Optimal experimental design
for the well and the ill (-posed problems)

André Bardow
Institute of Technical Thermodynamics, RWTH Aachen University
Acknowledgements

• P. Domagalski, V. Göke, X. Liu, H.-J. Koß
• E. Kriesten, S. Kossack, W. Marquardt
• M. Voda, F. Casanova, B. Blümich

• M. Ottens

Funding

• German Science Foundation (DFG)
• Delft Research Centers for Life-Sciences and for Sustainable Processes
Design for the well and the ill

- Design of diffusion experiments
- Design for ill-posed problems: The METER approach
- OED for physical property models
What is optimal experimental design?
Pimp my diffusion experiment

take an old car…

make it faster, nicer,… – better!

Fig. 1. Schematic drawing of the diffusimeter. 1—l麟 of light source; 2—tube to admission air currents in the light path; 3—stirring motor mount; 4—stirring motor; 5—sensor of temperature controller; 6—hanger; 7—cell frame support; 8—special window holder; 9—beam support; 10—water bath support; 11—water bath; 12—access hole to reach inside of the beams; 13—ways; 14—leveling collar to check any changes in the position of beam relative to that of beam support.

take an old experiment…

make it faster, nicer,… – better!
## Current diffusion experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Temperature range</th>
<th>Pressure range</th>
<th>Experiment time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diaphragm cell</td>
<td>Good (0.5–1%)</td>
<td>Moderate (T \leq 400 \text{ K})</td>
<td>Wide(^\dagger)</td>
<td>Days</td>
</tr>
<tr>
<td>Capillary</td>
<td>Moderate High (0.2%)</td>
<td>Wide range Near ambient</td>
<td>Narrow(^{\dagger\dagger})</td>
<td>Days</td>
</tr>
<tr>
<td>Open</td>
<td>High (0.2%)</td>
<td>Near ambient</td>
<td>Narrow(^{\dagger\dagger})</td>
<td>Days</td>
</tr>
<tr>
<td>Closed</td>
<td>Moderate ((~ 1%))</td>
<td>Wide range (T \leq 600 \text{ K})</td>
<td>Moderate(^\dagger)</td>
<td>(&lt; 1 \text{ hour})</td>
</tr>
<tr>
<td>Conductance</td>
<td>Moderate (1–2%)</td>
<td>Wide range (T \leq 700 \text{ K})</td>
<td>Wide(^{\dagger\dagger})</td>
<td>(&lt; 1 \text{ hour})</td>
</tr>
<tr>
<td>Taylor dispersion</td>
<td>Moderate High (\leq 0.1%)</td>
<td>Near ambient</td>
<td>Narrow</td>
<td>1–2 days</td>
</tr>
<tr>
<td>NMR</td>
<td>High (\leq 0.1%)</td>
<td>Near ambient</td>
<td>Narrow</td>
<td>1–2 days</td>
</tr>
<tr>
<td>Gouy</td>
<td>High (\leq 0.1%)</td>
<td>Near ambient</td>
<td>Narrow</td>
<td>1–2 days</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>High (\leq 0.1%)</td>
<td>Near ambient</td>
<td>Narrow</td>
<td>1–2 days</td>
</tr>
</tbody>
</table>

Wakeham, Measurement of the transport properties of fluids, IUPAC 1991
set free variables $\mathbf{x}$ such that information on parameter $\theta$ is maximized

maximize curvature of parameter estimation objective

OED objective for parameter estimation

$$\max_{\mathbf{x}} \frac{\partial^2 \phi}{\partial \theta^2}$$

parameter estimation objective

$\phi^*$

error level

$\phi(\mathbf{x}^{(1)})$

$\phi(\mathbf{x}^{(2)})$

$\theta^*$

$\theta$
Model-based design of $\mu$-diffusion chip

⇒ Diffusion experiments on the micro-scale

H-cell microchannel

(P. Domagalski, M. Ottens, AB, submitted)
Unpublished slides deleted
### Update: diffusion experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Temperature range</th>
<th>Pressure range</th>
<th>Experiment time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diaphragm cell</td>
<td>Good (0.5–1%)</td>
<td>Moderate (T ≤ 400 K)</td>
<td>Wide†</td>
<td>Days</td>
</tr>
<tr>
<td>Capillary</td>
<td>Moderate</td>
<td>Wide range</td>
<td>Narrow††</td>
<td>Days</td>
</tr>
<tr>
<td>Conductance</td>
<td>High (0.2%)</td>
<td>Near ambient</td>
<td>Narrow††</td>
<td>Days</td>
</tr>
<tr>
<td>Taylor dispersion</td>
<td>Moderate</td>
<td>Wide range (T ≤ 600 K)</td>
<td>Moderate†</td>
<td>&lt; 1 hour</td>
</tr>
<tr>
<td>NMR</td>
<td>Moderate</td>
<td>Wide range (T ≤ 700 K)</td>
<td>Wide††</td>
<td>&lt; 1 hour</td>
</tr>
<tr>
<td>Gouy</td>
<td>High (≤ 0.1%)</td>
<td>Near ambient</td>
<td>Narrow</td>
<td>1–2 days</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>High (≤ 0.1%)</td>
<td>Near ambient</td>
<td>Narrow</td>
<td>1–2 days</td>
</tr>
</tbody>
</table>

**Microfluidic H-Cell**

- Moderate
- To be explored

**Experiment time**
- Days
- < 1 hour
- 1–2 days

**Note:** Moderate, To be explored, minutes.
Can we do even better?

- diffusion coefficients are concentration-dependent
- usually many experiments to determine this dependency
- here: one Raman-experiment plus 2 NMR data points

⇒ whole dependence from minimal data
⇒ but requires to solve ill-posed problem

(E. Kriesten et al., *Fluid Phase Equilibria*, 2009)
Tikhonov regularization for function estimation

**Estimation problem**

\[
\min_r \| c - \tilde{c} \|
\]

s.t. \( c(t) - c_0 = \int_0^t r(t)dt \)

\[ g(t) = \int_T K(t, s; d) f(s)ds \]

**Regularized problem**

\[
\min_r \| c - \tilde{c} \| + \lambda \| r \|
\]

s.t. \( c(t) - c_0 = \int_0^t r(t)dt \)

\[
\min_f \sum_{i=1}^n \frac{1}{\sigma_i^2} \left( g(t_i) - \int_T K(t_i, s; d) f(s)ds \right)^2 + \lambda \| Lf \|_{L^2}^2
\]

lack of fit ("least-squares")

smoothness penalty

bias-variance trade-off crucial for success

bias

variance

data error

regularization error

regularization \( \lambda \)
Optimal Experimental Design: Idea

set free variables $\mathbf{x}$ such that information on parameter $\theta$ is maximized

maximize curvature of parameter estimation objective

$\text{OED objective}$

$$\max_{\mathbf{x}} \frac{\partial^2 \phi}{\partial \theta^2}$$

- focus on variance only
- bias contribution is missing
- not applicable

"estimate the wrong solution with less uncertainty"
Minimum Expected Total Error (METER)

Optimal design:
find settings for experiment that allow best determination of unknown function

classical approach (Box-Lucas school): minimize parameter variance only

approach proposed for ill-posed problems (Box-Draper school):
minimize

\[
E\left(\left\| f - \hat{f} \right\|_{L_2}^2\right)
= \left\| f - K^+ (\lambda) K f \right\|_{L_2}^2
+ \sigma^2 \text{trace} \left( K^+ (\lambda) \left( K^+ (\lambda) \right)^T \right),
\]

with \( K^+ (\lambda) = \left( K^T K + \lambda L^T L \right)^{-1} K^T \). “regularized Fisher information matrix”

Example: Numerical differentiation

Application: Batch reactor

- Determine reaction rate $r(t)$ from concentration measurements $c(t)$

$$c(t) = \int_0^1 r(t')dt' \quad \Rightarrow \quad r(t) = \frac{dc(t)}{dt}$$

- Numerical differentiation of experimental data

⇒ ill-posed problem

- Solution by finite differences

$$\hat{r}(t_i) = \frac{c(t_i) - c(t_{i-1})}{\Delta t}$$

- Time step $\Delta t$ too large → approximation error
- Time step $\Delta t$ too small → amplification of noise

Determine best time step in optimal design

Application: Batch reactor
$r(t) = \exp(-t)$

$t = [0,10]$
Comparison with Numerical Computations

- predicted behaviour is found in practice
- small difference from approximate bias term
Comparison with Classic Criteria

\[ r(t) = \exp(-t) \]
\[ t = [0,10] \]

**E-optimal design:** max min eigenvalue of Fisher matrix

- standard criteria favor maximal time step
- predictions are even qualitatively in error
Discussion

- *METER* criterion captures bias-variance trade-off crucial in the solution of ill-posed problems

- classical methods are even qualitatively wrong

- extension to nonlinear problems is straightforward by linearization and local analysis as in standard theory

- regularization parameter $\lambda$ incorporates naturally

- main drawback: initial guess of unknown function $f$ required $\Rightarrow$ robust METER design formulations
Predictive physical property models

Which species should be tested to develop predictive property models?
Experimental design stages

Stage I: Model discrimination
- propose model candidates
- experiment design
- data analysis
- suitable model structure

Stage II: Parameter estimation
- suitable model structure
- experiment design

Integrate choice of test species in experimental design
**Example: Multicomponent Diffusion**

<table>
<thead>
<tr>
<th>Fick</th>
<th>Maxwell-Stefan</th>
<th>Thermo</th>
</tr>
</thead>
</table>
| \[
\begin{bmatrix}
D_{11}(x) & D_{12}(x) \\
D_{21}(x) & D_{22}(x)
\end{bmatrix}
\] | \[
\begin{bmatrix}
D_{11}(x) & D_{12}(x) & D_{13}(x) \\
D_{21}(x) & D_{22}(x) & D_{23}(x) \\
D_{31}(x) & D_{32}(x) & D_{33}(x)
\end{bmatrix}
, \Gamma(x)
\] | measured | theory and predictive models | poorly known |
Example: Multicomponent Diffusion

Fick
\[
\begin{bmatrix}
D_{11}(x) & D_{12}(x) \\
D_{21}(x) & D_{22}(x)
\end{bmatrix} = f \left( \begin{bmatrix}
\frac{D_{12}(x)}{-D_{21}(x)} & \frac{D_{13}(x)}{-D_{23}(x)} \\
\frac{D_{21}(x)}{D_{12}(x)} & \frac{D_{23}(x)}{D_{13}(x)}
\end{bmatrix}, \Gamma(x) \right)
\]

Maxwell-Stefan
\[D_{11}^{x_1 \to 0} = \left( \frac{x_2}{(D_{12}^0)^{x_2}(D_{12}^{x_3 \to 1})^{x_3}} + \frac{x_3}{(D_{13}^0)^{x_3}(D_{13}^{x_2 \to 1})^{x_2}} \right)^{-1} = D_{11}^{x_1 \to 0}\]

Thermo

Predictive models for
- Wesselingh-Krishna (WK)
- Kooijman-Taylor (KT)
- Rehfeldt-Stichlmair (RS)
- Vignes-Krishna-van Baten (DKB)
- Darken-Krishna-van Baten (VKB)

Diagram from Rehfeldt et al. (2007)

(Bardow et al., Fluid Phase Equilibria. 2009, Liu et al, Fluid Phase Equilibria, 2011)
Classical OED for Model Discrimination

Question: At which concentration should we measure?

bad choice \( x^{(1)} \)

since

\[
y^M(x^{(1)}, \theta^M) = y^H(x^{(1)}, \theta^H)
\]

Hunter-Reiner criterion (1965):

maximize difference between model responses

\[
\max_x \left[ y^M(x, \theta^M) - y^H(x, \theta^H) \right]^2
\]

here: Buzzi-Ferraris & Forzatti (1983)
Question: Which **chemical system** should be selected?

Cyclohexane+\textbf{1,4-Dioxane}+Toluene

Cyclohexane+\textbf{n-Hexane}+Toluene

Look for molecule that allows for maximum model discrimination
Systematic method to develop new or novel products.

Define needs

Generate ideas to meet these needs

Screen and select the best ideas

Decide how the product is manufactured

Cussler and Moggridge, 2001
Optimal Experiment Design as Product Design Problem

Define needs

Generate ideas to meet these needs

Combine Molecular Building Blocks

Screen and select the best ideas

Group-Contribucion Methods & Objective function

Computer-aided molecular design

Experiment and model analysis
Combined OED-CAMD work process

(Bardow, Kossack, Kriesten, Marquardt, Computer-Aided Chemical Engineering, 2008)
MEXA-CAMD for Diffusion Case Study

Binary Input data

physical property model

\[
D_{ij}^0 = \sqrt{D_{ij}^0 D_{ji}^0} = D_{ij}^{x_i \rightarrow 1}
\]

\[
D_{ik}^0 D_{jk}^0 D_{ji}^0 D_{ij}^0 = D_{ij}^{x_i \rightarrow 1}
\]

\[
\frac{D_{ik}^0}{x_i + x_j} + \frac{D_{jk}^0}{x_i + x_j} = D_{ij}^{x_i \rightarrow 1}
\]

\[
x_i + x_j = D_{ij}^{x_i \rightarrow 1}
\]

experiment model

\[
\hat{D}_{i}^{x_i \rightarrow 0} = \sum_{j=1; j \neq i}^{n_e} \left( \frac{x_j}{(D_{ij}^0)^{1/4}} \right) D_{ij}^{x_i \rightarrow 0}
\]

\[
f^m(\psi^m, d^m) = \hat{D}_{i}^{x_i \rightarrow 0}
\]

\[
\psi^m = [D_{ij}^{0,m}, D_{ji}^{0,m}, D_{ik}^{0,m}, D_{ki}^{0,m}, D_{jk}^{0,m}, D_{kj}^{0,m}]
\]

\[
da^m = [x_i^m, x_j^m, x_k^m]
\]
Continuous property targeting in OED

\[ \hat{D}_{x_i} \to 0 = f^m(\psi^m, d^m) \]

\[ \psi^m = [D_{ij}^0, D_{ji}^0, D_{ik}^0, D_{ki}^0, D_{jk}^0, D_{kj}^0] \]

\[ d^m = [x_i^m, x_j^m, x_k^m] \]

Step 1: Relaxed MEXA-OED problem

\[ \max_{\psi, d} \sum_{m=1}^{M} \sum_{m'=m+1}^{M} (f^m - f^{m'})^T V^{-1} (f^m - f^{m'}) \]

continuous search in property space → standard NLP problem

Step 2: CAMD

\[ \psi^{OPT} \to \text{CAMD} \to \text{real mixtures } M, \psi_M \]
CAMD targets from OED

**Goal:** Find **third component** in **Cyclohexane + Toluene + ?** allowing for optimum model discrimination

⇒ find $\psi^{OPT} = [D_{13}^0, D_{31}^0, D_{23}^0, D_{32}^0]$

Problem can be further reduced using Wilke-Chang

⇒ find $\psi^{OPT} = \left[ \frac{\sqrt{M_3}}{\eta_3}, V_3 \right]$

targets for CAMD algorithm
Mixture OED as Product Design Problem

Define needs

Generate ideas to meet these needs

Screen and select the best ideas

Computer-aided molecular design

Experiment and model analysis

Software package used: ICAS (Gani et al.)
Includes Group-contribution methods and rules to generate only meaningful molecules
**Goal:** Find **third component** in Cyclohexane+Toluene+? maximal discrimination of property model candidates

<table>
<thead>
<tr>
<th>Component name</th>
<th>M (g/mol)</th>
<th>η (cP)</th>
<th>V_m (cm³/mol)</th>
<th>MD objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-methyl-cyclopentene</td>
<td>82.14</td>
<td>0.27</td>
<td>119.57</td>
<td>130</td>
</tr>
<tr>
<td>methyl-cyclopentane</td>
<td>84.16</td>
<td>0.41</td>
<td>120.69</td>
<td>46</td>
</tr>
<tr>
<td>cyclohexene</td>
<td>82.14</td>
<td>0.54</td>
<td>108.04</td>
<td>22</td>
</tr>
</tbody>
</table>

factor six in experimental effort between top 3 molecules
**Goal**: Find **third component** in Cyclohexane + Toluene + ?

maximal discrimination of property model candidates

---

- Application to NMR measurements and molecular simulations
  - none of the existing models works in a robust fashion
  - theoretically-based model derived and validated

CAMD for OED

- add test molecules and mixtures to experimental design
- integration CAMD into OED work process
- extends to all OED problems

⇒ reduction in experimental effort
⇒ quantitative assessment of model validity
Summary

- OED leads non-intuitive designs reducing experimental effort by orders
- software appearing for 1D PDAE systems
- work on ill-posed problems has started
- seamless integration of mixture selection possible

⇒ OED mandatory in model-based experimentation
⇒ comprehensive methodological toolbox available