

# Design of Experiments in Holistic CPES Testing

Lecture at ERIGrid Summerschool  
Risø, Denmark, August 28 2018

Cornelius Steinbrink  
OFFIS – Institute for Information Technology  
R&D Division Energy  
Oldenburg, Germany



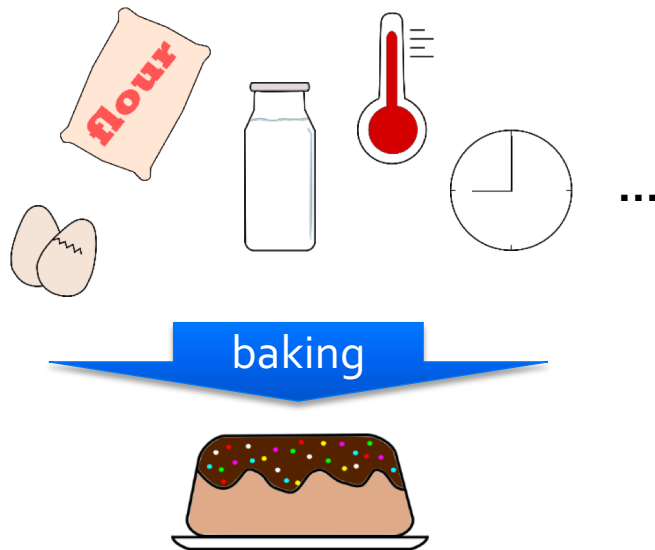
# What will we learn today?

## Agenda

- ▶ The **What** and **Why** of *statistical design of experiments*
- ▶ History
- ▶ Vocabulary
- ▶ Steps to planning and conducting experiments
- ▶ Classical vs modern DOE
- ▶ DOE & HTD
- ▶ Hands-on work with mosaik (co-simulation)

# What is the problem?

How do I get the perfect cake?



The perfect cake...

- ▶ ... look?
- ▶ ... texture?
- ▶ ... taste?

- ▶ Math view:

$$y_1, \dots, y_n = f_{cake}(x_1, \dots, x_n) + \varepsilon$$

- ▶ Find  $x_1, \dots, x_n$  to optimize  $y_i$
- ▶ Simple analytical solution?
- ▶ → No, this is the real world
- ▶ Cake formula unknown or too complex
- ▶ Too many ingredients and/or quality factors
- ▶ Fluctuation/error  $\varepsilon$  unknown

# We are not bakers but engineers!

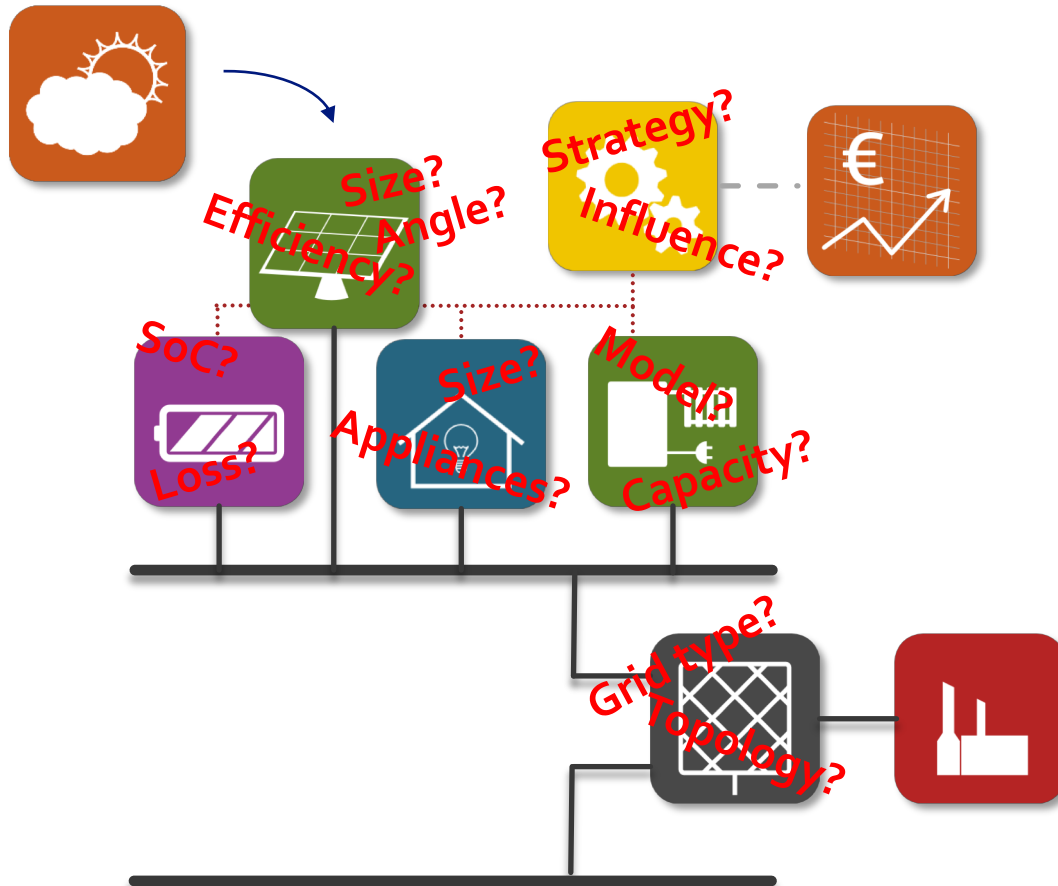
This must be easier, right?



What would be a good home energy management system?

# We are not bakers but engineers!

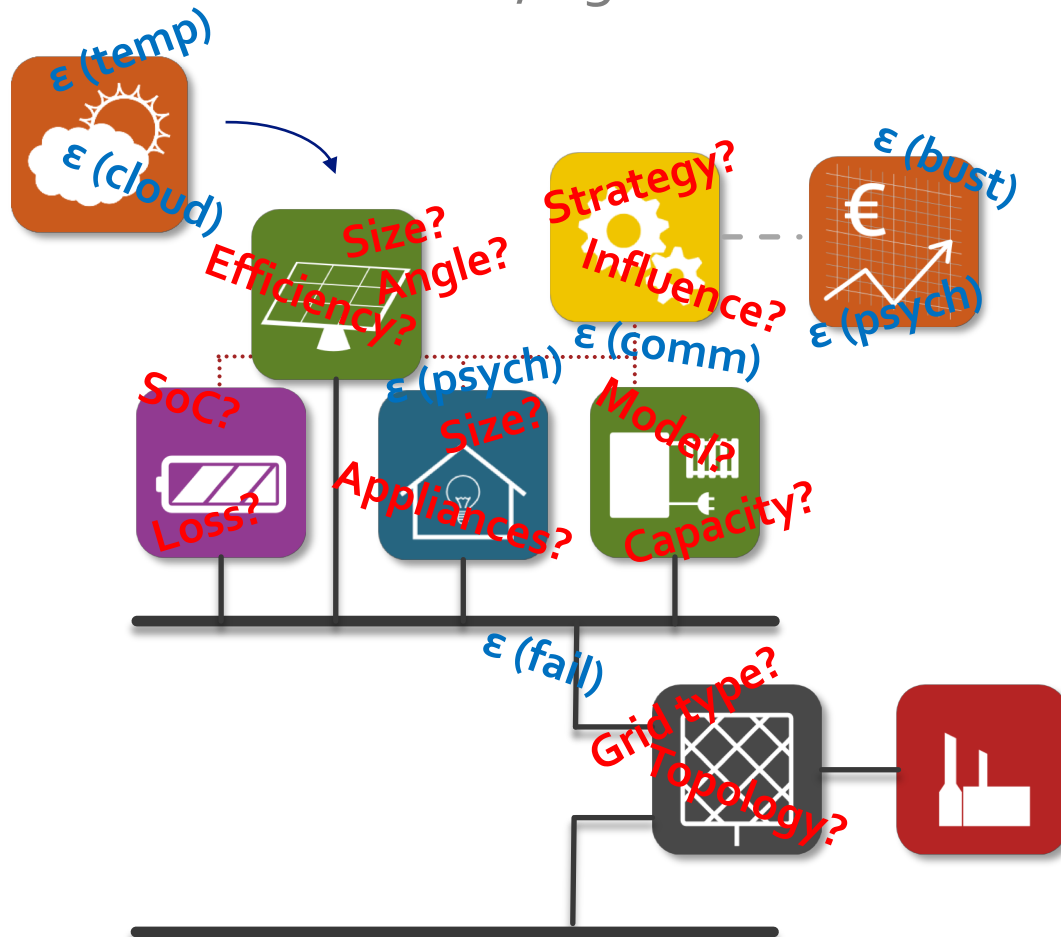
This must be easier, right?



Various parameters  
that can be tuned

# We are not bakers but engineers!

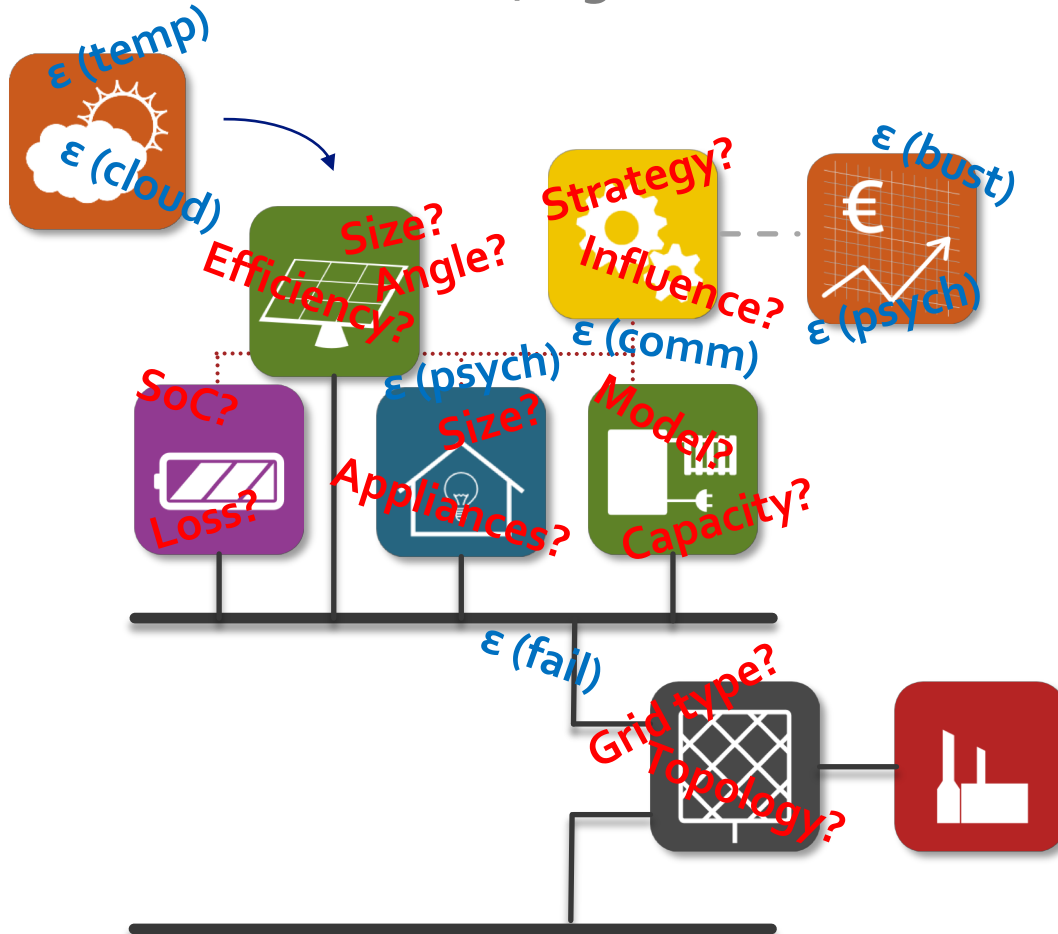
This must be easier, right?



Various sources of  
fluctuation or  
failure

# ► We are not bakers but engineers!

This must be easier, right?

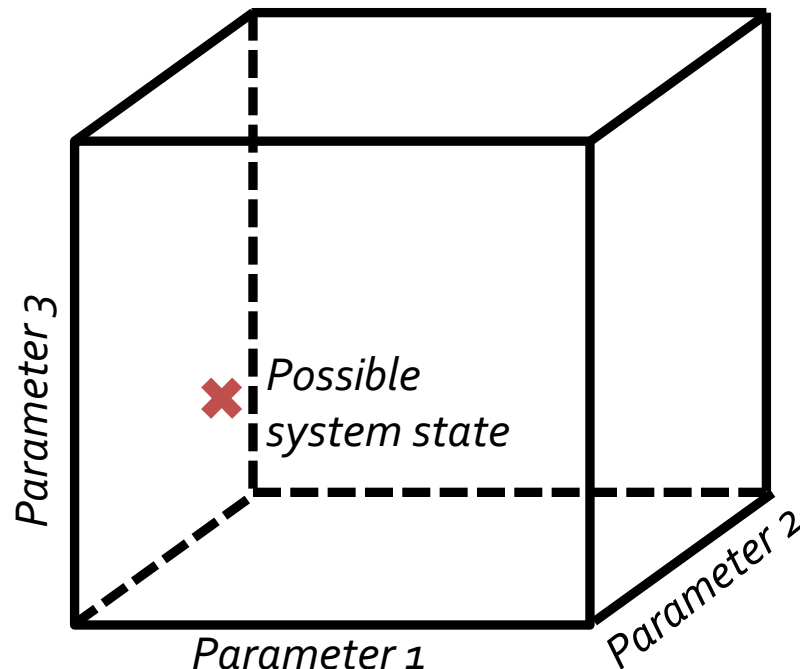


We can simulate, but...

- ... too complex for analytical solution
- ... too complex to model monolithically  
→ black-box system
- ... too many possible combinations to test them all

# The parameter space

The different possible states/designs of our system



- ▶ Usually more than three parameters
- ▶ Potentially continuous parameters

Parameters that influence our system span up a space that...

- ▶ ... is likely high-dimensional
- ▶ ... likely contains an infinite number of combinations

→ No one can bake that many cakes!

(or run infinite smart grid experiments or simulations, for that matter)

# So what do we want?

(except cake)

Get the **maximal amount of information** from a **limited set of experiments/simulations**

and for that we need

**Statistics**

more precisely

**Design of Experiments (DOE)**

# History lesson

## The origins of DOE

- ▶ First documented cases in 18<sup>th</sup> century
- ▶ Modern foundation in 1920s by Ronald A. Fisher
  - ▶ Repetition
  - ▶ Randomization
  - ▶ Blocking
  - ▶ Confounding/orthogonality
  - ▶ Analysis of Variance
- ▶ Main purpose: agriculture
- ▶ Broader adoption in 1950s
  - ▶ Clinical trials
  - ▶ Manufacturing



[http://www.swlearning.com/quant/kohler/stat/biographical\\_sketches/Fisher\\_3.jpeg](http://www.swlearning.com/quant/kohler/stat/biographical_sketches/Fisher_3.jpeg)

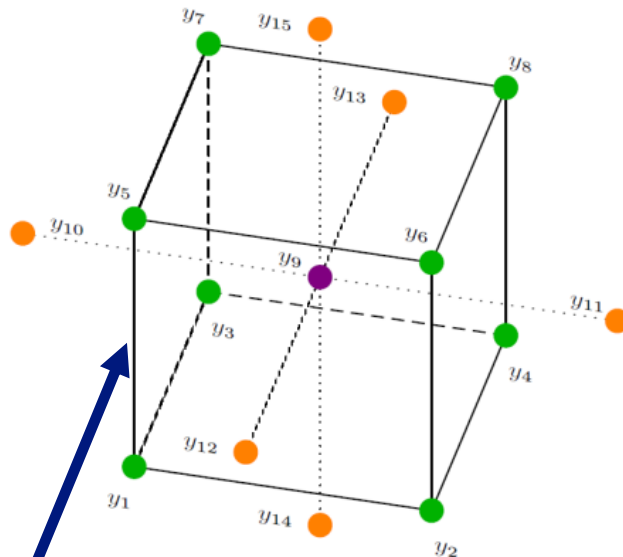
# History lesson

## The origins of DOE

- ▶ George E. P. Box ~ 1960s
  - ▶ Expanded on Fisher's work (Fisher's son in law)
  - ▶ Experiments with small animals and poison gas (in the war)
- ▶ Genichi Taguchi ~ 1940s
  - ▶ Quality standards for engineering
  - ▶ Methods flawed; still big impact on quality of Japanese production
- ▶ Multidimensional sampling methods in 1970s
- ▶ Application and adoption of DOE for computer simulation around 2000s
  - ▶ Jack P. C. Kleijnen: "Design and Analysis of Simulation Experiments", 2007

# Vocabulary

How are things called in DOE?



Different types of factors

- ▶ **Treatment factor:** Direct interest to experiment
- ▶ **Nuisance factor:** Not of interest, but not negligible

Factors can also be:

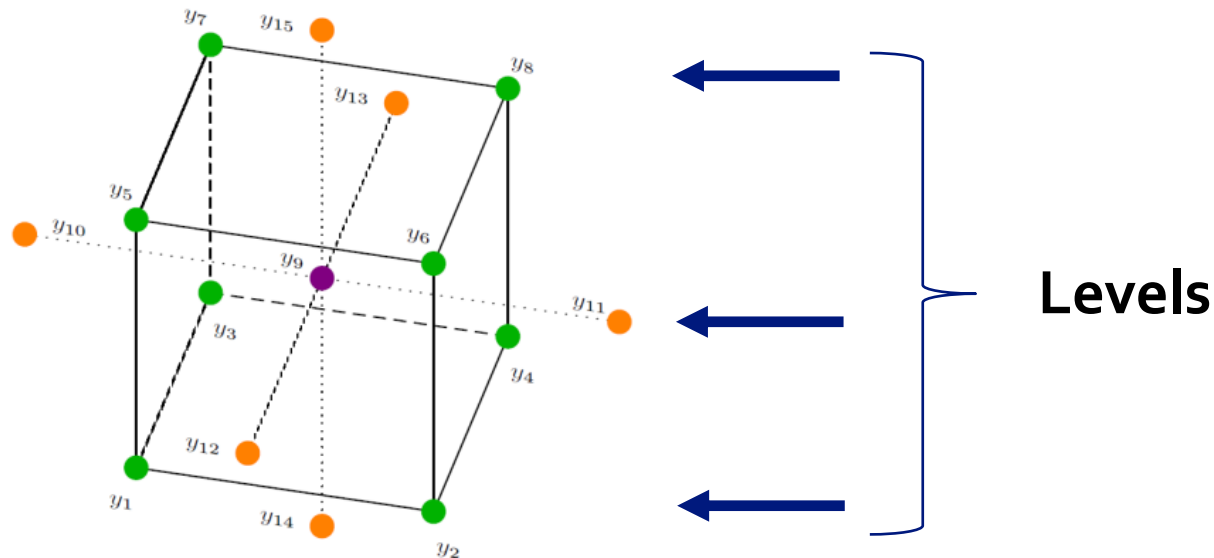
- ▶ Controllable vs uncontrollable
- ▶ Quantitative vs qualitative
- ▶ Known vs unknown

**Factor:** Parameter/variable that influences the system

- ▶ DOE is about finding out the influence of factors

# Vocabulary

How are things called in DOE?

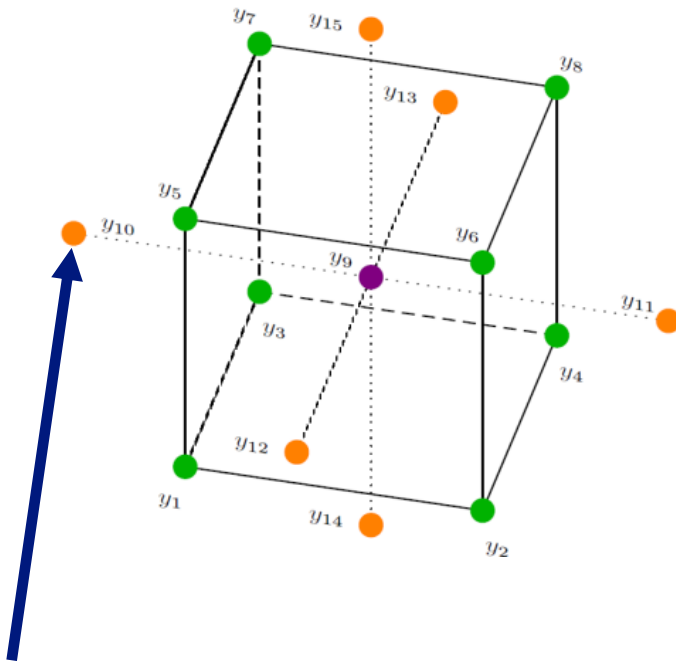


The **levels** of a factor describe its discretization in the experiment

- How many different values of a factor are tested in the experiment?

# Vocabulary

How are things called in DOE?

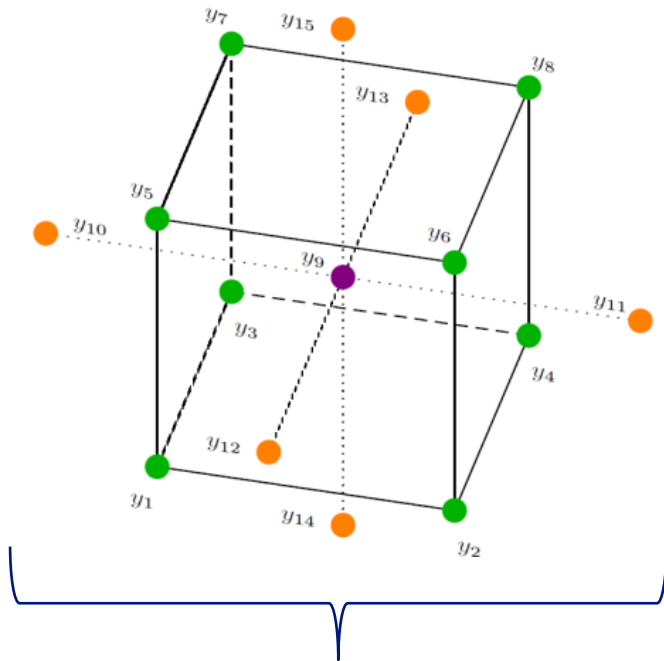


**Treatment:** The combination of one value per factor

- Can be translated into one experiment run

# Vocabulary

How are things called in DOE?



Experiment plans are often documented via tables:

<i>A</i>	<i>B</i>	<i>C</i>
—	—	—
+	—	—
—	+	—
+	+	—
—	—	+
+	—	+
—	+	+
+	+	+

(Example right:

3 factors,  
2 levels,  
8 treatments;

Example left:

3 factors,  
5 levels,  
15 treatments)

**Experiment plan:** The number of treatments that is applied

- For efficient experimentation the plan should follow an established **Design** (i.e. sampling strategy)

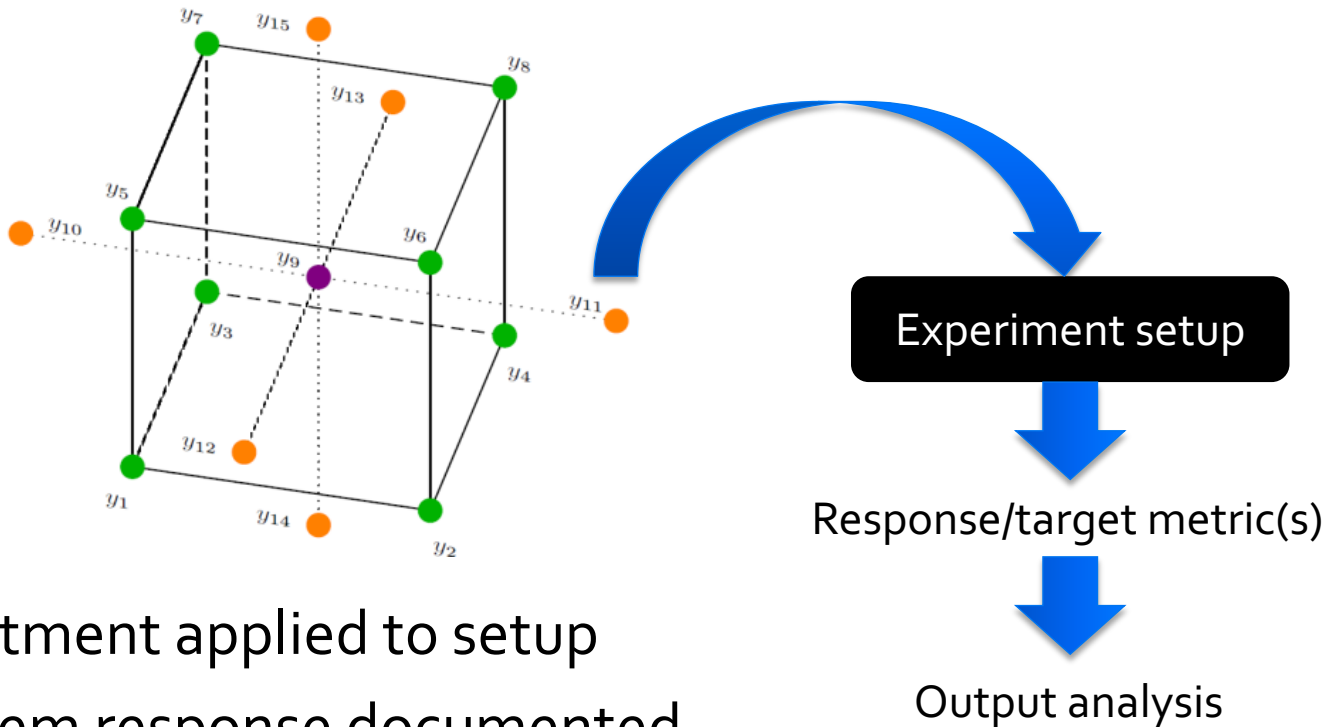
# Vocabulary

How are things called in DOE?

- ▶ Designs are typically created to avoid **confounding** of factors
  - ▶ Confounded factors make it impossible to distinguish between their respective effects
- ▶ The order of applied treatments should be **randomized**
  - ▶ Accounting for existing but unknown factors
- ▶ Known nuisance factors can be accounted for with **blocking**
  - ▶ One block of treatments per level of nuisance factor
  - ▶ Randomized repetition of treatments between blocks

# Vocabulary

How are things called in DOE?



- ▶ Treatment applied to setup
- ▶ System response documented
- ▶ Analysis methods are selected, e.g.:
  - ▶ Analysis of Variance (ANOVA),
  - ▶ Description models (polynomial regression),
  - ▶ ...

# Steps in experiment planning

How do we have to consider?

## 1. Know what you want to find out!

- ▶ Purpose of Investigation!
  - ▶ E.g.: Characterization of X
  - ▶ ... Characterization of X depending on *what*?

## 2. Identify your factors

- ▶ “I’m interested in the effect of this” → Treatment factor
- ▶ “I’m not interested but shouldn’t exclude this” → Nuisance factors
- ▶ → Requirements for test realization (factors must be accessible)!

# Steps in experiment planning

How do we have to consider?

## 3. Do I know enough about the system for planning?

- ▶ Expecting nonlinearities?
- ▶ Are factor interactions important?
- ▶ Which levels to test?
- ▶ → Maybe conduct **screening** experiments

## 4. Establish experiment plan

- ▶ Exposing nonlinear effects? → Adjust factor levels
- ▶ Main effects?
- ▶ Factor interactions? → Adjust **design resolution**
  - ▶ (We'll get to that later)

# Steps in experiment planning

How do we have to consider?

## 5. Make sure you're doing it right

- ▶ There are control methods along the way
  - ▶ Correctly chosen design? → e.g. via correlation matrix
  - ▶ Correctly chosen regression model? → e.g. via half-normal plot
  - ▶ Prediction accuracy? → e.g. via residual plots

## 6. Analysis

- ▶ Are there fluctuations in your system?
- ▶ → Different analysis requirements of physical and simulation experiments!

# Two types of experiments

Different requirements

## Physical experiments

- ▶ Fluctuations
- ▶ Typically fewer treatments
- ▶ Small number of factors
- ▶ Typical goals:
  - ▶ Understand main effects
  - ▶ Worst-case analysis
  - ▶ ...

## Simulation experiments

- ▶ Deterministic
- ▶ Many runs possible
- ▶ Possibly many factors
- ▶ Typical goals:
  - ▶ Explore nonlinearity
  - ▶ Create metamodel
  - ▶ Identify malfunction risks
  - ▶ ...

# “Classical” DOE

For physical experiments

*“Nonlinearities tend to be over- and factor interactions underestimated”*

- ▶ Often 2-level designs (depicted as + and -)
- ▶ *Full-factorial designs*: All level combinations for all factors
  - ▶ Not economically feasible for many factors!
- ▶ *Fractional-factorial designs* (a subset of treatments)

Resolution	Implication
III	Main effects confounded with 2-factor interactions. Typically only useful for screening.
IV	Useful for identifying main effects.
V	Identification of main effects and 2-factor interactions.
VI / V+	Theoretically less confounding, but no practical information gain.

# “Classical” DOE

For physical experiments

- ▶ Fractional-factorial designs can be created via *aliasing*
- ▶ Holds information about resolution/confounding

A	B	C		
-	-	-		
+	-	-		
-	+	-		
+	+	-		
-	-	+		
+	-	+		
-	+	+		
+	+	+		

- ▶ 4 factors, 2 levels, 8 treatments
- ▶ D is confounded/aliased with ABC
- ▶ → Resolution IV design

Available Factorial Designs (with Resolution)

	Factors														
Run	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
4	Full	III													
8		Full	IV	III	III	III									
16			Full	V	IV	IV	IV	III	III	III	III	III	III	III	
32				Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV	
64					Full	VII	V	IV	IV	IV	IV	IV	IV	IV	
128						Full	VIII	VI	V	V	IV	IV	IV	IV	

<http://blog.minitab.com/blog/applying-statistics-in-quality-projects/design-of-experiments-fractionating-and-folding-a-doe>

# “Classical” DOE

For physical experiments

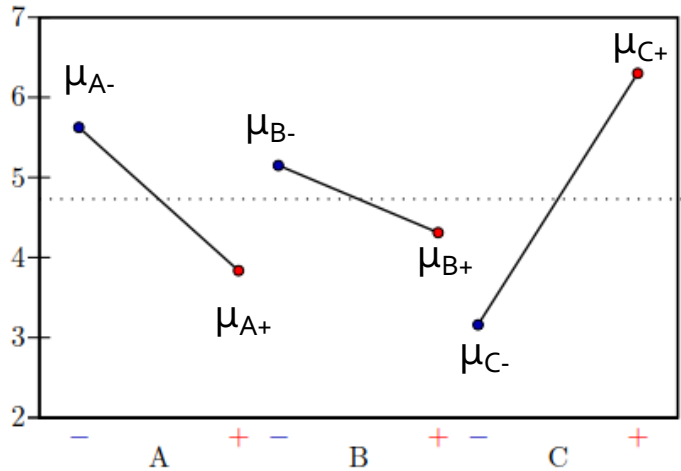
“I have to do all of this by hand?!” – No.

- ▶ Several established methods for constructing designs
- ▶ Tools that implement them
  - ▶ Minitab, MATLAB, R, ...
- ▶ 2-level designs: Yates, Plackett-Burman, ...
- ▶ More levels (nonlinearity): Central-Composite, Box-Behnken, ...
- ▶ Common approach: start screening with Resolution III design and expand to IV design if needed (“folding”)
- ▶ You have to *understand* confounding/resolution to avoid wrong conclusions!

# “Classical” DOE – Analysis

For physical experiments

Siebertz et al. „Statistische Versuchsplanung“ 2010.



- ▶ Stability against noise?  
(Significance?)
- ▶ → We need a statistical test!

- ▶ ANOVA: Among the most common analysis methods in classical DOE
- ▶ One-way ANOVA: Do different levels of a factor have different effects?
  - ▶ Employing the *F-test*
  - ▶ For two levels equivalent to *t-test*
- ▶ Two-way ANOVA: Include factor interactions

# “Classical” DOE – Analysis

For physical experiments

A	...	Resp
-		6
+		13
-		8
+		9
-		4
+		11
-		3
+		7

- ▶  $H_0: \mu_{A-} = \mu_{A+}$
- ▶ Calculate group means and total mean:
  - ▶  $\mu_{A-} = \frac{6+8+4+3}{4} = 5.25, \mu_{A+} = \frac{13+9+11+7}{4} = 10$
  - ▶  $\mu = \frac{5.25+10}{2} = 7.625$
- ▶ Sum of squared differences between groups (SSB):
  - ▶  $SSB = n_{A-}(\mu_{A-} - \mu)^2 + n_{A+}(\mu_{A+} - \mu)^2 = 45.125$
  - ▶ Degree of freedom  $f_B$  is one less than number of groups (=1)
  - ▶ Mean square value:  $MS_B = SSB/f_B = 45.125$

# “Classical” DOE – Analysis (cont’d)

For physical experiments

A	...	Resp
-		6
+		13
-		8
+		9
-		4
+		11
-		3
+		7

- ▶ Sum of squared differences within groups (SSW):
  - ▶  $SSW = \sum (x_{i,A-} - \mu_{A-})^2 + \sum (x_{i,A+} - \mu_{A+})^2 = 34.75$
  - ▶ Degree of freedom  $f_W$  is groups\*(samples-1) = 6
  - ▶ Mean square value:  $MS_W = SSW/f_W = 5.792$
- ▶ F-ratio  $F = \frac{MS_B}{MS_W} \approx 7.791$
- ▶ Check with F-distribution for our degrees of freedom  $F_{(1,6)}$ 
  - ▶  $F_{(1,6)}$  for 5% significance level is 5.987 (< 7.791)
  - ▶  $H_0$  is rejected with  $\alpha=0.05$  (p-value)
  - ▶ Changing A has an effect on the system (probably)!

# “Classical” DOE – Analysis (cont’d)

For physical experiments

- ▶ Variance explained by factor changes compared to overall variance → The **higher the ratio**, the **more likely** that the **factor is important**
- ▶ Only the very basic of ANOVA
  - ▶ Checking for influence of other factors, factor interactions and blocks
  - ▶ Accounting  $\alpha$ - and  $\beta$ -risk (falsely rejecting  $H_0$  or falsely accepting  $H_0$ )
  - ▶ Eliminating insignificant factors from model and adjusting model
  - ▶ ...
- ▶ Again: There are programs to do this for you
  - ▶ But you have to understand the basics 😊

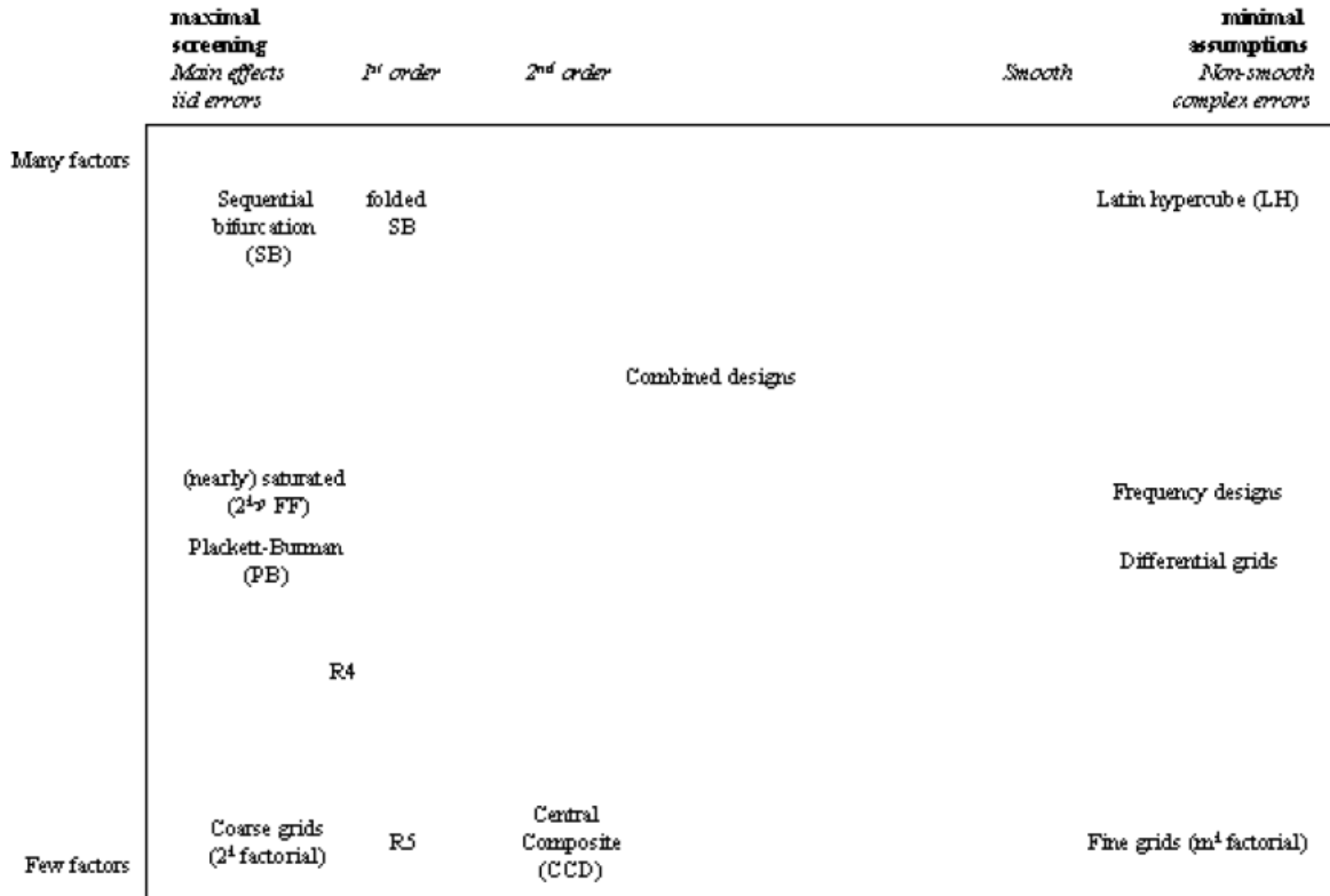
# “Modern” DOE

For computer experiments

- ▶ In simulation we can do *more*. – So we should!
- ▶ Our “old” designs still may be applicable here
  - ▶ If we are only interested in main effects
  - ▶ BUT: simulation can provide a more rigorous understanding of our system!
- ▶ General idea: establish response surface/metamodel for black box system
- ▶ Select design depending on...
  - ▶ Number of factors
  - ▶ Assumptions about system behavior

# “Modern” DOE

For computer experiments

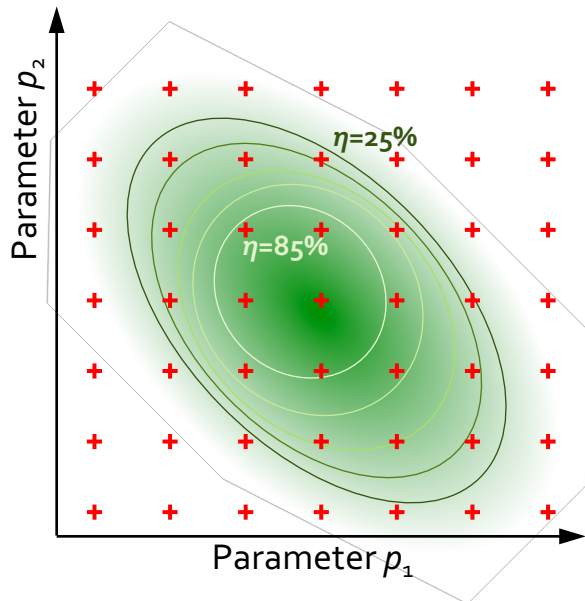


Kleijnen et al. “A User’s Guide to the Brave New World of Designing Simulation Experiments” 2003

# “Modern” DOE

For computer experiments

New feature:  
Space-filling designs



- ▶ Latin-Hypercube Sampling
  - ▶ Stratification of sampling space
- ▶ Low-discrepancy sequences
  - ▶ Sobol, Hammersley, ...
- ▶ Differences in space-filling property and orthogonality
  - ▶ Optimization potential for different system characteristics
  - ▶ ... and different metamodel types

No stochasticity, but uncertainty!

... But that's another story (Uncertainty Quantification)

# DOE for holistic testing?

How does it all fit together?

## How does DOE fit into holistic testing?

→ Methodologies have different focus but serve the same purpose: **Reproducibility**

- ▶ HTD: Documentation, structure and “common language”
- ▶ DOE: Efficient experimentation and statistical significance

# DOE for holistic testing?

How does it all fit together?

## ► Purpose of Investigation

- Characterization: Explore system and its fluctuation/uncertainty → Explore **p-value**
- Validation/Verification: Check system's fluctuation/uncertainty against reference (**Quality attributes**) → Challenge **p-value**

► **Treatment factors**  $\leftrightarrow$  **Variability attributes** (TC), **Inputs** (TS)

► **Response**  $\leftrightarrow$  **Target metrics** (TC), **Target measures/outputs** (TS)

► **Nuisance factors**  $\leftrightarrow$  **Sources of uncertainty, other parameters**

# DOE for holistic testing?

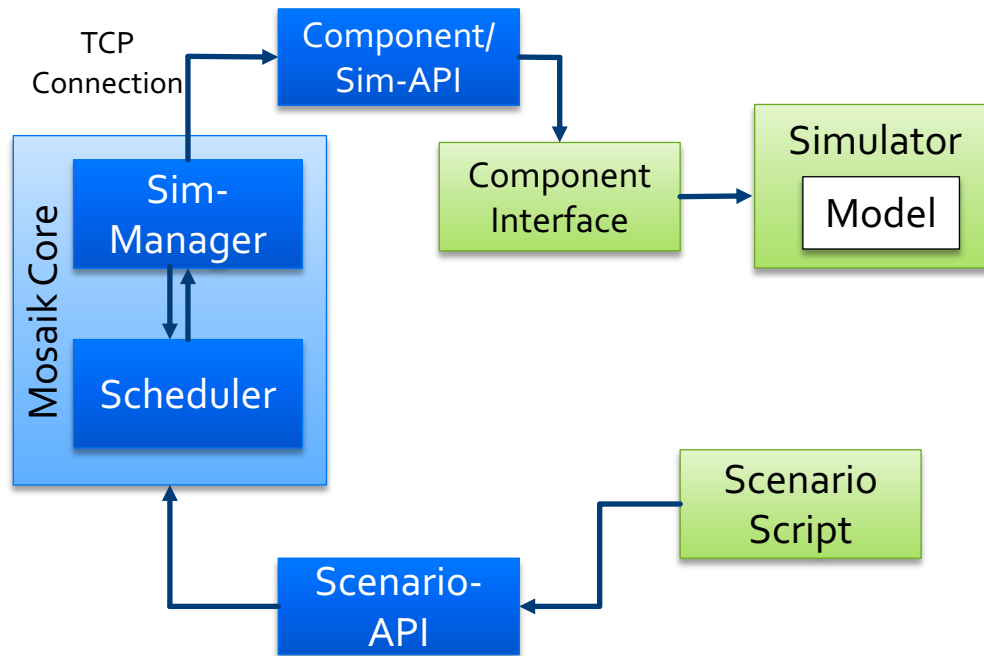
How does it all fit together?

- ▶ Test plan/design  $\leftrightarrow$  Test design, System evolution
- ▶ Screening  $\leftrightarrow$  Qualification Strategy (TS interdependence)
- ▶ Control methods  $\leftrightarrow$  Qualification Strategy (iterative refinement)
- ▶ HTD is a guideline to structure/refine/document your DOE considerations

# Applying DOE

With mosaik – co-simulation software

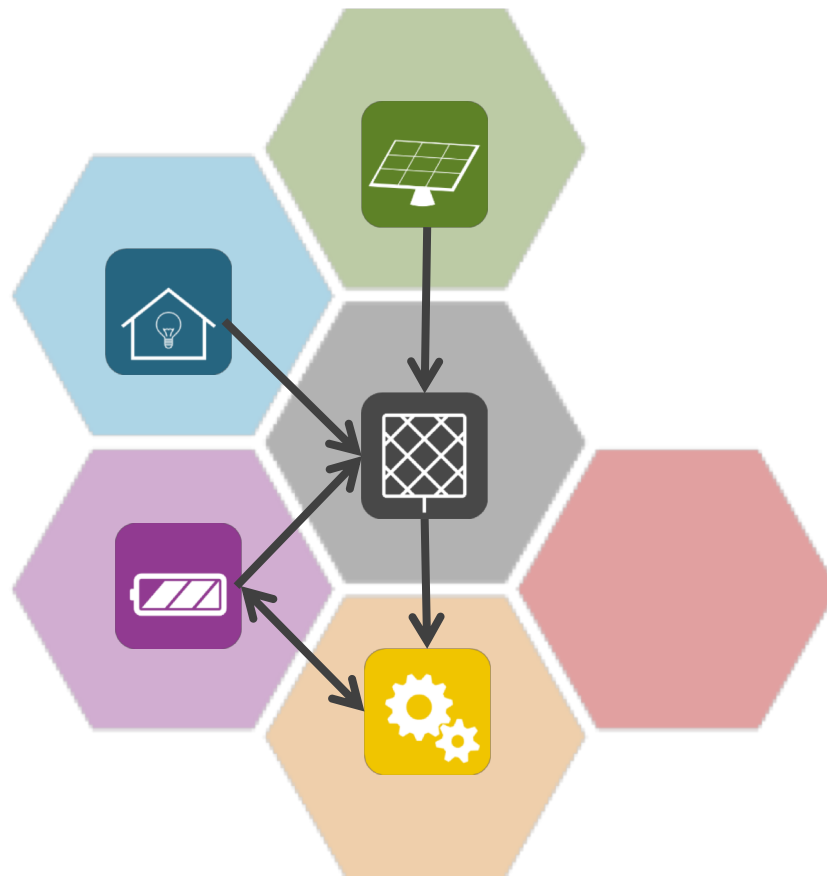
- ▶ Independent simulation tools coupled
- ▶ Co-simulation of holistic system



- ▶ Users can...
  - ▶ Integrate new components
  - ▶ Use components in **simulation scenarios**

# Applying DOE

Example: Home Energy Management System



# Hands-on (1)

Now it's your turn

- ▶ Task 1: Validate effect on battery and control parameters on HEMS
  - ▶ Battery storage capacity, range 15 – 25
  - ▶ Controller change rate, range 0.3 – 0.7
  - ▶ Response: Self consumption index
  - ▶ Season: summer
- ▶ Assume linear behavior → full-factorial 2-level design
- ▶ Plot effect per factors
- ▶ Get significance via one-way ANOVA
  
- ▶ Optional: Include blocking for seasons
  - ▶ One block summer, one block winter

# Hands-on (1) – Helpful code

Now it's your turn

## ▶ ANOVA

- ▶ From `scipy` import `stats`
- ▶ `F, p = stats.f_oneway(group1_results, group2_results)`
- ▶ You may try `pyDOE` for the test design, but there might be bugs depending on the versioning of `pyDOE` and `numpy`
  - ▶ ... and 2-factor, 2-level designs are easy to create anyway

# Hands-on (2)

Now it's your turn

- ▶ Task 2: Explore HEMS response to battery parameter changes (characterization)
  - ▶ Battery storage capacity, range 15 – 25
  - ▶ Battery charge capacity, range 2 – 7
  - ▶ Response: Self consumption index
  
- ▶ Simulation system without fluctuation
- ▶ Employ space-filling sampling
- ▶ Fit metamodel of your choice

# Hands-on (2) – Helpful code

Now it's your turn

## ► Sobol sequence

- `import sobol_seq (install first)`
- `sobol_seq.i4_sobol_generate(n_factors, n_samples)`

## ► Metamodel suggestion: Kriging

- `from sklearn.gaussian_process import GaussianProcessRegressor`
- `from sklearn.gaussian_process.kernels import RBF, ConstantKernel as C`
- `kernel = C(1.0, (1e-3, 1e3)) * RBF(10, (1e-2, 1e2))`
- `gp = GaussianProcessRegressor(kernel=kernel, n_restarts_optimizer=9)`
- `gp.fit(input, response)`
- `prediction = gp.predict(evaluation_grid)`