Abstract:

Single ion activities for pH determination: Combining robotics and artificial intelligence to tackle an unsolved fundamental problem

Co-supervisor 1: Prof. Dr. Matthias Driess, TU Berlin

Co-supervisor 2: Prof. Dr. Karsten Reuter, FHI Berlin

The pH is a widely applied concept in all fields of catalysis defined as the negative decadic logarithm of single ion activity of H^+. However, in general, single ion activities are thermodynamically not defined and are a conventionally immeasurable quantity. Therefore, the pH of simple solutions such as seawater, blood, cell plasma, and the most common electrolytes such as aqueous NaOH or HCl is not precisely defined. The challenge one faces when trying to measure a single ion activity is that always the precise knowledge of single ion activities is required to measure another single ion activity. To overcome this challenge, a combination of two recently emerging fields, robotics and artificial intelligence, should be used. The robotics should be utilized to automatically measure a large set of data that depends on a single ion activity and an additional factor (e.g., another single ion activity or a liquid junction potential). Artificial intelligence should then be used to isolate the single ion activity from this data set.

Extended description version of the project:

Single ion activities for pH determination: Combining robotics and artificial intelligence to tackle an unsolved fundamental problem

Co-supervisor 1: Prof. Dr. Matthias Driess, TU Berlin

Co-supervisor 2: Prof. Dr. Karsten Reuter, FHI Berlin

1. Overall goal of the project

The pH is a widely applied concept in chemical and biologic transformations, catalysis as well as in the synthesis of the respective catalysts. Nevertheless, the pH values of simple solutions such as 1 M NaOH, blood, or seawater are not precisely defined, as the pH definition is based on the single-ion activity of H^+. In general, single ion activities are thermodynamically not defined and are a conventionally immeasurable quantity. Therefore, the official definition of the pH scale provided by IUPAC in 2002 (Buck et al. *Pure Appl. Chem.* **2002**, *74*, 2169–2200) only covers the pH range 1-13 (ionic strengths below 0.1 mol/kg), where the Debye-Hückel (DH) theory is still suitable for the calculation of single ion activities. In this project, we aim to address this since a century existing pH definition problem by combining two emerging fields, robotics (Prof. Driess) and artificial intelligence (Prof. Reuter).

2. State of the art

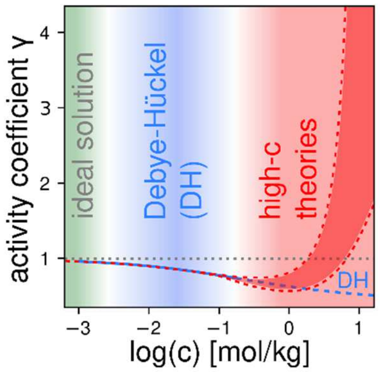

Figure 1: Activity coefficient *γ* of water within the Debye-Hückel (DH) model. Starting at concentrations *c* ~ 0.1 mol/kg DH theory becomes increasingly inaccurate and various high-*c* theories predict strongly non-ideal behavior.

The concept of a single ion activity $a = \gamma\ c$ generalizes the dependence on the ion concentration c for non-ideal solutions where the activity coefficient $\gamma \neq 1$ (Burgot, *Ionic Equilibria in Analytical Chemistry* **2012**, Springer Science & Business Media). A range of theories exist to describe γ, comprising notably Debye-Hückel (DH) theory for the low-concentration limit and various approximate theories like the Davies model for higher concentrations (Debye and Hückel, *Physik Z.* **1923**, 24, 185; Davies, *J. Chem. Soc.* **1938**, 2093-2098; Guggenheim, *Phil. Mag.* **1935**, 19,588; Pitzer, *J. Phys. Chem.* **1973**, 77(2), 268–277). Figure 1 plots the prototypical behavior of γ for water at room temperature and 1:1, monovalent electrolyte solutions as predicted by these theories, demonstrating the expected dramatic, but only approximately known non-ideal behavior ($\gamma \neq 1$) of electrolyte solutions at higher concentrations.

As stated above, the IUPAC approach to define pH relies on DH theory which only holds for ionic strengths below 0.1 mol/kg (Buck et al. *Pure Appl. Chem.* **2002**, *74*, 2169–2200, cf Fig. 1). For solutions with higher ionic strengths, up to date, no precise answer can be given on their pH value, as there is no accurate theory available to establish the single ion activities. Thus, this includes the most relevant intermediate concentration regime around c ~ 1 mol/kg, in which the picture of describing an electrolyte through its single ion activities should generally still be useful. This regime covers prominent aqueous electrolytes such as seawater, blood, the cytoplasm of a human cell, or alkaline and acidic solutions used in water splitting. With respect to the latter, it is important to realize that due to mass transport limitations ion concentrations >1 mol/kg (mean ion separation < 4 water molecules) can easily be reached at the electrodes even in formally much lower concentrated electrolytes. In some cases, this lack of a precise value of the pH of commonly applied aqueous solutions can lead to significant problems. An example is alkaline water splitting, where a literature review has found that the pH of 1 M KOH has been reported to be anything between 13.5-14 (Hausmann et al. *ACS Energy Lett.* **2021**, *6*, 3567–3571). For the hydrogen evolution reaction performed with a state-of-the-art catalyst, if a pH of 14 instead of 13.5 is used to determine the energy loss with the Nernst equation, an error of 60% follows, as the thermodynamics of this reaction are directly dependent on a_{H+}.

Lacking reliable theories for the intermediate concentration regime, measurements appear as the straightforward alternative. However, as mentioned, the problem here is that always single ion activities must be known to measure another single ion activity. This can be illustrated with an example most relevant for the pH. The IUPAC pH definition relies on the reversible hydrogen electrode (RHE), which is pH/a_{H+} dependent. This pH dependence is described by the Nernst equation,

$$E_{TD} = E^{\ominus} - \frac{RT}{zF} lnQ, \tag{1}$$

where E_{TD} is the thermodynamic (half-)cell redox potential under the experimental conditions in V, $E^{\ominus}$ the standard (half-)cell redox potential in V, R the universal gas constant in J mol^{-1} K^{-1}, T the temperature in K, z the number of transferred electrons, F the Faraday constant in s A mol^{-1}, and Q the reaction quotient. For the RHE ($H_2O + e^- \rightarrow 0.5H_2 + OH$) under a constant H_2 pressure of 1 bar this Nernst equation becomes

$$E_{TD} = E^{\ominus} + \frac{RT}{F} ln(a_{H+}). \tag{2}$$

So, if $E^{\ominus}$ is known and E_{TD} can be measured precisely, this equation directly connects a measurable quantity to a single ion activity. However, the E_{TD} does not only rely on the pH dependency of the reversible hydrogen electrode, but also on something called the liquid junction potential. Yet, to calculate the liquid junction potential, the knowledge of pH/a_{H+} and the other constituting single ion activities are needed. Therefore, one always measures both quantities at once and never isolated (Bates, R. G. *Determination of pH*; 1973). Thus, to measure pH/a_{H+} one must separate the two contributions. Unfortunately, this is by definition not possible using a thermodynamic law as those require the knowledge of the quantities one intends to measure. In addition, it is at present unclear at which concentrations the separation of mean activities into single ion activities might become conceptually problematic, e.g. when specific ion-ion interactions are not transferable, as remarked already by Brönsted (*J. Am. Chem. Soc.* **1922**, 44(5), 877–898). Therefore, a different ansatz and new methods are required to solve this century old problem.

3. Specific aims and how they may be reached.

The aim of this project is to develop a new approach for the determination of single ion activities and thus pH values of solutions with an electrolyte concentration above 0.1 mol/kg. As direct measurements of such single ion activities are not possible by thermodynamic methods, a combination of two recently emerging fields, robotics and artificial intelligence, should be used to instead determine a maximally consistent and robust set of single ion activities across the periodic table. In this regard, in the first work package (**WP1**), automatized robotics will establish an infrastructure to automatically gather a large data set of maximum information content. In the second work package (**WP2**), artificial intelligence (AI) is used to drive the automatized lab set up of WP1 and infer the ion activities from the obtained data.

WP1: Currently, robots that can perform electrochemistry measurements fully automatically are being reported in academia (e.g., Cebi *et al. Procedia Manuf.* **2021**, *54*, 203–208) and are commercialized (e.g., FASTER: https://research.csiro.au/lithium/research-areas/faster-robotic-electrochemistry/). Herein, we want to collect a large number of mean ion activities (a_{mean}) of salt pairs such a H^+ and Cl^- (a physically measurable quantity defined as $a_{\text{mean}} = (a_{H+} \times a_{Cl-})^{0.5}$), and electrical potentials that depend on a single ion activity together with a liquid junction potential, e.g., reversible hydrogen electrode measurements. Both quantities can be measured using a simple two electrode setup with a common reference electrode. The automatic process merely must involve the fabrication of electrolytes with different salts and at different salt concentrations.

The obtained mean ion activities and electrical potentials will depend on the single ion activities of cations and anions in the electrolyte. For the mean ion activity coefficients, one will always measure two single ion activities at the same time. A large set of anions (A1, A2, A3, A4, A5...) should be measured for one cation (C1), e.g., C1+A1, C1+A2, C1+A3, C1+A4.... Then, for each anion a large set of cations should be measured, e.g., C1+A1, C2+A1, C3+A1.... Furthermore, for the same cation-anion pairs, the electrical potential in a reversible hydrogen setup should be measured, which will yield values that include the single ion activity of either the cation or anion together with the liquid junction potential. From this strongly interdependent data, in WP2, AI will infer the single ion activity coefficients.

WP2: The scope of this WP is the development of a methodology for analyzing and guiding the automated experiments from WP1, with the aim of inferring a robust and consistent set of single ion activities from the measured mean ion activities and electrical potentials. For data analysis, a *Bayesian Inference* (BI) framework will be used. This BI model will simultaneously infer the activities of all ions in the full concentration range of interest, as well as model correlations between measurements. BI is particularly suited for this project, as it allows for sequential data acquisition and provides reliable uncertainty estimates for the quantities of interest. Furthermore, Bayesian methods can use prior information about activities and correlations to improve the convergence of the predictions. In the context of this project, such prior information could come from analytical models like DH or from first-principles calculations (see below). Based on the BI model, an AI approach will be developed to steer the data acquisition process. In particular, the uncertainty estimates for individual predictions and correlations between different predictions will be used to decide which experiment (i.e. which ion pair at which concentration) to perform next. Here, strategies from active and reinforcement learning will be explored, with the aim of obtaining a reliable and data-efficient process. Ultimately, the combined AI/BI framework will provide estimates of single ion activities which are maximally consistent across all investigated ion-pairs and concentrations and are thus suitable for a general definition. Just as importantly, any remaining uncertainty in the high-concentration regime (beyond measurement errors) can provide important insights into when the concept of a single ion activity fundamentally breaks down.

As activities are directly related to the chemical potential of ionic species in solution, atomistic modelling in combination with free energy methods can be used to provide theory-derived single ion activities or at least proxies for these. This is in particular true as the strongly non-ideal behavior typically sets in when ions are separated from each other by less than 4 water molecules (>1 mol/kg), a regime that can be assessed even with computationally expensive *ab initio* methods. Therefore, alongside the direct involvement with the experimental project, first-principles simulations shall be performed for selected ions and ion pairs dissolved in water at high ion concentrations. Here, purely enthalpic contributions to the activity can be trivially assessed by comparing the average energetics at different concentrations e.g. along molecular dynamics runs, a method that is frequently used e.g. to assess the stability of adsorbates at solid-liquid interfaces. However, chemical potential estimates necessitate in principle free energy simulations, which – as a result – shall be performed as well. We anticipate, that these studies will help to clarify the difference between free energy and enthalpy methods in liquid water environments, which is of general importance for the theoretical community. Simultaneously, the computed single ion activities provide valuable priors for the BI approach described above that allow to minimize the number of practically required measurements (i.e. maximum data efficiency).

Facilities: The Driess group is well equipped with electrochemical set-ups and all required chemicals for the fabrication of the electrolytes. During the project, this equipment will be used together with robotics to create an automated setup. The Reuter group disposes over sufficient computing resources at the centralized Max Planck Computing and Data Facility to carry out all planned modeling and calculations.