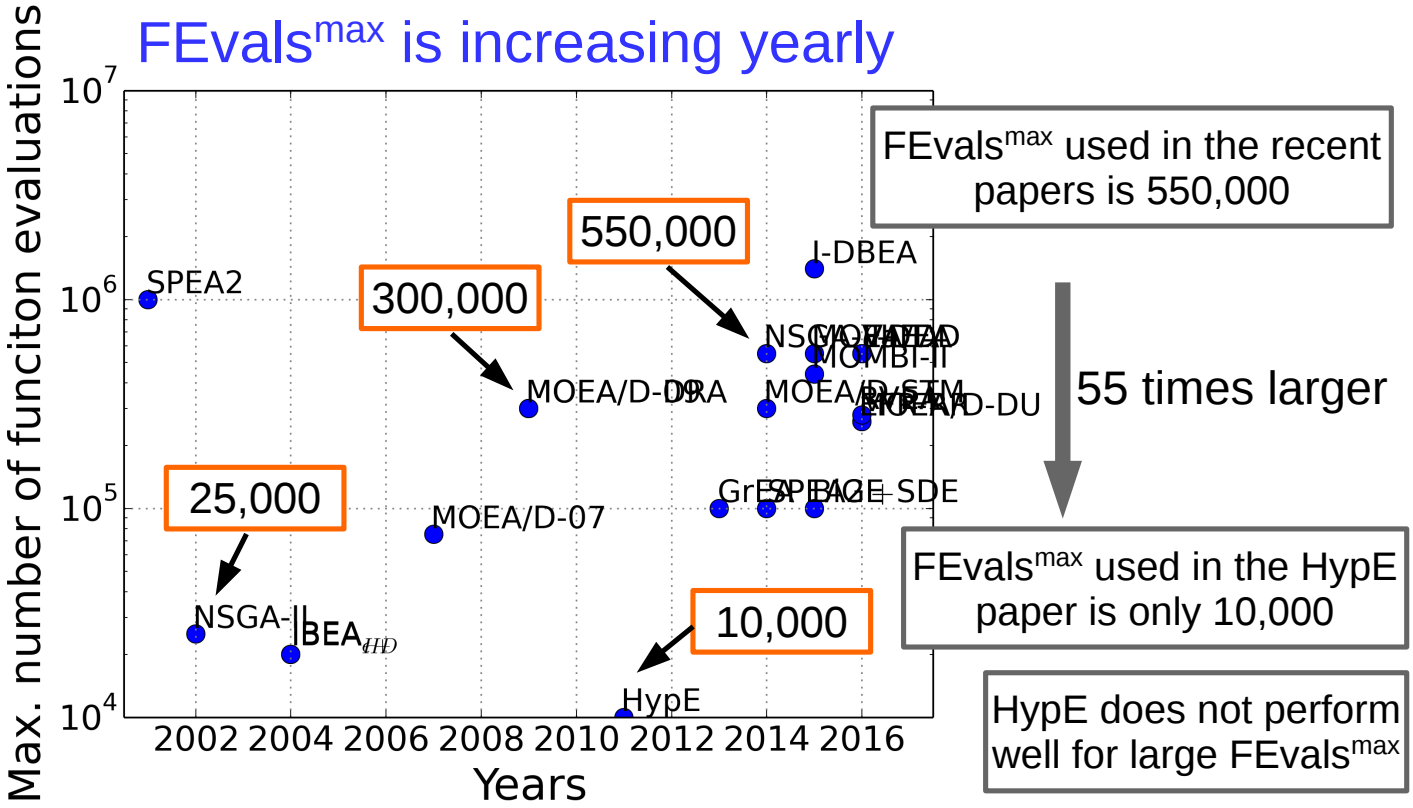
The "best" MOEA is dependent on functions, M, and FEvals

- Some classical MOEAs (e.g., IBEA$_{\epsilon+}$, HypE) perform well
- The performance of some latest MOEAs are not so good

Two reasons

- Reason1: The small setting of FEvals$^{\max}$ $(50,000)$
  - I do not think FEvals$^{\max} = 50,000$ is small
  - In practice, it is difficult to set FEvals$^{\max} \geq 50,000$
- Reason2: The unbounded external archive (UEA) was used

Reason1: The small setting of FEvals$^{\max}$ $(50,000)$



Why does HypE perform well in our benchmarking study?

- Many articles report the poor performance of HypE
- This is because of the large setting of FEvals$^{\max}$

Latest MOEAs were designed for optimization for FEvals$^{\max} = 50,000$

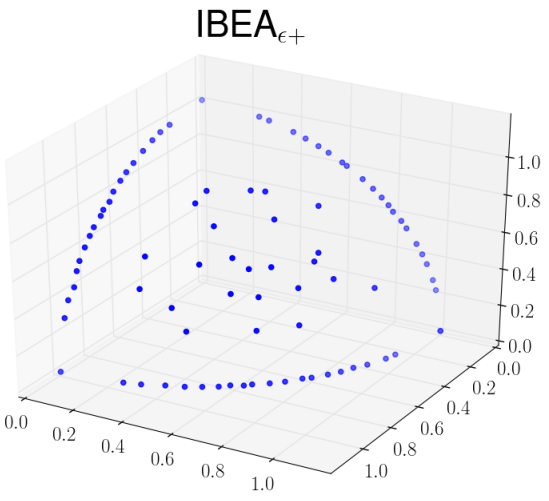
- The anytime performance of latest MOEAs is poor?

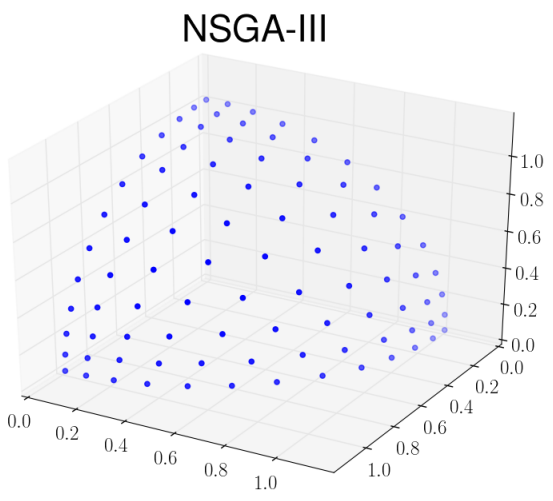
Reason2: Incorporation of the UEA is beneficial for classical MOEAs?

The distribution of solutions in the population and UEA

Final pop. scenario

Solutions obtained by IBEA are biased to specific regions

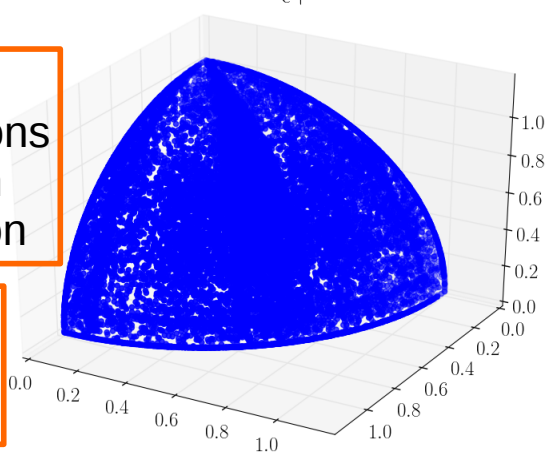


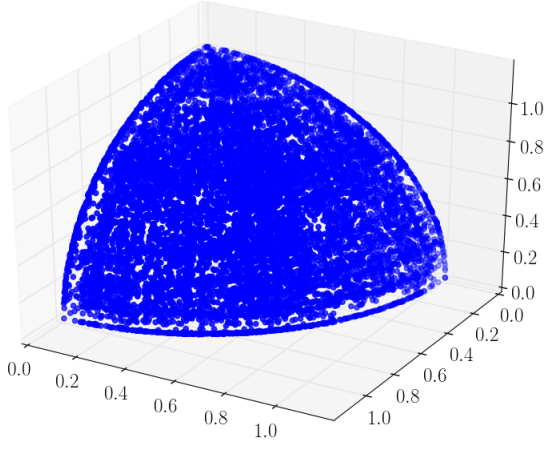


UEA scenario

IBEA can generate well-distributed solutions but cannot maintain them in the population

This problem is addressed by incorporating the UEA





Our performance evaluation method vs. the COCO software [Hansen 16]

Unbeautiful tricks were used in our benchmarking study

- Computational cost for postprocessing is very expensive
- About 1 month was taken using a 6-core Xeon machine
- *Beautiful* tricks for postprocessing of MaOPs are nessesarry

	The COCO software	Our approach
Functions	BBOB-biobj [Tusar 16]	WFG [Huband 06]
# of objectives M	2	$2, 3, 4, 5, 6$
Scalability to dim. D	Yes	No
UEA update	On-line	Off-line
HV calculation	Immediately	Periodically $1000, 2000, \ldots$
Performance profile	Exact	Inexact
$\mathrm{FEvals}^{\max}$	$1 \sim 10,000,000$	$1 \sim 50,000$

Conclusion

The anytime performance of the 21 MOEAs was exhaustively investigated under the Unbounded External Archive (UEA) scenario

- In many cases, the performance of MOEAs for MOPs and MaOPS was evaluated under the final population scenario

Our results are significantly different from the results reported in previous studies under the final population scenario

- The performance of some recently proposed MOEAs is not as good as some classical MOEAs
- RVEA, IBEA$_{\epsilon+}$, HypE, BiGE, and some MOEA/Ds

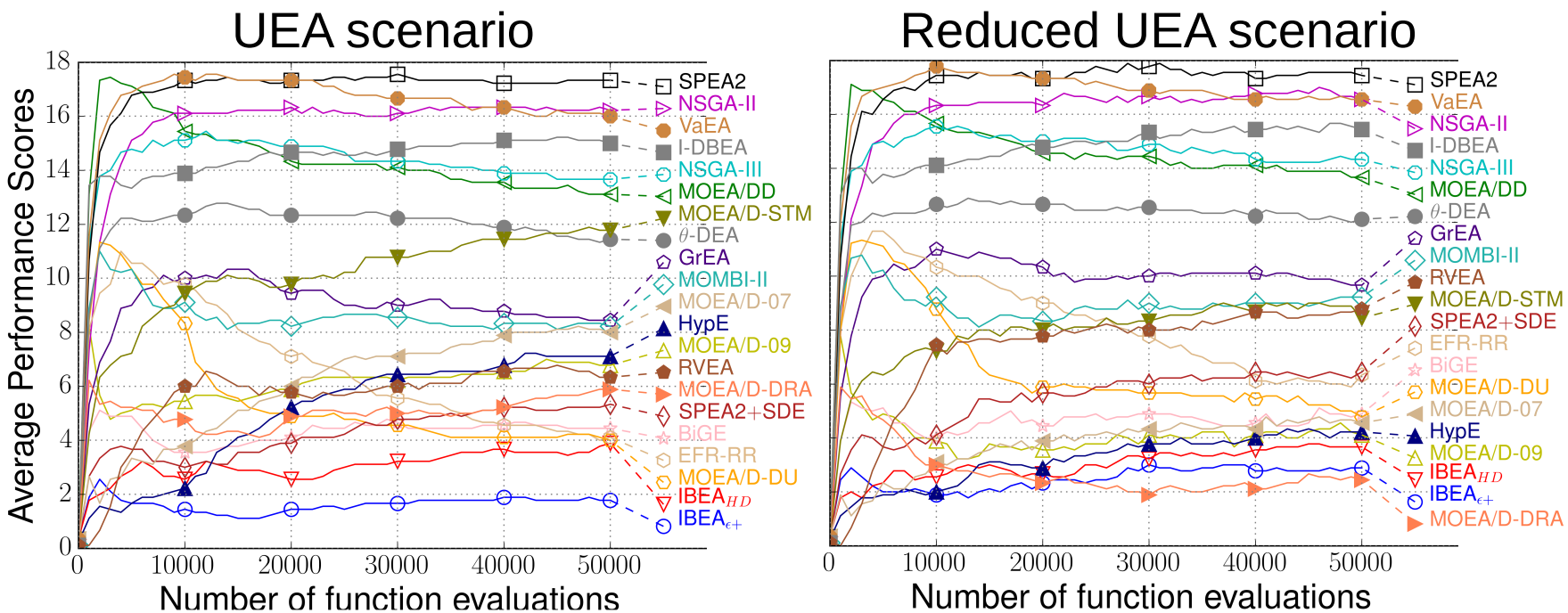
Many future work

- Benchmarking with FEvals$^{\max} > 50,000$
- Benchmarking with other problems with $M \geq 7$
  - Constrained, scaled, real-world MOPs
- Comparing MOEAs with tuned control parameters

UEA scenario vs. Reduced UEA scenario

How are the results affected depending on optimization scenarios?

The anytime performance based on the APS values on the nine WFGs with $M = 5$ under the UEA and reduced UEA scenarios



- Relative performance of some MOEA/D variants are improved under the reduced UEA scenario
- The relative performance of dominance based MOEA are almost same under both optimization scenarios

Why we used the normalized WFG functions

"MOEA" = a complex algorithm that consists of multiple components

- Parent selection methods
- Environmental selection methods
- Variation operators
- Normalization strategies, etc.

We want to focus on the parent and environmental selection methods

- Variation operators are not main algorithmic components
- Normalization strategies are not main algorithmic components
- We want to remove the effect of variation operators and normalization strategies

On computational cost for postprocessing for the UEA and reduced UEA scenarios

We tested 21 MOEAs for 9 functions with $M \in \{2, 3, 4, 5, 6\}$

- TOTALLY, about 1 month was taken
- NOT, for one algorithm for one problem instance

In general, real-world problems require the execution of a simulation

- It takes a long time to evaluate the solution
  - Aerodynamic wing design problem: 11 minutes [Ong 03]
  - Integrated circuit design problem: 10 12 minutes [Liu 14]
  - Car rear design problem: 1 hour [Le 13]
- Computational cost for postprocessing can be ignored for real-world problems

Descriptions of experimental results

The traditional description

- Algorithm1 is better than Algorithm2

Our description

- Algorithm1 is better than Algorithm2 at FEvals = Y

Ideal description

- Algorithm1 is X times better than Algorithm2 at FEvals = Y