

**Abstract:****Understanding Catalyst Reactions by Time-Resolved Spectroscopy and Mathematic Modelling****Co-supervisor 1: PD Dr. Marcus Weber, ZIB, CoE Math+, SALSA, CRC1114****Co-supervisor 2: Dr. Marius Horch, FUB, UniSysCat**

Complex catalytic systems are characterized by chemical and non-chemical processes that proceed on multiple time and length scales. Understanding these processes as well as their spatial and temporal coupling requires time-resolved spectroscopic techniques that yield real-time insights into molecular structure and dynamics. Extracting the vast amount of information encoded in sequential spectroscopic datasets is far from trivial, though, due to kinetic complexity and a high number of intermediates that may be partly invisible or indistinguishable by the spectroscopic technique used. In addition, datasets from time-resolved spectroscopy often require extensive processing, which also holds the risk of including human bias in the data interpretation. The proposed project will address these challenges by establishing mathematic strategies for the unbiased analysis of time-resolved spectroscopic datasets. Combining non-negative matrix factorization with physical boundary conditions, these strategies will help deducing catalytic mechanisms from spectroscopic experiments. Guided by two groups with a background in chemical mathematics and advanced spectroscopy, the successful candidate will work in a highly interdisciplinary environment and in tight collaboration with other partners within UniSysCat.

## Extended description version of the project:

### Understanding Catalyst Reactions by Time-Resolved Spectroscopy and Mathematic Modelling

**Co-supervisor 1: PD Dr. Marcus Weber, ZIB, CoE Math+, SALSA, CRC1114**

**Co-supervisor 2: Dr. Marius Horch, FUB, UniSysCat**

#### 1. Overall goal of the project

At its very core, catalysis is concerned with the rates of chemical reactions and strategies for their acceleration. As a consequence, the development of new catalytic strategies requires insights into the time-evolution of model catalysts as well as a thorough mechanistic understanding of the reactions they catalyse. Time-resolved spectroscopic techniques yield valuable information on molecular structure and dynamics that is vital for reaching these goals. Extracting this information from complex and often extensive experimental datasets represents a formidable task, though. The proposed project will address this challenge by **establishing powerful and unbiased mathematical strategies for analysing time-resolved spectroscopic data of complex catalytic systems.**

To this end, **the project will focus on selected model catalysts** that (A) raise interesting questions about the nature, sequence, and kinetics of the probed reaction steps, (B) can be studied by time-resolved spectroscopic techniques, and (C) yield versatile types of datasets that can serve as test and training cases for developing powerful and transferrable mathematical strategies. Possible model catalysts, for which extended time series of spectra are available, include a **Ru-Hbpp complex** catalysing the oxidation of water and different types of enzymes that catalyse the reversible cleavage of molecular hydrogen, so-called **hydrogenases**. Methodologically, we will focus on **vibrational spectroscopy**, i.e. resonance Raman (RR) and infrared (IR) techniques.

#### 2. State of the art

Spectroscopic studies within UniSysCat have yielded extensive time series of data for different catalysts and reactions. While all these datasets represent valuable targets and test cases for the strategies to be established in the proposed project, we will focus on the two model systems mentioned above and the relevant mathematical approaches in the following.

**Ru-Hbpp** is a homo-binuclear ruthenium complex that **catalyses the oxidation of water**. Different mixed-valence intermediates as well as oxygen and water adducts are supposedly formed during this reaction. The complex and its catalytic reaction with water has been studied by **time-resolved RR spectroscopy**, but analysis of the data has been proven to be difficult, *inter alia* because individual intermediates exhibit different levels of resonance enhancement or may even be invisible by RR spectroscopy. Due to these challenges, an established NMF approach (*vide infra*) did not yield conclusive insights into the catalytic mechanism and the kinetics of elementary steps.

**Hydrogenases catalyse the cleavage of molecular hydrogen**, typically at a binuclear active site containing iron (and nickel). This unique metal cofactor also contains carbon monoxide and cyanide ligands, which represent ideal probes for **IR spectroscopic studies**. Various IR spectroscopic datasets are available that reflect different types of relaxation reactions of hydrogenases, e.g. from a reduced towards an oxidized state, from a light-induced towards a dark-adapted state, or from a vibrationally excited back to the ground state. These data contain information about catalytic transformations and molecular properties of the involved intermediates, but a proper understanding is difficult, *inter alia* due to the occurrence of chemically different intermediates with similar spectroscopic signatures.

**The outlined challenges can be addressed with different mathematical tools**, some of which will be employed and extended in the proposed project. In the sequential measurement of RR or IR signals, a total spectrum is recorded at each point in time. Each of these spectra can be understood as a weighted sum of the “pure” spectra of the species involved. Mathematically, this can be expressed in form of a matrix equation

$$M = \chi S,$$

where  $M$  lists the measured signal intensities per wavelength, row by row,  $\chi$  contains the weightings of the "pure" spectra per point in time and  $S$  contains the "pure" spectra [1]. Since all matrices involved are non-negative for a given  $M$ , the problem of determining the weights and the pure spectra is a **non-negative matrix factorization problem (NMF)**. This problem has no unique solution, and even the number of species is not pre-defined. Approaches that reduce the number of possible solutions exclusively with the help of *mathematical conditions* on the matrices are currently much discussed [2,3]. So far, however, they have not provided fully satisfying solutions. On the other hand, approaches that require precise *physical knowledge* of the kinetics (which is coded in the matrix  $\chi$ ) or precise *knowledge of the "pure" spectra* of the species involved ( $S$ ) lead to results that depend very much on possibly invalid assumptions about the investigated process.

[1] **K. Fackeldey, J. Röhm, A. Niknejad, S. Chewle, M. Weber:** Analyzing Raman Spectral Data without Separability Assumption. *J Math Chem*, 59:575–596, 2021.

[2] **M. Weber:** Implications of PCCA+ in Molecular Simulation. *Computation*, 6(1):20, 2018.

[3] **R. Luce, P. Hildebrandt, U. Kuhlmann, J. Liesen:** Using Separable Nonnegative Matrix Factorization Techniques for the Analysis of Time-Resolved Raman Spectra. *Appl Spectrosc.* 2016 Sep;70(9):1464-75.

### 3. Specific aims and how they may be reached:

The following specific **Key Objectives** will be addressed by the project:

1. Establish mathematical strategies for **interpreting sequential IR spectra** of catalysts. One key challenge here is the analysis of spectra reflecting intermediates that are spectroscopically identical but kinetically different, or *vice versa*. This situation violates the typically assumed linear independence of the columns of  $\chi$  and the rows of  $S$  in the NMF approach.
2. Improve mathematical strategies for **analysing Raman spectroscopic data – also in the resonant limit**, which is particularly important when studying catalytic sites. The NMF approach usually requires that the weights of the pure spectra add up to one, i.e., that the matrix  $\chi$  has the row sum 1. This condition is not met if resonance enhancement varies among intermediates, e.g. if spectroscopically invisible species (no enhancement) are involved.
3. Introduce a **generalized, adaptable approach for understanding sequential data** from spectroscopy or other time-resolved experiments. *Inter alia*, this requires the development of strategies that work around the strict positivity condition for the matrices in the NMF approach, which is too rigorous for analysing real-world data with noise.

To address the Objectives, a **twofold approach** will be used: (I) Mathematical constraints that are incompatible with the properties of real-world spectra (*vide supra*) will be softened, as exemplified in the Work Packages (*vide infra*). (II) Softening mathematical constraints will inevitably lead to an increased number of possible solutions for the optimization problem. By including alternative constraints, a reformulated NMF approach will yield a **limited number of physically and chemically meaningful solutions**. To give some examples: (A) The kinetic model on which the experiment is based can be read from the matrix  $\chi$  [4]. Partial physical knowledge about the kinetic model can, thus, exclude certain  $\chi$ -matrices from the set of possible solutions. (B) Individual known species or specific initial conditions in the experiment can be taken into account by specifying parts of  $S$  or  $\chi$  or by selecting the starting point of the optimization. (C) The matrix  $\chi$  can be used to estimate the eigenvalues of a Markov operator whose eigenvalues correspond to characteristic time scales of the process [5]. Together with prior kinetic knowledge, this allows excluding further  $\chi$  matrices. The mathematical modelling of these (and further) physical conditions will be part of the research.

The project is divided into three synergistic **Work Packages**:

1. Strategies for **analysing sequential IR spectra** will be developed using **hydrogenase** as a model catalyst. Various reaction sequences will be studied to (1) gain detailed insights into involved intermediates and elementary steps and (2) develop mathematical tools for analysing different types of challenging datasets. For instance, the linear-independence condition from Objective 1 could be circumvented by a reformulated NMF approach that requires linear independence for one of the two matrices  $\chi$  or  $S$  only.

2. Advanced mathematical methods for **analysing time-resolved (resonance) Raman spectra** will be developed using **Hbpp** as a model catalyst. As for the first Work Package, this part of the project will (1) facilitate the understanding of the catalytic processes reflected by the specific spectroscopic data and (2) yield a suitable mathematical toolkit for similar studies within UniSysCat. For instance, the normalization criterium from Objective 2 may be softened by requiring an upper bound for the sum of the entries of  $\chi$  only.
3. Integrating strategies from Work Packages 1 and 2, a **generalized approach for the analysis of sequential (spectroscopic) data** will be established. This part of the project will likely include the analysis of data obtained from other challenging catalytic systems studied in UniSysCat, thereby contributing to both ongoing research and mathematical method development. One of the general key challenges – the positivity condition from Objective 3 – may be reformulated as a penalty term to the objective function.

As the project will focus on the development of mathematical tools for addressing the Objectives, the ideal candidate should have a solid mathematical background, but a thorough understanding of chemical and physical phenomena is also a prerequisite for a successful completion of the project. We think that a physicist or theoretical chemist would be a particularly suitable candidate for translating physicochemical expert knowledge into mathematical modelling. Experimental work or quantum chemical calculations are not key aspects of the project but may be included if helpful. The two supervisors can provide guidance regarding the mathematical and physicochemical aspects of the project, and necessary computational and laboratory resources are available in their groups.

The planned project will involve collaborations with the groups of P. Hildebrandt and I. Zebger. We expect further fruitful interactions with other (spectroscopic) groups within UniSysCat.

[4] **P.J. Schmid**: Dynamic mode decomposition of numerical and experimental data. *Journal of Fluid Mechanics* 656(1):5–28, 2010.

[5] **M. Sarich, R. Banisch, C. Hartmann, C. Schütte**: Markov State Models for rare Events in Molecular Dynamics. *Entropy* 16:258-286, 2014.