Big Data Summer

A summer school of the BiGmax Network

Platja d'Aro, Spain, September 9 – 13, 2019

# Clean data acquisition in oxidation catalysis

*Annette Trunschke*

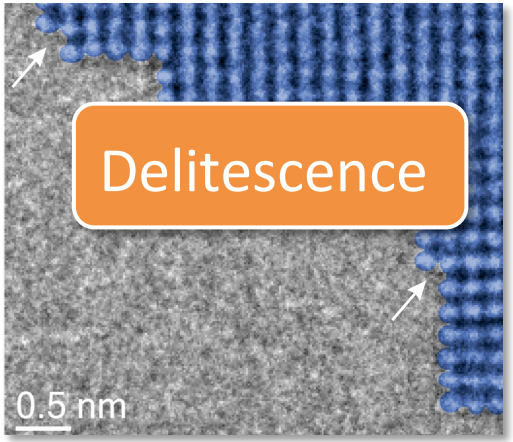
*Department of Inorganic Chemistry, Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Germany*

# Data: Raw material to generate understanding

- Data will have to be
  - Well documented by meta data
  - Stored in a findable way
  - Documented and stored in a re-useable form
  - Clean enough to be interoperable with later and external analysis
  - Complete enough to allow reconstruction of the experiment

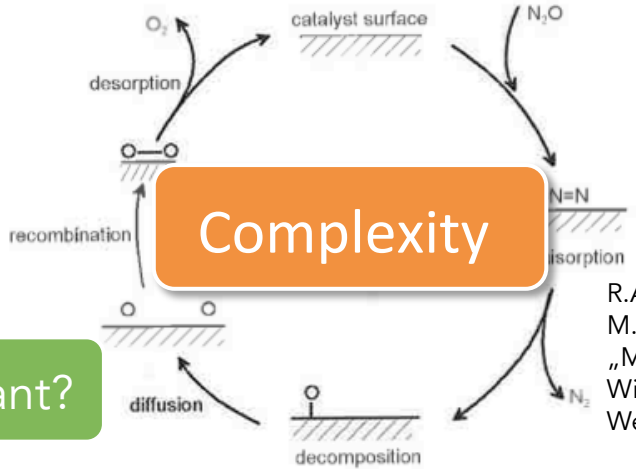
# Challenges and chances – types of data

## Active site




*Journal of Catalysis* **2015**, *326*, 560-573.

## Catalytic cycle / mechanism




R.A. van Santen, M. Neurock, „Mol. Het.Catal.", Wiley-VCH, Weinheim 2006

Theory

*

What is important?

Big* data / AI

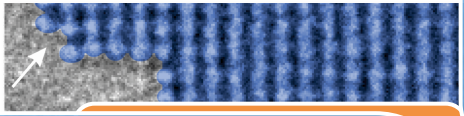
Complexity

Dynamics

*but with much less data compared to YouTube ($10^9$ GB)

# Challenges and chances – types of data

## Active site



Chracterization of activated and and spent catalysts

- Bulk and surface analysis of powders
- Synthesis conditions including raw materials and history

## Model investiagtions

- Surface science analysis
- DFT calculations

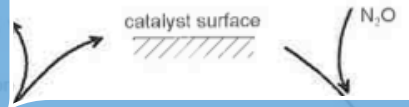
*

Theory

Operando spectroscopy

- Simultaneous measurement of chemical/physical properties and functional properties in one experiment

## Catalytic cycle / mechanism

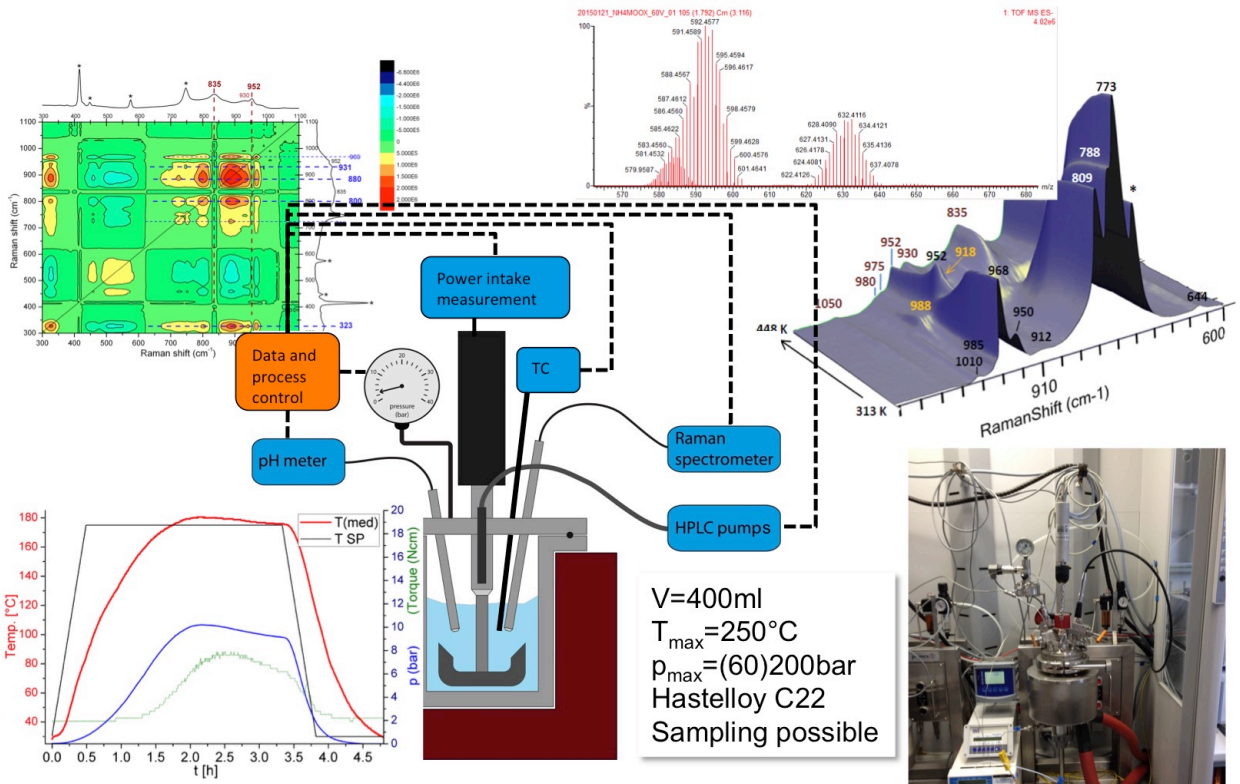



Reaction kinetics

- Process conditions and reactor type
- Analitical methods of concentration determination
- Method of calculations applied in activity/selectivity determination

*but with much less data compared to YouTube ($10^9$ GB)

# Reproducible synthesis



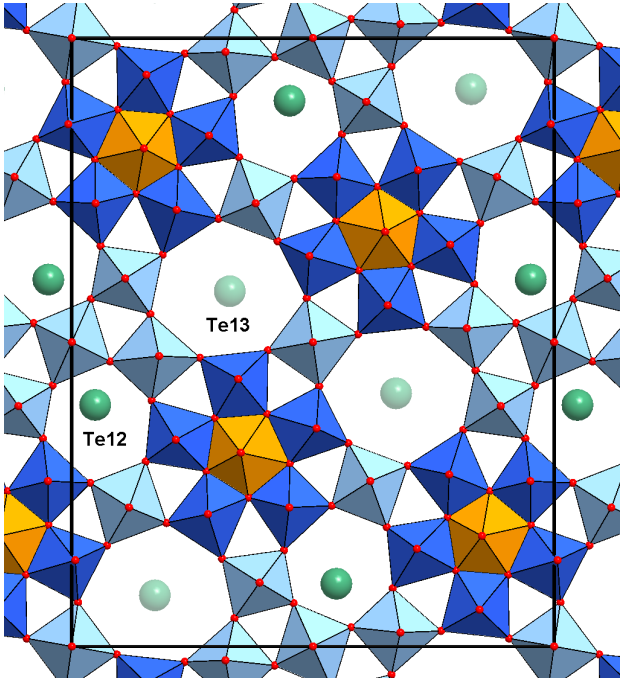

HPM-PT-040 (Premex Reactor GmbH)
Kaiser Optics Raman Spectrometer RXN1

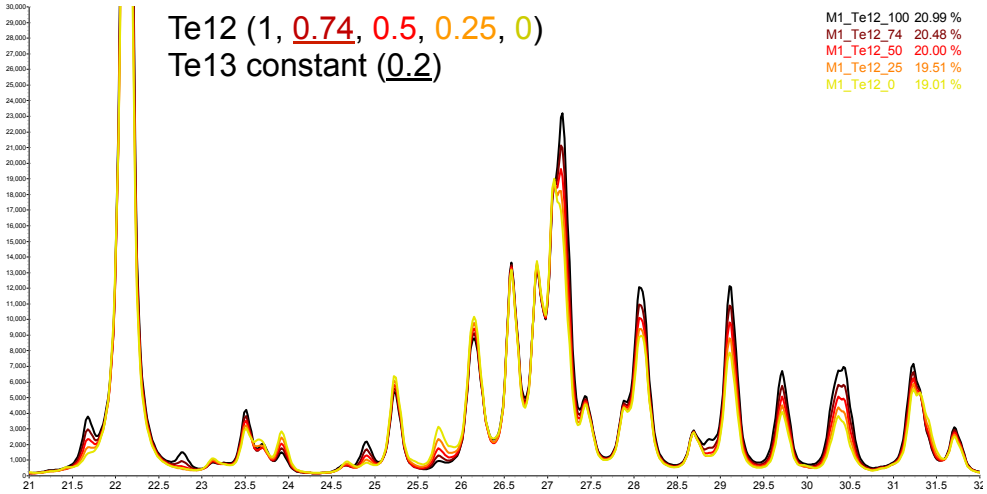


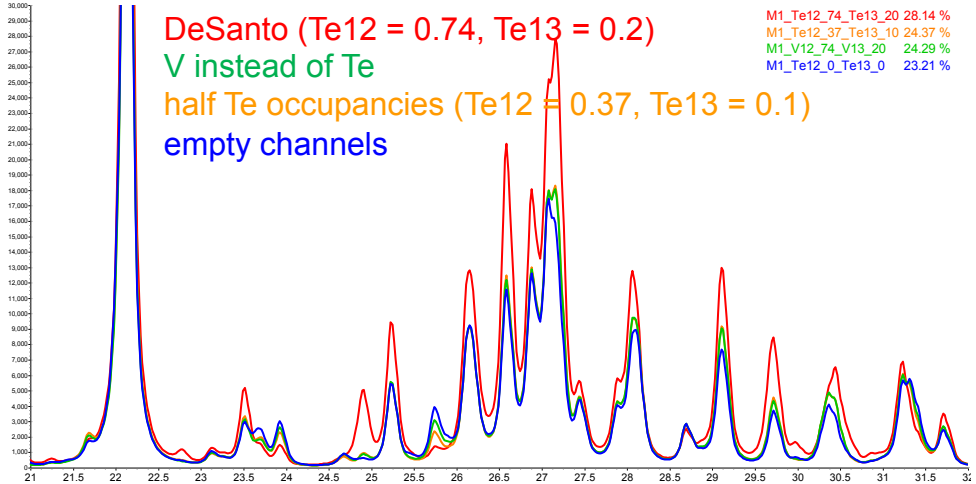
- Automated synthesis reactors deliver synthesis metadata
- Difficulties in automation:
  - Separation and washing steps
  - Activation procedures

*Zeitschrift für anorganische und allgemeine Chemie* 2014, *640*, 2730-2736.

# Reproducible synthesis



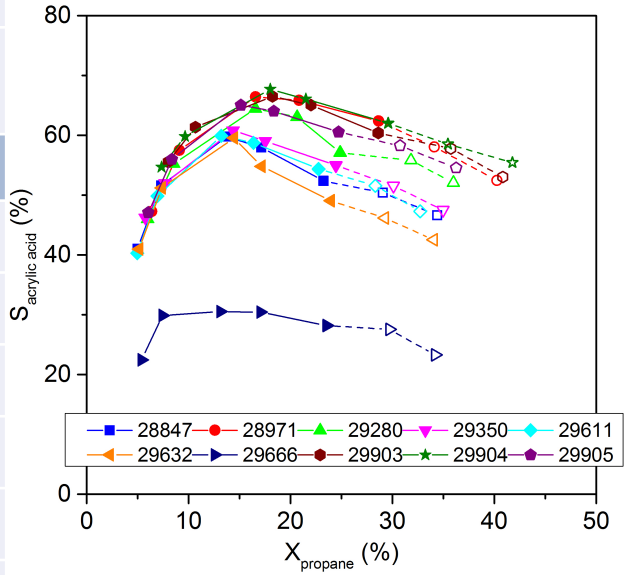



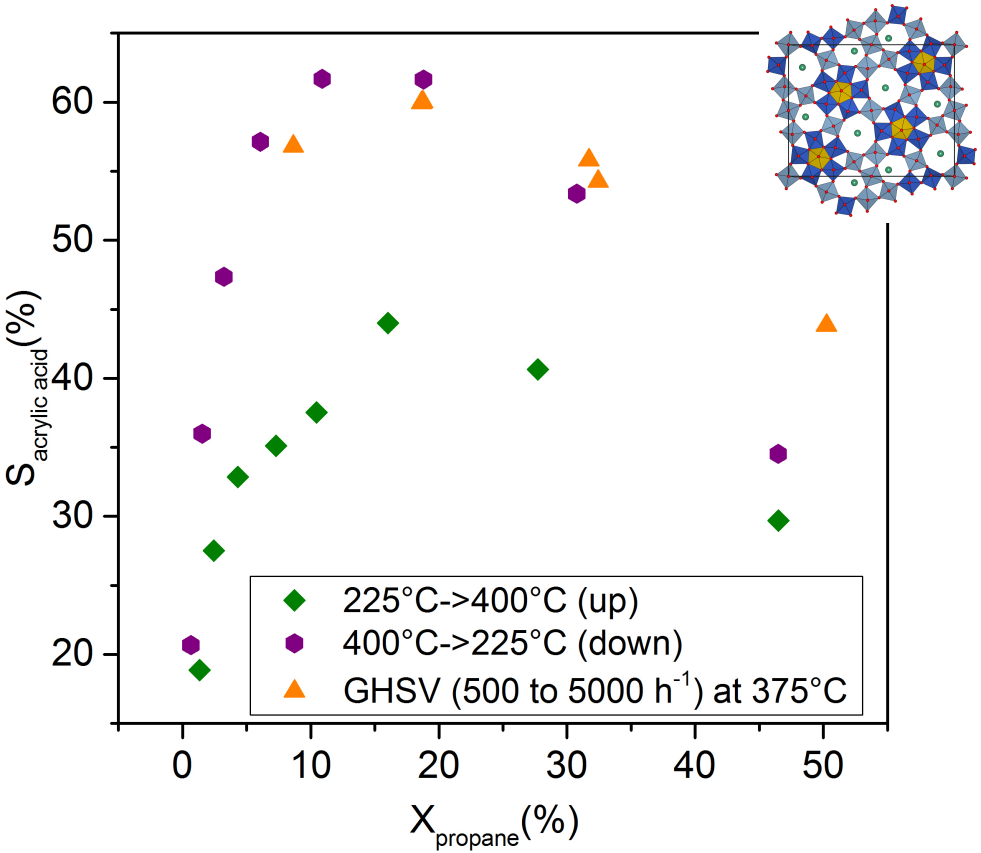

Simulations: Frank Girgsdies

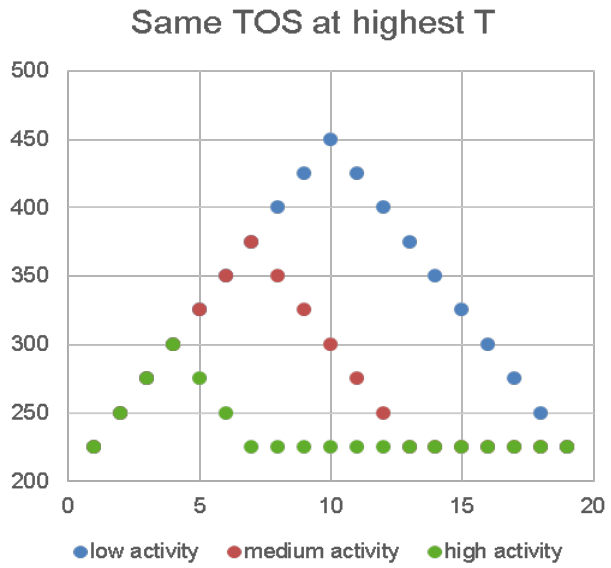
# “Reproducible” synthesis

| SN | m (g) | XRD lattice parameter | | | XRD | XRF | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Before testing | a | b | c | M1 phase (%) | Mo | V | Te | Nb |
| 29200 | 130.11 | Precursor | | | - | 0.4474 | 0.0978 | 0.0813 | 0.0997 |
| 29258[A] | 1.69 | 21.1682790 | 26.6586529 | 4.0158877 | 97.69 | 0.4579 | 0.0994 | 0.0837 | 0.0873 |
| 29350[B] | 8.59 | 21.1667252 | 26.6579321 | 4.0164262 | 97.74 | 0.4546 | 0.0979 | 0.0835 | 0.0913 |
| 29425[A] | 1.71 | 21.1713859 | 26.6656641 | 4.0171937 | 97.57 | 0.4566 | 0.1004 | 0.0840 | 0.0877 |
| 29611[B] | 8.10 | 21.1723662 | 26.6627860 | 4.0170173 | 97.96 | 0.4575 | 0.0995 | 0.0840 | 0.0874 |
| 29632[B] | 8.34 | 21.1715227 | 26.6655441 | 4.0171018 | 97.66 | 0.4522 | 0.1006 | 0.0829 | 0.0924 |
| 29666[B] | 8.30 | 21.1699864 | 26.6621774 | 4.0164017 | 97.89 | 0.4522 | 0.1010 | 0.0828 | 0.0924 |



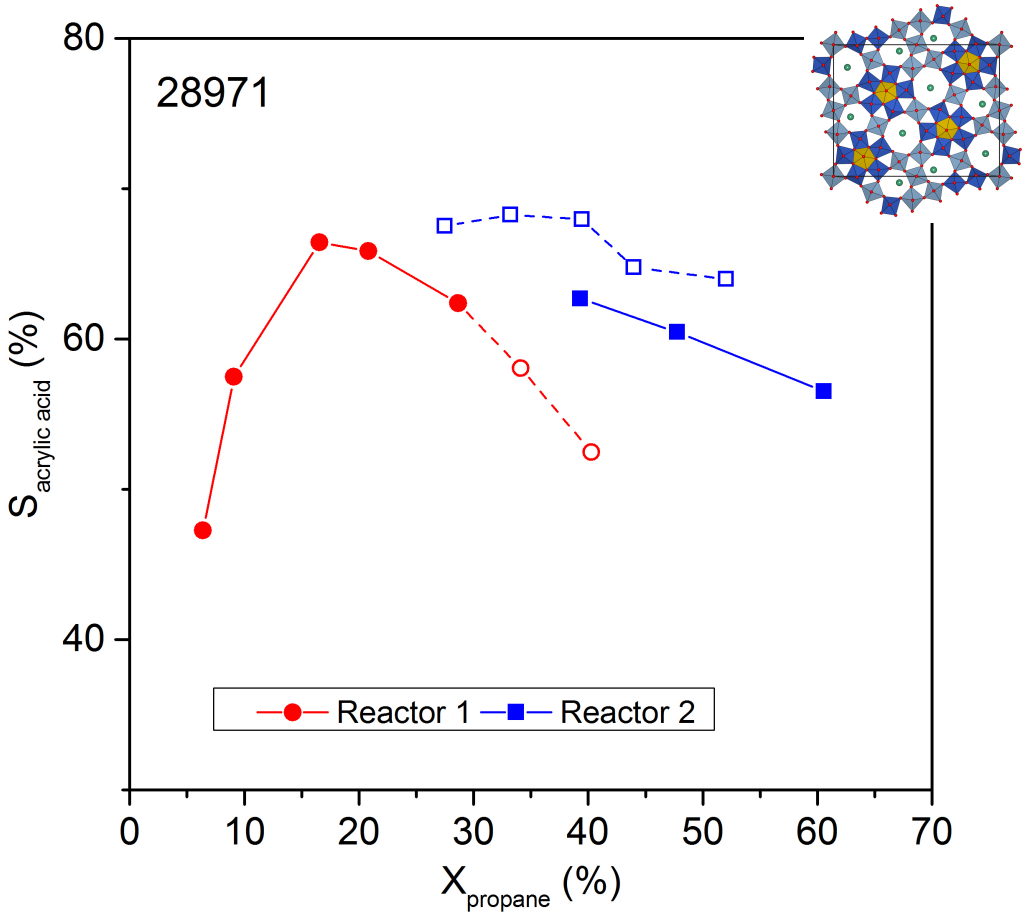
# Sequence of catalytic test experiment




- Sequence matters due to the chemical memory of catalysts
- Catalysts needs to be equilibrated at first

# Heat and mass transport phenomena




- Pressure drops, carbon balance,…

# Calculation of activity

Conversion *X*

$$X_i = \frac{n_{i,0} - n_i}{n_{i,0}}$$

Reaction rate *r*

$$r = \frac{1}{\nu_i} \frac{dn_i}{m_{cat} dt}$$

Turn over frequency

$$TOF\left[s^{-1}\right] = \frac{N_i}{N_{sites} \cdot t}\left[s^{-1}\right]$$

- Nature and number of active sites usually unknown
- Difficult / impossible to measure TOF in heterogeneous catalysis !!!

# Calculation of selectivity

| | key compound consumed | sum of products |
|---|---|---|
| stochiometric factors | formally correct, requires high knowledge level $S_{cf}$ | $S_{pf}$ |
| number of carbon atoms | $S_{cn}$ | most convenient, requires minimum of knowledge $S_{pn}$ |

$$S_{cf}(\text{product}) = \frac{\frac{\nu(\text{key compound})}{\nu(\text{product})} \cdot c(\text{product})_{out}}{c(\text{key compound})_{in} - c(\text{key compound})_{out}}$$

$$S_{pn}(\text{product}) = \frac{\frac{n_{\text{carbon atoms}}(\text{product})}{n_{\text{carbon atoms}}(\text{key compound})} \cdot c(\text{product})}{\sum \frac{n_{\text{carbon atoms}}(\text{product})}{n_{\text{carbon atoms}}(\text{key compound})} \cdot c(\text{product})}$$

- Preferred raw data is the concentration as a function of time

# Developing quality control

THEORY DEPARTMENT
Fritz-Haber-Institut, MPG

BasCat
UniCat BASF JointLab

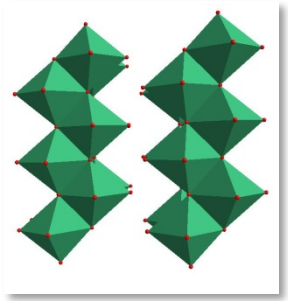
Data Science Project in Oxidation Catalysis

Generation of a Reference Data Set

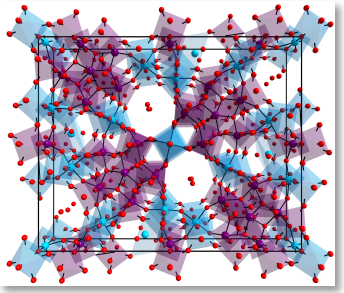
1. Definition of clear rules to test and characterize catalysts in form of an experimental handbook
2. Synthesis of 10 well-known catalysts in selective oxidation that show diversity
3. Testing (C2-C4 oxidation) and characterization of the catalysts according to the workflow described in the handbook involving activated and spent catalysts
4. Data evaluation – finding a continuous function $P(\mathbf{d})$ that generalizes the discrete data sets $\{P_i, \mathbf{d}_i\}$

Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M., Big Data of Materials Science: Critical Role of the Descriptor. *Physical Review Letters* **2015,** *114*, 105503.

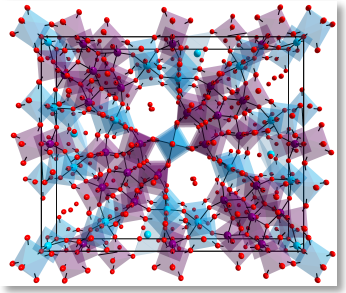
# Catalyst platform for training in ethane, propane, *n*-butane oxidation



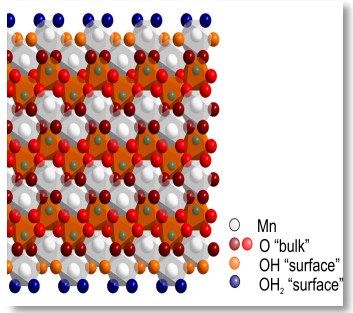
$MoO_3$



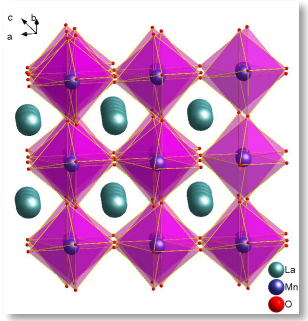
Mo-V oxide



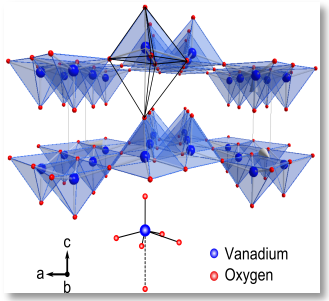
Mo-V-Te-Nb oxide



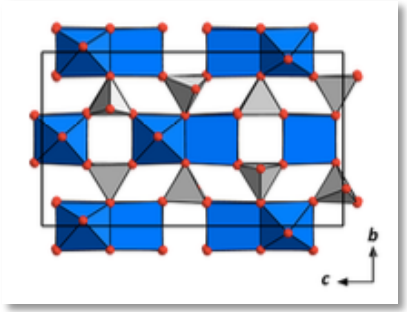

$MnWO_4$



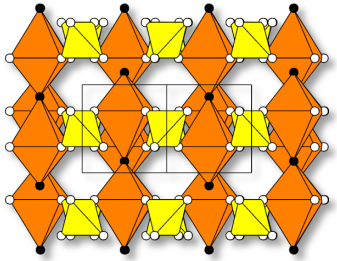

$SmMnO_3$



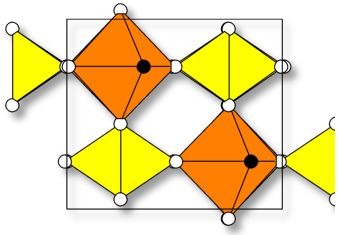

$V_2O_5$



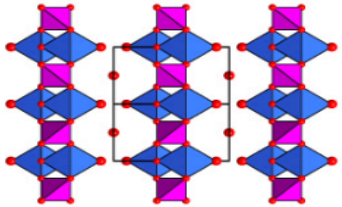
$(V^{IV}O)_2P_2O_7)$



$\alpha_{II}$-$VOPO_4$



$\beta$-$VOPO_4$



$VOPO_4*2H_2O$

10 g of each, pressed and sieved

# Kinetic data

| | Catalyst preparation | | | | Characterization | | | | | Catalysis | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Synthesis | | Activation | | XRD | | | | ... | *n*-butane oxidation | | | | ... |
| Catalyst ID | Synthesis technique | ... | $T_{max}$ [°C] | ... | ICSD | *a* [Å] | ... | ... | | Feed | *r* (300°C) [mmol $g^{-1}$ $h^{-1}$] | *r* (300°C) [mmol $m^{-2}$ $h^{-1}$] | $E_a$ [kJ $mol^{-1}$] | |
| 1 | HT | | 650 | | | | | | | Standard | | | | |
| | | | | | | | | | | Less $O_2$ | - | - | - | |
| | | | | | | | | | | ODH | - | - | - | |
| | | | | | | | | | | Fuel rich | - | - | - | |
| | | | | | | | | | | 5% water | - | - | - | |
| | | | | | | | | | | 10% water | - | - | - | |
| | | | | | | | | | | 20% water | - | - | - | |
| 2 | P | | | | | | | | | Standard | | | | |
| | | | | | | | | | | Less $O_2$ | - | - | - | |
| | | | | | | | | | | ODH | - | - | - | |
| | | | | | | | | | | Fuel rich | - | - | - | |
| | | | | | | | | | | 5% water | - | - | - | |
| | | | | | | | | | | 10% water | - | - | - | |
| | | | | | | | | | | 20% water | - | - | - | |
| ... | | | | | | | | | | | | | | |

- Single rate not useful
- Broad parameter space and different reactions increase the number of data
- How to optimize selectivity? For example:

$$\frac{1}{r_{C_3H8}} = f(r_{CO_2})$$

# Chemical and physical properties

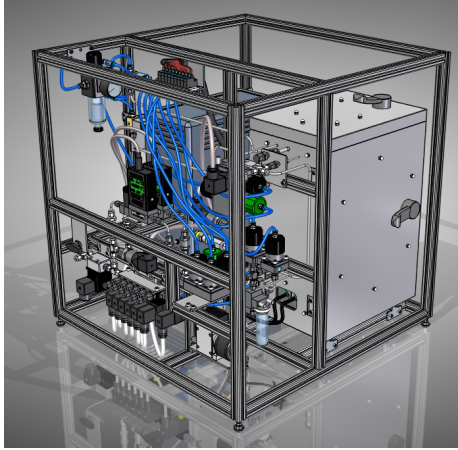
| Property | Physical value | Methods |
|---|---|---|
| Bulk crystal structure (ICSD#) | Lattice constants *a*, *b*, *c* [Å], $\alpha$, $\beta$, $\gamma$ [°]<br>Unit cell volume *V* [$nm^3$]<br>crystallographic density $\rho_{crystal}$ [$g/cm^3$] | XRD, TEM |
| Bulk chemical composition | $c_M$, $c_O$ [at%] | XRF, ICP-OES, EDX |
| Surface composition | $c_M$, $c_O$ [at%]<br>surface active site density [$nm^{-2}$] | XPS |
| Binding energy, surface oxidation state | B.E. [eV], $M^{n+}$ | XPS |
| Valence band onset, M nd onset, *E* secondary electron cutoff, work function, maximum band bending at the surface "s" (surface potential barrier), change in electron affinity | $E_{VB}$, $E_{nd}$, $E_{cutoff}$, $\Phi$, $eV_s$, $\Delta X$ [eV] | XPS, NAP-XPS, ResPES |
| Absorption edge energy | O K-Edge, V $L_3$-Edge [eV] | NEXAFS |

# Chemical and physical properties

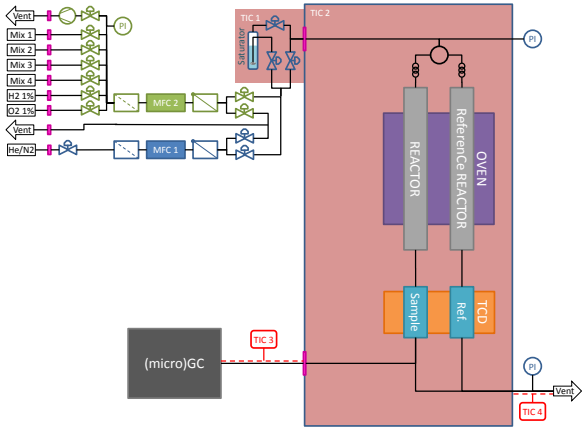
| Property | Physical value | Methods |
|---|---|---|
| Specific surface area (BET)<br>Total pore volume, mesopore volume | $S_{BET}$ [$m^2/g$]<br>$V_{total}$, $V_{mp}$ [$m^3/g$] | Nitrogen adsorption |
| Heat of adsorption<br>Ads. capacity of reactants ($C_nH_{2n+2}$, $O_2$) | $\Delta H_{ads}$ [kJ $mol^{-1}$]<br>$n$ [mol/g] | Microcalorimetry |
| Conductivity, real part of permittivity $\varepsilon'$,<br>imaginary part of permittivity $\varepsilon''$,<br>Apparent activation energy of conduction | $\sigma$ [S/m], $\varepsilon'$, $\varepsilon''$,<br>$E_c$ [eV] | MCPT |
| Optical edge energy | $E_{edge}$ [eV] | UV/Vis |
| Oxidation/reduction equivalents | Oxygen defect density [$V_o^{\cdot\cdot}/nm^2$] | TPR/TPD |
| Desorption temperature | $T$ [K] | TPD |
| $^{18}O$ exchange temperature as function of feed, M-O stretching frequencies | $T$ [K]<br>$\nu$ [eV] | Raman |
| Activity CO oxidation | $T_{10}$, $T_{50}$ [K] | Fixed bed reactor |

# Hardware

Pierre Kube



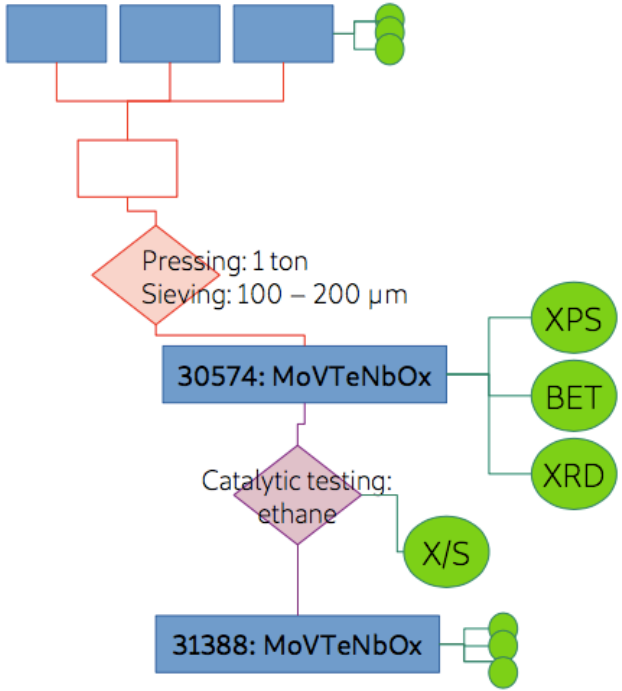
- Kinetic reaction experiments in steady state and pulse operation with temperature-programmed reactions / desorption experiments allows to investigate the very same specimen without the need to transfer the material
- Advanced programing operation of the central unit allows the generation of experiment trains in recipe style probing relevant properties in consecutive order
- Kinetic experiments with pre-programmed parameters are interlaced with activation steps and determination steps of redox/acid-base properties; Fully automated operation allows the execution of long-term experiment series (typically up to 100 hours)



# Data management

Peter Kraus

- **Tree based structure**
  - Sample numbers as nodes
  - Treatments as paths
- **Samples**
  - Provenance: synthesis, pre-treatment, parent species
  - Characterisation: "standard methods"
  - Performance: catalytic testing
  - Tags: "MoVTeNbOx" or "Clean Data"
  - Spent samples get a new sample number
- **Treatments**
  - Steps in synthesis, pre-treatment
  - Processes with attached data: Catalytic testing, BESSY XPS, Conductivity, Microcalorimetry...



# Data management

Peter Kraus

- **NoSQL database**
  - JSON-based structures for structure and metadata
    - Human- and machine- readable
    - Non-rigid, extensible
  - Allows "binary" and textfile attachments
    - CSV files have to be annotated
  - MongoDB a good initial choice

- **October:** Collect all available data for 30574
  - Estimate overall size of the archive
  - First implementation of data structure

```
{
    "id": ,
    "sn": 30574,
    "provenance": {
        "parents": [29345],
        "treatment": [
            {"calcination": {"T": , "unit": "°C"}},
            {"pressing": {"pressure": 1, "unit": "ton"}},
            {"sieving": {"size": [100, 200], "unit": "µm"}},
            ],
        "data": [
            {"id": [...], "format": "Pico Technology logs"}
            ]
        },
    "XRD": {
        "a": {"value": 21.1733, "error": 0.0011, "unit": "Å"},
        "b": {"value": 26.6574, "error": 0.0017, "unit": "Å"},
        "c": {"value": 4.01639, "error": 0.00018, "unit": "Å"},
        "V": {"value": 2266.9, "error": 0.2, "unit": "Å³"},
        "crystal system": "orthorhombic",
        "space group": {"symbol": "Pba2", "number": 32}
        "data": [
            {"id": [...], "format":, "hash": }
            ]
        }
}
```

# Conclusion

- Clean experimental data reveal the complexity of describing “the function” of a material
- The handbook and the resulting clean data will serve as input base for discussion in the community
- In addition to metadata storage, some compulsory procedures need to be established and control mechanisms need to be introduced
- Minimum requirements, conditions, descriptive metadata, and benchmark standards will be defined while ensuring the necessary flexibility and freedom of research