The HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY (HJIC)
formerly (until 2012) the Hungarian Journal of Industrial Chemistry

The HJIC is an international periodical that focuses on results of fundamental and applied research in the field of

- Biotechnology
- Chemical Engineering Science
- Chemical Processes
- Energetics
- Environmental Chemistry
- Environmental Engineering & Technology
- Industrial Management
- Material Science
- Mechanical Engineering
- Mechatronics
- Process & System Engineering
- Recycling

in the form of original papers, technical reports, reviews, communications, and conference proceedings written in English. HJIC is abstracted by Chemical Abstracts, indexed by Hungarian Scientific Bibliographic Database (MTMT), and archived by REAL-J Repository of the Hungarian Academy of Sciences.

EDITORIAL OFFICE: UNIVERSITY OF PANNONIA, P.O. BOX 158, VESZPRÉM, 8201 (HUNGARY)
Tel.: +36 (88) 624-746, e-mail: hjic@almos.uni-pannon.hu; web: hjic.mk.uni-pannon.hu
Felelős szerkesztő: Dr. Szilágyi Róbert Károly, Ph.D. Nyelvi lektor: Mr. Matthew Chen
Kiadja: Pannon Egyetem, 8200 Veszprém, Egyetem u. 10.
Elektronikus terjesztés: De Gruyter Open (formerly Versita), Warsaw, Poland
Levélcím: H-8201 Veszprém, Postafiók 158, Tel.: (88) 624-000
Felelős kiadó: a Pannon Egyetem, Mérnöki Kar dékánja (Prof. Dr. Szalai István, Ph.D.)

TABLE OF CONTENTS

DE GRUYTER OPEN

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
Vol. 44(1) pp. 1–28 (2016)
hjic.mk.uni-pannon.hu
DOI: 10.1515/hjic-2016-0001

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

MASS ACTION AND CONSERVATION OF CURRENT

ROBERT S. EISENBERG*

Department of Molecular Biophysics and Physiology, Rush University Medical Center, Chicago IL 60612, USA

The law of mass action does not force a series of chemical reactions to have the same current flow everywhere. Interruption of far-away current does not stop current everywhere in a series of chemical reactions (analyzed according to the law of mass action), and so does not obey Maxwell's equations. An additional constraint and equation is needed to enforce global continuity of current. The additional constraint is introduced in this paper in the special case that the chemical reaction describes spatial movement through narrow channels. In that case, a fully consistent treatment is possible using different models of charge movement. The general case must be dealt with by variational methods that enforce consistency of all the physical laws involved. Violations of current continuity arise away from equilibrium, when current flows, and the law of mass action is applied to a non-equilibrium situation, different from the systems considered when the law was originally derived. Device design in the chemical world is difficult because simple laws are not obeyed in that way. Rate constants of the law of mass action are found experimentally to change from one set of conditions to another. The law of mass action is not robust in most cases and cannot serve the same role that circuit models do in our electrical technology. Robust models and device designs in the chemical world will not be possible until continuity of current is embedded in a generalization of the law of mass action using a consistent variational model of energy and dissipation.

Keywords: law of mass action, current continuity, chemical device design, variational methods, ion channels

1. Introduction

The law of mass action is used widely, nearly universally, in chemistry to describe chemical reactions. The law of mass action does not automatically conserve current, as is clear from the mathematics of a simple case, chosen to illustrate the issues involved (*Fig.1*). If current is not conserved in a theory, charges accumulate that cannot accumulate in the real world. In the real world, tiny charge accumulation - much less than one percent - produces forces that change predictions of the theory a great deal. Indeed, in the real world tiny charge accumulation produces forces that destroy biological membranes and thus living systems, forces large enough to ionize atoms, creating the ionized plasma of electrons, we call a spark, or lightning, that can destroy laboratory apparatus, if not the laboratory itself.

Turning to mathematics for proof. Readers should distrust a claim of conflict between long established laws and so I turn to mathematics for proof to show the conflict unambiguously, hopefully without argument. Consider the chemical reaction $X \rightleftharpoons Y \rightleftharpoons Z$: the currents in the two reactions $X \rightleftharpoons Y$ and $Y \rightleftharpoons Z$ are not the same, $I_{XY} \neq I_{YZ}$. Rather, the difference in current is shown to be $I_{XY} - I_{YZ} = z_X k_{xy} \mathrm{F}[X] - z_Y k_{yx} \mathrm{F}[Y] - z_Y k_{yz} \mathrm{F}[Y] + z_Z k_{zy} \mathrm{F}[Z]$. The difference is not zero in general circumstances, nor robust. The difference can be zero only under special circumstances (details are in *Eqs.(6–11)* below).

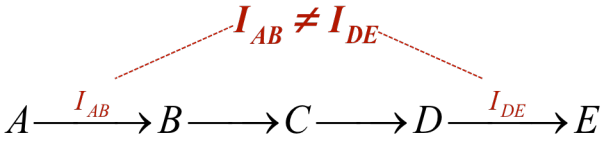

Figure 1. Current flow is not the same in a series of classical reactions.

The Appendix shows that the current imbalance $I_{XY} - I_{YZ}$ between the reactions quickly produces enormous forces. Enormous forces are not observed and we conclude that the law of mass action cannot be generally true, although it can be useful (and true) under special circumstances, as also shown in the Appendix, and of course in the special circumstance of equilibrium, when current does not flow.

Law of mass action is an incomplete truth. The 'law' of mass action seems to be an example of an " … incomplete truth [that] may become ingrained and taken as the whole truth … [thereby confusing] … what is *only sometimes* true with what is *always* true…" (slight rewording of Richard Feynman, p. 15–61 of Ref. [1]). The 'laws' of science are usually learned early in our careers before we have refereed grants and papers, before our critical skills are honed. These scientific laws

*Correspondence: beisenbe@rush.edu

are honored because of their historical role, as much as their logical importance. The laws are often residues of revolutions that once gave us new knowledge. Old ideas can have a life of their own, a momentum that is hard to change when knowledge expands. It is easy to continue to use old ideas uncritically even after they have been overtaken by new knowledge.

Law of conservation of current is different: it is universal. Conservation of current is as universal as any law in science. It is a law that has not aged as the law of mass action has. Maxwell's equations and conservation of current have a present role and validity at least as important as their historical one. They are said to be exact at any single time and over any distance scale that can be observed, from sub-atomic, and much smaller, to interstellar, from times much shorter than those of atomic motion to the years it takes light to travel from stars and nearby galaxies. Their daily use in high energy accelerators involving inconceivably small distances and brief times, their use in the microwave and faster devices of modern technology, and their use in interstellar astronomy are convincing practical proofs of the nearly universal validity of these equations of electrodynamics.

The role of Maxwell's equations depends on the subtle idea of *displacement current in a vacuum, current not carried by the movement of mass*, and this idea is easy to lose sight of in chemical applications where other forms of displacement current involving the properties of matter (usually called 'polarization' in the chemistry literature, along with quite different phenomena, e.g. charge distribution in a carbonyl bond or peptide linkage) receive more attention. The displacement current needs to be incorporated into chemical models, in my view, if it is to satisfy both conservation of mass and conservation of current. Both material displacement current accompanying polarization phenomena, and vacuum displacement current need to be incorporated. On the time-scale of atomic motions, 10^{-16} s, both displacement currents are substantial, and must be incorporated if conservation of current is to be enforced.

An important question is how the law of mass action has produced such useful chemistry, over so many years, if it does not conserve current flow? The answer is that chemistry has not been interested in non-equilibrium systems of molecules with current flow, as much as it has been interested in the molecules themselves. Chemists make molecules, not currents, with the notable exception of electrochemists, and there the difficulties in dealing with long-range electric fields have long been recognized [2–4].

In classical biochemistry, the law of mass action has been used in another context, more biological, with less attention to the virtues of consistency and invariant physical parameters [5]. A senior biochemist recently said, "The art of biochemistry is to study enzymes under conditions that give insights into biological function, even if the rate constants do not fit results over a range of conditions. The art is to choose experimental conditions in which biochemical reactions are well behaved and rate constants resemble those in real biological systems, so results are useful in understanding how living systems work."

The law of mass action was developed to deal with isolated systems, originally with perfect gases [6–10], and allowed chemists to deal with molecular and atomic reactions before physicists became convinced that atoms existed. But isolated chemical reactions must contact the outside world to pass signals and interact with it, just as electronic systems contact the outside world and pass signals through inputs and outputs. Biological systems contact surrounding solutions and cells through ion channels and transporters. The law of mass action was not designed to contact the outside world. It was designed to help chemists build and understand individual molecules. Signals and connections with the outside world almost always involve electricity because charge flows through the contacts that connect chemical reactions with the outside world. The contacts are usually the boundary conditions of mathematical models.

Engineering deals with systems that are not isolated. Its devices contact the world through power supplies and inputs and outputs. Devices have outputs that follow inputs according to simple rules, for example, the output of an amplifier follows the input according to the gain. The input-output rules are derived from Kirchhoff's current law, i.e. continuity of current in one dimension, and Kirchhoff's voltage law. Biology deals with systems that are not isolated. Living systems usually have inputs and outputs and are driven by concentration gradients that are power supplies. Biological systems interact with surrounding solutions, cells, and tissues. The *law of mass action must be extended to deal with inputs and outputs and flow of electrical current* if theories and simulations of nanodevices (technological or biological) are to be useful under more than one set of conditions.

Ionic solutions satisfy three conservation laws, conservation of mass, conservation of charge, and conservation of current. Chemistry uses mass conservation almost everywhere, in the form of the law of mass action [11–13] (see *Eqs.(6–11)* below). Physics uses conservation of charge and current whenever it deals with electricity [1, 14–21]. The flow of charge is continuous in Maxwell's equations, without loss in series circuits,

$$\nabla \cdot \mathbf{I} = 0 \tag{1}$$

The current $\mathbf{I}$ is simple in a vacuum and in a vacuum that also contains particles, e.g. electrons in a vacuum tube. In the latter case, the total current is what I call the Maxwell current $\mathbf{I}_{\text{Mxw}}$ because this is the invariant introduced by Maxwell to allow the light waves of his equations to propagate forever through a vacuum.

$$\mathbf{I}_{\text{Mxw}} = \mathbf{I}_{\text{particles}} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \tag{2}$$

Current in matter is more complex than in a vacuum tube because it includes complex movements of charged particles (and charge inside particles) as well. The complex movement of particles must be described by coupled field equations because the movements are driven by many forces, e.g. mechanical, convectional, thermal, diffusional, as well as electrical. The issues involved are illustrated in *Fig.2* [5] and the following extensive discussion. I then advocate using variational treatments of forces and movements because variational methods *automatically* enforce consistency with the field equations of all the forces.

The material displacement current accompanying polarization $\mathbf{I}_{\text{Pol}}$ is customarily treated differently from the rest of the movement of particles in a tradition started by Faraday and reinforced by Maxwell who did not recognize the existence of permanent, i.e. fixed charge [26]. The polarization of matter is the distortion of the distribution of charge produced by the electric field (roughly analogous with the distortion of the oceans produced by the gravitational field of the moon, that produces tides on our beaches twice a day).

Material displacement current $\mathbf{I}_{\text{Pol}}$ can be separated from the complex movements of particles in matter by two properties:

(1) Material displacement current is transient. If an electric field is applied to matter, $\mathbf{I}_{\text{Pol}}$ flows as the spatial distribution of charge changes, i.e. as it polarizes, but eventually, flow ceases and $\mathbf{I}_{\text{Pol}} \rightarrow 0$, as $t \rightarrow \infty$ even if the local electric field remains constant.

(2) If an electric field is applied, and then turned off, the charge $\mathbf{Q}_{\text{Pol}} = \int_0^\infty \mathbf{I}_{\text{Pol}}\, dt$ that flows when the field is turned on is equal to the charge that flows when the field is turned off, when measured for long enough time periods.

The total displacement current $\mathbf{I}_{\text{Dis}}$ is often isolated and identified by these features in experiments involving transients. In experiments involving sinusoidal applied fields, $\mathbf{I}_{\text{Dis}}$ is usually recognized by its ninety degree phase shift.

$$\mathbf{I}_{\text{Dis}} = \varepsilon_r \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}$$

$$\overbrace{\mathbf{I}_{\text{Dis}}}^{\substack{\textbf{Total}\\ \text{Displacement}\\ \text{Current}}} = \overbrace{\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}}^{\substack{\textbf{Vacuum}\\ \text{Displacement}\\ \text{Current}}} + \overbrace{\mathbf{I}_{\text{Pol}}}^{\substack{\textbf{Material}\\ \text{Displacement}\\ \text{Current}}} \quad ; \quad \overbrace{\mathbf{I}_{\text{Pol}}}^{\substack{\textbf{Material}\\ \text{Displacement}\\ \text{Current}}} = \overbrace{\left(\varepsilon_r - 1\right)\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}}^{\text{Susceptibility}} \tag{3}$$

The vacuum displacement current is a large fraction (say half) of the displacement current on the time-scale of atomic motions, less than 10^{-15} s.

All current produces a magnetic field, according to Maxwell's extension of Ampere's law

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{I} = \mu_0 \left(\mathbf{I}_{\text{Mxw}} + \mathbf{I}_{\text{Pol}}\right) = \mu_0 \left(\mathbf{I}_{\text{particles}} + \mathbf{I}_{\text{Dis}}\right)$$
$$= \mu_0 \left(\mathbf{I}_{\text{particles}} + \overbrace{\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}}^{\substack{\text{Vacuum Displacement}\\ \text{Current}}} + \overbrace{\left(\varepsilon_r - 1\right)\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}}^{\text{Susceptibility}} \right) \tag{4}$$

The divergence operator $\nabla \cdot \mathbf{f}$ evaluates the conservation of flow in the vector field $\mathbf{f}$ it acts on (as its derivation from integral relations shows nicely [27])

$$\nabla \cdot \left(\nabla \times \mathbf{B}\right) = \nabla \cdot \left(\mu_0 \mathbf{I}_{\text{particles}} + \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \left(\varepsilon_r - 1\right)\mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right)$$
$$= \mu_0 \left(\nabla \cdot \mathbf{I}_{\text{Mxw}} + \nabla \cdot \mathbf{I}_{\text{Pol}}\right) = 0 \tag{5}$$
$$\text{so} \qquad \nabla \cdot \mathbf{I} = 0$$

where we use the vector identity that the divergence of a curl is always zero. $\mathbf{B}$ is the magnetic vector field; μ_0 is the magnetic constant, the magnetic 'permeability' of a vacuum; and ε_0 is the corresponding 'electrostatic constant', the permittivity of free space. Note that $\mu_0\, \varepsilon_0 = c^{-2}$, where c is the velocity of light. Magnetism only arises from current flow. It is mysterious that magnetic charges (monopoles) do not exist, i.e. $\nabla \cdot \mathbf{B} = 0$.

Maxwell's equations *Eqs.(1)* and *(5)* guarantee that current is exactly the same everywhere in a series of two terminal devices. If parameters or geometry are changed, electrical forces and potentials change automatically to ensure the same current flows everywhere (in a series system involving Maxwell's equations). Interruption of current anywhere in a series circuit interrupts current flow everywhere, even signals carried by currents far away from the interruption.

The law of mass action needs to be extended to deal with displacement current so it can enforce global continuity of current flow. If a particular chemical reaction is at equilibrium, and no current flows, conservation of charge can be enforced easily on the law of mass action. But current flow presents a different situation because the law of mass action must then be modified to deal with the global properties of the electric field. If we are dealing with technological or biological nanodevices, chemical reactions are connected to the outside world and the law of mass action must be extended to enforce continuity of current flow everywhere under all conditions.

Biological systems involve both chemical reactions and charge. Biological systems are always embedded in ionic solutions, and nearly always involve chemical reactants and enzymes with electrical charge, even if (like water), their net charge is zero. Substrates of reactions catalyzed by enzymes are usually charged and are nearly always dissolved in complex solutions containing the ions Na^+, K^+, and Cl^- and often Ca^{2+}. Trace concentrations of calcium are often used by biological systems as controls that turn biological function on or off. Reactants are almost always charged

in biological systems. Reactants almost always flow and carry current in biological systems. Equilibrium and death are nearly synonymous in life. Biological systems must satisfy conservation of current along with conservation of matter. We face a problem when we try to apply both conservation laws together: The Law of Mass Action does not conserve current.

'*The Law of Mass Action does not conserve current*' seems an unlikely statement and so mathematical proof is needed more than verbal argument. A simple sequence of chemical reactions was examined and the questions "Does the law of mass action conserve current? Is current the same everywhere in the series of reactions? Do the potentials change automatically so current is always the same everywhere in a series circuit?" were asked.

The same questions can then be asked of whatever series of reactions are of interest. Sometimes the answer will be that current does not flow or does not matter and the law of mass action can be easily modified to conserve current. In some symmetrical reactions, the answer will be that current and mass are both conserved, as shown in the Appendix. More often, the answer will be that current and mass are not both conserved, as shown in more detail in the Appendix. In that case, the law of mass action must be extended to maintain continuity of current.

Eq.(6) shows the reactions we use to illustrate the problem:

$$X \underset{k_{yx}}{\overset{k_{xy}}{\rightleftharpoons}} Y \underset{k_{zy}}{\overset{k_{yz}}{\rightleftharpoons}} Z \qquad (6)$$

The reactions of *Eq.(6)* were chosen to show the problem in the simplest case. We define 'law of mass action' for this paper as just *Eqs.(6)* and *(9)* with the symbols defined below. The rate constants in *Eqs.(6)* and *(9)* are taken as constants and thus of course are uncorrelated and independent of each other. If the law was robust, the rate constants found experimentally under one set of conditions would be found under another set of conditions. Such is sometimes the case, but not very often [5, 28].

Generalizations of rate constants are sometimes made but, as discussed below in the section *'How to extend the law of mass action?',* generalizations in the literature [11, 13] deal with 'chemical correlations'[4] and the non-ideality of some types as opposed to current flow, to the best of my knowledge. They do not allow the law of mass action defined here by *Eqs.(6)* and *(9)* to globally satisfy the conservation of current *Eq.(1)*. The generalizations are well designed to deal with systems close to equilibrium with some types of non-ideality, but the generalizations do not discuss current flow, and include little[3] or no[4] discussion about the non-idealities produced by the ionic atmosphere (in the equilibrium case [29]), and changes of shape of the ionic atmosphere (in the non-equilibrium case [30–37]).

Current flow is important in most applications of the law of mass action. It is almost always present if the reaction is part of a device that communicates with the outside world. The global nature of the electric field (illustrated in *Fig.1*) allows remote devices and boundary conditions to change local atomic flows, an effect not necessarily present at equilibrium. In a non-equilibrium system, electrical forces and potentials change everywhere - automatically as a result of the equations of the electric field - to ensure the same current flows in all places (in a series system). Indeed, interruption of current in a series of reactions stops current anywhere, even far away.

Of course, there are special cases in which mass action can by itself conserve current flow exactly and those in fact may be the cases where it has proven most quantitatively useful, particularly when extended to deal with non-ideality [11, 13]. In addition, reactions may conserve current approximately. In other cases, like in *Eq.(6)*, the reaction will not conserve current, even approximately. Each case needs to be studied separately. The Appendix provides more detail.

The central fact - that applies to any chemical reaction, not just to *Eqs.(6)* and *(9)* - is that the global realities of the electric field need to be embedded in the atomic-scale treatment of the reaction, particularly when current flows.

Proof

The current flow in the reactions of *Eq.(6)* is easily shown to be:

$$\begin{aligned} I_{XY} &= \mathrm{F} z_X \cdot k_{xy} [\![X]\!] - \mathrm{F} z_Y \cdot k_{yx} [\![Y]\!] \\ I_{YZ} &= \mathrm{F} z_Y \cdot k_{yz} [\![Y]\!] - \mathrm{F} z_Z \cdot k_{zy} [\![Z]\!] \end{aligned} \qquad (7)$$

Units for net current [38] I_{XY} are $(C/dm^3)/s$ = C dm^{-3} s^{-1} and $(moles/dm^3)/s$ = moles dm^{-3} s^{-1} for unidirectional flux J_{xy}. Double brackets like $[[X]]$ indicate activities, i.e. the generalization of concentration (number density) needed in biological solutions, as discussed below. Units for rate constants are $(moles/(dm^3\ s))/(moles/dm^3)$ = 1/s. The valences, i.e. charges on one molecule of each reactant are z_X or z_Y. F is Faraday's constant.

In general, the conservation of current law illustrated in *Eq.(5)* is violated

$$I_{XY} \neq I_{YZ}, \qquad (8)$$

for a range of concentrations, rate constants, or charges.

The main result of *Eqs.(7)* and *(8)* and the Appendix is simple: *the law of mass action itself conserves current only under special symmetric circumstances*. The law of mass action does not automatically change the electrical potentials to ensure that the current flow in a series circuit is the same everywhere.

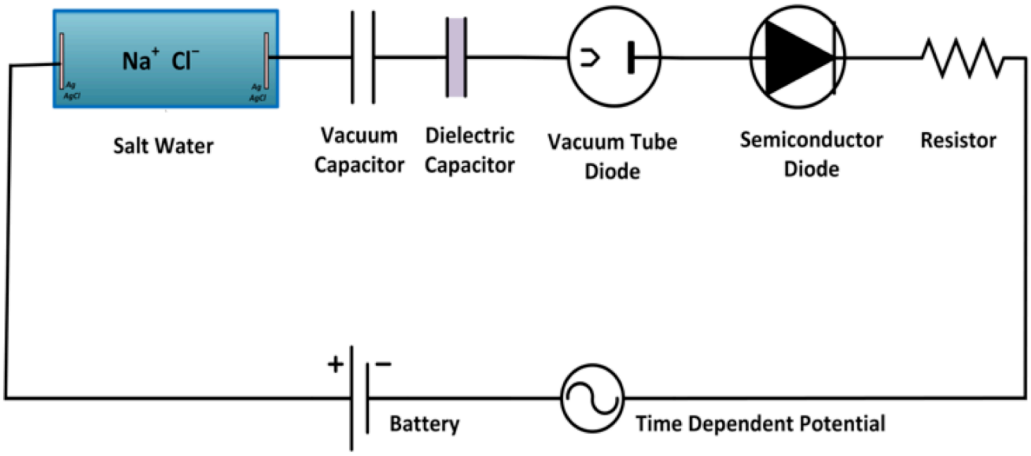

Figure 2. A demonstration that current in one physical system being quite different from current in another. Charge flow is not simply the physical movement of particles of definite charge (and mass). Current in a vacuum capacitor, in an ionic solution, and in a wire are all quite different. Current is not just the movement of ions, electrons or protons. Note that the electric charge in different devices varies on very different scales from subatomic (in dielectrics) to atomic (in diodes, ionic conductors, etc.) to macroscopic. Thus, treatments of charge and current must be multi-scale (redrawn from Ref. [5]).

Details of Proof. Rate equations equivalent to chemical reactions involve *net flux* like $\bar{J}_{XY}$. The chemical reactions in *Eq.(6)* define the net fluxes in units of moles/(dm^3 s),

$$\begin{aligned} \bar{J}_{XY} &= k_{xy}[\![X]\!] - k_{yx}[\![Y]\!] \\ \bar{J}_{YZ} &= k_{yz}[\![Y]\!] - k_{zy}[\![Z]\!] \end{aligned} \tag{9}$$

with the definitions of rate constant k_{xy}, for example

$$k_{xy} = \frac{J_{xy}}{[\![X]\!]} = \frac{J_{xy}}{-\frac{d}{dt}[\![X]\!]_{J_{yx}=0}} \tag{10}$$

Remark. Unidirectional fluxes like J_{xy} are conventionally measured by tracers - originally radioactive isotopes, now usually fluorescent probes - flowing *into* an acceptor solution in which the concentration of tracer is zero. The concentration in the acceptor solution of the substance is not zero in most cases, although the concentration of the tracer is (nearly) zero. Note that the activity $[[X]]$ depends on the concentration of *all of the other ions* in a significant way in non-ideal solutions [5, 34, 39–65]. In non-ideal solutions, changes in the concentration of one substance $[Y]$ change the activity of another substance $[[X]]$ and also change its flow.

The proof of *Eq.(7)* is completed by writing net fluxes like $\bar{J}_{XY} = J_{xy} - J_{yz}$ as the difference of unidirectional fluxes. Net flux is converted into net current like I_{XY} using the proportionality constant Fz (where z is the species in question) between flux and current:

$$\begin{aligned} I_{XY} &= \mathrm{F}z_X \cdot k_{xy}[\![X]\!] - \mathrm{F}z_Y \cdot k_{yx}[\![Y]\!] \\ I_{YZ} &= \mathrm{F}z_Y \cdot k_{yz}[\![Y]\!] - \mathrm{F}z_Z \cdot k_{zy}[\![Z]\!] \end{aligned} \tag{11}$$

Obviously, $I_{XY} \neq I_{YZ}$. This ends the proof.

It seems unlikely that conservation of current and conservation of mass are in conflict in the law of mass action, so mathematics is needed to show that current is not conserved. I treat a simple but widely used case, hoping that this will be more convincing than words and easier to extend to other reaction schemes. *The methods used can be applied to any series of reactions to see if they satisfy both conservation of current and conservation of mass*. In general, an additional constraint and equation is needed to enforce the global continuity of current flow. The additional constraint is introduced in the special case where the chemical reaction describes spatial movement through narrow channels. In that case, a fully consistent treatment is possible using a variety of models for charge movement. The general case where chemical reactions describe covalent bond changes is not worked out in detail here. In my view, the general case must be dealt with by variational methods that enforce consistency of all the physical laws. Energetic variational methods appropriate for this purpose (in dissipative systems like ionic solutions) are described.

The Appendix shows that charge imbalance predicted by the law of mass action would likely have noticeable effects. A small charge imbalance quickly produces potentials and forces that destroy membranes (~0.3 V), molecules in a liquid (~2 V, H_2O, with AglAgCl electrode), molecules in a gas (~480 V for H_2O gas), and atoms in a gas (~14 eV for N_2). These effects would destroy experiments and are not found in the laboratory, which would also soon be destroyed. We

conclude that theories that predict charge imbalance because they do not conserve current flow are not acceptable approximations in the world of experiments or life.

2. Current Flow in Complex Systems

We now discuss current flow in some detail. As we discuss current flow through salt water, vacuum capacitors, dielectric capacitors, vacuum tube diodes, semiconductor diodes, resistors, wires and batteries we will see the bewildering number of ways in which the electric field changes to ensure that current flow is always continuous.

Current is an abstraction that is conserved. Physicists teach that conservation of charge and current are universally true, exact from very small to very large scales and very small times to very large times [68]. What physicists often do not teach clearly is that charge and current are abstract ideas. Many students believe that current is always carried by particles and so conservation of mass (of charged particles) implies conservation of charge and current. This is not true. Current is carried between the plates of a vacuum capacitor as a displacement current that does not involve the movement of particles [1, 14–16]. Current flow in semiconductors is carried by mathematical fictions called quasi-particles that move according to classical physics [17–21], not involving solutions of Schrödinger's equation. This is an important property of semiconductors [69–70], crucial to their use.

An essential idea of electricity and magnetism as explained in textbooks [14–16] is that *current is continuous* (without loss in a series circuit) *no matter what the physical nature of the current*. Electrical potentials and forces change automatically to guarantee continuity of current flow under all conditions, in experiments and in the equations of electromagnetism. I found Saslow's [14] treatment of continuity of current particularly clear and useful.

Fig.2 tries to show this idea in a concrete way that is easy to build in a lab. 'Completing the circuit' implies that current in every device is the same. Continuity of current - the same as Kirchhoff's current law in a one dimensional sequence of reactions - states that time-varying currents are the same in any series of devices even if they exhibit very different physics and different constitutive laws, even if they involve chemical reactions (*Fig.1* and *Eq.(6)*).

Ionic Conductor. The ionic conductor in *Fig.2* is a cylinder containing NaCl. Here, current flow (at a frequency say of 1 Hz in a 0.02 M solution) is almost entirely the physical movement of charged particles, of ions, say sodium and chloride ions, and follows simple constitutive laws (when concentrations are < 0.02 M and flows are not large enough and do not last long enough to change concentrations). These ions are hard spheres. The finite size of these spheres is significant in the ionic mixtures found everywhere in biology and in general at concentrations greater than say 0.02 M. The finite size makes constitutive equations (valid at all concentrations of mixtures of different types of ions) much more difficult than the classical constitutive equations for quasi-particles that are points. Finite size implies saturation effects - space cannot be filled more than once - and these imply that 'everything interacts with everything else'.

Numerical difficulties in dealing with spheres are substantial. *Spheres must be computed in three dimensions because spheres do not exist in one and two dimensions*. That is to say, objects with a single radius have very different surface to volume ratios in one, two and three dimensions and so fill space very differently. Phenomena in which spheres fill a significant fraction of three-dimensional space are not easily approximated in one or two dimensions. Computation of the forces that prevent overlap of spheres is difficult because those forces are strong and vary significantly in three dimensions. Bottom line: non-ideal solutions remain a challenge, as documented below.

Vacuum capacitor. Let's move on to a *vacuum capacitor*, in which the space between the two plates is completely empty of matter (as it would be in outer space, for example). The current flow through this capacitor is just as real as the movement of ions of NaCl in the cylinder even though no particles or spheres are present, and no mass moves at all.

The displacement current between the plates of the vacuum capacitor is a property of the electric field itself, as explained in textbooks of electricity and magnetism [1, 14–16] and is described by the exact and simple constitutive equation for the vacuum $i_{\text{displacement}} = C(\partial V/\partial t)$, where the displacement current $i_{\text{displacement}}$ (amps) is strictly proportional to the capacitance C (farads) and the time rate of change $\partial V/\partial t$ of the voltage across the vacuum capacitor. Unlike other constitutive equations, the constitutive equation for vacuum current is exact, valid to some eighteen significant figures [15].

This displacement current induces a magnetic field just as current carried by ions produces a magnetic field. Indeed, without displacement current in a vacuum, Maxwell's equations do not allow sunlight to propagate through the vacuum of space. With this exact expression for displacement current, light propagation is a solution of the Maxwell equations, and in fact the speed c of propagation of light can be computed from measurements of electrical and magnetic constants, entirely independent of measurements of light itself, by Maxwell's remarkable formula $c = 1/\sqrt{\mu_0 \varepsilon_0}$. Light propagates according to Maxwell's equations over astronomical distances, so we know that the constitutive equation that calculates that speed must be accurate to many significant figures.

Dielectric capacitor. In the dielectric capacitor in *Fig.2* (filled with for example the plastic polytetrafluoroethylene = PTFE = Teflon), current flow is more complex, and involves the effect of an applied electric field on the spatial distribution of the electric charge intrinsic to the atoms, molecules, and substances of the dielectric. ('Intrinsic' here means the distribution

of charge present when there is no applied electric field). Note that the electric charge in a dielectric, an ionic solution, a protein or nucleic acid, for that matter, varies significantly on all scales from subatomic to macroscopic.

The properties of dielectrics cannot be described in detail here because they vary so much with material, voltage, and time. But some properties need to be emphasized. Intrinsic charges do *not* move from plate to plate in a real capacitor. The current that flows from plate to plate and within the dielectric is a dielectric displacement current not carried by the movement of mass (any substantial distance). The current in a real capacitor is an abstraction. It is the sum of the vacuum displacement current and the material displacement current I_{Pol} (in the dielectric) produced by the distortion of intrinsic charges. The material displacement current describes the movement of a *charge on a non-linear time-dependent spring (with damping) that eventually stops moving* after being perturbed by an external electric field for a finite time. Calculating this material displacement current involves solving a non-equilibrium time-dependent version of the Schrödinger equation, a difficult task, involving a macroscopic number of atoms and millisecond timescales in cases of biological interest.

Ideal dielectrics do not approximate ionic solutions. Scientists have avoided the difficulties of solving the Schrödinger equation in a macroscopic system like a dielectric, or ionic solution, by using approximations. They discuss an ideal dielectric with properties independent of field strength and independent of time. The approximation over a wide range of electric field strengths is satisfactory in most materials.

The approximation of a time-independent ideal dielectric is poor over a wide range of times. For example, most solutions of ions in water need effective dielectric coefficients to describe the polarization charge induced by permanent charges. This dielectric charge changes - after a step electric field is applied - from about 2 to about 80 as time progresses, from zero to 10^{-15} sec to say 10^{-5} sec and the change in dielectric charge depends on the substance [71]. There is not a universal constitutive law or approximation for the time dependence of dielectric properties. A factor of 40 change with time is not small. Multi-scale analysis must deal with this multi-scale problem because atomic motion has macroscopic consequences on all time-scales.

The time dependence of real dielectrics needs to be dealt with in biological applications: ideal dielectrics do not approximate the properties of ionic solutions in which biology occurs. The time dependence of polarization charge in proteins has been extensively studied [72–74] and varies over the whole time-scale from atomic to macroscopic. It often has slow components, 10^{-3} sec or slower.

Biology is controlled by atomic structures that move significantly in 10^{-15} sec but have dramatic effects on biological functions between 10^{-5} sec and 10 sec later. The time dependence of these induced polarization charges appears in simulations (with atomic resolution) as changes in the orientation and induced polarization of water molecules, and in the distribution of ions, as well as in the distribution in proteins of permanent charges, permanent dipoles, and induced dipoles.

Atomic-scale simulations must last a long time. Atomic-scale simulations seeking agreement with experimental data - available for a wide range of solutions [71], and proteins [72–74] - must last long enough to account for the discrete charge movements that produce an effective dielectric coefficient of 80, as measured experimentally. Otherwise, substantial changes in the electric field will not be seen in the simulation, even if those changes in the electric field have great biological significance.

Electric fields in biology are strong. A nerve signal is a change in an electric field of the greatest biological significance. Information transfer in the nervous system, coordination of contraction in muscle, including the coordination of cardiac contraction that allows the heart to pump blood, are all directly controlled by the action potential, the biological name for the nerve signal. Electric fields during a biological action potential are slow (say 10^{-3} sec), and macroscopic in scale, transmitting 10^{-3} meters, propagating one meter (in humans), and they are remarkably multi-scale. Handfuls of atoms in a channel protein control the macroscopic propagation and the nerve signal itself is current carried by ions (that are single atoms) moving through those proteins. The electric field of the action potential is strong, typically some 0.1 V across a 2×10^{-9} meter-thick membrane or a 3×10^{-10} meter-long selectivity filter (EEEE group in a calcium channel), 5×10^{7} to 3×10^{8} V/m. These electric fields distort the intrinsic distribution of charge within the dielectric on many scales of time and distance, including electrons inside molecules and atoms. They reorient polar molecules that have an intrinsic asymmetric distribution of charge. Charges only move a small amount in dielectrics in response to the applied electric field - reminiscent of the sloshing of tides in an ocean and on beaches on earth created by the moon's gravitational field - *and they eventually return to their resting position when the field is turned off* but those small movements of charge produce large effects because the electric field is so strong.

Dielectric currents $\mathbf{I}_{Pol}$ share some of the properties of vacuum displacement currents but they do not follow a universal exact constitutive law, not even approximately. It is important not to confuse dielectric $\mathbf{I}_{Pol}$ and vacuum displacement currents $\varepsilon_0\ \partial\mathbf{E}/\partial t$. It is important to realize dielectric displacement currents can be large and produce large effects in biology, as they do in semiconductors.

Non-linear components of dielectric current are important in biology. Non-linear components of dielectric displacement current have large effects important in biology (e.g. in muscle [75], where they were first discovered, and nerves [76]) where they control the opening and closing of the channels that produce the action potential [77] and some enzyme

functions [78], as well. These non-linear dielectric (i.e. displacement) currents can be recorded as 'gating' currents because of the Shockley-Ramo theorem [79], which is a restatement of continuity of current. The dielectric currents involve a tiny fraction of all the charges in and near to an ionic channel, far less than one percent (see Appendix "Size of Effects" near *Eq.(28)*), but continuity of current and the strength of the electric field make gating currents easily measurable in the hands of skilled experimentalists.

Any electric field is extraordinarily strong compared to diffusion. Tiny displacement currents can be recorded routinely in many laboratories because the electric field is so strong. If it were not so strong, these tiny concerted movements of atoms would be lost amongst the Brownian thermal motion. Extraordinarily small changes in net charge are enough to guarantee that currents are the same in all elements of a series circuit because the capacitances (ratio of charge to potential) involved are often some 10^{-17} farads, see Appendix. Continuity of current flow is guaranteed by changes in electrical forces and potential resulting from very small amounts of charge.

Concentration effects are very small. On the other hand, a one percent deviation in density of mass has a tiny effect on diffusion and chemical reactions, which is hardly noticeable. Diffusion forces are tiny perturbations in the electric force field and energies involved in diffusion are tiny perturbations in the total energy of charged systems. Electric forces and energies are not small parts of the total energy and it is not wise - to say the least - to treat them as perturbations of uncharged systems. Small changes in charge distribution produce large changes in flow because flow tends to be an exponential function of the spatial profile (even when the profile does not have a large, single, symmetrical peak [80] of potential energy) and the profile of energy is a sensitive function of permanent and dielectric charge, ionic conditions, etc. [81]. It is not wise [82] to neglect the effect of permanent charges of the channel protein on the shape and size of the electric field as is done in many continua [83–105] and rate models of gating and permeation of ion channels, and most treatments of enzyme function and catalysis.

Current is indeed an abstraction. In each of the devices in *Fig.2*, charge follows different laws, because it has different physical properties, sometimes carried by charged particles, sometimes produced by the rate of change in electric field (displacement current), sometimes as a result of the 'movement' of quasi-particles, and sometimes as a consequence of the 'movement' of electrons in a macroscopically delocalized quantum state of a wire.

Current always flows without loss in each device. This abstract property of current is always true, because the potentials change automatically to ensure continuity of current. However, the 'laws' describing current as a function(al) of time and potential (for example) depend on the physical nature of the charge and its movement. Different devices have different relationships between current, voltage, and time. These different 'constitutive' equations are described in engineering literature in great detail.

The different constitutive equations combine with Kirchhoff's current law to describe current flow from one place and one device to another. Together, the equations describe the universal fact that interrupting current flow in one chemical reaction (of a series of reactions) will interrupt current flow in every other reaction (in that series) even if the interruption is meters away from a chemical reaction being studied on the atomic scale (scale = 1 Å). Together the equations describe the experimental fact that electric fields, forces, and potentials automatically rearrange so that Kirchhoff's current law and continuity of current flow are always present exactly on all scales no matter what is 'carrying' the current. Indeed, *the physics of current flow changes to accommodate the continuity of current*.

One is reminded of the power of the electric field when one unplugs a computer. Interruption of a circuit meters away from the diodes of the computer's power supply stops the flow of quasi-particles - holes and semi-electrons - across atomic-scale junctions of semiconductor diodes, often of the PN variety. The electric potential changes so strongly in response to the interruption (because the power input of the computer power supply stores a great deal of charge) that the electric field exceeds the dielectric strength of air. The *automatic change in the electric field - needed to maintain continuity of current flow - is enough to 'change the physics' of the system*. Electrons are stripped off the atoms of air resulting in a plasma, a spark. We should be frightened of sparks and their electric fields. Sparks start fires, sometimes in upholstery or drapes.

Vacuum tubes. We consider charge movement in a vacuum tube diode as the next device in *Fig.2*. Vacuum tubes control the flow of electron current by changes in their internal electric fields and were called 'valves' in the UK for that reason. Vacuum tubes, semiconductors and even some open ion channels follow simple constitutive laws of rectifiers, as described in textbooks of electronic devices or ion channels.

In vacuum tubes, current is indeed carried by a stream of isolated charged particles, electrons with a definite mass and charge, moving through a vacuum, interacting only through their electric fields. At 1 Hz, essentially all the current in a vacuum diode is carried this way. Current through a diode is not proportional to the voltage across the diode because the electric fields within the diode change shape, despite the simple physics of conduction. The fields change shape as the voltage across the diode is changed because of screening and shielding. The electric fields within the tube are different at small and large potentials. The different internal electric fields change current flow, creating rectification. The electric field creates a large barrier in one direction so current in that direction is small; the electric field creates a small barrier in the other direction, and current in that other direction is large.

Rectification is of historical interest because vacuum tubes allowed the early detection of radio waves in American homes in the 1920's, as valves did in the UK. The demand for portable radios led to solid-state diodes, then solid-state 'triodes', field effect transistors, integrated circuits, and our modern digital world [106]. Vacuum tube diodes had certain difficulties - they were big (10^{-2} m at their smallest), hot, greedy consumers of power, cost about $3 in the 1950's, and electrically unstable: they drifted continuously. They were quickly replaced with semiconductor diodes that cost less than $\$10^{-10}$ each, do not drift significantly, and can be as small as 10^{-8} meters nowadays.

We turn now to the semiconductor diode that operates on a very different scale from the vacuum diode. The rectification of both diodes - that is their function - depends on essentially the same physics, namely the shape of the electric field *and its change in shape with the direction of current flow*. The electric field of charge must be respected on both the scale of vacuum diodes and that of semiconductor - 10^{-2} meters of vacuum tubes, 10^{-9} meters of semiconductor diodes - indeed on all scales because the electric field has such large effects on all scales. The immediate implication is that theory and simulation must calculate the electric field on all scales, using an explicitly multi-scale analysis, since it seems unlikely that any one type of simulation or theory can span atomic to biological to laboratory distances and times, let alone the interstellar scales on which the laws of electricity are known to be valid.

Semiconductor diodes. Current in semiconductor diodes is carried by quasi-particles [70], called holes and 'electrons' (better named quasi-electrons, or semi-electrons in my view). These quasi-electrons and holes are defined because they interact much more simply than the totality of real electrons and lattice of atoms in semiconductors [21, 70, 107–109]. They simplify the quantum mechanical many-body problem into the much simpler motion of imagined quasi-particles. Current is carried in semiconductors by rearrangements of charge in the entire lattice of atoms that make up the semiconductor. Fortunately in terms of our technology and daily life, current in germanium and silicon can be approximated by classical theories that deal with one quasi-electron or one quasi-hole at a time as they move in mean fields, without requiring a solution to the Schrödinger equation at all [17–21].

As a textbook puts it eloquently (p. 68 of Ref. [69]): "Electron is a quasi-particle consisting of a real electron and an exchange correlation hole … a cloud of effective charge of opposite sign due to exchange and correlation effects arising from interactions with *all* other electrons of the system. Hole is a quasi-particle, like the electron, but of opposite charge; it corresponds to the absence of an electron from a single particle state that lies just below the Fermi level". The motion of these quasi-particles is described by mean field models, evaluated both by simulations [17, 21] and theories [17–21], e.g. the Poisson drift-diffusion equations [110], often called *PNP* (for Poisson Nernst Planck) in biophysics and nanotechnology [56, 61, 111–117]. *PNP* is of particular importance because it is used widely, nearly universally, to design and understand the devices of our semiconductor technology, from transistors to computer chips. *PNP* is used rather widely in nanotechnology these days and increasingly in membrane biophysics, mathematical biology, and electrochemistry, including battery technology, as well as the technology of cement.

Rectification in biological membranes. Early treatments of rectification in biological membranes, then called ionic conductances, ([83]; Appendix of [118]) in fact drew heavily on Mott's (nearly) contemporaneous treatments of rectification in diodes [119], but none of these were consistent: they assumed the electric field, instead of computing it. The charge in the system did not produce the assumed electric field if substituted into Poisson's equation. Goldman [83], then a graduate student of K.S. Cole, recognized the problem, but did not know how to remedy it, nor did these workers [83, 118] understand that electric fields in diodes or open channels were not constant in any sense, including space (numerous personal communications to me from K.S. Cole, 1960–1962, and A.L. Hodgkin, 1960–1995). Mott soon realized [120] that the change in shape of electric fields was a crucial phenomenon in semiconductor diodes, updating the constant field assumption of his original paper [119]. Biophysicists were evidently not aware of the evolution of Mott's understanding [121].

Semiconductor physicists understood that consistent treatments produced electric fields that varied greatly with conditions, as the fields change to force continuity of current [122]. *Indeed, transistor design would not be possible* [21, 106, 123–124] *if computational electronics had assumed constant fields the way some physicists did early on* [119], but not for long [120]. In biophysics, the importance of computing the fields from the charges (so the fields and charges were consistent) was not realized (as far as I know) until much later [81, 111–112]. Until then, and even after, Nernst-Planck equations were used in biophysics *without including the ionized acid and base groups of the channel protein*, i.e. without including the permanent charge of the protein [85–105, 121]. These ionized groups render polymers into ion exchangers [125]. The equivalent doping profiles convert semiconductors into devices. Leaving out ionized groups "is like studying a galaxy without stars", as a prominent physical chemist once told me. The distribution of permanent charge creates a baseline electric field. Changes in shape of the electric field with concentration, amino acid composition of the channel protein, etc. are responsible for many of the important properties of these systems. Unfortunately, the importance of changes of shape in the electric field, and of the fixed charges involved, were and still are often ignored.

Charge carriers in semiconductors can be really quite strange from a physical point of view. Charge

carriers in silicon and germanium do not exist outside the lattice of the semiconductor as distinct entities. They do not exist in the same sense that Na^+ and Cl^- ions exist, but are mathematical representations, with lifetimes sometimes as short as milliseconds. Quasi-particles are the second derivative of a Fermi surface of silicon and germanium semiconductors, under particular conditions. They are defined, as mentioned above, to allow easy classical analysis compared to the intractable quantum mechanical many-body problem of a macroscopic semiconductor.

Much of the success of our semiconductor, digital, and video technology is due to the accuracy of the constitutive *PNP* equations describing holes and semi-electrons. They robustly describe the characteristics of semiconductor devices of many different types, *with very different input-output relations,* as different as an exponentiator and a logarithmic convertor. *PNP* is so useful because almost all the devices of our digital technology work under a restricted set of conditions in which flows are crucial but are of the special type, well described by quasi-particles moving in a mean field. Treatment of distortions of the electric field of all the atoms in macroscopic devices is not needed. Almost all of the devices of electronics use power supplies to maintain different voltages at different locations far from the PN junctions of the device itself (and thus require a global treatment of the electric field). These voltages perturb the distribution of velocities of charged particles so the distribution exhibits net flow [80, 126]. The slight perturbation is enough to imply the *PNP* equations and those provide enough non-linearity to make amplifiers, switches, and the full set of logical circuits necessary to make a computer [21, 107–109]. In fact, in this case the *PNP* equations can be solved analytically and exactly to give intuitive, pleasingly simple formulae for current flow, once the shape of the electric field is known [80].

PNP equations describe a wide variety of current voltage relations and devices. The *PNP* equations are the constitutive laws used to describe semiconductor devices as their current voltage relations change drastically (with voltage, for example) from that of a linear amplifier, to a switch, exponentiator, multiplier or even logarithmic converter [21, 107–109]. Non-linear input-output relations, as varied as these, enable a rich variety of devices.

Non-linear input-output relations as diverse as these are not described easily - or described at all for that matter - in most areas of physics and chemistry. All the non-linear devices in a computer are actually mathematical solutions of the *PNP* equations in a complex silicon structure built to have the particular spatial distribution of permanent charge ('doping') that produces the desired properties of the device, i.e. the input-output relations.

The predictive power of *PNP* is very important in the design of robust semiconductor devices - that do not fail even when used many millions of times a second in computers that contain a trillion transistors. In fact, *the intrinsic physical properties of semiconductors are adjusted by their designers so PNP remains a good description* [127–130]*, even though the size of the device has been dramatically decreased*. The concentrations of fixed charge dopants, geometries, and recently even dielectric coefficients are adjusted by semiconductor engineers in their successive iterations of Moore's law [131–133], so *PNP* remained a good description as performance increased by factors of a billion or so, over 50 years.

Evidently, *reliable design is more important than raw performance*. It seems more important for the designers (and marketplace) that an equation describes behavior robustly and accurately over a range of conditions than that the device be as fast or small as possible [127–130]. Evolutionary selection in biology also seems to choose robustness over efficiency in many cases. Devices of nanotechnology need to be similarly robust, I believe, before they will be used extensively.

PNP is not enough, however, when ionic solutions are involved. A great deal of effort has been spent applying *PNP* equations to electrochemical systems [56, 61, 111–117] hoping they might serve as adequate robust constitutive equations, but that is not the case. The non-ideality of ionic solutions, arising in large measure from saturation effects produced by the finite size of ions, demands more powerful mathematics than the partial differential equations of *PNP* used in computational electronics.

Resistor. The next device we discuss in *Fig.2* is a *resistor*, which in some ways is the easiest to describe because current is proportional to voltage with a single proportionality constant over a wide range of voltages, times, and conditions. A resistor follows Ohm's constitutive law with a resistance independent of potential over a wide range from 10^{-5} V to say 100 V, and for values of resistance from 10^{-1} Ω to 10^{8} or 10^{9} Ω. The range of validity of Ohm's law is an enormous help in circuit design. *Circuit models involving resistors, capacitors, inductors, and operational amplifiers are transferrable*. They behave as real devices behave without changes in parameters. Largely for that reason, designs are inexpensive and robust.

Despite the simplicity of resistors, the actual current carrier in a carbon resistor is unclear, at least to me. No one cares very much I suspect because the device works nearly perfectly. The carrier of charge does not matter very much. What matters is the constitutive law that describes the relation between current, voltage, and time. The constitutive law should satisfy conservation of mass, charge, and current.

It is instructive to write the constitutive law for a resistor Ohm's law for only particle current, using a conservation of particle (mass) formulation, and then write it again for particle plus displacement current from one terminal to another. If you apply a step function of current (or potential for that matter) to the purely particle formulation, a paradox arises. The potential changes but the particle current flowing into the resistor from the left is exactly equal to the particle current

flowing out on the right *at all times*. Why does the potential change if there is no accumulation of charge?

The paradox can be resolved in two (nearly) equivalent ways:

(1) The constitutive equation of Ohm's law can be used with the extended definition of current that includes the displacement current. In this case, there is continuity of generalized current, but there is *NOT* continuity of particle current at all times. The transient accumulation of particle current provides the charge that changes the potential.

(2) Alternatively, the circuit model of the resistor in *Fig.2* can be changed to have an explicit capacitor in parallel with it. In this case, the charge accumulates on the capacitor, and the resistor itself can have continuity of flux of particles at all times and follow Ohm's law using the current/flux of particles (and not the displacement current). The charge accumulating on the capacitor creates and changes the electric field.

Maxwell himself repeatedly used capacitors in this spirit to understand the role and significance of displacement current. Sections 102, 125, 199 and Chapter 8 are some examples in Ref. [134]. We (following the insight and advice of Wolfgang Nonner) have used capacitors as a low resolution, consistent way to connect permanent charges (specifically, ionized side chains in a channel protein away from the pore) and electrical potentials in the pore of a channel protein, e.g. the potassium channel [135].

Current flow in wires can also be strange. The current carriers in a wire are delocalized electrons in the simple case of a single solid conductor of metal, and follow the simplest constitutive law of all over long(ish) time-scales, say times longer than 10^{-5} sec. Nevertheless, most of the electronics in our digital technology operate over time-scales much shorter than that.

Now we confront the importance of the time variable. *The physical nature of current flow depends on time-scales, even in wires* [14, 16]. The range of time-scales in our technology is enormous, from more than one second to less than 10^{-9} sec. Over shorter time-scales ($< 10^{-4}$ s) the wires that must be used are often twisted pairs [136], each made itself of many very fine wires. Without twisting, these pairs of wires do not allow successful connections to the internet because rapidly changing signals are not carried reliably by single wires [137]. *Over those short time-scales, currents flow outside wires,* guided by the conductor, to be sure, but outside the conductor nonetheless [16, 137]. The twisting of wires is a necessity if they are to carry signals robustly and reliably so we can use them in our video devices and smartphones, even in old-fashioned hard-wired telephones [136] that only need to amplify audio signals heard by adult humans (< 10,000 Hz).

The physical nature of current depends on time-scale. The physical nature of current in almost any system depends on the time-scale, and differs at different times as much as it differs in different devices. Constitutive equations depend on time. Different devices have different constitutive equations with different time dependences. Again, current is an abstraction, a different abstraction at different times in one device, as well as different in different devices.

Charge movement in batteries. Batteries are present in *Fig.2* both as an isolated device and as the Ag||AgCl interface between wires and NaCl solution in the conducting cylinder previously discussed. I hesitate to describe a constitutive equation for the flow of charge in batteries in general because the flow is so very complex, different in different devices, and important for the practical daily use of batteries [138–139], and its interaction with surface charges is also subtle and important [140]. It is enough to say here that current flow through electrochemical systems is carried by different electrochemical systems differ dramatically and change in dramatically with time, frequency, composition and concentration of ionic solutions, as well as electrical potential, and current flow. Over short time-scales, e.g. 10^{-6} sec - that are still long compared to the time-scales important in computers - current flow from the Ag wire and AgCl electrode material into the NaCl solution is entirely displacement current lagging behind voltage. However, over the longer time-scales characteristic of biological systems (greater than say 0.1 sec), the current is carried by a complex combination of Ag^+ and Cl^- ions with negligible displacement current.

We conclude that current is indeed an abstraction with different physical meanings in different systems and over different time-scales. No one can visualize and no one knows - at least I do not know anyone who knows - why or even how this abstraction can be so perfectly conserved under all conditions and over all scales, from Angstroms to meters, from femtoseconds to seconds. It is truly amazing to think of the changes in electric forces needed to accommodate and enforce continuity of current in salt water, vacuum capacitors, dielectric capacitors, vacuum tube diodes, semiconductor diodes, resistors, wires and batteries from atomic time-scale 10^{-16} sec and distances of 10^{-11} meters to the biological scales of seconds and meters from inside atoms (pp. 8–9 of Ref. [1]) to intercontinental distances (in submarine cables) and interstellar space.

3. General Remarks about Current Flow

The discussion of *Fig.2* leads to some general remarks about current flow. The interplay of conservation of matter and conservation of charge and current is different in each system, according to the constitutive law of the system. The idea of a current flow is an abstraction built to accommodate the physics of each system, while maintaining the main feature of electrodynamics, the exact conservation of current, at any one time, on any scale. Conservation of mass is

certainly also followed in these systems, but in a more relaxed way. Significant deviations are allowed for a time. Mass can accumulate for a time without catastrophic results.

Accumulation of mass does not have dramatic effects. In the systems considered here, the accumulation of mass does not usually have dramatic effects. A one percent accumulation of mass has a small effect on chemical potentials, although of course exceptions can occur. There is no general law for the accumulation of matter. The effect depends on the constitutive law, and its interplay with conservation of matter, charge, and current.

Accumulation of charge has dramatic effects. The accumulation of charge is different. A one percent accumulation of charge has a huge effect. As Feynman memorably mentions at the very beginning of his textbook [1], *one percent excess of charge in a person at arm's length produces a large enough force to lift the earth!*

The accumulation of charge follows an exact law: the displacement current is the sum of all the other charge flows and it changes the rate of change of the electric field in a precise way so total current is conserved precisely. Continuity of current is obeyed and so such enormous forces do not occur. If continuity of current were not obeyed exactly, enormous forces would soon develop (see Appendix). Such forces are of course incompatible with life or laboratory experiments. Fortunately, accumulated charge is easier to deal with in theory than accumulation of mass: the vacuum displacement current does not depend on constitutive laws, it simply depends on the time derivative of potential and so can be calculated.

Accumulated charge is much simpler than accumulated mass. Accumulated charge has universal properties, *independent of the physical nature of the charge*. Particle and quasi-particle currents that accumulate at a junction change the time derivative of electric potential so the electrical potential carries away a displacement current. This displacement current is *exactly equal to the sum of the other currents flowing into the junction* without known error, to about one part in 10^{18}. Continuity of current is exact *independent of constitutive laws* if current is re-defined to include displacement current. The displacement current (and equivalently $\partial V/\partial t$ 'take up the slack' so that Kirchhoff's current law (using the extended definition of current) is exact in one dimensional systems like a sequence of chemical reactions. No charge accumulates at all beyond that defined by the integral of the displacement current. The electrical forces and $\partial V/\partial t$ change so the displacement current is exactly equal to the sum of the other currents, and continuity of generalized current is exact.

Accumulation of charge is special because it is universal. The precise linkage between potential change, charge accumulation, and displacement current is a special feature of electromagnetism because it is universal. It is a property of a vacuum, the constitutive equation of a vacuum, if poetic license is allowed. In this sense charge is more fundamental and universal than mass, a fact which certainly came to me as a surprise [141]. Then I learned that charge is constant at all velocities according to the 'Lorentz (relativistically) invariant', whereas mass is not. Mass depends on velocity and charge does not. Special relativity seems to make charge a more fundamental physical property than mass.

The reader may have difficulty visualizing the interactions that enforce conservation of the abstraction 'charge/current' in all these devices with all these properties over the entire time-scale. I certainly do. *However, current flow and charge are conserved on and between all scales under all conditions,* even if we cannot visualize how that manages to be so. Experiments demonstrate that fact. Current does flow continuously without loss in a circuit. Consider a battery feeding a circuit. If a wire is cut far from the battery, current flow stops everywhere. The chemical reaction in the battery is disrupted on an atomic scale, by the (lack of) current flow meters away. Those of us living in colder climates have seen the effects of starting a car with another car's battery. We have learned to be careful because even a twelve volt battery can produce dangerous sparks in air, even though air has nearly infinite resistance ($> 10^{11}$ Ω for dry air), and is in that way, nearly a vacuum. Abstract current is conserved 'exactly' even if we cannot visualize how that happens.

Science often contains mysteries that cannot be visualized - consider Maxwell's attempts to visualize his equations as properties of an ether. Science often poses questions that cannot be answered. Why is there no magnetic charge? Why is the charge on an electron 1.6×10^{-19} C? *Why is charge independent of velocity in special relativity when mass, distance, and even time are not?* Why are physical laws invariant when locations S move at constant velocity $\partial s/\partial t$ - mass, distance, - or at constant acceleration $\partial^2 s/\partial t^2$ - general relativity - but not when other time derivatives of location are constant, like a constant third derivative $\partial^3 s/\partial t^3$ or linear combinations of $\partial^n s/\partial t^n$, perhaps even fractional derivatives? As practical people, scientists cannot afford to just wait while we wonder about such things. Scientists wonder a bit *and* then move on, hoping our successors can do better than we have.

Biologists and engineers in particular cannot afford to linger on mysteries they do not understand. So many of those mysteries in biology have turned out to be caused by low resolution of our instruments, unable to resolve crucial structures. Think of Thomas Henry Huxley looking at the shortening of the striations of muscle [142] that were not understood until one of his grandsons (Andrew Huxley) studied them many years later working in Cambridge, UK [143]. Think of Lee de Forest using vacuum tubes without understanding how they work. Biologists and engineers cannot afford to wait to understand everything. They must isolate the mysteries and move on to study other things. Here, we move on to discuss devices and the theories and simulations used to understand them.

Biological implications of continuity of current. It is important to note that the continuity of current law has important biological implications in systems more general than a series of chemical reactions. The continuity of current law implies the cable equations (called the telegrapher's equations in mathematics literature), see derivation from the three dimensional theory in [144–147] and pp. 218–238 of [148]. The cable equation [149] is the foundation of the Hodgkin-Huxley model [150–154] of the action potential of nerve and muscle fibers. Kirchhoff's current law links the atomic properties of ions, the molecular properties of ion channels, and the centimeter-scale spread of current and potential that create the propagating action potential in nerve fibers meters in length.

In short or round(ish) cells, or in organelles like mitochondria, continuity of current forces coupling between multiple pathways of current crossing membranes, *even if the currents are carried by different ions, or by electrons, through different structures* nanometers apart in the membrane of the finite cells or organelles.

The flux-coupling characteristic of active transport systems - including the coupled flows in chemiosmotic systems that perform oxidative phosphorylation or photosynthesis - might arise in this manner. *Coupling of flows of charges, whether electrons or ions, is an unavoidable consequence of the GLOBAL conservation of charge and current,* of Kirchhoff's current law GLOBALLY enforced in three dimensions, and not a consequence of local chemical interactions, just as coupling of membrane currents with axial currents in a nerve fiber is an unavoidable consequence of Kirchhoff's current law, not of local chemical reactions.

It is interesting to compare the incorrect chemical theory of nerve propagation by Nobel Laureate A.V. Hill [155] with the correct electrical theory of the then undergraduate [156–157], later Nobel Laureate [150–154], Alan Hodgkin. Kirchhoff's current law in the form of the cable equation [149] was the key to Hodgkin's understanding. The classical voltage clamp experiments were designed to remove difficult terms and isolate membrane terms in the cable equation - personal communication, A.L. Hodgkin, 1961 - that today we know describe ion channels, opening, closing and conducting [121, 158–159].

The cable equation links movement of atoms inside channel proteins to macroscopic current flow that produces nerve propagation of the macroscopic electrical potential, the nerve signal that spreads meters. Macroscopic potentials modify atomic movements involved in gating and conduction. Atomic movements create macroscopic electrical potentials.

Equations of the electric field are true on all scales and so allow a unique linkage between models of atomic motion, protein behavior, and macroscopic propagation of electrical signals. I suspect linkage equations of this type - valid on all scales - will be needed to make any multi-scale analysis robust and transferrable, if it reaches from atoms to meters, from femtoseconds to minutes as models of nanodevices must.

Models, devices, effective parameters, and transferrable theories. Parameters of models or devices can often be chosen so an incomplete theory or simulation describes a system under one set of conditions but not another. Experiments often show that rate constants must be adjusted dramatically as conditions change, and the adjustments can rarely be predicted ahead of time by theory.

Chemistry and biology are filled with examples of non-transferable models. Chemical reactions follow rate equations, but the rates are not constant and not independent of one another as conditions change, even though traditional theory assumes they should be [5]. Biology describes enzymes with one set of parameters but finds those change when conditions change and attributes that, somewhat mysteriously, to 'allosteric effects' and conformation changes.

Non-transferable theories have limited use. Biology and much of chemistry works under a wide range of conditions and so incomplete theories with effective parameters have limited use. Even if sensible or valid, theories (and simulations) with effective parameters like these are not accurate enough to design robust devices. By leaving out something important, those theories or simulations omit an energy term that almost certainly varies with conditions. The resulting effective parameters change in large and unpredictable ways.

Incomplete theories and simulations are not very useful over a range of experiments and conditions. Incomplete theories are not likely to be transferable (from one set of conditions to another) in the language of the chemistry literature. Devices designed from incomplete theories or simulations are unlikely to be robust or work well under a range of conditions. Biological systems analyzed with non-transferable theories (or simulations) are unlikely to be realistic in general because biological systems usually work in a range of ionic concentrations different from those used in the laboratory.

Simulations must deal with trace Ca^{2+}. Biological systems usually work in mixtures with a range of Ca^{2+} concentrations, in which the Ca^{2+} concentration has important practical effects, often turning systems on or off or controlling their rate monotonically. Simulations and theories in biology have limited use until they are calibrated so we can be sure they actually are correct under the range of conditions and Ca^{2+} concentrations the biological systems use. Simulating Ca^{2+} activity in pure solutions is a challenging problem [160]. Simulating Ca^{2+} activity in biological mixtures within the 10^{-8} to 20 M concentration range that are physiological has not been attempted to date. 10^{-8} M concentrations of Ca^{2+} are found inside most cells, while 20 M concentrations of Ca^{2+} are found in and near ion channels, nucleic acids, and enzyme active sites, where the chemistry of life is catalyzed and controlled.

Biological and chemical sciences will benefit enormously if theories and simulations can be made transferable, using one set of parameters to describe systems under a range of conditions, as many physical and most engineering theories and simulations do. I believe the law of mass action must be extended to conserve current before theories and simulations can be made transferable from condition to condition and from physics to chemistry to biology using only mathematics.

How to extend the law of mass action so it conserves current? An obvious way to extend the law of mass action is to include activities and electrical potential in the rate constant to 'right the rates' by making them (typically exponential) functions of potential [2–4, 11, 13]. This in fact has been done for a very long time in the study of reactions at the electrodes of electrochemical cells and recently in other ways in the treatment of the formation of concrete [161–163]. The Butler-Volmer and Tafel equations [2] include electrical potential in rate constants in an empirical way with limited [3–4] but real success.

Success is limited I suspect because difficulties of embedding an electrical potential in rate constants are formidable if current flows. We must 'fix the fields' so they are global and depend on current flow everywhere. Otherwise, they cannot conserve charge flow and support continuity of current as required by Maxwell's equations: *The way to 'right the rates' is to fix the fields, everywhere*.

4. Additional Perspectives.

Some additional general remarks may be helpful:

(1) A thermodynamic treatment is clearly impossible since the goal is to calculate large fluxes and currents that do not occur in a thermodynamic system at equilibrium by definition without flows.

(2) A rate treatment of the frictional treatment of flux over a large potential barrier is needed. The classical Brownian motion problem of Kramers [164–165] is a necessary step forward, even if it is inconsistent because it does not compute the potential barrier from the charges in the system. Kramers' treatment need not be restricted to large barriers. A simple expression for rates over any shape barrier is available [80] and needed [166] because so few barriers are both symmetrical and large as required in classical high barrier approximations.

(3) The rate constants over one barrier must depend on the electrical potential in distant locations. Otherwise, interrupting current flow in a distant location cannot interrupt current locally. This requirement implies that the electrical potential must be determined by a global equation like Poisson's equation, including boundary conditions far from the individual chemical reactions and barriers. The barriers of Kramers' model are variable and not constant as conditions change, including conditions far from the barrier itself. The fields and energy landscapes change and it is indeed their change that allows continuity of current, as illustrated in *Eq.(1),* to be satisfied.

(4) The rate constants over one barrier are likely to depend on concentrations in other places because the solutions containing the reactants are not ideal. A general characteristic of non-ideal solutions is that 'everything depends on everything else'. More specifically, the activity of one reactant (the free energy per mole) depends on the concentrations of other species in practice, as well as in principle.

(5) A general theory of all non-equilibrium processes is not likely to be useful: a general theory has to describe too much. A general theory must include hydrodynamic behavior of considerable complexity, since aqueous solutions are fluids satisfying the Navier-Stokes equations of fluid mechanics. A general theory would also include explosions since they occur with regrettable frequency at electrodes of electrochemical cells, when H_2 gas is generated (inadvertently) by an overvoltage.

Fortunately, non-equilibrium processes in biology and much of technology occur in ionic solutions in which atomic motion is heavily damped. That damping ensures that the distribution of velocities is a displaced Maxwellian [126], as it is in semiconductor devices [21, 167–168]. The displaced Maxwellian exhibits non-zero mean velocity and so allows flux and current through the system from power supplies to outputs. That flux and current is enough to produce the very non-linear devices of our digital (semiconductor) technology [17–21], and non-linear phenomena like the propagating signal of the nervous system, the action potential [152–154].

Boundary conditions are needed to extend the law of mass action to deal with the outside world. Equilibrium statistical mechanics and thermodynamics were designed to avoid the complexities of boundary conditions using the 'thermodynamic limit' to allow analysis. But when current flows, interactions with the outside world, are unavoidable the thermodynamic limit is not appropriate, and boundary conditions describing those interactions are needed.

It is difficult if not impossible to deal in general with boundary conditions for the laws of mass action, and chemical reactions, because of the wide variety of chemical interactions and physical geometries captured in a phase space of very high dimension. Boundary conditions are easier to deal with in three dimensional physical space, where the law of mass action is widely, nearly universally, used to describe ion motion through narrow channels [80, 121, 169–170].

Current in channels. Current flow through a 'hole in a (insulating) wall' is a subject of great significance because the hole in the wall allows for control of the current. Holes in proteins, membranes, and channels in field effect transistors are all nanovalves providing essential functions to a large fraction of biology and semiconductor technology, and of great

interest in electrochemistry. Indeed, biological channels are nearly picovalves allowing a handful of atoms to control macroscopic flow. It is hard to imagine a smaller valve that could control macroscopic flow. We consider models of current flow through such holes and show how boundary conditions can be applied to nanovalves and systems of this type.

We do not consider simulations here because they do not deal with the essential features of these systems. Simulations have considerable difficulties in computing the actual currents through systems controlled by a handful of atoms. The currents are macroscopic phenomena, occurring in the world of milliseconds to minutes, and spatial dimensions from say 10^{-9} to 1 meter. The currents are controlled by atomic-scale structures in biology and near atomic-scale (10^{-8} meter) structures in semiconductors and electrochemistry. Changing a few atoms changes macroscopic currents as is shown in biophysical experiments (usually involving site directed mutagenesis) every day. The macroscopic currents are driven by chemical and electrochemical potentials involving large numbers (>10^{15}) atoms in most cases. Simulations must then compute macroscopic-scale inputs and outputs while preserving atomic-scale spatial resolution of the controlling atoms. These issues must be all dealt with at once, because they all exist at once in the systems of interest, and they must be dealt with accurately, because valves typically depend on the balance of nearly equal forces, e.g. electrostatic and diffusion. Each force must be quite accurately calculated because it is the difference that controls function. It will be some time before simulations can surmount these problems and be used to make practical devices [171–173]. Meanwhile, we use the mesoscopic approach, which has been so fruitful in computational electronics [18, 21, 122] where atomic-scale simulations are rarely if ever used. The key to the mesoscopic approach is the choice and treatment of correlations. Not all correlations can be handled (see the infinite series in [174] which of course has not been shown to converge).

We consider the Poisson equation and use the treatment in Barton [175] (p. 168) to illustrate the issues in connecting a nanovalve to the outside three dimensional world.

$$\nabla^2\phi(\mathbf{r}) = -\rho(\mathbf{r}) \tag{12}$$

An integral ('Kirchhoff') representation of an equation inside the nanovalve is $K_{vol}(\mathbf{r}) + K_{surf}(\mathbf{r})$, where $K_{vol}(\mathbf{r})$ involves the usual free-space Green's function $G_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r}) = 1/4\pi(\hat{\mathbf{r}} - \mathbf{r})$.

$$K_{vol}(\mathbf{r}) = \int_{vol} d\hat{V}_{vol}\rho(\hat{\mathbf{r}})G_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r})$$
$$\text{where } G_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r}) = \frac{1}{4\pi}\frac{1}{\hat{\mathbf{r}} - \mathbf{r}} \tag{13}$$

$$K_{surf}(\mathbf{r}) = \int_{surf} d\hat{S}_{surf}\,\partial_n\phi(\hat{\mathbf{r}})\hat{G}_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r}) - \int_{surf} d\hat{S}_{surf}\,\phi(\hat{\mathbf{r}})\partial_n\hat{G}_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r}) \tag{14}$$

Here we use Barton's notation for the normal derivative $\partial_n\phi(\hat{\mathbf{r}})$ of the potential (for example) as a function of the location of the source $\hat{\mathbf{r}}$. Inside the nanovalve the volume $K_{vol}(\mathbf{r})$ and surface terms $K_{surf}(\mathbf{r})$ are combined to give the solution $\phi(\mathbf{r})$ of Poisson's equation as shown in *Eq.(12)*

$$\begin{aligned} &K_{vol}(\mathbf{r}) + K_{surf}(\mathbf{r}) = \phi(\mathbf{r}) \quad \mathbf{r} \text{ is inside the nanovalve} \\ &K_{vol}(\mathbf{r}) + K_{surf}(\mathbf{r}) = 0 \qquad \mathbf{r} \text{ is outside the nanovalve} \end{aligned} \tag{15}$$

The free-space Green's function $G_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r})$ is of course the average of the free-space potential in atomic-scale simulations. The special properties of nanovalves depend a great deal on the properties of their surfaces because nanovalves are so small. Structural biology determines the surface of the protein nanovalve and the amino acids forming that surface. Physics determines the surface Green's function $\partial_n\hat{G}_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r})$ and thus the surface normal derivative $\partial_n\phi(\hat{\mathbf{r}}\,|\,\mathbf{r})$ and the surface charge.

Determining these properties is the goal of analysis of specific nanovalves and does not concern us here. It is enough to mention that these properties can be determined from experiments analyzed by the mathematical solution [176] of the appropriate inverse problem. The spatial distribution of structure and permanent structure can be determined from measurements of current-voltage relations under a wide range of ionic conditions, concentrations, and voltages ± 6 k_BT/e (150 mV). The large amount of accurate data allows for an accurate solution of the inverse problem.

Connection of the nanovalve to the external world is what concerns us here. In mathematical language, the problem reduces to the Kirchhoff representation of the end of the channel, written from *Eq.(14)* by isolating the surface of the ends from the rest of the structure.

$$K_{end}(\mathbf{r}) = \int_{end} d\hat{S}_{end}\,\partial_n\phi(\hat{\mathbf{r}})\hat{G}_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r}) - \int_{end} d\hat{S}_{end}\phi(\hat{\mathbf{r}})\partial_n\hat{G}_{free}(\hat{\mathbf{r}}\,|\,\mathbf{r}) \tag{16}$$

In general, this problem can be complex involving interactions of all sorts between the interior of the nanovalve or channel and the external world. Indeed, in some biological channels (calcium channels) this may be important (although so far little studied). In general, however, nanovalves are devices designed to work reasonably robustly and independently of the world

around them. Robust devices need to be transferable from one place to another and so the complex interactions are minimized by the design and evolution of the systems. If we oversimplify to make the point clearly: nanovalves are exceedingly narrow where they allow control but widen dramatically outside that region so 'resistance' to flow is concentrated in the narrow region. Control is robust, available under a wide range of surrounding conditions.

In the nanovalves of semiconductor technology, buffer regions isolate the nanovalve and allow it to have robust properties. The connection to the outside world is through a buffer region of a semiconductor separating the metallic contact from the nanovalve itself. In the nanovalves of biology, the buffer regions are the antechambers of the channel and the surrounding ionic baths between the Ag|AgCl wire (or salt bridge) and the ion channel. These regions are designed (or evolved) so the current in the buffer region is 'Ohmic' independent of time during the function of the nanovalve. These regions are designed to minimize the layers of charge of *Eq.(16)* that create undesirable complex behavior not easily controlled by the valve itself.

The connection of the nanosystem to the outside world is the current flowing in and out of the channel, and of the potential at the ends of the channel. The charges flowing in and out of the channel is not equal because of transient-charge storage ('capacitive' properties, famously voltage-dependent and non-linear), which is large and significant in semiconductor valves [21, 107, 177–179]. In biological channels, voltage-dependent charge storage phenomena are present as well, where they are called 'gating currents' [75–76, 180]. These non-linear displacement currents flowing in channels and associated structures are small but they are controllers of biological function of great importance, and link the motions of handfuls of atoms to macroscopic function. These currents are small because they are produced by motions of a small number of charges compared to the total number of charges in the system. They can be measured because continuity of current guarantees that charge movements arising in conformation changes must also flow in the electrodes and circuits connected to them [181].

The currents flowing in and out of the nanovalves are the connections to the outside world and need to be included in the analysis of the nanovalve itself. These are the currents that are continuous. The currents are the same everywhere in a series-connected system like those described in *Eq.(6)*. The rate constants of mass action models of particle movement can be connected to the external world using a theory that accounts for current flow everywhere, along with diffusion, and perhaps migration as well.

Rate constants for nanovalves. Analytical expressions for the rate constants for movement of charged particles in channel structures can be derived in quite a general way, starting with Langevin equations for the thermal motion of ions [80, 126, 182–184]. It is necessary to use the full Langevin equation (including second derivatives with respect to location) if the treatment is to allow for two boundary conditions (i.e. electrochemical potential on the inside and also the outside of the channel) and macroscopic flux. (If one uses only the Smoluchowski, high-friction version of the Langevin equation, with only first order spatial derivatives, the distribution of velocities of particles has mean zero and no net flux). Trajectories can be doubly conditioned, allowing separate boundary conditions for the two sides of the channel, and the resulting multiple integrals can be performed analytically, somewhat surprisingly, to give the expressions simulated in Ref. [166], derived in Ref. [126], as shown in Ref. [80].

$$\mathbf{L}\text{eft} \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \mathbf{R}\text{ight} \tag{17}$$

where

$$J_k = \overbrace{l \cdot k_f \cdot C_k\left(\mathit{Left}\right)}^{\text{Unidirectional Efflux } J_{out}} - \overbrace{l \cdot k_b \cdot C_k\left(\mathit{Right}\right)}^{\text{Unidirectional Influx } J_{in}} \tag{18}$$

The rate constants are conditional probabilities derived using the theory of stochastic processes from the properties of doubly conditioned Brownian trajectories [126, 182–184].

$$J_k = \overbrace{\underbrace{C_k\left(L\right)}_{\substack{\text{Source}\\ \text{Concentration}}} \underbrace{\left(\frac{D_k}{l}\right)}_{\substack{\text{Diffusion}\\ \text{Velocity}}} \underbrace{\text{Prob}\left\{R \mid L\right\}}_{\substack{\text{Conditional}\\ \text{Probability}}}}^{\text{Unidirectional Efflux}} - \overbrace{C_k\left(R\right)\left(\frac{D_k}{l}\right)\text{Prob}\left\{L \mid R\right\}}^{\text{Unidirectional Influx}} \tag{19}$$

$\uparrow$ Channel Length

or

$$J_k = \overbrace{l \cdot k_f C_k\left(\mathit{Left}\right)}^{\text{Unidirectional Efflux } \boldsymbol{J}_{out}} - \overbrace{l \cdot k_b C_k\left(\mathit{Right}\right)}^{\text{Unidirectional Influx } \boldsymbol{J}_{in}} \tag{20}$$

where

$$k_f \equiv \frac{J_{out}}{C_k\left(\mathit{Left}\right)} = k\left\{\mathit{Right}\middle|\mathit{Left}\right\} = \frac{D_k}{l^2}\text{Prob}\left\{\mathit{Right}\middle|\mathit{Left}\right\} = \frac{D_k}{l^2}\frac{\exp(z_k \mathrm{F} V_{trans}/\mathrm{R}T)}{\frac{1}{l}\int_0^l \exp(z_k \mathrm{F}\phi(\zeta)/\mathrm{R}T)\,d\zeta}; \tag{21}$$

$$k_b \equiv \frac{J_{in}}{C_k\left(\mathit{Right}\right)} = k\left\{\mathit{Left}\middle|\mathit{Right}\right\} = \frac{D_k}{l^2}\text{Prob}\left\{\mathit{Left}\middle|\mathit{Right}\right\} = \frac{D_k}{l^2}\frac{1}{\frac{1}{l}\int_0^l \exp(z_k \mathrm{F}\phi(\zeta)/\mathrm{R}T)\,d\zeta}.$$

R is the gas constant, F is Faraday's constant, T is the absolute temperature, and V_{trans} is the electrical potential across the channel not including potential drops outside the channel.

The coupling to the long-range fields and flows is through the expressions for the electrical potential $\phi(x)$ because the regions outside the channel are decently ohmic [185–189]. First order dependence on concentration in the bath can be described by changes in the concentrations on the left $C_k(L)$ and the right $C_k(R)$. The changes in concentration can often be described adequately this way because they are slow and small. The potential is computed in all space (channel and surrounding baths) by a consistent theory, of one flavor or another. *PNP* [20, 56, 81, 113–114] deals with the motion of point charged particles; ***EnVarA*** [190] deals with spherical particles that can diffuse, migrate or flow by convection, *steric-PNP* [191] is an approximate version much easier to compute, and *PNP-Fermi* [192–194] deals with finite size by enforcing a Fermi distribution that prevents over filling and accounts for saturation of space (by spheres), and no doubt there are many other appropriate models in the vast literature (which includes semiconductor applications, ionic solutions, ion channels, and formation of concrete). These models produce potential profiles that automatically change with conditions so rate constants change and produce currents that are continuous and satisfy Maxwell equations.

Nanovalves are obviously a small subset of the applications of the law of mass action, but it is some comfort to see how consistent analysis can be done in this case so the laws of mass action and continuity of current can both be satisfied. The general case is much harder. It is difficult to grasp all the dimensions of chemical reactions at this stage. One can reach in that direction by studying specific chemical reactions, where simple representations (involving one-dimensional reaction paths) are enough to describe important phenomena.

Variational methods can extend the law of mass action. If the goal is to build transferable systems, so we can build robust devices, as in electronic technology, we must use mathematics that allows interactions between charges and fields, currents and fluxes and flows of solvent, extending from atomic to macroscopic scales. Variational methods are designed to deal with systems with multiple forces and flows in which interactions are unavoidable and complex. In these systems, interactions must be included in all analyses. Otherwise, theories have more adjustable parameters than can be determined experimentally and still cannot deal with a range of conditions because interactions change with conditions in ways too complex for ordinary theories.

Theories of ionic solutions have difficulties. Sadly, theories of ionic solutions seem to have these difficulties Theories of ionic solutions need a large number of adjustable parameters and still cannot describe biological solutions, for example. The first step of analysis seems to be the identification of properties of single ions from measurements of solutions that always contain at least two types of ions, cation, and anion, because of electrical neutrality. The identification of properties of single ions remains 'elusive' even in Hünenberger and Reif's [195] six hundred-page paper containing more than two thousand references. Hünenberger and Reif's title itself characterizes single ion solvation *as elusive even in the infinitely dilute solutions they consider*. The ionic solutions needed to sustain life (and used in much of electrochemical technology) are much more concentrated and have many more and stronger interactions among solutes; interactions between solutes and water; and interactions with far-field boundary conditions. They do not resemble infinitely dilute solutions [39–45]. References [34, 39, 41–42, 50–51, 55, 62, 196–198] draw particular attention to the difficulties and remind readers that almost all biology and electrochemistry occurs in solutions more concentrated than 0.1 M, often in solutions much more concentrated. Solutions tend to be most concentrated where they are most important, near electrodes (in electrochemistry) and in and near nucleic acids, binding sites of proteins, enzyme active sites [199], ion transporters and ion channels [200–204]. Almost all of biology occurs in ionic mixtures and involves flow and so is described particularly poorly by existing theories and simulations [56, 65, 196–197, 205–206]. The sad limitations of our understanding of ionic mixtures, like those in which all of biology occurs, are not widely known, and so embarrassing that many do not want to know of them, as I did not for many years. It is necessary then to document the frustration by quotations from leading workers in that field. The classical text of Robinson and Stokes [40] is still in print and widely used. It is a book not noted for emotion that still gives a glimpse of its authors' feelings of frustration (p. 302).

> "In regard to concentrated solutions, many workers adopt a counsel of despair, confining their interest to concentrations below about 0.02 M,... " [207].

In a recent comprehensive treatment [51] of non-ideal properties of solutions, the editor Werner Kunz says (p. 11 of [52]):

> "It is still a fact that over the last decades, it was easier to fly to the moon than to describe the free energy of even the simplest salt solutions beyond a concentration of 0.1 M or so."

New mathematical tools are needed to resolve a stalemate that has existing since the 1920's. The powerful tools of variational calculus automatically deal with interactions that vary dramatically with conditions. If the mathematics does not deal with interactions, those interactions will not be computed correctly and will wreak havoc with theories that are based on algebraic descriptions of interactions or theories based on the law of mass action with constant rate constants.

Energy Variational Approach. The energy variational approach ***EnVarA*** is *defined* by the Euler-Lagrange process [208] as generalized by Liu, and colleagues, into an energy/dissipation functional. The generalized functional combines two variations yielding a single set of Euler-Lagrange equations (in Eulerian coordinates of the laboratory) using push-back and pull-forward changes of variables [190, 209–212]. In this way, ***EnVarA*** can deal with dissipation (friction) and ionic solutions.

EnVarA describes conserved energy using the classical Hamiltonian variational principle of least action described in textbooks of mechanics. It deals with friction using the Rayleigh dissipation principle described in textbooks of irreversible thermodynamics. When combined, *these principles allow energy to be degraded into entropy as matter and charge flow in frictional materials* [213], like electrolyte solutions.

Ionic solutions experience friction because an ionic solution is a condensed phase essentially without empty space. Ice floats on water, so liquid H_2O has greater density than solid H_2O and presumably less space between its atoms than the solid. Atoms and molecules in a condensed phase cannot move without colliding. Collisions randomize originally correlated motions and make them into the kind of randomized motion that we call heat [214]. The variance of the displacement is what we call temperature. The macroscopic names for the conversion of translational to randomized zero-mean (nearly) Brownian motion, are dissipation and friction.

Energetic variational methods are particularly useful because they allow multi-scale derivation of partial differential equations (and far-field boundary conditions) from physical principles when multiple fields are involved, like convection, diffusion, steric exclusion, and migration in an electric field. Energetic variational principles have recently become available for systems involving friction [209, 211, 215], that is to say, for systems involving ionic solutions [117, 190, 210, 216]. Energetic variational principles combine the full power of the Navier-Stokes equations (a) with either a Lennard-Jones representation of finite-size ions or (b) with a density functional theory of ionic solutions built from Rosenfeld's density functional theory of liquids [190, 217].

Computations must be done in three dimensions because spheres do not exist in one and two dimensions. These theories and their simplifications [191, 218–219] are difficult to compute in three dimensions because of the steeply singular forces used to ensure that atoms do not overlap. Overlap must not be permitted because spheres cannot overfill space: space can be saturated with spheres. Saturation effects are a main cause of non-ideality, particularly in the extremely crowded conditions in and near enzyme active sites, ionic channels, nucleic acids, and the working electrodes of electrochemical cells where twenty molar solutions are not uncommon [199]. (As a rule of thumb, ions are crowded and electric fields are largest [220–222] where they are most important in technology and biology).

If saturation is described by a Fermi-like distribution - as recently derived for spheres of unequal sizes [211, 223] - some of these difficulties can be attacked [192–194]. A fourth order partial differential equation can be written [192] which is easily integrated in three dimensions, after it is reduced to a pair of second order partial differential equations (with carefully defined boundary conditions) and computed with appropriate numerical methods. But it is still not clear how best to apply any of these methods to chemical reactions (*Fig.1*) described by the law of mass action with rate constants extended to be functions or functionals and not constants.

We now have the tools, and we now see the goal - a global treatment combining conservation of mass in chemical reactions described by the law of mass action with conservation of charge flow and current described by Kirchhoff's current law everywhere.

Coda:
Our grasp must be sure,
but our reach should exceed our grasp,
as we do our science.

We must grasp both charge/current conservation and mass action before we can produce robust theories (or simulations) of chemical reactions in ionic solutions that successfully use one set of parameters over a range of conditions, and include the global properties of electric fields.

Correct calculations are needed because there is no engineering without numbers and accurate computations [171–173]. Calculations from theories and simulations - of electronic-, ionic-, or biologically inspired devices - must be checked and calibrated against known results. *Otherwise devices built from those calculations will not work* [56, 65, 196–197, 205–206]. If theories and simulations of electrical devices were not robust, if parameters had to be changed as conditions changed, our electronic technology would be severely limited, to say the least.

Chemical reactions involving current flow must be within our theoretical grasp before we can develop transferable theories. Then we can build devices that perform as expected, as electronic devices usually do. Only then can we expect exponential growth in molecular engineering whether biological or technological. It seems no coincidence that exponential growth in electronic technology came after scientists had a secure grasp of global electrostatics and the *PNP* equations of electrodiffusion in semiconductors. Let us hope that energetic variational methods can grasp ions and chemical reactions in water as well and as successfully as *PNP* has grasped the useful properties of holes and (semi-)electrons in silicon and germanium.

5. Conclusion

The law of mass action is used widely, nearly universally, in chemistry to describe chemical reactions.

The law of mass action does not automatically conserve current, as is clear from the mathematics of a simple case, chosen to illustrate the issues involved. If current is not conserved in a theory, charges accumulate that cannot accumulate in the real world. In the real world, tiny charge accumulation - much less than one percent - produces forces that change predictions of the theory a great deal. Indeed, in the real world tiny charge accumulation produces forces large enough to destroy biological membranes and thus living systems, forces large enough to ionize atoms, create a plasma of electrons (like sparks or lightning) and thus make experiments impossible in normal laboratory settings.

The mathematics in this paper shows that the law of mass action violates conservation of current when current flows if the rate constants depend only on the potential (chemical and electrical) in one location and its immediate vicinity. The same difficulty arises when rate models of the Markov type deal with the movement of charge. The implication is that such models cannot deal with current through an open channel, with the gating properties of channels if the gating mechanism is charged, or with the gating current produced by that mechanism, for example.

The essential issue is that rate constants are LOCAL functions of potential (at one place or in a small region) so they cannot know about current flow far away or at boundary conditions. If current is interrupted far away, local chemical reactions obviously change, but rate constant models show no change. Consider, for example, what happens in a battery when current flow is interrupted far from the battery.

These problems can be fixed by computing the potential (chemical and electrical $\phi(x)$) GLOBALLY and using the neighborhood values of potential that are the result of the GLOBAL calculation. Rate constants in this case have simple physical interpretations and simple expressions involving only one integration, explained in *Eq.(19)*.

$$J_k = \overbrace{\underbrace{C_k(L)}_{\text{Source Concentration}} \underbrace{\left(\frac{D_k}{l}\right)}_{\text{Diffusion Velocity}} \underbrace{\text{Prob}\{R|L\}}_{\text{Conditional Probability}}}^{\text{Unidirectional Efflux}} - \overbrace{C_k(R)\left(\frac{D_k}{l}\right)\text{Prob}\{L|R\}}^{\text{Unidirectional Influx}}$$
$$\text{Channel Length}$$

$$\text{Prob}\{R_{ight}|L_{eft}\} = \frac{\exp(z_k F V_{trans}/RT)}{\frac{1}{l}\int_0^l \exp(z_k F\phi(x)/RT)\,dx};$$

$$\text{Prob}\{L_{eft}|R_{ight}\} = \frac{1}{\frac{1}{l}\int_0^l \exp(z_k F\phi(x)/RT)\,dx}.$$

Of course, the rate model and the global potentials interact and must be solved together so they are consistent. Variational methods guarantee such consistency. Energetic variational methods are needed if the systems are dissipative. Ions in water and channels are dissipative systems because they are condensed phases with little empty space. When ions move, they collide with atoms and dissipate energy in the form of heat.

Acknowledgement

Fred Cohen, Eduardo Rios, and Brian Salzberg provided most useful criticisms and suggestions of an earlier version of this paper. Many thanks! Ardyth Eisenberg edited the manuscript with vigor and love, for which I am most grateful, as it motivates me every day.

REFERENCES

[1] Feynman, R.P.; Leighton, R.B.; Sands, M.: The Feynman Lectures on Physics, Mainly Electromagnetism and Matter (Addison-Wesley Publishing Co., New York) 1963 Vol. 2, p. 592 www.feynmanlectures.caltech.edu/II_toc.html

[2] Bard, A.J.; Faulkner, L.R.: Electrochemical Methods: Fundamentals and Applications (2nd ed.; John Wiley & Sons, New York) 2000

[3] Gutman, E.M.: Can the Tafel equation be derived from first principles? *Corrosion Sci.*, 2005 **47**(12), 3086–3096 DOI 10.1016/j.corsci.2005.05.0055

[4] Gutman, E.M.: Corrigendum to "Can the Tafel equation be derived from first principles?" *Corrosion Sci.*, 2006 **48**(11), 3886 DOI 10.1016/j.corsci.2006.08.001

[5] Eisenberg, B.: Shouldn't we make biochemistry an exact science? *ASBMB Today,* 2014 **13**(9, October), 36–38

[6] Rowlinson, J.S.: The Perfect Gas (Macmillan, New York) 1963 p. 136

[7] Rice, S.A.; Gray, P.: Statistical Mechanics of Simple Fluids (Interscience Wiley, New York) 1965 p. 582

[8] McQuarrie, D.A.: Statistical Mechanics (Harper and Row, New York) 1976

[9] Barratt, J.-L.; Hansen, J.-P.: Basic concepts for simple and complex liquids (Cambridge University Press) 2003 p. 296

[10] Hansen, J.-P.; McDonald, I.R.: Theory of Simple Liquids (3rd ed.; Academic Press, New York) 2006 p. 428

[11] Baird, J.K.: A Generalized Statement of the Law of Mass Action, *J. Chem. Ed.,* 1999 **76**(8), 1146

[12] Lund, E.: Guldberg and Waage and the law of mass action, *J. Chem. Ed.,* 1965 **42**(10), 548

[13] Koudriastev, A.B.; Jameson, R.F.; Linert, W.: Law of Mass Action (Springer, New York) 2001 p. 328

[14] Saslow, W.M.: Electricity, Magnetism, and Light (Academic Press, New York) 2002 p. 800

[15] Zangwill, A.: Modern Electrodynamics (Cambridge University Press, New York) 2013 p. 977

[16] Joffe, E.B.; Lock, K.-S.: Grounds for Grounding (Wiley-IEEE Press, New York) 2010 p. 1088

[17] Jacoboni, C.; Lugli, P.: The Monte Carlo Method for Semiconductor Device Simulation (Springer Verlag, New York) 1989 pp. 1–356
[18] Lundstrom, M.: Fundamentals of Carrier Transport (2nd ed.; Addison-Wesley, New York) 2000
[19] Markowich, P.A.; Ringhofer, C.A.; Schmeiser, C.: Semiconductor Equations (Springer-Verlag, New York) 1990 p. 248
[20] Selberherr, S.: Analysis and Simulation of Semiconductor Devices (Springer-Verlag, New York) 1984 pp. 1–293
[21] Vasileska, D.; Goodnick, S. M.; Klimeck, G.: Computational Electronics: Semi-classical and Quantum Device Modeling and Simulation (CRC Press, New York) 2010 p. 764
[22] Thomson, J.J.: Notes on recent researches in electricity and magnetism: intended as a sequel to Professor Clerk-Maxwell's Treatise on electricity and magnetism (Clarendon Press) 1893
[23] Thomson, J.J.: XXXI. On the effect of electrification and chemical action on a steam-jet, and of water-vapour on the discharge of electricity through gases *The London, Edinburgh, and Dublin Phil. Mag. J. Sci.,* 1893 **36**(221), 313–327
[24] Thomson, J.J.: Nobel Lecture: Carriers of Negative Electricity. In Nobel Media AB 2014: 1906 www.nobelprize.org/nobel_prizes/physics/laureates/1906/thomson-lecture.html
[25] Abraham, M.; Becker, R.: The Classical Theory of Electricity and Magnetism (Blackie, Glasgow, UK) 1932, and subsequent Dover reprints p. 303
[26] The existence of permanent charge as the main source of the electric field was not recognized, at least in the UK and at Cambridge, until J.J. Thomson discovered the electron. Thomson wrote a book [22] (long after Maxwell's death) identified " … as a sequel" [22] to Maxwell's Treatise (see title page of Thomson's treatise). A search of the PDF file shows the book does not contain the word charge. Soon after publishing his Maxwell's Treatise - without mentioning charge - Thomson discovered charge in its most permanent invariant form. He discovered the electron as a ray in a vacuum [23, 24]. Each electron had a permanent invariant charge of 1*e*. After the discovery of the electron, it was clear that permanent charge is the main source of the electric field. Induced charge (the field-dependent component of what chemists call 'polarization') was seen to be a consequence of the field, to be computed as an output of a model of the field, produced by the displacement of matter, more or less akin to gravity's displacement of the oceans that produces tides. The induced charge contributes to the resulting field, to be sure, but it is not the main source, any more than the tides are the main source of the gravitational interactions of the earth and moon. Maxwell, Faraday, and Thomson's 'sequel to Maxwell' held the opposite view. The pre-electron according to Thomson, along with Faraday, and probably Maxwell, viewed induced charge somewhat mysteriously (in my opinion), as almost the same thing as the electric field and in some sense as its source. Physicists today give permanent charge a prominent place and view the relation between charge and field as the essence of electrodynamics, as one imagines Thomson did after his discovery of the electron. The many textbooks that follow Abraham and Becker [25] document this view.
[27] Schey, H.M.: Div, grad, curl, and all that (Norton) 1973.
[28] Eisenberg, B.: Mass Action and Conservation of Current, 2015 arXiv 1502.07251
[29] Chazalviel, J.-N.: Coulomb screening by mobile charges (Birkhäuser, New York) 1999 p. 355
[30] Fuoss, R. M.; Onsager, L.: The Kinetic Term in Electrolytic Conductance, *J. Phys. Chem., B,* 1958 **62**(10), 1339–1340
[31] Fuoss, R.M.; Onsager, L.: Thermodynamic Potentials of Symmetrical Electrolytes, *Proc. Natl. Acad. Sci. USA,* 1961 **47**(6), 818–825
[32] Fuoss, R.M.; Onsager, L.: The conductance of symmetrical electrolytes. I. Potential of total force. *J. Phys. Chem. B,* 1962 **66**(9), 1722–1726
[33] Justice, J.-C.: Conductance of Electrolyte Solutions. I*n Comprehensive Treatise of Electrochemistry Volume 5 Thermodynamic and Transport Properties of Aqueous and Molten Electrolytes* (Conway, B.E.; Bockris, J.O.M.; Yaeger, E., Eds. Plenum: New York) 1983 pp. 223–338
[34] Laidler, K.J.; Meiser, J.H.; Sanctuary, B.C.: Physical Chemistry (4th ed.; Brooks Cole, Belmont CA) 2003 p. 1060
[35] Dufreche, J.F.; Bernard, O.; Durand-Vidal, S.; Turq, P.: Analytical theories of transport in concentrated electrolyte solutions from the MSA, *J. Phys. Chem. B,* 2005 **109**(20), 9873–9884
[36] Van Damme, S.; Deconinck, J.: Relaxation effect on the Onsager coefficients of mixed strong electrolytes in the mean spherical approximation, *J. Phys. Chem. B,* 2007 **111**(19), 5308–5315
[37] Roger, G. M.; Durand-Vidal, S.; Bernard, O.; Turq, P.: Electrical conductivity of mixed electrolytes: Modeling within the mean spherical approximation, *J. Phys. Chem. B,* 2009 **113**(25), 8670–8674
[38] Unidirectional fluxes like J_{xy} use lower case subscripts. Net fluxes like $\bar{J}_{XY}$ use upper case subscripts and are marked with an over bar. Currents like I_{XY} use upper case subscripts because they are net currents, and are not unidirectional. Unidirectional currents are not defined or used in this paper.

[39] Kraus, C.A.: The present status of the theory of electrolytes, *Bull. Amer. Math. Soc.*, 1938 **44**, 361–383
[40] Robinson, R.A.; Stokes, R.H.: Electrolyte Solutions (2nd ed.; Butterworths Scientific Publications, also Dover books, 2002, London) 1959 p. 590
[41] Zemaitis, J.F. Jr.; Clark, D.M.; Rafal, M.; Scrivner, N.C.: Handbook of aqueous electrolyte thermodynamics (Design Institute for Physical Property Data, American Institute of Chemical Engineers, New York) 1986
[42] Pitzer, K.S.: Thermodynamics (3rd ed.; McGraw Hill, New York) 1995 p. 626
[43] Barthel, J.; Krienke, H.; Kunz, W.: Physical chemistry of electrolyte solutions: Modern aspects (Springer, New York) 1998
[44] Vlachy, V.: Ionic effects beyond Poisson-Boltzmann theory *Annual Rev. Phys. Chem.,* 1999 **50**(1), 145–165
[45] Jacobsen, R.T.; Penoncello, S.G.; Lemmon, E. W.; Span, R.: Multi-parameter equations of state. In *Equations of State for Fluids and Fluid Mixtures* (Sengers, J.V.; Kayser, R.F.; Peters, C.J.; White, H.J. Jr., Eds. Elsevier: New York), 2000; p. 849–882
[46] Levin, Y.: Electrostatic correlations: from plasma to biology, *Reports Prog. Phys.,* 2002 **65**(11), 1577
[47] Fawcett, W.R.: Liquids, solutions, and interfaces: from classical macroscopic descriptions to modern microscopic details (Oxford University Press, New York) 2004 p. 621
[48] Lee, L.L.: Molecular thermodynamics of electrolyte solutions (World Scientific Singapore) 2008
[49] Dan, B.-Y.; Andelman, D.; Harries, D.; Podgornik, R.: Beyond standard Poisson-Boltzmann theory: ion-specific interactions in aqueous solutions, *J. Phys. Cond. Matt.,* 2009 **21**(42), 424106
[50] Kontogeorgis, G.M.; Folas, G.K.: Thermodynamic models for industrial applications: from classical and advanced mixing rules to association theories (John Wiley & Sons) 2009 p. 721
[51] Kunz, W.: Specific ion effects (World Scientific Singapore) 2009 p. 348
[52] Kunz, W.; Neueder, R.: An Attempt at an Overview, In *Specific Ion Effects* (Ed.: Kunz, W.; World Scientific Singapore) 2009; p. 11–54
[53] Li, B.: Continuum electrostatics for ionic solutions with non-uniform ionic sizes *Nonlinearity* 2009 **22**(4), 811
[54] Luo, Y.; Roux, B.T.: Simulation of osmotic pressure in concentrated aqueous salt solutions *J. Phys. Chem. Lett.,* 2009 **1**(1), 183–189
[55] Fraenkel, D.: Simplified electrostatic model for the thermodynamic excess potentials of binary strong electrolyte solutions with size-dissimilar ions, *Mol. Phys.,* 2010 **108**(11), 1435–1466
[56] Eisenberg, B.: Crowded Charges in Ion Channels. In *Advances in Chemical Physics*, (Ed. Rice, S.A.; John Wiley & Sons, Inc.: New York) 2011 pp. 77–223 arXiv 1009.1786v1
[57] Gee, M.B.; Cox, N.R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P.E.: A Kirkwood-Buff derived force field for aqueous alkali halides, *J. Chem. Theory Comp.,* 2011 **7**(5), 1369–1380
[58] Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N.F.A.: Kirkwood-Buff coarse-grained force fields for aqueous solutions, *J. Chem. Theory Comp.,* 2012 **8**(5), 1802–1807
[59] Wei, G.; Zheng, Q.; Chen, Z.; Xia, K.: Variational multi-scale models for charge transport, *SIAM Rev.,* 2012 **54**(4), 699–754
[60] Liu, J.-L.; Eisenberg, B.: Correlated ions in a calcium channel model: A Poisson-Fermi theory, *J. Phys. Chem. B,* 2013 **117**(40), 12051–12058
[61] Boda, D.: Monte Carlo simulation of electrolyte solutions in biology: in and out of equilibrium *Annual Rev. Comp. Chem.,* 2014 **10**, 127–164
[62] Fraenkel, D.: Computing excess functions of ionic solutions: the smaller-ion shell model versus the primitive model. 1. Activity coefficients, *J. Chem. Theory Comp.,* 2015 **11**(1), 178–192 DOI: 10.1021/ct5006938
[63] Fraenkel, D.: Computing excess functions of ionic solutions: the smaller-ion shell model versus the primitive model. 2. Ion-size parameters, *J. Chem. Theory Comp.,* 2015 **11**(1), 193–204 DOI: 10.1021/ct500694u
[64] Eisenberg, B.: Life's solutions. A mathematical challenge. 2012 arXiv 1207.4737
[65] Eisenberg, B.: Interacting ions in biophysics: real is not ideal. *Biophys. J.,* 2013 **104**, 1849–1866
[66] Reynaud, S.; Lambrecht, A.: Casimir forces 2014 arXiv 1410.2746
[67] Jaffe, R.L.: The Casimir Effect and the Quantum Vacuum, 2005 arXiv hep-th/0503158v1
[68] Even the Casimir effect of quantum physics [66] is seen as a property of Maxwell's equations by some [67].
[69] Kaxiras, E.: Atomic and Electronic Structure of Solids (Cambridge University Press, New York) 2003 p. 676
[70] Kittel, C.: Solid-State Physics, Eighth Edition (8th ed.; Wiley, New York) 2004 p. 704
[71] Barthel, J.; Buchner, R.; Münsterer, M.: Electrolyte data collection Vol. 12, Part 2: Dielectric properties of water and aqueous electrolyte solutions (DECHEMA, Frankfurt am Main) 1995
[72] Oncley, J.: Electric moments and relaxation times of protein molecules, *J. Phys. Chem.,* 1940 **44**(9), 1103–1113

[73] Oncley, J.: The electric moments and the relaxation times of proteins as measured from their influence upon the dielectric constants of solutions, in *Proteins, Amino Acids and Peptides* (Cohn, E.J.; Edsall, J.T. Eds., Reinhold Publishing Co., New York) 1943

[74] Oncley, J.: The investigation of proteins by dielectric measurements, *Chem. Rev.,* 1942 **30**(3), 433–450

[75] Schneider, M.F.; Chandler, W.K.: Voltage dependent charge movement in skeletal muscle: a possible step in excitation-contraction coupling, *Nature,* 1973 **242**(5395), 244–246

[76] Armstrong, C.M.; Bezanilla, F.: Currents related to movement of the gating particles of the sodium channels, *Nature,* 1973 **242**(5398), 459–461

[77] Vargas, E.; Yarov-Yarovoy, V.; Khalili-Araghi, F.; Catterall, W.A.; Klein, M.L.; Tarek, M.; Lindahl, E.; Schulten, K.; Perozo, E.; Bezanilla, F.; Roux, B.: An emerging consensus on voltage-dependent gating from computational modeling and molecular dynamics simulations, *J. Gen. Physiol.,* 2012 **140**(6), 587–594

[78] Lacroix, J.; Halaszovich, C. R.; Schreiber, D.N.; Leitner, M.G.; Bezanilla, F.; Oliver, D.; Villalba-Galea, C.A.: Controlling the activity of a phosphatase and tensin homolog (PTEN) by membrane potential, *J. Biol. Chem.,* 2011 **286**(20), 17945–17953

[79] Nonner, W.; Peyser, A.; Gillespie, D.; Eisenberg, B.: Relating microscopic charge movement to macroscopic currents: the Ramo-Shockley theorem applied to ion channels, *Biophys. J.,* 2004 **87**, 3716–3722

[80] Eisenberg, B.: Mass Action in Ionic Solutions, *Chem. Phys. Lett.,* 2011 **511**, 1–6

[81] Eisenberg, R.S.: Computing the field in proteins and channels, *J. Membr. Biol.,* 1996 **150**, 1–25. arXiv 1009.2857

[82] Eisenberg, R.S.: From structure to function in open ionic channels, *J. Membr. Biol.,* 1999 **171**, 1–24

[83] Goldman, D.E.: Potential, impedance and rectification in membranes *J. Gen. Physiol.* 1943 **27**, 37–60

[84] Hodgkin, A.; Huxley, A.; Katz, B.: Ionic Currents underlying activity in the giant axon of the squid, *Arch. Sci. Physiol.,* 1949 **3**, 129–150

[85] de Levie, R.; Moreira, H.: Transport of ions of one kind through thin membranes, *J. Membr. Biol.,* 1972 **9**(1), 241–260

[86] de Levie, R.; Seidah, N.G.; Moreira, H.: Transport of ions of one kind through thin membranes. II. Non-equilibrium steady-state behavior, *J. Membr. Biol.,* 1972 **10**(2), 171–192

[87] de Levie, R.; Seidah, N.G.: Transport of ions of one kind through thin membranes III. Current-voltage curves for membrane-soluble ions, *J. Membr. Biol.,* 1974 **16**(1), 1–16

[88] Simons, R.: The steady and non-steady state properties of bipolar membranes *Biochim. Biophys. Acta – Biomembranes,* 1972 **274**(1), 1–14

[89] Coster, H.G.L.: The double fixed charge membrane, *Biophys. J.* 1973 **13**, 133–142

[90] Levitt, D.G.: General continuum analysis of transport through pores. II. Nonuniform pores, *Biophys. J.,* 1975 **15**(6), 553–563

[91] Levitt, D.G.: General continuum analysis of transport through pores I. Proof of Onsager's reciprocity postulate for uniform pore, *Biophys. J.* 1975 **15**(6), 533–551

[92] Levitt, D.G.: Electrostatic calculations for an ion channel. I. Energy and potential profiles and interactions between ions, *Biophys. J.,* 1978 **22**, 209–219

[93] Levitt, D.G.: Electrostatic calculations for an ion channel. II. Kinetic behavior of the gramicidin A channel, *Biophys. J.,* 1978 **22**(2), 221–248

[94] Levitt, D.G.: Comparison of Nernst-Planck and reaction-rate models for multiply occupied channels, *Biophys. J.,* 1982 **37**, 575–587

[95] Buck, R.P.: Kinetics of bulk and interfacial ionic motion: microscopic bases and limits for the Nernst-Planck equation applied to membrane systems, *J. Membr. Sci.,* 1984 **17**(1), 1–62

[96] Cooper, K.; Jakobsson, E.; Wolynes, P.: The theory of ion transport through membrane channels, *Prog. Biophys. Molec. Biol.,* 1985 **46**, 51–96

[97] Levitt, D.G.: Strong electrolyte continuum theory solution for equilibrium profiles, diffusion limitation, and conductance in charged ion channels, *Biophys. J.,* 1985 **52**, 575–587

[98] Levitt, D.G.: Interpretation of biological ion channel flux data. Reaction rate versus continuum theory, *Ann. Rev. Biophys. Chem.,* 1986 **15**, 29–57

[99] Cooper, K.E.; Gates, P.Y.; Eisenberg, R.S.: Surmounting barriers in ionic channels *Quart. Rev. Biophys.,* 1988 **21**, 331–364

[100] Cooper, K.E.; Gates, P.Y.; Eisenberg, R.S.: Diffusion theory and discrete rate constants in ion permeation, *J. Membr. Biol.,* 1988 **109**, 95–105

[101] Levitt, D.G.: Continuum model of voltage-dependent gating. Macroscopic conductance, gating current, and single-channel behavior, *Biophys. J.,* 1989 **55**(3), 489–498

[102] Levitt, D.: General continuum theory for a multi-ion channel. Application for a multi-ion channel, *Biophys. J.,* 1991 **59**, 278–288

[103] Levitt, D.: General Continuum theory for a multi-ion channel, *Biophys. J.,* 1991 **59**, 271–277

[104] Roux, B.: The membrane potential and its representation by a constant electric field in computer simulations, *Biophys. J.,* 2008 **95**(9), 4205–4216

[105] Gumbart, J.; Khalili-Araghi, F.; Sotomayor, M.; Roux, B.: Constant electric field simulations of the membrane potential illustrated with simple systems, *Biochim. Biophys. Acta* 2012 **1818**(2), 294–302
[106] Riordan, M.; Hoddeson, L.: Crystal fire (Norton, New York) 1997
[107] Pierret, R.F.: Semiconductor device fundamentals (Addison Wesley, New York) 1996
[108] Shur, M.: Physics of semiconductor devices (Prentice Hall, New York) 1990 p. 680
[109] Plummer, J.D.; Deal, M.D.; Griffin, P.B.: Silicon VLSI technology: fundamentals, practice, and modeling (Prentice Hall, New York) 2000 p. 817
[110] Jerome, J.W.: Analysis of charge transport. mathematical theory and approximation of semiconductor models (Springer-Verlag, New York) 1995 pp. 1–156
[111] Eisenberg, R.; Chen, D.: Poisson-Nernst-Planck *(PNP)* theory of an open ionic channel, *Biophys. J.,* 1993 **64**, A22
[112] Eisenberg, R.S.: Atomic biology, electrostatics and ionic channels. In *New Developments and Theoretical Studies of Proteins* (Ed.: Elber, R.; World Scientific: Philadelphia) 1996, Vol. 7, p. 269–357 arXiv 0807.0715
[113] Bazant, M.Z.; Thornton, K.; Ajdari, A.: Diffuse-charge dynamics in electrochemical systems, *Phys. Rev. E,* 2004 **70**, 021506
[114] Coalson, R.D.; Kurnikova, M.G.: Poisson-Nernst-Planck theory approach to the calculation of current through biological ion channels, *IEEE Trans. Nanobiosci.,* 2005 **4**(1), 81–93
[115] Burger, M.: Inverse problems in ion channel modeling, *Inverse Problems* 2011 **27**(8), 083001
[116] Wu, H.; Lin, T.-C.; Liu, C.: Diffusion limit of kinetic equations for multiple species charged particles, *Arch. Rational Mech. Anal.,* 2015 **215**(2), 419–441
[117] Hsieh, C.-Y.; Hyon, Y.; Lee, H.; Lin, T.-C.; Liu, C.: Transport of charged particles: entropy production and maximum dissipation principle, 2014 arXiv 1407.8245v1
[118] Hodgkin, A.L.; Katz, B.: The effect of sodium ions on the electrical activity of the giant axon of the squid, *J. Physiol.,* 1949 **108**, 37–77
[119] Mott, N.F.: The theory of crystal rectifiers, *Proc. Roy. Soc. A,* 1939 **171**, 27–38
[120] Mott, N.F.; Gurney, R.W.: Electronic processes in ionic crystals (Oxford University Press, New York City, USA) 1948
[121] Hille, B.: Ion Channels of Excitable Membranes (3rd ed.; Sinauer Associates Inc., Sunderland) 2001 pp. 1–814
[122] Shockley, W.: Electrons and Holes in Semiconductors to applications in transistor electronics (van Nostrand, New York) 1950 pp. 558
[123] Van Roosbroeck, W.: Theory of flow of electrons and holes in germanium and other semiconductors, *Bell System Techn. J.,* 1950 **29**, 560–607
[124] Shurkin, J. N.: Broken genius: The rise and fall of William Shockley, creator of the electronic age (Macmillan, New York) 2006
[125] Helfferich, F.: Ion exchange (1995 reprint, McGraw Hill reprinted by Dover, New York) 1962 p. 640
[126] Eisenberg, R.S.; Kłosek, M.M.; Schuss, Z.: Diffusion as a chemical reaction: Stochastic trajectories between fixed concentrations, *J. Chem. Phys.,* 1995 **102**, 1767–1780
[127] Dennard, R.H.; Gaensslen, F.H.; Rideout, V.L.; Bassous, E.; LeBlanc, A.R.: Design of ion-implanted MOSFET's with very small physical dimensions, *Solid-State Circuits, IEEE J.,* 1974 **9**(5), 256–268
[128] De Loach, B.C. Jr.: The IMPATT story, *Electron Dev. IEEE Trans.,* 1976 **23**(7), 657–660
[129] Critchlow, D.L.: MOSFET scaling-the driver of VLSI technology, *Proc. IEEE,* 1999 **87**(4), 659–667
[130] Sah, C.-T.: Evolution of the MOS transistor - from conception to VLSI, *Proc. IEEE,* 1988 **76**(10), 1280–1326
[131] Moore, G.E.: Cramming more components onto integrated circuits, *Electronics Mag.,* 1965 **38**, 114–117
[132] Moore, G.E.: In *Lithography and the future of Moore's law* (Ed.: Allen, R.D.; SPIE: Santa Clara, CA, USA) 1995 pp. 2–17
[133] Lundstrom, M.: Applied Physics Enhanced: Moore's Law Forever?, *Science,* 2003 **299**(5604), 210–211
[134] Maxwell, J.C.: A Treatise on Electricity and Magnetism (reprinted 1954) (3rd ed.; Dover Publications, New York) 1891 Vol. 1-2
[135] Fonseca, J.E.; Boda, D.; Nonner, W.; Eisenberg, B.: Conductance and concentration relationship in a reduced model of the K^+ channel, *Biophys. J.,* 2010 **98**, 117a
[136] Bell, A.G.: Telephone-circuit, 1881
[137] Stolle, R.: Electromagnetic coupling of twisted pair cables, *Selected Areas in Communications, IEEE J.,* 2002 **20**(5), 883–892
[138] Kirshenbaum, K.; Bock, D.C.; Lee, C.-Y.; Zhong, Z.; Takeuchi, K.J.; Marschilok, A.C.; Takeuchi, E.S.: In situ visualization of $Li/2VP_2O_8$ batteries revealing rate-dependent discharge mechanism, *Science,* 2015 **347**(6218), 149–154
[139] Dudney, N.J.; Li, J.: Using all energy in a battery, *Science,* 2015 **347**(6218), 131–132
[140] Saslow, W.M.: Voltaic cells for physicists: two surface pumps and an internal resistance, *Am. J. Phys.,* 1999 **67**(7), 574–583

[141] No universal precise linkage exists between mass accumulation, rate of change of chemical potential, and flux, for example. The linkage for mass depends on the details of the constitutive equation, and on all components of an electrolyte, if the system is a non-ideal ionic solution, like those that life requires.
[142] Huxley, T.H.: Crayfish, an introduction to the study of biology (Ed. Paul, K.; Trench, London) 1884 p. 371
[143] Huxley, A.F.: Looking back on muscle. In *The Pursuit of Nature* (Ed.: Hodgkin, A.L.; Cambridge University Press: New York) 1977 pp. 23–64
[144] Peskoff, A.; Eisenberg, R.S.; Cole, J.D.: Matched asymptotic expansions of the Green's function for the electric potential in an infinite cylindrical cell, *SIAM J. Appl. Math.,* 1976 **30**(2), 222–239
[145] Peskoff, A.; Eisenberg, R.S.: The time-dependent potential in a spherical cell using matched asymptotic expansions, *J. Math. Biol.,* 1975 **2**, 277–300
[146] Eisenberg, R.S.; Johnson, E.A.: Three dimensional electrical field problem in physiology, *Prog. Biophys. Mol. Biol.,* 1970 **20**, 1–65
[147] Barcilon, V.; Cole, J.; Eisenberg, R.S.: A singular perturbation analysis of induced electric fields in nerve cells, *SIAM J. Appl. Math.,* 1971 **21**(2), 339–354
[148] Kevorkian, J.; Cole, J.D.: Multiple Scale and Singular Perturbation Methods (Springer-Verlag, New York) 1996 pp. 1–632
[149] Hodgkin, A.L.; Rushton, W.A.H.: The electrical constants of a crustacean nerve fiber, *Proc. Roy. Soc. (London) Ser. B,* 1946 **133**, 444–479
[150] Hodgkin, A.L.; Huxley, A.F.: A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol.,* 1952 **117**, 500–544
[151] Hodgkin, A.L.: Chance and design (Cambridge University Press, New York) 1992 p. 401
[152] Huxley, A.: Sir Alan Lloyd Hodgkin 5 February 1914 - 20 December 1998, *Biographical Memoirs Fellows Roy. Soc.*, 2000 **46**, 221–241
[153] Huxley, A.F.: Kenneth Stewart Cole, *Biographical Memoirs Fellows Roy. Soc.*, 1992 **38**, 98–110 books.nap.edu/html/biomems/kcole.pdf
[154] Huxley, A.F.: From overshoot to voltage clamp *Trends Neurosci*. 2002 **25**(11), 553–558
[155] Hill, A.V.: Chemical Wave Transmission in Nerve (Cambridge University Press) 1932 p. 74
[156] Hodgkin, A.L.: Evidence for electrical transmission in nerve: Part II *J. Physiol*. 1937 **90**(2), 211–232
[157] Hodgkin, A.L.: Evidence for electrical transmission in nerve: Part I *J. Physiol*. 1937 **90**(2), 183–210
[158] Neher, E.: Ion channels for communication between and within cells Nobel Lecture, December 9, 1991. In *Nobel Lectures, Physiology or Medicine 1991–1995* (Ed.: Ringertz, N.; World Scientific Publishing Co: Singapore) 1997 pp. 10–25
[159] Sakmann, B.; Neher, E.: Single Channel Recording (2nd ed.; Plenum, New York) 1995 p. 700
[160] Saxena, A.; García, A.E.: Multi-site ion model in concentrated solutions of divalent cations ($MgCl_2$ and $CaCl_2$): osmotic pressure calculations, *J. Phys. Chem. B,* 2014 **119**(1), 219–227
[161] Jensen, M.M.; De Weerdt, K.; Johannesson, B.; Geiker, M.R.: Use of a multi-species reactive transport model to simulate chloride ingress in mortar exposed to NaCl solution or sea-water, *Comp. Mat. Sci.,* 2015 **105**, 75–82
[162] Liu, Q.-F.; Yang, J.; Xia, J.; Easterbrook, D.; Li, L.-Y.; Lu, X.-Y.: A numerical study on chloride migration in cracked concrete using multi-component ionic transport models, *Comp. Mat. Sci.,* 2015 **99**(3), 396–416 DOI 10.1016/j.commatsci.2015.01.013
[163] Hosokawa, Y.; Yamada, K.; Johannesson, B.; Nilsson, L.-O.: Development of a multi-species mass transport model for concrete with account to thermodynamic phase equilibriums, *Mat. Struct.,* 2011 **44**, 1577–1592
[164] Hänggi, P.; Talkner, P.; Borokovec, M.: Reaction-rate theory: Fifty years after Kramers. *Rev. Modern Phys.,* 1990 **62**, 251–341
[165] Fleming, G.; Hänggi, P.: Activated barrier crossing: applications in physics, chemistry and biology (World Scientific, River Edge, New Jersey) 1993
[166] Barcilon, V.; Chen, D.; Eisenberg, R.S.; Ratner, M.: Barrier crossing with concentration boundary conditions in biological channels and chemical reactions, *J. Chem. Phys.*, 1993 **98**, 1193–1211
[167] Hess, K.: Advanced theory of semiconductor devices (IEEE Press, New York) 2000 p. 350
[168] Bløtekjær, K.; Lunde, E.B.: Collision integrals for displaced Maxwellian distribution, *Phys. Stat. Solidi (B)*, 1969 **35**(2), 581–592
[169] Hille, B.: Ionic Selectivity, saturation, and block in sodium channels. A four barrier model, *J. Gen. Physiol.*, 1975 **66**, 535–560
[170] Hille, B.: Transport across cell membranes: carrier mechanisms, Chapter 2. In *Textbook of Physiology* (2nd ed.; Eds.: Patton, H.D.; Fuchs, A.F.; Hille, B.; Scher, A.M.; Steiner, R.D. Saunders: Philadelphia) 1989 Vol. 1, pp. 24–47
[171] Maginn, E.J.: From discovery to data: What must happen for molecular simulation to become a mainstream chemical engineering tool, *AIChE J.,* 2009 **55**(6), 1304–1310
[172] Eisenberg, B.: Multiple Scales in the Simulation of Ion Channels and Proteins, *J. Phys. Chem. C,* 2010 **114**(48), 20719–20733

[173] Post, D.E.; Votta, L.G.: Computational Science Demands a New Paradigm, *Phys. Today,* 2005 **58**, 35–41

[174] Schuss, Z.; Nadler, B.; Eisenberg, R.S.: Derivation of Poisson and Nernst-Planck equations in a bath and channel from a molecular model, *Phys. Rev. E. Stat. Non-linear Soft Matter Phys.,* 2001 **64**(3 Pt 2), 036116

[175] Barton, G.: Elements of Green's functions and propagation: potentials, diffusions, and waves (Oxford Clarendon Press, New York) 1989 p. 465

[176] Burger, M.; Eisenberg, R.S.; Engl, H.: Inverse problems related to ion channel selectivity, *SIAM J. Applied Math.,* 2007 **67**(4), 960–989

[177] Sze, S.M.: Physics of Semiconductor Devices (John Wiley & Sons, New York) 1981 p. 838

[178] Gray, P.R.; Hurst, P.J.; Lewis, S.H.; Meyer, R.G.: Analysis and design of analog integrated circuits (4th ed.; John Wiley, New York) 2001 p. 875

[179] Ieong, M.; Doris, B.; Kedzierski, J.; Rim, K.; Yang, M.: Silicon device scaling to the sub-10-nm regime, *Science,* 2004 **306**(5704), 2057-2060 DOI 10.1126/science.1100731

[180] Bezanilla, F.: How membrane proteins sense voltage, *Nat. Rev. Mol. Cell. Biol.,* 2008 **9**(4), 323–332

[181] Eisenberg, B.; Nonner, W.: Shockley-Ramo theorem measures conformation changes of ion channels and proteins, *J. Comp. Electr.,* 2007 **6**, 363–365

[182] Nadler, B.; Schuss, Z.; Singer, A.: Langevin trajectories between fixed concentrations, *Phys. Rev. Lett.,* 2005 **94**(21), 218101

[183] Singer, A.; Schuss, Z.: Brownian simulations and uni-directional flux in diffusion, *Phys. Rev. E,* 2005 **71**, 026115

[184] Singer, A.; Schuss, Z.; Nadler, B.; Eisenberg, R.S.: Models of boundary behavior of particles diffusing between two concentrations. In *Fluctuations and Noise in Biological, Biophysical, and Biomedical Systems II: series Vol. 5467* (Eds.: Abbot, D.; Bezrukov, S. M.; Der, A.; Sanchez, A. SPIE Proc.: New York) 2004 Vol. 5467, pp. 345–358

[185] Peskoff, A.; Bers, D. M.: Electrodiffusion of ions approaching the mouth of a conducting membrane channel, *Biophys. J.,* 1988 **53**, 863–875

[186] Hall, J.E.: Access resistance of a small circular pore, *J. Gen. Physiol.,* 1975 **66**(4), 531–532

[187] Aguilella-Arzo, M.; Aguilella, V.; Eisenberg, R.S.: Computing numerically the access resistance of a pore, *Eur. Biophys. J.,* 2005 **34**, 314–322

[188] Luchinksy, D.G.; Tindjong, R.; Kaufman, I.; McClintock, P.V.E.; Eisenberg, R.S.: Self-consistent analytic solution for the current and the access resistance in open ion channels *Physical Review E* 2009 **80**, 021925 DOI 10.1103/PhysRevE.80.021925

[189] Luchinsky, D.G.; Tindjong, R.; Kaufman, I.; McClintock, P.V.E.; Peter, V.E.; Khovanov, I.A.; Eisenberg, B.: Observation of "remote knock-on", a new permeation-enhancement mechanism in ion channels, *Biophys. J.* 2014 **106**(2 Suppl. 1), 133A

[190] Eisenberg, B.; Hyon, Y.; Liu, C.: Energy variational analysis EnVarA of ions in water and channels: field theory for primitive models of complex ionic fluids, *J. Chem. Phys.,* 2010 **133**, 104104

[191] Horng, T.-L.; Lin, T.-C.; Liu, C.; Eisenberg, B.: PNP equations with steric effects: a model of ion flow through channels, *J. Phys. Chem. B,* 2012 **116**(37), 11422–11441

[192] Liu, J.-L.; Eisenberg, B.: Analytical models of calcium binding in a calcium channel, *J. Chem. Phys.,* 2014 **141**(7), 075102

[193] Liu, J.-L.; Eisenberg, B.: Poisson-Fermi model of single ion activities in aqueous solutions, *Chem. Phys. Lett.,* 2015 **637**, 1-6 DOI 10.1016/j.cplett.2015.06.079

[194] Liu, J.-L.; Eisenberg, B.: Numerical methods for a Poisson-Nernst-Planck-Fermi model of biological ion channels, *Phys. Rev. E,* 2015 **92**(1), 012711

[195] Hünenberger, P.; Reif, M.: Single-ion solvation. Experimental and theoretical approaches to elusive thermodynamic quantities (Royal Society of Chemistry, London) 2011 p. 690

[196] Eisenberg, B.: Ionic interactions are everywhere, *Physiology,* 2013 **28**(1), 28–38

[197] Eisenberg, B.: Life's solutions are not ideal, 2011 arXiv 1105.0184v1

[198] Sengers, J.V.; Kayser, R.F.; Peters, C.J.; White, H.J. Jr.: Equations of state for fluids and fluid mixtures (Experimental Thermodynamics) (Elsevier, New York) 2000 p. 928

[199] Jimenez-Morales, D.; Liang, J.; Eisenberg, B.: Ionizable side chains at catalytic active sites of enzymes, *Eur. Biophys. J.,* 2012 **41**(5), 449–460

[200] Boda, D.; Nonner, W.; Valisko, M.; Henderson, D.; Eisenberg, B.; Gillespie, D.: Steric selectivity in Na channels arising from protein polarization and mobile side chains, *Biophys. J.,* 2007 **93**(6), 1960–1980

[201] Gillespie, D.: Energetics of divalent selectivity in a calcium channel: the ryanodine receptor case study, *Biophys. J.,* 2008 **94**(4), 1169–1184

[202] Boda, D.; Valisko, M.; Henderson, D.; Eisenberg, B.; Gillespie, D.; Nonner, W.: Ionic selectivity in L-type calcium channels by electrostatics and hard-core repulsion, *J. Gen. Physiol.,* 2009 **133**(5), 497–509

[203] Giri, J.; Fonseca, J.E.; Boda, D.; Henderson, D.; Eisenberg, B.: Self-organized models of selectivity in calcium channels, *Phys. Biol.,* 2011 **8**(2), 026004

[204] Gillespie, D.: A review of steric interactions of ions: Why some theories succeed and others fail to account for ion size. *Microfluid. Nanofluid.,* 2015 **18**(5), 717-738 DOI 10.1007/s10404-014-1489-5

[205] Eisenberg, B.: A leading role for mathematics in the study of ionic solutions, *SIAM News,* 2012 **45**(9), 11–12

[206] Eisenberg, B.: Ionic interactions in biological and chemical systems: a variational approach, *Trans. Farad. Soc.,* 2013 **160**(1), 279–296 arXiv 1206.1517v2

[207] Note that almost all biology and electrochemistry occurs in much more concentrated solutions.

[208] Evans, L.C.: An introduction to stochastic differential equations (American Mathematical Society) 2013 p. 150

[209] Xu, S.; Sheng, P.; Liu, C.: An energetic variational approach to ion transport, *Comm. Math. Sci.,* 2014 **12**(4), 779–789 arXiv 1408.4114

[210] Wu, H.; Lin, T.-C.; Liu, C.: On transport of ionic solutions: from kinetic laws to continuum, 2014 arXiv 1306.3053v2

[211] Forster, J.: Mathematical modeling of complex fluids (Master's Thesis, University of Wurzburg, Wurzburg, Germany) 2013

[212] Ryham, R.J.: An energetic variational approach to mathematical modeling of charged fluids, charge phases, simulation and well posedness (Ph.D. Thesis, Pennsylvania State University, State College) 2006

[213] Xu, X.; Liu, C.; Qian, T.: Hydrodynamic boundary conditions for one-component liquid-gas flows on non-isothermal solid substrates, *Comm. Math. Sci.,* 2012 **10**(4), 1027–1053

[214] Brush, S.G.: The kind of motion we call heat (North Holland, New York) 1986

[215] Hyon, Y.; Kwak, D.Y.; Liu, C.: Energetic variational approach in complex fluids: Maximum dissipation principle, *Disc. Cont. Dyn. Syst. (DCDS-A),* 2010 **26**(4), 1291–1304

[216] Mori, Y.; Liu, C.; Eisenberg, R.S.: A model of electrodiffusion and osmotic water flow and its energetic structure, *Phys. D: Nonlin. Phenom.* 2011 **240**(22), 1835–1852.

[217] Hyon, Y.; Fonseca, J.E.; Eisenberg, B.; Liu, C.: Energy variational approach to study charge inversion (layering) near charged walls, *Disc. Cont. Dyn. Syst. (DCDS-B),* 2012 **17**(8), 2725–2743

[218] Lin, T.C.; Eisenberg, B.: A new approach to the Lennard-Jones potential and a new model: PNP-steric equations, *Comm. Math. Sci.,* 2014 **12**(1), 149–173

[219] Lee, C.-C.; Lee, H.; Hyon, Y.; Lin, T.-C.; Liu, C.: New Poisson-Boltzmann type equations: one-dimensional solutions, *Nonlinearity,* 2011 **24**(2), 431

[220] Suydam, I.T.; Snow, C. D.; Pande, V.S.; Boxer, S.G.: Electric fields at the active site of an enzyme: direct comparison of experiment with theory, *Science,* 2006 **313**(5784), 200–204

[221] Hildebrandt, P.: More than fine tuning, *Science,* 2014 **346**(6216), 1456–1457

[222] Fried, S.D.; Bagchi, S.; Boxer, S.G.: Extreme electric fields power catalysis in the active site of ketosteroid isomerase, *Science,* 2014 **346**(6216), 1510–1514

[223] Liu, J.-L.: Numerical methods for the Poisson-Fermi equation in electrolytes, *J. Comp. Phys.,* 2013 **247**, 88–99

APPENDIX
When does mass action conserve current?
What are the effects of discontinuity in current flow?

Size of discontinuity of current flow.

The difference in current in two sequential chemical reactions is shown in *Eq.(22)*. The difference is the discontinuity of current, the violation of Kirchhoff's law of continuity of current flow. The difference can be zero only in special circumstances. The difference is not zero in general circumstances, nor robustly.

$$\frac{I_{XY} - I_{YZ}}{\mathrm{F}} = z_X \cdot k_{xy}\left[X\right] - z_Y \cdot k_{yx}\left[Y\right] - z_Y \cdot k_{yz}\left[Y\right] + z_Z \cdot k_{zy}\left[Z\right] \tag{22}$$

When are both conservation laws satisfied? When I_{XY} - I_{YZ} = 0 in *Eqs.(22–25)*, the law of mass action is consistent with Kirchhoff's current law and conservation of mass and conservation of charge/current are all satisfied.

Special cases. Units of current I_{XY} - I_{YZ} here are C s^{-1} dm^{-3}.

Special Case A: If all *concentrations are set equal to one*, the currents (in the special case with a tilde) equal

$$\frac{\tilde{I}_{XY} - \tilde{I}_{YZ}}{\mathrm{F} \cdot 1 \frac{\mathrm{mole}}{\mathrm{liter}}} = z_X \cdot k_{xy} - z_Y \cdot k_{yx} - z_Y \cdot k_{yz} + z_Z \cdot k_{zy}; \quad \text{concentrations} = 1 \ \frac{\mathrm{mole}}{\mathrm{liter}} \tag{23}$$

*Special Case A**: If we also set all *charges equal to one*, along with *concentrations equal to one*,

$$\frac{\hat{I}_{XY} - \hat{I}_{YZ}}{\mathrm{F} \cdot 1 \frac{\mathrm{mole}}{\mathrm{liter}}} = k_{xy} - k_{yx} - k_{yz} + k_{zy}; \ \text{concentrations} = 1 \ \frac{\mathrm{mole}}{\mathrm{liter}}; \quad z_X = z_y = z_Z = 1 \tag{24}$$

In this special case of *Eq.(24)*, labelled *A**, asymmetry (net difference) in rate constants determines the discontinuity of current, the violation of Kirchhoff's current law.

Special Case B: Alternatively, we can set all *rate constants* and all *concentrations equal to one*,

$$\frac{\widehat{I}_{XY} - \widehat{I}_{YZ}}{\mathrm{F} \cdot 1 \frac{\mathrm{mole}}{\mathrm{liter}} \frac{1}{\mathrm{sec}}} = z_X - z_Y - z_Y + z_Z; \ \text{concentrations} = 1 \ \frac{\mathrm{mole}}{\mathrm{liter}}; \ \text{rate constants} = \ 1 \ \frac{1}{\mathrm{sec}} \tag{25}$$

In this special case of *Eq.(25)*, labelled *B*, asymmetry (net difference) of charges (valences) determines the discontinuity of current, the violation of Kirchhoff's current law.

Asymmetry of parameters violates conservation of charge/current and produces discontinuity in current from device to device, i.e. it produces accumulation of charge, with sizable effects, as shown next and are to be expected, given the strength of the electric field as discussed in the text.

Size of effects. To estimate the effect on electrical potential *V*, we need to know the size of the system. Imagine a spherical capacitor of radius *R*. Its capacitance to ground - or coefficient of the self-energy, if one prefers the phrase in the chemical literature, is $C = Q/V = 4\pi\varepsilon_r\varepsilon_0 R$ or numerically $4\pi \times 8.85 \times 10^{-12}$ [farad/meter] $\varepsilon_r R = 1.11 \times 10^{-10}$ $\varepsilon_r R$ [farad], where ε_r is the relative dielectric coefficient, about 80 in water solutions over longish times (say greater than 10^{-5} s). Then, a 1 nm radius capacitor with dielectric coefficient 80 has a capacitance of 8.9×10^{-18} farads.

Small charges produce large voltages in such a tiny capacitor. Even the charge on just one ion (1.6×10^{-19} C) would produce 18 mV, large enough (compared to the thermal potential of 25 mV) to have a noticeable (~50%) effect in theories and simulations, because exp (–18/25) = 0.49 (components of rates often vary exponentially according to $\exp(-V/k_B T)$). A unit discontinuity in current I_{XY} - I_{YZ} in *Eqs.(23-25)* lasting for a second would produce a voltage of $V = Q/\mathrm{C} = (1/\mathrm{F})/(8.9 \times 10^{-18}) = 1.14 \times 10^{12}$ V.

Of course, 1 second is a long time for current to flow. If current flowed on a biological time-scale, for 1 ms in a structure 1 nm in radius, with dielectric coefficient 80, the electrical potential would be much less, 'only' $10^9/\varepsilon_r$ V, somewhat less than 10^7 V at low frequencies in water. Current flow of even a microsecond, would produce nearly ten thousand volts.

Rate of change of potential. We can also look at the effect on the rate of change of potential. The discontinuity of current is connected to the rate of change of potential by a version of Coulomb's law:

$$\frac{\partial V}{\partial t} = \frac{1}{\mathrm{C}} I \tag{26}$$

If we apply this formula to the discontinuity of current in the special case of *Eq.(24)* labelled *A**, we can estimate how quickly that discontinuity of current would change the potential:

$$\frac{\partial V}{\partial t} = \frac{\hat{I}_{XY} - \hat{I}_{YZ}}{\mathrm{F} \cdot 1(\mathrm{mole/liter})} \cdot \frac{\mathrm{F}}{\mathrm{C}} = \frac{\hat{I}_{XY} - \hat{I}_{YZ}}{\mathrm{F} \cdot 1(\mathrm{mole/liter})} \cdot \frac{\mathrm{F}}{4\pi\varepsilon_r\varepsilon_0 \mathrm{R}} = \mathrm{F}\frac{k_{xy} - k_{yx} - k_{yz} + k_{zy}}{4\pi\varepsilon_r\varepsilon_0 \mathrm{R}} \tag{27}$$

For a capacitor R = 1 nm with $\varepsilon_r = 80$ and capacitance of 8.9×10^{-18} farads (see above):

$$\frac{\partial V}{\partial t} = 1.1 \times 10^{17} \times \mathrm{F}\left(k_{xy} - k_{yx} - k_{yz} + k_{zy}\right) \text{ in V/s} \tag{28}$$

In other words, the breakdown voltage (~0.2 V) of membranes and proteins would be reached in $1.1 \times 10^{-22}/(k_{xy} - k_{yx} - k_{yz} + k_{zy})$ seconds. The breakdown voltage for matter in general (say 10^6 V) would be reached very quickly. We conclude that *failure of the law of mass action to conserve current is likely to have noticeable effects*.

DE GRUYTER OPEN

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
Vol. 44(1) pp. 29–32 (2016)
hjic.mk.uni-pannon.hu
DOI: 10.1515/hjic-2016-0003

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

SEPARATION OF PROCESS WASTEWATER WITH EXTRACTIVE HETEROGENEOUS-AZEOTROPIC DISTILLATION

ANDRÁS JÓZSEF TÓTH,* ÁGNES SZANYI, ENIKŐ HAAZ, AND PÉTER MIZSEY

Department of Chemical and Environmental Process Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3., Budapest, 1111, HUNGARY

The application of vapour-liquid equilibria-based separation alternatives can be extraordinarily complicated for the treatment of process wastewaters containing heterogeneous-azeotropic. Despite dissimilar successfully tested methods for separation, there is possibility to get better distillation method by enabling the separation of more and more specific process wastewater. Extractive heterogeneous-azeotropic distillation (EHAD) is a new advance in treatment of fine chemical wastewater showing special features to cope with the treatment of highly non-ideal mixtures. This method combines the worth of heterogeneous-azeotropic and extractive distillations in one apparatus without addition of any extra materials. The study of the separations of ternary component process wastewater from the fine chemical industry shows both in the modelled and experimental results that EHAD can be successfully applied. The measured and modelled compositions at extreme purities, that is, close to 0 % or 100 %, can be different because of the inaccuracies of the modelling. This highlights the paramount importance of the experiments if special extra-fine chemicals with almost no impurities, e.g. of pharmacopoeial quality are to be produced by special distillation technique. This study expands the application of EHAD technique, this new field is the separation of process wastewaters.

Keywords: process wastewater, non-ideal mixtures, extractive heterogeneous-azeotropic distillation

1. Introduction

The words of Dr. James R. Fair are still true: "Distillation, king in separation, will remain as the workhorse separation device of the process industries. Even though it is old in the art, with a relatively mature technology support base, it attracts research and professional interest. Without question, distillation will sail into the future with clear skies and a strong wind. It will remain the key separation method against which alternate methods must be judged" from Ref. [1]. The above quote clearly explains why distillation should be investigated continuously [2–4].

In our work, the separation of process wastewater is studied. Such mixtures are quite common in pharmaceutical industry, where the treatment of generally azeotropic mixtures and then fine chemical quality should be achieved; while simultaneous, it is challenging to balance the energetic and environmental consequences. Corresponding non-ideal process wastewaters commonly include different varieties of aldehydes, alcohols, ethers and esters in aqueous media. These compounds already foretoken these problems may resurrect during vapuor-liquid equilibria-based treatment. The separation of a water, isopropyl acetate, ethyl acetate and ethanol process wastewater from the pharmaceutical industry [5] has drawn attention to the need for the improvement of the vapour-liquid equilibria-based separation techniques that includes the worth of using different treatment solutions.

A new hybrid treatment alternative, the so-called extractive heterogeneous-azeotropic distillation (EHAD), has been introduced [6–9] that associates the advantages of heterogeneous-azeotropic and extractive distillations (*Fig.1*). The heterogeneous-azeotropic opportunity presumes that water component is current in the process wastewater and moreover limited immiscibility may exist. Heteroazeotropic distillation utilizes deviations in volatility and liquid-liquid phase

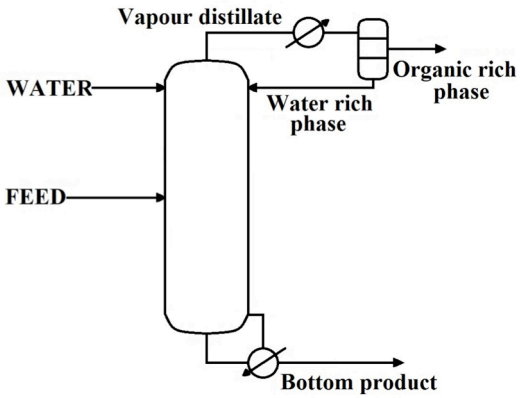

Figure 1. Scheme of the EHAD technique [9].

*Correspondence: andras86@kkft.bme.hu

split by connecting a column and phase separator/decanter [10–12].

EHAD differs from heterogeneous-extractive distillation [13] because no new azeotropic mixture is created, videlicet the entrainer is water in the case and it is already current in the process wastewaters to be treated. Furthermore, relative volatility changing and extractive effect of entrainer agent is absolutely utilised and there is no rectifying section in the column [9]. The capability of the extractive heterogeneous-azeotropic distillation can be demonstrated with a case study. This was carried out for a separation method of 9 devices, extractors and distillation columns [14].

If fine purity is required, e.g. in biofuel or pharmacopoeial applications further unit operations might be needed, e.g. rectification again. However, it can be observed that within the range of extreme purities (close to 0 % or 100 %) the modelling is not reliable enough and the experiments cannot be omitted. The applicability and effectiveness of extractive heterogeneous distillation has been already tested on many different highly non-ideal mixtures with simulations and experiments verifying the accuracy of modelling [8–9]. The strategy clearly shows that EHAD is a forceful method for separation of heterogeneous azeotropic mixtures containing water.

2. Experimental

In spite of the clear guidelines for the applicability of EHAD [8], further application areas are going to be investigated. On the other hand, the separation of actual process wastewater should be interpreted explaining the unit operation and identifying new possible application areas.

2.1. Simulations

For the sake of our aims, a process wastewater was selected: ethyl acetate (EtAc), methyl alcohol (MeOH) and water (H_2O). Flowsheet simulations were achieved with ChemCAD in advance on experiments in order to reduce the necessary number of measurements and to look up promising separation alternatives. Moreover, cooling and heating demands, mass- and waste-flow rates and ideal reflux ratio were also determined. The UNIQUAC equilibrium method is applied [15–18] for the computation of vapour-liquid equilibria of non-ideal mixtures. If binary pairs exist without the availability of UNIQUAC data, the UNIFAC method was applied [19]. Liquid-liquid equilibria were also considered in the form of a vapour-liquid-liquid equilibria (VLLE) model [20]. UNIFAC and UNIQUAC equilibrium models have VLL equilibria options.

The flowsheet of EHAD and intricacies of process wastewater are demonstrated in *Fig.2* computed with ChemCAD. There are one heterogeneous and one homogeneous azeotropes, all in all two binary type [20–23].

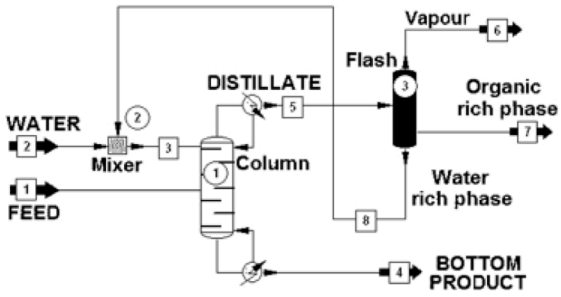

Figure 2. Separation of the process wastewater in flowsheet simulator environment.

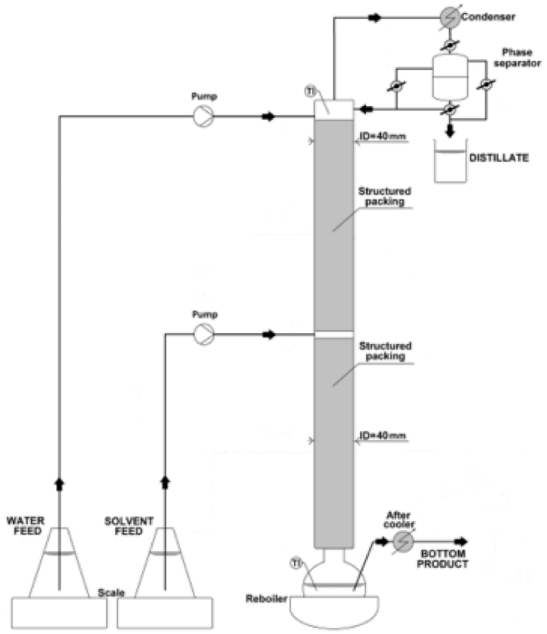

Figure 3. Scheme of EHAD laboratory column.

2.2. Experiments

To save material (amount of sample) and experimental time, first of all, comprehensive flowsheet simulations should be implemented. The results of computer simulations foretoken and prepare the measurements. The measurements are achieved on the distillation column of laboratorial size demonstrated in *Fig.3*.

The major specifications of the distillation apparatus are succeeding: internal width of 2 cm structured packing. The theoretical plates of the column are measured with methyl-alcohol – water mixture, that is 10. The feed (F) is not forewarmed, it is injected on the fifth plate of the apparatus and flow value is kept at 0.06 kg h^{-1}. The entrainer is injected on the tenth plate of the tower and flow value is kept at 0.31 kg h^{-1}. The apparatus warming is directed with a 0.300 kW heating crater. The overhead product, after condensation goes to the decanter, or phase separator. The upper phase is measured and the lower phase is pumped as reflux into the apparatus. The concentration of feed, distillate (D), and waste output (W) were measured using a GC-14B type Shimadzu gas chromatographer with CP-SIL-5CB capillary column.

Table 1. Results of EHAD technique: simulated and measured data (in mass percent)

	F	simulation		experiment	
		D	W	D	W
H_2O	4	3.6	94.7	5.3	94.6
MeOH	26	0.1	5.2	0.2	5.4
EtAc	70	96.3	0.1	94.5	b.d.l.

3. Results and Analysis

We summarise the measured and calculated results in *Table 1*. The comparison also shows the accuracy of EHAD in the ternary case selected for the given study.

The water addition ($x_{F,water}$) and the feed content of ternary mixture ($x_{F,solvent}$) are presented in *Fig.4*, along with the theoretical waste output (x_W) and distillate (x_D) content as well. The two-phase region is found under the dashed curve. Operation of this column is shown by the operating lines.

3.1. Modelling of High Purity Separation

The concentrations 0% and 100% are two extreme numbers, since these do not exist in practical separation technology. We can approach these numbers with ever more advanced and expensive technologies, but we can never arrive at them. Moreover, the problem of extreme concentrations poses two questions: (i) do we have sophisticated enough analytical techniques to do measurements in such regions, and (ii) can we calculate/model these regions accurately/reliably?

Positive answers are crucial in the case of high purity distillation. Modelling is important since it is used for the reduction of solution space and to find promising alternatives. Modelling can therefore be applied to the preparation and selection of the experiment(s). This philosophy is also applied to high purity rectification problems. In our ternary example we arrive at such an extreme concentration range where our analytics is not capable of following the separation. But the modelling shows some promise here.

Another case studies [6–8] from the fine chemical industry highlights that modelling is not reliable in those regions that are close to the extreme numbers of 0% and/or 100%. A demonstrative example for such a crucial situation is the production of ethanol of pharmacopoeial purity. The ethanol produced of this purity may contain contaminants only in concentrations of a few ppm. There are alternative processes for the production of such extremely pure ethanol involving rectification. However, modelling alone can be misleading and it might show such solution alternatives that prove to be unacceptable if they are experimentally tested. On the other hand, the modelling results are definitely uncertain within such extreme purity ranges. The modelled results in all likelihood cannot be supported by experiments. These examples show that if rectification-based process alternatives are investigated the experiments yield reliable results, since modelling is unreliable at purities close to the extreme concentration ranges.

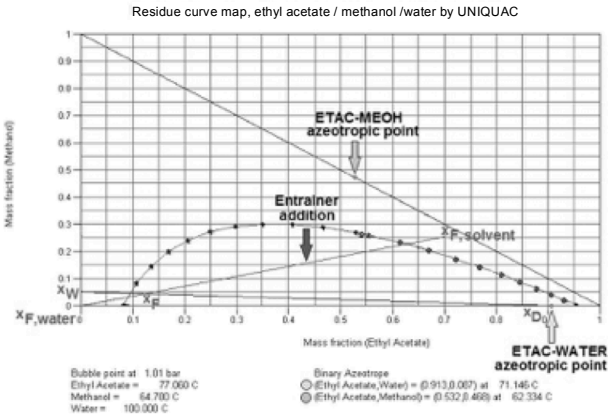

Figure 4. Calculated equilibria and operating lines of EHAD.

4. Conclusion

EHAD separation technique improves in the separation of heterogeneous azeotropic mixtures. It opens an alternative solution for distillation specialists. Treatment of difficult process wastewaters can be solved cost-effective and easily with application of EHAD.

Comparison of modelling techniques and experiments shows good agreement; however, there might be contradictions among measured and modelled data if high purity products are required. Our results obtained as a solution to fine chemical industrial and other, e.g. pharmacopoeial purity distillation problems show that close to 0% and/or 100% concentrations the experiments and modelling can provide unreliable results underlining the paramount importance of experiments and proper analytics.

SYMBOLS

b.d.l.	below detection limit
D	distillate
EHAD	extractive heterogeneous-azeotropic distillation
EtAc	ethyl acetate
F	feed
MeOH	methyl-alcohol
VLE	vapour-liquid equilibrium
VLLE	vapour-liquid-liquid equilibria
W	waste output
$x_{F,solvent}$	feed concentration
$x_{F,water}$	entrainer concentration
x_D	distillate concentration
x_W	waste output concentration

Acknowledgement

The research was supported by the OTKA 112699 grant.

REFERENCES

[1] Kister, H.: Distillation Design (McGraw-Hill Education, New York) 1992
[2] Szabó, L.; Németh, S.; Szeifert, F.: Separation of multicomponent mixtures, *Hung. J. Ind. Chem.*, 2011 **39**(2), 295–300
[3] Toth, A.J.; Gergely, F.; Mizsey, P.: Physico-chemical treatment of pharmaceutical wastewater: distillation and membrane processes, *Per. Pol. Chem. Eng.*, 2011 **55**(2), 59–67 DOI 10.3311/pp.ch.2011-2.03
[4] Mizsey, P.; Toth, A.J.: Application of the principles of industrial ecology for the treatment of process waste-waters with physicochemical tools, *Indust. Ecol.*, 2012 **1**(1), 101–125
[5] Mizsey, P.; Szanyi, A.; Raab, A.; Manczinger, J.; Fonyo, Z.: Intensification of a solvent recovery technology through the use of hybrid equipment, *Comp. Aided Chem. Eng.*, 2002 **10**(1), 121–126 10.1016/S1570-7946(02)80048-7
[6] Szanyi, A.; Mizsey, P.; Fonyo, Z.: Novel hybrid separation processes for solvent recovery based on positioning the extractive heterogeneous-azeotropic distillation, *Chem. Eng. Proc.*, 2004 **43**(3), 327–338 DOI 10.1016/S0255-2701(03)00132-6
[7] Szanyi, A.; Mizsey, P.; Fonyo, Z.: Optimisation of nonideal separation structures based on extractive heterogeneous azeotropic distillation, *Ind. Eng. Chem. Res.*, 2004 **43**(26), 8269–8274 DOI 10.1021/ie0495751
[8] Szanyi, A.; Mizsey, P.; Fonyo, Z.: Separation of highly non-ideal quaternary mixtures with extractive heterogeneous-azeotropic distillation, *Chem. Biochem. Eng. Q.*, 2005 **19**(2), 111–121
[9] Szanyi, A.: Separation of non-ideal quaternary mixtures with novel hybrid processes based on extractive heterogeneous-azeotropic distillation (PhD Dissertation, BME, Budapest) 2005
[10] Franke, M.; Gorak, A.; Strube, J.: Design and optimisation of hybrid separation processes, *Chem. Eng. Tech.*, 2004 **76**(3), 199–210 DOI 10.1002/cite.200406150
[11] Skiborowski, M.; Harwardt, A.; Marquardt, W.: Conceptual design of distillation-based hybrid separation processes, *Annu. Rev. Chem. Biomol. Eng.*, 2013 **4**(1), 45–68 DOI 10.1146/annurev-chembioeng-061010-114129
[12] Skiborowski, M.; Harwardt, A.; Marquardt, W.: Conceptual Design of Azeotropic distillation processes (in Eds.: Gorak, A.; Sorensenm, E.; Distillation: Fundamentals and principles, Academic Press, Aachen, Germany) 2014, Chapter 8, p. 321
[13] Wijesinghe, A.M.J.C.: Development of industrial complexes of special rectification techniques for solvent recovery (PhD dissertation, Lomonosov Institute of Fine Chemical Engineering, Moscow) 1985
[14] Raab, A.: Separation of highly non-ideal mixtures for solvent recovery (MSc Thesis, BME, Budapest) 2001
[15] Abrams, D.S.; Prausnitz, J.M.: Statistical thermodynamics of liquid mixtures: A new expression for the excess gibbs energy of partly or completely miscible systems, *AIChE J.*, 1975 **21**(1), 116–128 DOI 10.1002/aic.690210115
[16] Egner, K.; Gaube, J.; Pfennig, A.: GEQUAC, an excess Gibbs energy model describing associating and non-associating liquid mixtures by a new model concept for functional groups, *Fluid Phase Equilib.*, 1999 **158–160**, 381–389 DOI 10.1016/S0378-3812(99)00137-5
[17] Klamt, A.; Krooshof, G.J.P.; Taylor, R.: COSMOSPACE: Alternative to conventional activity-coefficient models, *AIChE J.*, 2002 **48**(10), 2332–2349 DOI 10.1002/aic.690481023
[18] Wiśniewska-Goclowska, B.; Malanowski, S.X.K.: A new modification of the UNIQUAC equation including temperature dependent parameters, *Fluid Phase Equilib.*, 2001 **180**(1–2), 103–113 DOI 10.1016/S0378-3812(00)00514-8
[19] Fredenslund, A.; Jones, R.L.; Prausnitz, J.M.: Group-contribution estimation of activity coefficients in non-ideal liquid mixtures, *AIChE J.*, 1975 **21**(6), 1086–1099 DOI 10.1002/aic.690210607
[20] Akita, K.; Yoshida, Y.: Phase-equilibria in methanol-ethyl acetate-water system, *J. Chem. Eng. Data*, 1963 **8**(1), 484–490 DOI 10.1021/je60019a003
[21] Gmehling, J.; Menke, J.; Krafczyk, J.; Fischer, K.: Azeotropic data (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany) 1994
[22] Gmehling, J.; Onken, U.; Rarey-Nies, J. R.: Vapor-liquid equilibrium data collection (Dechema, Virginia, USA) 1978
[23] Marsden, C.: Solvents And Allied Substances Manual With Solubility Chart (Cleaver-Hume and Elsevier, London) 1954

DE GRUYTER OPEN

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
Vol. 44(1) pp. 33–38 (2016)
hjic.mk.uni-pannon.hu
DOI: 10.1515/hjic-2016-0004

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

VAPOUR PRESSURE OF ETHANOL AND 1-PROPANOL BINARY MIXTURES

MISIRKHAN TALIBOV[1] AND JAVID SAFAROV[1,2]

[1] Department of Heat and Refrigeration Techniques, Azerbaijan Technical University, H. Javid Ave. 25, AZ1073 Baku, AZERBAIJAN
[2] Institute of Technical Thermodynamics, University of Rostock, Albert-Einstein Str. 2, D-18059 Rostock, GERMANY

The vapour pressure of binary mixtures containing ethanol and 1-propanol were investigated at temperatures ranging from 274.15 to 443.15 K using two different setups with static methods. The measured values were fitted to a Clausius-Clapeyron type relationship. The heat of evaporation of mixtures was determined from the vapour-liquid equilibria data.

Keywords: vapour pressure, ethanol, 1-propanol, static method, pressure transmitters, Clausius-Clapeyron equation

1. Introduction

Investigation of thermodynamic properties of pure liquids and their mixtures is important in various fields of science, chemical engineering, economy and industry. Aliphatic alcohols are commonly applied in chemical, biological, and medical uses as solvents for fats, oils, resins, paints, and nitrocellulose with regard to the manufacture of goods from perfumes to brake fluids [1]. In addition, the studied solutions of ethanol (C_2H_5OH) and 1-propanol (C_3H_7OH) are also used as heat transfer fluids in heat reservoirs, solar heating systems, oxygenates in fuels, and cryogenic power generation systems [2]. For the design and modelling of such applications, the determination of flow in pipes, heat transfer, and mass transfer operations requires the knowledge of thermophysical properties. Density, vapour pressure, speed of sound, viscosity, and heat capacity often need to be defined for these purposes.

This work is a continuation of our previous publications in the field of thermophysical properties of alcohol and their solutions [3–6]. Hereby, the vapour pressure of binary solutions of (1-x) C_2H_5OH + x C_3H_7OH were investigated. The vapour pressure data of binary solutions of ethanol and 1-propanol at different temperatures and concentrations were determined. After the analysis of the literature using “ThermoLit” from NIST, we concluded that only a few vapour pressure values for these systems have been reported to date [7–11].

*Correspondence: javid.safarov@uni-rostock.de

Early studies by Parks and Schwenk [7] reported the vapour pressure of a (1-x) C_2H_5OH + x C_3H_7OH mixture at 298.15 K using glass apparatus and the differential method. A good commercial grade ethanol (w = 99.9 %) and "refined" commercial 1-propanol (w = 99.34 %) were used during the preparation of solutions. Later, Udovenko and Frid [8] investigated the vapour pressure of the same mixture, but within a higher temperature range (323.15 to 353.15 K) using the dynamic method. The vapour liquid equilibria (VLE) of these systems were analysed using a refractometer. The activity coefficients γ of both pure components were calculated.

A series of studies in the early 90s included the work of Zielkiewicz [9], who studied the vapour pressure at 313.15 K using the static method. Dried ethanol and 1-propanol were used during the preparation of solutions. The temperature and pressure were controlled within ±0.001 K and ±0.004 kPa, respectively. Binary samples were prepared by weighing within an uncertainty of ±0.0005 mole fractions. Solution preparations were carried out using the dry nitrogen process. Furthermore, Pradhan *et al.* [10] investigated the vapour pressures of ethanol and 1-propanol solutions at 303.15 K using the static method. For the fitting of obtained values a modified NRTL equation was used.

Quite recently, Cristino *et al.* [11] carried out high temperature VLE measurements for the system of ethanol and 1-propanol solutions within a temperature range of 403.2 to 423.2 K using a flow apparatus. Alcohols used during the preparation of solutions had a confirmed purity greater than 99.9 weight percent. The pressure was controlled using two pressure transducers within ranges of 0 – 0.4 (uncertainty of ±0.0002 MPa) and 0–1.7 MPa (uncertainty of ±0.0009 MPa). The

Table 1. Summary of the vapour pressure P literature investigations of a (1-x) C_2H_5OH + x C_3H_7OH mixture.

Reference	Method	Properties	Temperature (T in K)	Concentration (x mole fraction)	Uncertainty ΔP	Fitted density equation	Purity	Source
Parks [7] 1924	GA	P, T, ΔH	298.15	0.0000 to 1.0000			99.9% (Et) 99.34% (Pr)	CS
Udovenko [8] 1948	DM	P, T, γ	323.15 – 353.15	0.0000 to 1.0000		CC	ARG (Et) ARG (Pr)	R
Zielkiewicz [9] 1993	SM	P, T	313.15	0.0436 to 0.9291	±0.004 kPa			POCh
Pradhan [10] 1993	SM	P, T	303.15	0.0306 to 0.9700	±0.001 kPa		99.9% (Et) 99.6% (Pr)	AC
Cristino [11] 2015	FA	P, T, VLE	403.20 – 423.20	0.0017 to 0.9993	±0.0002 – ±0.0009 MPa	SAFT-VR	99.9% (Et) 99.9% (Pr)	P (Et) FS (Pr)

GA, glass apparatus; ΔH, heat of mixing; P, vapour pressure; T, temperature; x, mole fraction; Et, Ethanol; Pr, 1-Propanol; CS, commercial sample; DM, dynamic method; γ, activity coefficient; CC, Clapeyron-Clausius equation; ARG, analytical reagent grade; R, Reachim, USSR; SM, static method; POCh, Avantor Performance Materials Poland S.A.; AC, Aldrich Chemical; FA, flow apparatus; SAFT-VR, statistical associating fluid theory; VLE, vapour-liquid equilibrium; P, Panreac; FS, Fisher Scientific.

temperature was measured using a platinum resistance thermometer with an uncertainty of ±0.1 K. The statistical associating fluid theory for potentials of variable range (SAFT-VR) was used to model the systems and found to accurately reproduce the experimental data. Using this analytical method the uncertainty of solution preparation was ±0.0001 mole fractions.

The outcome of a literature survey summarised in *Table 1* is that only small temperature, pressure, and concentration intervals were investigated to date in addition to older literature examples decades ago that may have used out-dated measurement techniques.

In this work, the vapour pressures of binary (1-x) C_2H_5OH + x C_3H_7OH solutions were investigated using two highly accurate, fully automatic static experimental setups and ultrapure Merck quality chemicals.

2. Experimental

2.1. Samples and Measurements

Ultra-pure ethanol EMPLURA® (w = 99.995%, CAS No. 71-36-3, Art. Nr. 8.22262.2500) and 1-propanol Analyse EMSURE® ACS, Reag. Ph Eur (w = 99.995%, CAS No. 71-23-8, Art. Nr. 1009971000) were purchased from Merck Schuchardt OHG, Germany. The samples were used without further purification. They were carefully degassed in glass flasks with special vacuum leak-proof valves before measurements were taken. The water content is determined by Karl Fischer titration and was determined to be less than a mass fraction of 20 ppm.

2.2. Experimental Procedure

The vapour pressure measurements of binary solutions of (1-x) C_2H_5OH + x C_3H_7OH were measured using two high-accuracy static experimental seweups [12–14]. The glass cells were used for vapour pressure measurements lower than ambient pressure at temperatures from 274.15 to 323.15 K. The metal cell was used for the higher temperature range of 323.15–433.15 K using the static method [12–14]. The glass cell method consists of absolute and differential parts (if the vapour pressure is smaller than the uncertainty of the absolute cell, 30 Pa). The vapour pressure of the solution was always higher than the uncertainty of measurements between 274.15 and 323.15 K. The measurements within this temperature range were carried out only using the absolute cell of installation. The internal volume of the glass cell in absolute measurements is approximately 78.56 cm^3, and the volume of steel tube cells is 1 cm^3. The glass cell static method consists of a bolted-top cell in a water-bath kept at constant temperature (± 0.01 K) using a thermostat.

The vapour pressure was measured using a calibrated high accuracy sensor head [Type 615A connected to the signal conditioner Type 670A, MKS Baratron, USA] attached to the top of the cell of various Keller pressure transmitters: maximum pressure of 300,000 Pa with an uncertainty of ΔP = ±(400 to 1,500) Pa, maximum pressure of 1,000,000 Pa with an uncertainty of ΔP = ±(1,000 to 5,000) Pa and maximum pressure of 1,600,000 Pa with an uncertainty of ΔP = ±(2,000 to 8,000) Pa. The experimental uncertainty of the pressure in the absolute vapour pressure measurement using the glass cell is ±10–30 Pa.

The internal volume of the measurement cell is approximately 140 cm^3. Temperatures were measured using two different platinum resistance thermometers, PT-100. The second platinum resistance thermometer, PT-100, transfers the measured temperature in the computer via an Omega PT-104A Channel RTD Input Data Acquisition Module (Omega Engineering, Inc., USA) for the measuring of temperature, with an accuracy of ±0.001 K. Experiments were carried out starting from a low temperature (333.15 K) to a high temperature (433.15 K) at 10 K intervals.

Before the experiments, the measurement cells were washed with water, methanol and acetone and then all residual fluids were removed. This procedure requires approximately 2 to 3 h or more to reach the

Table 2. Experimental mole fraction x of 1-propanol, and vapour pressure P (in Pa) of a solution of $(1-x)C_2H_5OH + x\ C_3H_7OH$ [a]

Temperature	mole fraction of 1-propanol (x)							
(K)	0.0000[b]	0.0989	0.1918	0.4034	0.5935	0.7971	0.9038	1.0000[c]
274.15	1684	1490	1374	1116	897	672	530	515
278.15	2248	1980	1843	1498	1219	902	739	697
283.15	3155	2810	2620	2149	1761	1330	1102	1008
293.15	5842	5390	5023	4186	3442	2670	2260	2034
303.15	10458	9762	9180	7780	6521	5150	4402	3854
313.15	18054	16872	15880	13558	11502	9260	8030	7048
323.15	29356	27918	26430	22909	19560	15890	13873	12273
333.15	46796	44590	42200	36700	31777	26000	22992	20472
343.15	71902	68812	65321	57200	49784	41400	36903	32867
353.15	108174	103196	98000	86256	75504	63400	56954	50997
363.15	157911	150535	143157	126500	111128	94202	85257	76746
373.15	224798	214272	203797	180800	159521	136296	123958	113402
383.15	313786	298327	284000	252345	223518	192294	176009	161109
393.15	429264	407124	387678	345105	306612	265259	244002	223982
403.15	576481	545340	519543	462803	412430	358865	331402	305477
413.15	759512	718454	684376	610504	545107	476594	441754	408702
423.15	982342	932045	887923	792004	708954	622271	578714	539077
433.15	1254038	1191945	1135123	1012845	907984	800473	746309	702376
443.15	1582042	1505202	1432927	1278187	1147706	1015139	949123	893968

[a] Standard uncertainties u are $u(T) = 0.01$ K and $u(x) = 0.0001$ mole fractions and the combined expanded uncertainties U_c are $U_c(P) = 30$ Pa for $P < 0.1$ MPa, $U_c(P) = 1500$ Pa for $P < 3$ MPa, and $U_c(P) = 8000$ Pa for $P < 16$ MPa (level of confidence = 0.95); [b] The vapour pressure values of ethanol were taken from Ref. [15]; [c] The vapour pressure values of 1-propanol were taken from Ref. [16].

desired minimal pressure (20–30 Pa). Equilibration of the cells is a rapid process and a constant pressure in the stationary regime is reached within 15 minutes. Equilibrium pressure readings are performed in triplicate approximately 10 to 20 min intervals.

Specific quantities of ethanol and 1-propanol were evacuated, degassed in two separate flasks and connected using an adapter [12]. Ethanol flowed into a flask containing 1-propanol and the concentration of the solution was determined using the weight of the flask containing the solution on an electronic scale (Sartorius ED224S, Germany) with an uncertainty of 0.0001 g. A quantity of the solution was injected into the equilibrium cells up to approximately 50% of their volume.

The vapour pressures of the water, methanol, acetone, toluene, 1-butanol, *etc*. were measured as reference substances for testing both setups [12–14]. The experimental vapour pressure results were assessed to be reliable to within an average uncertainty of ±0.05% according to test measurements.

3. Results and Discussion

The measured experimental vapour pressures for an ethanol/1-propanol mixture within the temperature range of 274.15 to 433.15 K are listed in *Table 2*, and are also shown in *Fig.1*. The vapour pressures of pure alcohols were taken from Refs. [15–16].

The experimental vapour pressure results, P in Pa of investigated solutions were fit to the Antoine equation:

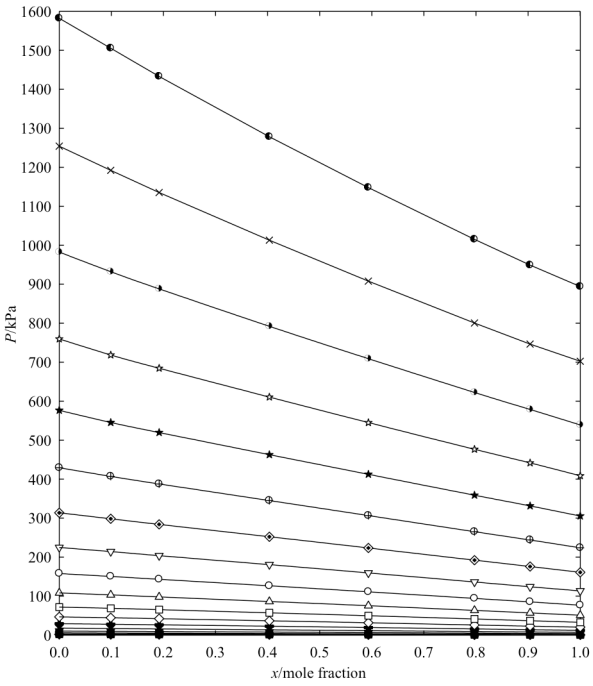

Figure 1. Plot of vapour pressure P (in kPa) of a $(1-x)\ C_2H_5OH + x\ C_3H_7OH$ solution mixture as a function of 1-propanol mole fraction x. ◆, 274.15 K; ■, 278.15 K; ▲, 283.15 K; ●, 293.15 K; ▼, 303.15 K; ✖, 313.15 K; ✚, 323.15 K; ◇, 333.15 K; □, 343.15 K; △, 353.15 K; ○, 363.15 K; ▽, 373.15 K; ◈, 383.15 K; ⊕, 393.15 K; ★, 403.15 K; ☆, 413.15 K, ◖, 423.15 K; x, 433.15 K; ◗, 443.15 K; lines fit to *Eqs.(3)* and *(4)*.

$$\ln (P) = A^A - B^A / (T/K + C^A) \qquad (1)$$

Table 3. Antoine parameters A^A, B^A, C^A and percent deviations ($\Delta P/P$ in %) as a function of 1-propanol mole fraction.

mole fraction	A^A	B^A	C^A	$\Delta P/P$
0.0000[a]	23.1773	3461.23	-54.3818	0.6234
0.0989	22.8524	3275.81	-63.4603	0.0742
0.1918	22.7353	3228.33	-65.9886	0.0689
0.4034	22.4745	3118.11	-72.4808	0.1652
0.5935	22.3425	3077.68	-76.1926	0.2204
0.7971	22.2692	3064.56	-79.8646	0.3263
0.9038	22.1582	3009.02	-84.6711	0.1983
1.0000[b]	22.7515	3373.18	-70.0769	0.8270

[a] from Ref. [15]; [b] from Ref. [16].

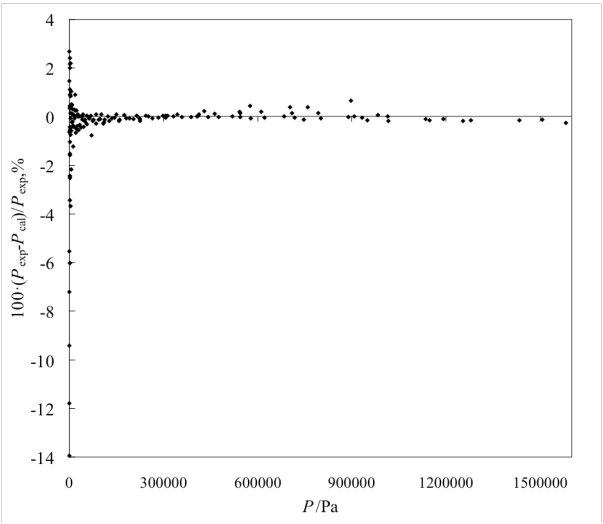

Figure 2. Deviation of experimental P_{exp} and calculated P_{cal} vapour pressure values versus pressure P using *Eqs.(3)* and *(4)* at various temperatures and mole fractions.

The fitted constants A^A, B^A, and C^A for the investigated solutions are summarised in *Table 3* with the standard mean deviation defined as follows:

$$\delta P/P = 100/n \cdot \sum_{i=1}^{n} \left[(P_{exp.} - P_{cal.})/P_{exp.} \right] \qquad (2)$$

From *Table 3*, it can be seen that coefficients A^A, B^A, and C^A exhibit non-trivial dependence from the mole fraction of 1-propanol. Fitting of these coefficients was a challenging task. Thus, we also used a Clausius–Clapeyron-type equation to obtain the vapour pressure results of the investigated solutions from mole fractions of 1-propanol:

$$\ln p = A^{CC} + \frac{B^{CC}}{T} + C^{CC} \ln T + D^{CC} T, \qquad (3)$$

where P is vapour pressure in Pa; T is the temperature in K; and A^{CC}, B^{CC}, C^{CC}, and D^{CC} are the coefficients of the equation, depending on the mole fraction of the solvent as follows:

Table 4. Clausius - Clapeyron equation fitting parameters a_i, b_i, c_i, and d_i from *Eqs.(3)* and *(4)*.

a_i	b_i	c_i	d_i
a_0 = 103.156	b_0 = -7994.80	c_0 = -12.3406	d_0 = 0.0098481
a_1 = 251.788	b_1 = -8366.78	c_1 = -42.1398	d_1 = 0.0527419
a_2 = 222.344	b_2 = -7727.52	c_2 = -36.8168	d_2 = 0.0438405
a_3 = -446.740	b_3 = 14403.70	c_3 = 74.8888	d_3 = -0.0951465

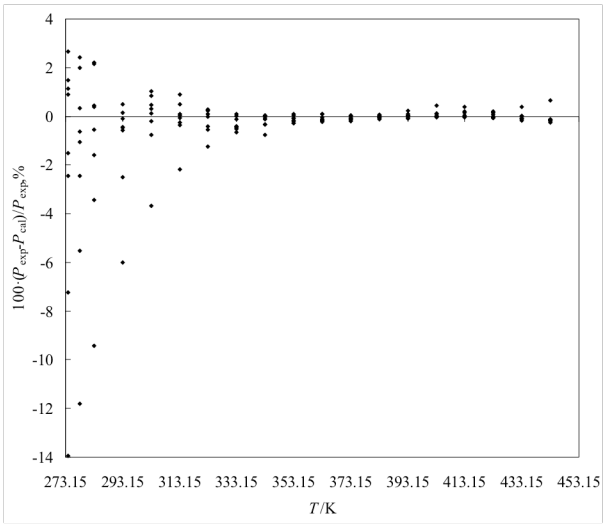

Figure 3. Deviation of experimental P_{exp} and calculated P_{cal} vapour pressure values versus temperature T using *Eqs.(3)* and *(4)* at various pressures P and mole fractions.

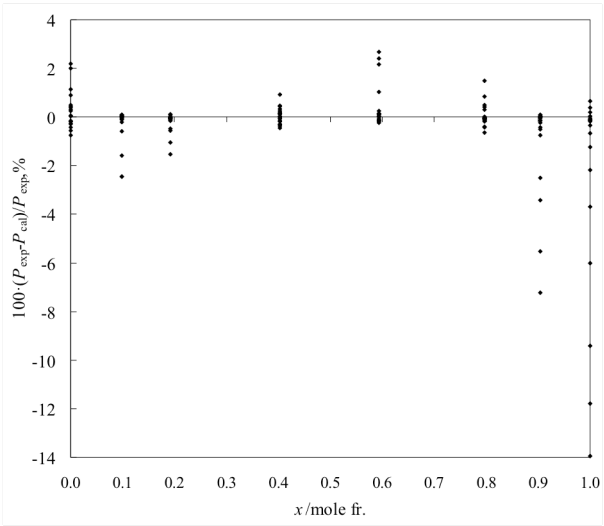

Figure 4. Deviation of experimental P_{exp} and calculated P_{cal} vapour pressure values versus mole fraction x using *Eqs.(3)* and *(4)* at various pressures P and temperatures T.

$$A^{CC} = \sum_{i=0}^{3} a_i x^j \quad B^{CC} = \sum_{i=0}^{3} b_i x^j \quad C^{CC} = \sum_{i=0}^{3} c_i x^j \quad D^{CC} = \sum_{i=0}^{3} d_i x^j \qquad (4)$$

The coefficients a_i, b_i, c_i, and d_i for the investigated ethanol/1-propanol mixtures are tabulated in *Table 4*. The uncertainty of fitting was approximately $u_r(\Delta P/P)$ = 0.7678. The plots of deviation of experimental P_{exp} and calculated P_{cal} vapour pressure values as a function of

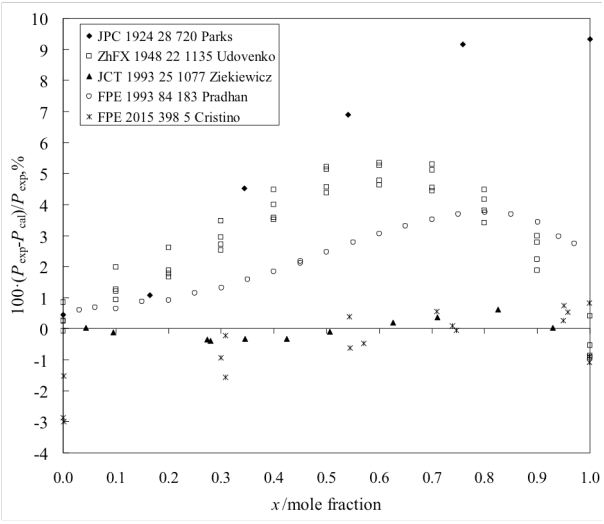

Figure 5. Deviation of experimental P_{exp} and literature P_{lit} vapour pressure values for the ethanol/1-propanol mixture versus 1-propanol mole fraction using *Eqs.(3)* and *(4)* at various pressures *P* and temperatures *T*.

pressure, temperature, and mole fraction using *Eqs.(3)* and *(4)* are shown in *Figs.2-4*, respectively.

The enthalpies of vaporisation, ΔH_{vap} in J mol^{-1}, for the (1-*x*) C_2H_5OH + *x* C_3H_7OH mixture at the four middle temperatures (293.15, 333.15, 373.15, and 423.15 within temperature ranges of 274.15–313.15 K, 313.15–353.15 K, 353.15–393.15 K, and 373.15–443.15 K, respectively) were defined using *Eq.(5)* from Ref. [12]:

$$\frac{d\ln P}{d\left(\frac{1}{T}\right)} = -\frac{\Delta H_v}{R} \tag{5}$$

If we plot ln(*P*) as a function of 1/T, we can define ΔH_v from the gradient of the line:

$$\Delta H_v = -R \cdot \frac{d\ln P}{d\left(\frac{1}{T}\right)} \ . \tag{6}$$

After the integration of *Eq.(6)* we can find

$$\ln p = \left(-\frac{\Delta H_v}{R}\right)\left(\frac{1}{T}\right) + \text{intercept} \tag{7}$$

$$\Delta H_v = RT(\text{intercept} - \ln P) \tag{8}$$

The calculated enthalpy of vaporisations ΔH_v in J mol^{-1} for the (1-*x*)C_2H_5OH + *x* C_3H_7OH mixture within the temperature range of 274.15–443.15 K are listed in *Table 5* and compared to the available literature results [7–11] shown in *Fig.5*.

When the measured values by Parks [7] at *T* = 298.15 K are compared to our values, we obtain ΔP = ±242 Pa or $\Delta P/P$ = ±5.24 % deviations. The maximum deviation is ΔP = 394 Pa at *x* = 0.759 mole fractions of 1-propanol. The Parks' values [7] are higher than our results and the vapour pressures of ethanol exhibit small deviations compared to ours and all other literature values presented in Ref. [15]. The vapour pressure of 1-propanol published in Ref. [7] exhibits a large deviation from ours and all other literature values presented in Ref. [16]. We hypothesise that the vapour pressure values of 1-propanol with high deviation from the literature were used during the analysis of concentration dependence in Ref. [7].

Table 5. Enthalpy of vaporisation, ΔH_v in kJ mol^{-1} for a (1-*x*) C_2H_5OH + *x* C_3H_7OH mixture at various temperatures.

x	293.15 K	333.15 K	373.15 K	423.15 K
0.0000	43.245	41.210	39.579	37.457
0.1574	43.905	41.300	39.542	37.576
0.2876	44.499	41.416	39.425	37.553
0.5351	45.768	41.989	39.707	37.718
0.7130	46.768	42.811	40.157	38.002
0.8699	47.809	43.815	40.965	38.613
0.9411	47.887	44.441	41.546	39.125
1.0000	47.908	45.052	42.330	39.963

The 44 data points of Udovenko and Frid [8] measured within the range of 323.15 – 353.15 K are mostly higher than our values. The average deviations of both sources are ΔP = ±242 Pa and $\Delta P/P$ = ±5.24% with maximum deviations of ΔP = 2952 Pa at *T* = 343.15 K and *x* = 0.5 mole fractions of 1-propanol.

The 11 data points of Zielkiewicz [9] at *T* = 313.15 K exhibit small deviations from our results with ΔP = ±33 Pa and $\Delta P/P$ = ±0.2631% mean deviation. The maximum obtained deviation in ΔP = -59 Pa at *x* = 0.2793 mole fractions of 1-propanol.

The next 22 data points of Pradhan *et al*. [10] are mostly higher than our values. The average mean deviation of this comparison is ΔP = ±143 Pa and $\Delta P/P$ = ±2.2378 %. The maximum obtained deviation in ΔP = 209 Pa and $\Delta P/P$ = ±3.5182 % at *T* = 303.15 K and *x* = 0.7002 mole fractions of 1-propanol.

The last 18 experimental values from the recent work of Cristino *et al*. [11] measured at high vapour pressure intervals of 304.2–967.4 kPa also exhibit small differences from our values as the mean deviation between two experimental sources is ΔP = ±5698 Pa and $\Delta P/P$ = ±0.9227 %. The maximum deviation of this comparison is ΔP = -21134 Pa at *T* = 413.2 K and *x* = 0.0002 mole fractions of 1-propanol.

4. Conclusion

Vapour pressure measurements for the binary mixture of ethanol and 1-propanol over a wide range of temperatures from 274.15 K to 468.15 K were studied. The Antoine and Clausius–Clapeyron equations were used to fit the experimental results. The enthalpies of vaporisation at four various temperatures were calculated. The available literature values were compared with measured values and small deviations were observed.

Acknowledgement

The research was supported by University of Rostock and Azerbaijan Technical University

REFERENCES

[1] Cano-Gómez, J.J.; Iglesias-Silva, C.A.; Ramos-Estrada, M.; Hall, K.R.: Densities and viscosities for binary liquid mixtures of ethanol + 1-propanol, 1-butanol, and 1-pentanol from 293.15 to 328.15 K at 0.1 MPa, *J. Chem. Engng. Data,* 2012 **57**, 2560–2567 DOI 10.1021/je300632p

[2] Kumagai, A.; Yokoyama, C.: Liquid viscosity of binary mixtures of methanol with ethanol and 1-propanol from 273.15 K to 333.15, *Int. J. Thermophys*., 1998 **19**, 3–13

[3] Aliyev, F.Sh.; Safarov, J.T.; Talibov, M.A.; Shahverdiyev, A.N.: Temperature and pressure dependence of density of the methanol and 1-propanol solutions, *J. Fundam. Sci. Azerbaijan Technical Uni*., 2008 **7**(3), 55–61

[4] Abdulagatov, I.M.; Aliyev, F.Sh.; Safarov, J.T.; Talibov, M.A.; Shahverdiyev, A.N.; Hassel, E.: High-pressure densities and derived volumetric properties (excess and partial molar volumes, vapour-pressures) of binary methanol + ethanol mixtures, *Thermochimica Acta,* 2008 **476**(1-2), 51–62 DOI 10.1016/j.tca.2008.07.011

[5] Aliyev, F.Sh.; Safarov, J.T.; Talibov, M.A.; Shahverdiyev, A.N.: (p,ϱ,T) properties of ethanol and 1-propanol solutions, *J. Fundam. Sci. Azerbaijan Technical Uni*., 2008 **7**(1), 35–39

[6] Safarov, J.T.; Aliyev, F.Sh.; Talibov, M.A.; Shahverdiyev, A.N.: Experimental densities of methanol + 1-propanol at temperatures from 298.15 to 423.15 K and at pressures up to 40 MPa, *Proc. 18th Eu. Conf. Thermophysical Prop*. (Pau, France) 2008 p. 45

[7] Parks, G.S.; Schwenk, J.R.: Some physical-chemical prop. of mixtures of ethyl and n-propyl alcohols, *J. Phys. Chem*., 1924 **28**(7), 720–729 DOI 10.1021/j150241a004

[8] Udovenko, V.V.; Frid, Ts.B.: Evaporation heat in binary mixtures: II. *Zh. Fiz. Khim*., 1948 **22**(9), 1135–1145

[9] Zielkiewicz, J.: Vapour + liquid equilibria in heptane + ethanol + propan-1-ol at the temperature 313.15 K, *J. Chem. Thermodyn.,* 1993 **25**(9), 1077–1082 DOI 10.1006/jcht.1993.1105

[10] Pradhan, A.G.; Bhethanabotla, V.R.; Campbell, S.W.: Vapour-liquid equilibrium data for ethanol-n-heptane-1-propanol and ethanol- n-heptane-2-propanol and their interpretation by a simple association model, *Fluid Phase Equil*., 1993 **84**, 183–206 DOI 10.1016/0378-3812(93)85123-4

[11] Cristino, A.F.; Morgado, P.; Galindo, A.; Filipe, E.J.M.; Palavra, A.M.F.; Nieto de Castro, C.A.: High-temperature vapour–liquid equilibria for ethanol-1-propanol mixtures and modeling with SAFT-VR, *Fluid Phase Equil*., 2015 **398**, 5–9 DOI 10.1016/j.fluid.2015.04.009

[12] Safarov, J.T.; Kul, I.; Talibov, M.A.; Shahverdiyev, A.N.; Hassel, E.: Vapour pressures and activity coefficients of methanol in binary mixtures with 1-hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide, *J. Chem. Engng. Data*, 2015 **60**(6), 1648–1663 DOI 10.1021/je501033z

[13] Talibov, M.A.; Safarov, J.T.; Shahverdiyev, A.N.; Hassel, E.: Vapour pressure of geothermal and mineral waters of Yardimli district of Azerbaijan, *Herald Kazan State Technol. Uni*., 2014 **17**(2), 114–118

[14] Safarov, J.T.; Ahmadov, B.; Mirzayev, S.; Shahverdiyev, A.N.; Hassel, E.: Vapour pressures of 1-butanol over wide range of temperatures, chemistry, *Bulgarian J. Sci. Ed*., 2015 **24**(2), 226–246

[15] Talibov, M.A.; Safarov, J.T.; Shahverdiyev, A.N.; Hassel, E.: Vapour pressure of ethanol in a wide range of temperature, *Azerbaijan Nat. Acad. Sci., Ser. Phys. Math. Techn. Sci*., 2015 **2**, 61–71

[16] Talibov, M.A.; Safarov, J.T.: Vapour pressure of 1-propanol in a wide range of temperature, *Azerbaijan Nat. Acad. Sci., Ser. Phys. Math. Techn. Sci*., 2016 **2**, in press

DE GRUYTER OPEN

HJHUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
Vol. 44(1) pp. 39–49 (2016)
hjic.mk.uni-pannon.hu
DOI: 10.1515/hjic-2016-0005

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

GROUP CONTRIBUTION METHOD-BASED MULTI-OBJECTIVE EVOLUTIONARY MOLECULAR DESIGN

GYULA DÖRGŐ[1] AND JÁNOS ABONYI* [1,2]

[1] Department of Process Engineering, University of Pannonia, Egyetem str. 10, Veszprém, H-8200, HUNGARY
[2] Institute of Advanced Studies Kőszeg, Chernel str. 14, Kőszeg, H-9730, HUNGARY

The search for compounds exhibiting desired physical and chemical properties is an essential, yet complex problem in the chemical, petrochemical, and pharmaceutical industries. During the formulation of this optimization-based design problem two tasks must be taken into consideration: the automated generation of feasible molecular structures and the estimation of macroscopic properties based on the resultant structures. For this structural characteristic-based property prediction task numerous methods are available. However, the inverse problem, the design of a chemical compound exhibiting a set of desired properties from a given set of fragments is not so well studied. Since in general design problems molecular structures exhibiting several and sometimes conflicting properties should be optimized, we proposed a methodology based on the modification of the multi-objective Non-dominated Sorting Genetic Algorithm-II (NSGA-II). The originally huge chemical search space is conveniently described by the Joback estimation method. The efficiency of the algorithm was enhanced by soft and hard structural constraints, which expedite the search for feasible molecules. These constraints are related to the number of available groups (fragments), the octet rule and the validity of the branches in the molecule. These constraints are also used to introduce a special genetic operator that improves the individuals of the populations to ensure the estimation of the properties is based on only reliable structures. The applicability of the proposed method is tested on several benchmark problems.

Keywords: computer-aided molecular design, multi-objective optimization, evolutionary algorithm, the Joback method, soft constraints

1. Introduction

The search for compounds exhibiting the desired physical and chemical properties is of significant industrial importance in the search for different chemicals and materials such as polymers [1, 2], blends [3], coatings, solvents, inert agents, heat transfer media [4], and drugs [5]. In the well-known technologies of the chemical, petrochemical, and pharmaceutical industries, the used medium for the given tasks has been developed via practical experience. For the improvement of these technologies or the design of a new process, every hypothetical molecule must be synthesized and tested to check the fulfillment of the design properties. This 'trial and error'-type method for the search of the appropriate agent with the defined properties is slow, inefficient and expensive, thus infeasible in the modern chemical industry and research. However, the problem is complex; the algorithmization can be carried out extensively with the use of property estimation methods and molecular structural feasibility operators. Algorithmic problems can be efficiently solved with the tools of process engineering. Over recent decades, the search for new compounds has resulted in a new area of process engineering, namely computer-aided molecular design (CAMD) [6]. The original task of the design of molecules is formulated as follows: given a set of desired properties, design a product that satisfies these needs. With the use of CAMD tools, the algorithmic approach to the same problem determines the search place: given a set of available structural groups for the satisfaction of desired properties, formulate a product from these sub-units that satisfies the targets. During the decomposition of a CAMD-based problem, two separate tasks can be derived. As can be seen in *Fig.1*, it can be divided into a forward problem, the prediction of a given property, based on the structural characteristics of a molecule; and a 'reverse' problem, the identification of a molecular structure for the satisfaction of target properties.

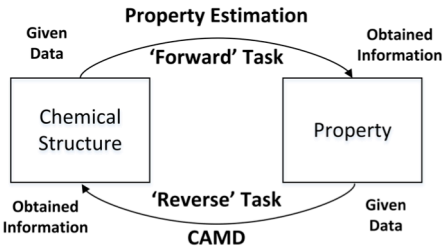

Figure 1. Molecular design is a 'reverse' property estimation task.

*Correspondence: janos@abonyilab.com

The forward problem, the estimation of properties, can be carried out by different methods; for example, polymers [1], solvents [7], surfactant solutions [8], refrigerants [9] and ionic liquids [10]. The limitations of any computer-aided molecular design techniques are closely related to the limitations of the property model being used [11]. The prediction of properties can be carried out with numerous types of methods, including group contribution methods (GC), quantitative structure-activity/property relationship methods (QSAR and QSPR), molecular modeling, empirical modeling and correlations, black box models like neural networks (NN) and the combination of these tools [12]. A novel method for property estimation is the COSMO-RS theory published by Klamt *et al*. that combines quantum chemistry and thermodynamics [13].

The 'reverse' problem, the design of candidate molecules with a given set of properties from a set of molecular sub-units is hardly diversified; the existing techniques were developed for specific molecules and applications. The known methods can be divided into two major groups [14].

The huge chemical search space is further complicated by the often competing target properties of the design process. A genetic algorithm is a promising method for the generation of new candidate molecules. Multi-objective optimization algorithms generate a set of optimal solutions. The Pareto fronts of these solutions simultaneously consider several design aspects. Since when solving the problem multiple target properties must be taken into consideration at the same time; the problem has been implemented in a well-established genetic algorithm-based multi-objective optimization environment, the Non-dominated Sorting Genetic Algorithm-II (NSGA-II). The search space is conveniently described by the occurrence of each fragment from a given set of available types of groups, and the "distance" of the properties from the target values is estimated by the Joback method. The feasibility of the molecule is tested by feasibility constraints for branching and the octet rule. These constraints are also used to introduce a special genetic operator that improves the individuals of the populations to ensure the estimation of the properties is based on only reliable structures. Thus as the result, an evolutionary approach for solving molecular design problems with descriptors of varying dimensionality has been developed, that moves effectively towards the Pareto optimal front.

In the present work the definition of the design problem is followed by a theoretical overview of the used property estimation method and of the nature of genetic algorithms paying special attention to NSGA-II. After the description of the different algorithms, proposed for the solution of the design task, the efficiencies of these approaches are examined through several benchmark problems, and the results are discussed extensively to determine improvements in the applicability of these algorithms.

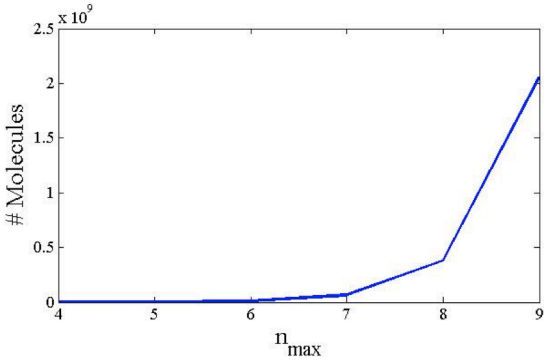

Figure 2. The combinatorics of group selection.

2. Methodology

2.1. Problem Formulation

In the first class, numerous candidate molecules are created randomly from a given set of groups. The number of candidate molecules that can be generated by selecting N groups from a set of K groups, allowing repetition and ignoring the order of selection can be determined by *Eq.(1)*.

$$C^{R}(K,N) = \frac{(K+N-1)!}{N!(K-1)} \quad (1)$$

The total number of candidate molecules that can be selected from a set of K groups is the sum of the results of *Eq.(1)* from N_1 until N_{max} as is given by *Eq.(2)* (N_1 is practically equal to at least 2, as no molecules consist of only one fragment). The sum of candidates is equal to

$$\sum_{N_1}^{N_{max}} C^{R}(K,N) = \sum_{N_1}^{N_{max}} \frac{(K+N-1)!}{N!(K-1)} \quad (2)$$

According to *Eq.(2)* the total number of candidate molecules can undergo a combinatorial explosion as can be seen in *Fig.2*.

In the second class for the solution of the 'reverse' problem, this increased number of candidates must be eliminated by some objective function that expresses the 'distance' from the target. An example of this approach can be a tree structure to mimic the chain of a molecule and reach feasible structures as was carried out in Refs. [12] and [14].

As in the literature, the proposed methods for the design of molecules are highly diversified. Lin *et al*. studied the design of metal catalysts [15], numerous articles can be found for the design of drugs [5, 16-18], Perdomo *et al*. designed and improved biodiesel fuel blends [3] and Kasat *et al*. summarized the applications of genetic algorithms in polymer science including polymer design [2].

To solve the design problem, several computational strategies have been used. Genetic algorithms (GA) are used in several publications [15, 19, 20], a combination of neural networks and genetic algorithms is used in [21] and linear programming is used in [12, 14].

2.2. Theoretical Methodologies

The definition of a chemical product design problem is based on the description of design constraints. A set of properties is specified as constraints with specified values with lower and/or upper boundaries. These properties are the *explicit property constraints* as their values can be determined directly by the application of some model calculation or experimentally. In the case of CAMD problems, explicit constraints are evaluated through property estimation methods, these can be, for example, critical properties, solubility indexes, normal boiling points, etc. However, property estimation methods have been significantly improved, there are products, for example, food, fragrances, health and safety products, and aesthetics that cannot be calculated with the use of these models, as these properties are based on subjective opinions or existing knowledge. In the case of these *implicit property constraints* (e.g. taste, aroma, color and health effects of products) the use of databases or the opinion of the designer can be implemented during the evaluation stage. During the basic CAMD process explicit constraints are taken into consideration, these relate mainly to physical properties, and implicit considerations are taken into account during the selection of available molecule fragments or compounds (e.g. no aromatic compounds are taken into consideration, or no halogens or cyanides are available) [6].

To understand the formulation of the solution to this explicit property constraint-based problem, the following information must be taken into consideration as the input information of the molecule design task (according to [12, 14]):

1. Set G of N_{max} groups of which the designed molecule can be composed
2. The boundaries for the specified properties to be satisfied: P_{lb}^{j}'s for the lower boundaries and P_{ub}^{j}'s for the upper boundaries, where $j = 1, 2, \ldots, m$, the specified properties
3. The lower (l_{ll}^{i}) and upper limits (l_{ul}^{i}), for the number of appearances of group i in the designed molecule ($i = 1, 2, \ldots, n$)
4. The property k can be estimated via a property estimation method as function f^{k} ($f = 1, 2, \ldots, m$), in the case of group contribution methods f^{k} can be written as $f^{k}(x_1, x_2, \ldots, x_n)$ (where $x_1, x_2, \ldots, x_n$ are the numbers of group types #1, #2, …, #n respectively).

The problem using the expressions above can be formulated as follows: i groups can be chosen from a given set of molecular subunits (G) considering the limits of l_{ll}^{i} and l_{ul}^{i}, to find all the possible molecular structures, while the property constraints given in *Eq.(3)* are satisfied (where $j = 1, 2, \ldots, m$).

$$P_{lb}^{j} \leq f^{j}(x_1, x_2, \ldots, x_n) \leq P_{ub}^{j} \qquad (3)$$

During the solution of the above-defined CAMD task, the generation and test method can seem to be inefficient as an enormous number of candidate molecules are created which finally turn out to be infeasible molecular structures.

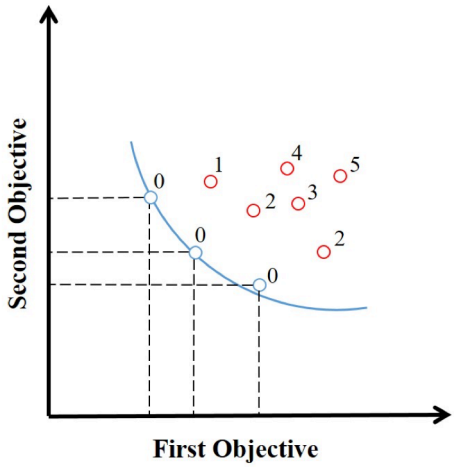

Figure 3. A Pareto front (The label of each solution refers to the number of other solutions that dominate them, non-dominating solutions are labeled with zero) [22].

As is familiar among financial and industrial problems, some properties need to be minimized and others maximized between the constraint values, while others need to be close to a specified value. This results in a multiple-objective optimization task with concurring targets. Our purpose is not to find a single solution, but a set of candidate molecules from the Pareto front. Pareto optimal solutions are those for which improvement in terms of one objective can only take place with the worsening of at least one other objective function. Pareto-ranking is the process of determining the rank of each solution through identifying the number of other solutions that dominate it (the number of solutions that are better than it in terms of every objective) [22]. A Pareto front can be seen in *Fig.3*.

In the present work feasibility constraints are implemented in the algorithm to filterout the resultant molecular structures in terms of feasibility. As in this approach the candidates are still filtered out after the property evaluation, the efficiency of the search can still seem to be inefficient. As the solution to this contradiction, a special genetic operator has been introduced that improves the individuals of the populations to ensure the estimation of the properties is based on only reliable structures, and the property evaluation is carried out on solely feasible molecules.

2.3. Property Estimation

As the new molecules created in the 'reverse' problem are evaluated via property estimation methods to verify the satisfaction of properties, the success of CAMD tasks depends on, to a large extent, the reliability of the estimation method being used. From the point of view of precision, the highly improved, detailed models seem to be tempting in terms of the application. However, the computational complexity of these models is increased

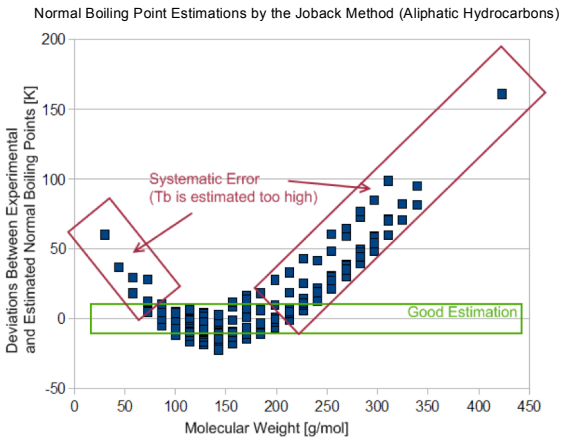

Figure 4. Deviations between predicted boiling points and experimental data [24].

as well. In terms of precision and complexity, group contribution methods are promising solutions, as the equation for estimation assumes a linear additivity dependence as presented in *Eq.(4)*, where θ_i^j is the group contribution value of group i for property j, θ_0^j is the offset value of property j and P_{est}^j is the estimated property value.

$$P_{est}^j(\bar{x}) = \theta_0^j + x_1\theta_1^j + x_2\theta_2^j + \ ... + x_n\theta_n^j \qquad (4)$$

The Joback method, also known as the Joback/Reid method, is proposed to estimate eleven important physical properties of pure materials. During the determination of group contribution values, a common set of structural groups was employed in the regression process. To obtain the minimum values, the minimization of the sum of the absolute errors found from the estimated and the experimental values was carried out. As not the square values, but the absolute values were minimized, the method provides an improved estimation for the majority of the cases, but estimates slightly higher error values for outliers [23]. The systematic deviations of the Joback method in the case of normal boiling points can be seen in *Fig.4*, where experimental data is taken from the Dortmund Data Bank.

The Joback method uses *Eqs.(5-15)* to predict the specific properties as follows:
Normal Boiling Point:

$$T_b[K] = 198 + \sum T_{b,i}x_i \qquad (5)$$

Melting Point:

$$T_m[K] = 122.5 + \sum T_{m,i}x_i \qquad (6)$$

Critical Temperature:

$$T_c[K] = T_b \frac{1}{\left[0.584+0.965\ \sum T_{c,i}x_i - \left(\sum T_{c,i}x_i\right)^2\right]} \qquad (7)$$

Critical Pressure (N_A is the number of atoms in the molecular structure):

$$P_c[\text{bar}] = \left[0.113 + 0.0032 \cdot N_A - \sum P_{c,i}x_i\right]^{-2} \qquad (8)$$

Critical Volume:

$$V_c[\text{cm}^3/\text{mol}] = 17.5 + \sum V_{c,i}x_i \qquad (9)$$

Heat of Formation (ideal gas, 298 K):

$$H^0[\text{kJ/mol}] = 68.29 + \sum H_i^0 x_i \qquad (10)$$

Gibbs Free Energy of Formation (ideal gas, 298 K):

$$G^0[\text{kJ/mol}] = 53.88 + \sum G_i^0 x_i \qquad (11)$$

Heat Capacity (ideal gas, parameters are valid from 273 K to approximately 1000 K):

$$C_P\left[\frac{\text{J}}{\text{molK}}\right] = \sum a_i x_i - 37.93 + [\sum b_i x_i + 0.210] \cdot T + \\ + [\sum c_i x_i - 3.91 \cdot 10^{-4}] \cdot T^2 + \\ + [\sum d_i x_i + 2.06 \cdot 10^{-7}] \cdot T^3 \qquad (12)$$

Heat of vaporization at normal boiling point:

$$\Delta H_{vap}[\text{kJ/mol}] = 15.30 + \sum H_{vap,i}x_i \qquad (13)$$

Heat of Fusion:

$$\Delta H_{fus}[\text{kJ/mol}] = -0.88 + \sum H_{fus,i}x_i \qquad (14)$$

Liquid Dynamic Viscosity (M_W is the molecular weight, the parameters are valid from the melting point up to 0.7 of the critical temperature):

$$\eta_L[\text{Pa} \cdot \text{s}] = M_w \cdot \exp\left(\left[\sum \eta_{a,i}x_i - 597.82\right]/T + \right. \\ \left. + \sum \eta_{b,i}x_i - 11.202\right) \qquad (15)$$

2.4. A Promising Approach for the Solution of the Design Task: Genetic Algorithms

Genetic Algorithms (GAs) are stochastic optimization methods that imitate natural selection. GAs provide not a single optimal solution to a problem, but several near-optimal solutions, which is the main advantage of evolutionary algorithms in the field of CAMD, because near-optimal solutions can be further processed later by the designer and the most promising ones can be selected for synthesis.

During the operation of a GA, a population of candidate solutions competes for survival, based on their resemblance to the target values. This resemblance is described by a normalized distance value between 0 and 1 and called the fitness. Candidate molecules are usually described by strings, and the components of these strings represent the 'genes' of the individual. The evaluation of the population is carried out therefore by

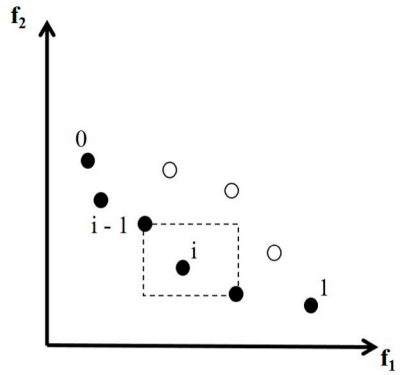

Figure 5. The crowding distance of the i[th] solution is the average side length of the cuboid (the front is marked with solid circles) [25].

the calculation of fitness and the surviving members have the chance to reproduce and propagate their genes, thus forming the next generation. This propagation is dependent on the genetic operators applied by the specific algorithm being used, the most common are crossover and mutation. The creation of next generations is continued until convergence is obtained (no considerable improvement is observed), or the maximum number of generations set by the user is reached [21].

2.4.1. The Non-dominated Sorting Genetic Algorithm-II (NSGA-II)

For the description of the NSGA-II algorithm, three innovations of the algorithm must be described first: the fast non-dominated sorting procedure, a fast crowded distance estimation method, and the crowded comparison operator [25, 26].

The *fast non-dominated sorting approach* of the NSGA-II is based on the calculation of three entities: the domination count, the number of solution, which dominates the given solution, and the set of solutions that the given solution dominates. In the first non-dominated front the domination count of all solutions is zero. After the determination of the first non-dominated front, each of the solutions dominated by its members is visited and his or her domination count is reduced by one. If the domination count of a solution becomes 0, then it becomes a member of the second non-dominated front. This algorithm is repeated until all fronts are determined.

Along with convergence to the Pareto optimal set of solutions, the maintenance of a healthy spread of solutions is required to avoid the problem of getting stuck in the area of a local Pareto optimum. To prevent this issue, the parameter of *crowding distance* is introduced. During the calculation of this parameter, the average distance between two points on either side of a particular solution along each objective is calculated. The overall crowding distance value is the sum of the individual crowding distance values along each objective. With the use of this parameter, the "density" of solutions in the search place can be calculated. A solution with a higher crowding distance value is less crowded by other solutions, in other words, the outlier solutions can be identified. Thus, the parameter is applicable for the maintenance of diversity. The crowding distance computation for two objectives is illustrated in *Fig.5*.

The goal of the genetic algorithm, to obtain a uniformly spread Pareto-optimal front, is reached with the help of the *crowded-comparison operator*. The operator guides the selection between two possible solutions as follows:

1. If the non-domination ranks of two solutions differ, the solution which dominates the other is preferred, in other words, whose domination index is less.
2. If the non-domination ranks of two solutions are equal (the two solutions are from the same front), then the solution of the less crowded region is preferred.

The main loop of the NSGA-II can be explained by understanding the operators described above. The randomly created initial parent population (with N members) is sorted based on the non-domination rank. The crossover *Eqs.(16-17)* and mutation *Eqs.(18-19)* operators are applied to create the next generation (with N members) [27]. The algorithm applies to the intermediate crossover, which creates two children from two parents: *parent1* and *parent2* (*child* and *parent* are vectors, $\bar{x}$, containing the results of the specific problems),

$$child1 = parent1 + rand \cdot ratio \cdot (parent2 - -parent1) \quad (16)$$

$$child2 = parent2 - rand \cdot ratio \cdot (parent2 - -parent1) \quad (17)$$

where *ratio* is a scalar between 0 and 1, and *rand* stands for a random number,

The applied Gaussian mutation adds a normally distributed random number to each variable,

$$child = parent + S \cdot rand \cdot (ub - lb) \quad (18)$$

$$S = scale \cdot \left(1 - shrink \cdot \frac{currGen}{maxGen}\right) \quad (19)$$

where *scale* is a scalar, that determines the standard deviation of the random number generated and *shrink* is a scalar between 0 and 1. As the optimization progresses, this shrink parameter decreases the mutation range. *currGen* and *maxGen* are the numbers of the current and maximal generations, respectively.

Since elitism is introduced, the creation of the first population differs from the creation of a subsequent one. The algorithm is described for the t^{th} generation.

A combined population (R_t) (with $2N$ members) is created by the summation of the parent population (P_t) and the population obtained by the use of crossover and mutation operators (Q_t). The population R_t is sorted according to non-domination and as all previous and current population members are included, elitism is ensured. Now solutions belonging to the first non-

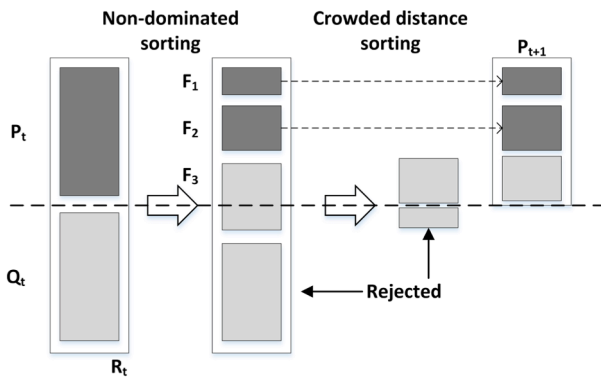

Figure 6. Graphical illustration of the NSGA-II procedure [25].

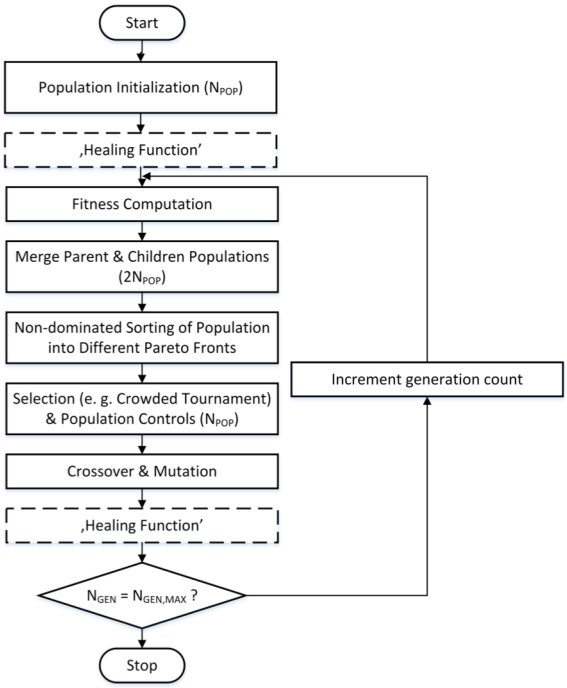

Figure 8. The algorithm of the genetic algorithm (the steps in dashed box are only implemented in Algorithm 5).

dominated front (F_1) are chosen for the next generation (P_{t+1}) (if the size of F_1 is smaller than N). This selection for the next generation is continued until the number of members from F_1 to F_i is larger than N. In these cases F_i is sorted based on the crowded-comparison operator and the best solutions are chosen to fill the empty slots of the new population. The NSGA-II procedure is illustrated in *Fig.6*.

2.5. Description of the Proposed Algorithm

The developed algorithm therefore needs to solve effectively the CAMD tasks based on the needs of the industrial and research work. By taking into consideration these needs, the optimal properties of the desired molecules can be defined, and this chemical information is essential for the defining of input parameters for the genetic algorithm. This schematic algorithm of the design process can be seen in *Fig.7*.

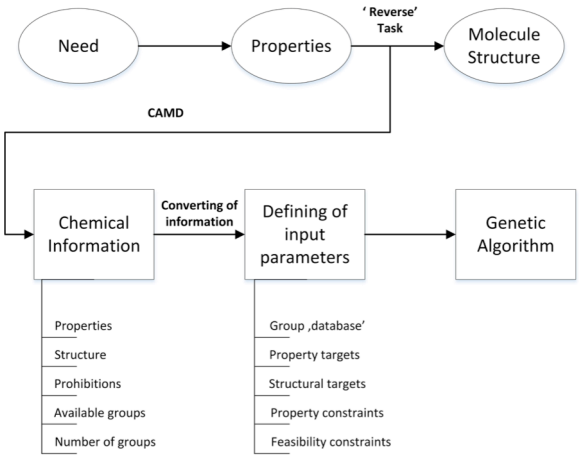

Figure 7. The algorithm of the design process.

The standard genetic algorithm was modified according to the task of the design of molecules. The chemical information is converted into input parameters. Thus, the type, minimum and a maximum number of available groups, the property and structural targets (e.g. acyclic, monocyclic, bicyclic structures), and the property and feasibility constraints (octet rule and a rule for branching) are defined. The algorithm of NSGA-II can be seen in *Fig.8*.

The efficiency of five different types of algorithm was tested in the present work. All the algorithms were implemented in MATLAB.

2.5.1. The 'base case' Algorithm 1

Only the type, the minimum and a maximum number of available groups, the target properties and the property constraints are defined. No structural targets or feasibility evaluations are implemented.

2.5.2. The 'octet rule' Algorithm 2

Besides the objectives and constraints of Algorithm 1, the octet rule is defined as target parameter. The octet rule is described in *Eq.(20)*,

$$\sum_i (2 - v_i) x_i = 2m \qquad (20)$$

where x_i, v_i are the number and valency, respectively, of groups of type i and m = 1, 0 or -1 for acyclic, monocyclic and bicyclic groups, respectively [28]. The valency parameter in this context means the number of available bonds on a group (thus the valency of double bonds counts as a single valency in this context).

2.5.3. The 'octet rule as a soft constraint' Algorithm 3

Besides the structure of Algorithm 2, the algorithm contains the octet rule (*Eq.(20)*) as a soft constraint to aspire the program to reach a feasible structure according to the octet rule. The soft constraint was defined as a curve similar to a reverse Gaussian distribution according to *Eq.(21)*, where $Res.Oct.$ is the result of the octet rule reordered to give 0 when the constraint is satisfied.

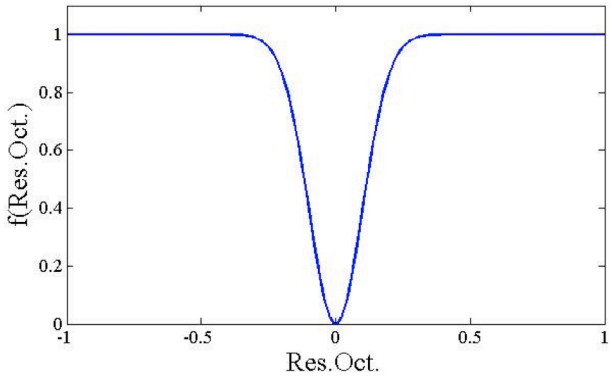

Figure 9. The curve of the used soft constraint.

$$f(Res.Oct.) = w \cdot (1 - e^{-\frac{(Res.Oct.)^2}{2\sigma^2}}) \qquad (21)$$

As can be seen in *Fig.9*, the curve of this function has a descending value near the $Res.Oct. = 0$ value, and is equal to 0 at exactly $x = 0$. The parameter w shows the 'weight', the constant value far from $Res.Oct. = 0$, and σ stands for the 'width' (thus the 'sharpness') of the function. A sharp function type ($w = 1$, $\sigma = 0.1$) was chosen as structural feasibility is not satisfied with mild ones.

2.5.4. The 'octet and branching rule' Algorithm 4

Only the octet rule cannot describe the structural feasibility. Two adjacent groups cannot be linked by more than one bond. The valency parameter in this context still means the number of available bonds on a group, as defined in the description of Algorithm 2. Given x_j groups of type j with valency v_j, a total of $x_j(v_j - 2) + 2$ attachments are available for bonding [28].

$$\sum_{i \neq j} x_i \geq x_j(v_j - 2) + 2 \qquad (22)$$

$$\sum_{i} x_i \geq x_j(v_j - 1) + 2 \qquad (23)$$

2.5.5. The 'healing function' Algorithm 5

The genetic algorithm is further improved with the introduction of a novel genetic operator called the healing function. The function is implemented in the algorithm at two points, after the initialization of the first population and after the selection, mutation and crossover steps are conducted in each generation. This function serves the population entities to be feasible with the help of the branching rule described in *Eq.(23)*. If the equation is not fulfilled then the difference from the optimal value shows the bonds needed in the molecule to reach the feasible structure. Using this value, the algorithm chooses as many available groups with one bond as needed for feasibility and completes the molecule with these groups. During the definition of the input parameters the number of available groups with one bond in the molecule is increased to ensure the availability of these groups. The number of branches and the groups with 3 or more bonds, must be significantly less than the ones with one bond available, e.g. 4 and 20 respectively, thus the termination of every chain in the molecule is ensured.

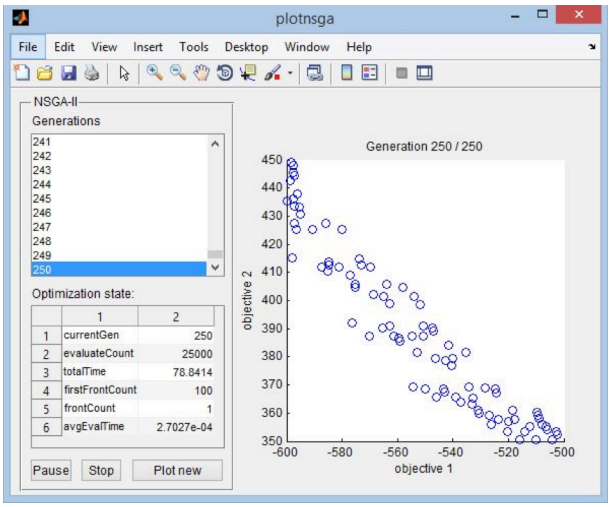

Figure 10. The plot window of NSGA-II.

2.6. The Evaluation of the Results

As the purpose of the current work is the development of an effective algorithm for the design of molecules obtaining target properties, the comparison of the results is an essential task to check the improvement. This efficiency inspection is carried out over two steps.

First a visual evaluation was carried out as the MATLAB implementation of the NSGA-II algorithm plots every generation as the script runs. The plot window of NSGA-II can be seen in *Fig.10*. Then the number of solutions is counted and, as the feasible structure is not ensured in every algorithm, the feasible ones based on the octet and branching rules are counted as well. The last generation in which the population of the Pareto front was changed was also determined.

3. Results and Discussion

The effect of the various user determined input parameters was examined and presented in several benchmark problems for the identification of different chemicals having the desired physical and chemical properties, as estimated by the multi-dimensional property model. The genetic algorithm worked with 250 generations containing 100 population members each. The effectivity of the different algorithms is compared through these design tasks.

3.1. Checking the Effectivity through the Search for Predefined Molecules

First the effectivity and applicability of the different algorithms were tested in terms of the search for different, predefined molecules. The properties of simple molecules were calculated *via* the Joback method and these values were set as targets to avoid the inaccuracy of the estimation method. The property constraints were set around these target values as given

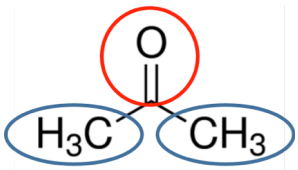

Figure 11. The structure of acetone fragmented according to the groups of the Joback method.

Table 1. The boiling and melting points of acetone.

	Experimental values	Estimated values
T_m [K]	178.25	173.50
T_b [K]	329.45	321.91

Table 2. The results of the search for acetone.

Algorithm	1	2	3	4	5
Solutions	2	2	1	1	2
(feasible)	(1)	(1)	(0)	(1)	(2)
$N_{gen,change}$	95	10	8	9	127
acetone	yes	yes	-	yes	-

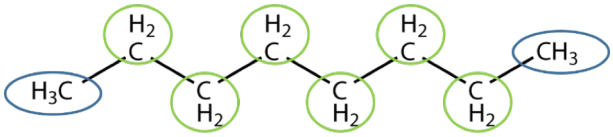

Figure 12. The structure of *n*-octane fragmented according to the groups of the Joback method.

Table 3. The boiling and melting points of octane.

	Experimental values	Estimated values
T_m [K]	216	179.92
T_b [K]	398	382.44

Table 4. The results of the search for *n*-octane.

Algorithm	1	2	3	4	5
Solutions	10	5	1	1	5
(feasible)	(2)	(1)	(0)	(1)	(5)
$N_{gen,change}$	191	58	17	207	47
n-octane	-	yes	-	yes	-

below. The search for two simple molecules, acetone, as its structure is quite simple, and *n*-octane were carried out to validate the algorithms.

3.1.1. Acetone

The structure of acetone can be represented by two methyl (blue circles) and one ketone (red circle) groups using the groups of the Joback method as can be seen in *Fig. 11*. The experimental and estimated boiling and melting points are presented in *Table 1*.

The input parameters for the design task were as follows:

Available groups:
$-CH_3$, $-CH_2-$, >CH-, >C<, -F and >C=O

Number of available groups:
0–4 for all the available groups
(in the case of Algorithm 5, the $-CH_3$, and -F numbers were set to 20 to allow healing)

Target properties:
T_m = 173.50 K, T_b = 321.91 K (estimated values)

Property constraints:
150 K < T_m < 200 K, 300 K < T_b < 350 K

Target molecule structure:
acyclic.

The results of different algorithms are presented in *Table 2*. Next to the number of solutions, the number of feasible solutions is presented in brackets and the value $N_{gen,change}$ stands for the last generation that changed the population of the Pareto front. The last row of the table shows if acetone is among the resultant structures.

Algorithm 5 seemed to be less effective in the light of the unsuccessful search for the structure of acetone, but if we consider that this algorithm could use up to 20 pieces of $-CH_3$ and -F groups, we can understand that the search place is significantly bigger than in the case of Algorithms 1-4.

3.1.2. *n*-Octane

The structure of *n*-octane can be represented by 2 methyl (blue circles) and 6 methylene (green circles) groups using the groups of the Joback method as can be seen in *Fig.12*. The experimental and estimated boiling and melting points are presented in *Table 3*.

The input parameters for the design task were as follows:

Available groups:
$-CH_3$, $-CH_2-$, >CH-, >C<, -F and >C=O

Number of available groups:
0–6 for all the available groups
(in the case of Algorithm 5, the $-CH_3$, and -F numbers were set to 20 to allow healing)

Target properties:
T_m = 179.92 K, T_b = 382.44 K (estimated values)

Property constraints:
150 K < T_m < 250 K, 350 K < T_b < 450 K

Target molecule structure:
acyclic.

The results of different algorithms are presented in *Table 4*, where the last row shows if *n*-octane is among the resultant structures. As in the case of the search for acetone, Algorithm 5 had a significantly bigger search place than Algorithms 1-4; thus, it could not find the structure of *n*-octane.

Table 5. The results of Case Study 1.

Algorithm	1	2	3	4	5
Solutions (feasible)	89 (14)	88 (14)	1 (0)	2 (2)	42 (42)
$N_{gen,change}$	250	250	219	186	250

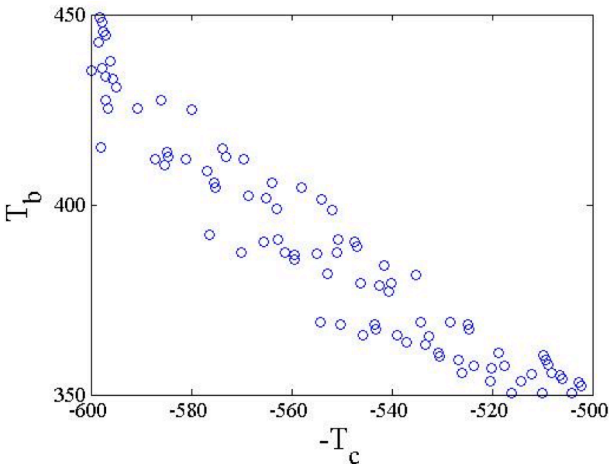

Figure 13. The Pareto Front of Algorithm 1 in Case Study 1.

3.2. Examples for Testing the Algorithms

In the following examples the effectivity of the proposed algorithms is tested via two benchmark problems. As concurring minimum and maximum search objectives are set as target variables, a Pareto front can be obtained, which represents the applicability of the genetic algorithms to solve CAMD tasks.

3.2.1. Case Study 1.

The input parameters for the design task were as follows:

Available groups:
$-CH_3$, $-CH_2-$, >CH-, >C<, -F and >C=O

Number of available groups:
0–4 for all the available groups
(in the case of Algorithm 5, the $-CH_3$ and -F numbers were set to 20 to allow healing)

Target properties:
$\max(T_c)$, $\min(T_b, T_m)$

Property constraints:
$500\ K < T_c < 600\ K$
$350\ K < T_b < 450\ K$
$100\ K < T_m < 200\ K$

Target molecule structure:
acyclic.

The results of Case Study 1 can be seen in *Table 5*. As the number of the last generation which changed the composition of the population is relatively high (250 is the maximum number of generations), the algorithms seem to evolve effectively towards the Pareto front. In the case of Algorithms 1 and 2, many of the found solutions proved to be infeasible, as expected since no feasibility constraints were involved in Algorithm 1 and no constraint for branching was available in Algorithm 2. The soft constraint of Algorithm 3 seemed to be ineffective. Algorithm 4 found only 2 feasible solutions, but these solutions were all feasible ones and they were close to the target properties. Algorithm 5 was very effective, although the search place was significantly bigger than in the case of Algorithms 1-4. The Pareto front of Algorithm 1 can be seen in *Figure 13*.

Table 6. The results of Case Study 2.

Algorithm	1	2	3	4	5
Solutions (feasible)	75 (0)	95 (5)	2 (0)	2 (2)	74 (74)
$N_{gen,change}$	250	250	223	197	250

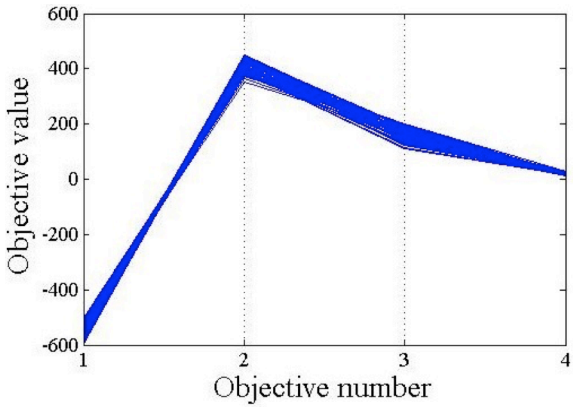

Figure 14. The results of Algorithm 5 in Case Study 2. Candidate molecules are represented by blue lines.

3.2.2. Case Study 2.

The input parameters for the design task were as follows:

Available groups:
$-CH_3$, $-CH_2-$, >CH-, >C<, -F, -Cl, -Br, -I, -OH, >C=O

Number of available groups:
0–6 for all the available groups
(in the case of Algorithm 5, the $-CH_3$, -F, -Cl, -Br, -I, and -OH numbers were set to 20 to allow healing)

Target properties:
$\max(T_c)$, $\min(T_b, T_m, P_c)$

Property constraints:
$500\ K < T_c < 600\ K$
$350\ K < T_b < 450\ K$
$100\ K < T_m < 200\ K$
$10\ bar < P_c < 30\ bar$

Target molecule structure:
acyclic.

In the case of Case Study 2 (*Table 6*), a wider searching range was available for the design task: 10 types of available groups with a maximum of six pieces of each (except for Algorithm 5, see the targets and constraints) and another target property was set, the minimization of critical pressure between the property constraints. The results were similar to Case Study 1, Algorithms 1 and 2 fundnd several infeasible results with the appropriate properties; Algorithm 3 seemed to be ineffective, Algorithms 4 and 5 provided reliable results, although Algorithm 5 still found more results according to the more pieces of available chain-terminating groups (groups with 1 valency). As can be seen from the results the improvement of the algorithms increases the number of feasible solutions significantly. The results of Algorithm 5 in Case Study 2 can be seen in *Figure 14*.

4. Conclusion

The design of molecules with specified properties has an increasing importance in the modern chemical industry. We proposed a multi-objective evolutionary optimization-based approach to take into account several objectives and constraints (e.g. financial aspects, toxicity). The algorithms generate a set of molecules arranged in a Pareto front related to the conflicting design targets calculated by the Joback method. To get reliable molecular structures we defined soft constraints based on the octet rule. The branching of the molecules was also tested and a healing function was designed to provide reliable results. The application of the proposed algorithms can be useful in the industry.

In the future, we are going to improve the algorithms with problem-specific genetic operators. These modifications seem to be promising for the significant increase in the search efficiency.

SYMBOLS

$C^{\mathrm{R}}(K,N)$	selection of N groups from a set of K groups
G	the set of n groups from which the designed molecule can be composed
$n_{\max}$	the available groups of the specified design task
x	the number of appearances of the specified group
P	a specified property value
P_{est}	estimated property value
P_{exp}	experimental property value
P_{lb}	the lower boundary of the specified property value
P_{ub}	the upper boundary of the specified property value
l_{ll}	the lower limit for the number of appearances of the specified group
l_{ul}	the upper limit for the number of appearances of the specified group
$x_1, \ldots, x_n$	the number of group type #1, …, #*n*, resp.
a_i^j	is the group contribution value of group *i* for property *j*
T_{b}	normal boiling point
T_{m}	normal melting point
T_{c}	critical temperature
P_{c}	critical pressure
V_{c}	critical volume
H^0	heat of formation
G^0	Gibbs free energy of formation
C_{P}	heat capacity
ΔH_{vap}	heat of vaporization
ΔH_{fus}	heat of fusion
η_{L}	liquid dynamic viscosity
ratio	a scalar between 0 and 1
rand	a generated random number
currGen	the number of the current generation
maxGen	the number of the maximal generation
scale	a scalar, that determines the standard deviation of the random number generated
shrink	scalar between 0 and 1. As the optimization progresses this parameter decreases the mutation range
R_t	A combined population (with 2*N* members)
P_t	parent population
Q_t	the population obtained by the use of crossover and mutation operators
F	the non-dominated front
v	the valency of the specified group. The valency parameter in this context means the number of available bonds on a group (thus the valency of double bonds counts as a single valency in this context).
m	the type of the designed molecule ($m = 1, 0$ or -1 for acyclic, monocyclic and bicyclic groups, respectively)
w	the 'weight' of the soft constraint curve (the constant value far from $x = 0$)
σ	the 'width' (the 'sharpness') of the soft constraint function curve

Acknowledgement

The research of J.A. has been supported by the National Research, Development and Innovation Office (NKFIH), project # OTKA 116674 entitled "Process mining and deep learning in the natural sciences and process development". G.D. was supported by the "A Pannon Egyetem tudományos műhelyeinek támogatása" TÁMOP-4.2.2.B-15/1/KONV-2015-0004 project.

REFERENCES

[1] Camarda, K.V.; Maranas, C.D.: Optimization in polymer design using connectivity indices, *Ind. Engng. Chem. Res.*, 1999 **38**(5), 1884–1892 DOI: 10.1021/ie980682n

[2] Kasat, R.B.; Ray, A.K.; Gupta, S.K.: Applications of genetic algorithms in polymer science and engineering, *Mat. Manufact. Proc.*, 2003 **18**(3), 523–532 DOI: 10.1081/AMP-120022026

[3] Perdomo, F.A.; Perdomo, L.; Millán, B.M.; Aragón, J.L.: Design and improvement of biodiesel fuel blends by optimization of their molecular structures and compositions, *Chem. Engng. Res. Design*, 2014 **92**(8), 1482–1494 DOI: 10.1016/j.cherd.2014.02.011

[4] Joback, K.G.: Computer aided molecular design (CAMD): Designing better chemical products. (Molecular Knowledge Systems Inc., Bedford, NH U.S.A.) 1998-2016 www.molecularknowledge.com

[5] Schneider, G.; Hartenfeller, M.; Reutlinger, M.; Tanrikulu, Y.; Proschak, E.; Schneider, P.: Voyages to the (un)known: Adaptive design of bioactive compounds, *Trends Biotechn.*, 2009 **27**(1), 18–26 DOI: 10.1016/j.tibtech.2008.09.005

[6] Gani, R.; Achenie, L.E.K.; Venkatasubramanian, V.: Introduction to CAMD in computer aided chemical engineering (Eds.: Luke, R.G.; Achenie, L.E.K.; Venkat, V.: Elsevier, Amsterdam, The Netherlands), 2002 Chapter 1, pp. 3–21 DOI: 10.1016/S1570-7946(03)80003-2

[7] Gani, R.; Jiménez-González, C.; Constable, D.J.C.: Method for selection of solvents for promotion of organic reactions, *Comp. Chem. Engng.*, 2005 **29**(7), 1661–1676 DOI: 10.1016/j.compchemeng.2005.02.021

[8] Camarda, K.V.; Bonnell, B. W., Maranas, C. D., Nagarajan, R.: Design of surfactant solutions with optimal macroscopic properties, *Comp. Chem. Engng.*, 1999 **23**(Supplement), S467–S470 DOI: 10.1016/S0098-1354(99)80115-X

[9] Sahinidis, N.V.; Tawarmalani, M.; Yu, M.: Design of alternative refrigerants *via* global optimization, *AIChE J.*, 2003 **49**(7), 1761–1775 DOI: 10.1002/aic.690490714

[10] McLeese, S.E.; Eslick, J.C.; Hoffmann, N.J.; Scurto, A.M.; Camarda, K.V.: Design of ionic liquids *via* computational molecular design, *Comp. Chem. Engng.*, 2010 **34**(9), 1476–1480 DOI: 10.1016/j.compchemeng.2010.02.017

[11] Gani, R.: Computer-aided methods and tools for chemical product design, *Chem. Engng. Res. Design*, 2004 **82**(11), 1494–1504 DOI: 10.1205/cerd.82.11.1494.52032

[12] Holenda, B.; Holenda, B.; Dallos, A.; Nagy, Á.; Friedler, F.; Fan, L.-T.: A combinatorial approach for generating environmentally benign solvents and separation agents, *Chem. Eng. Trans., Ser.*, 2003 **3**, 871–875

[13] Klamt, A.: Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena, *J. Phys. Chem.*, 1995 **99**(7), 2224–2235 DOI: 10.1021/j100007a062

[14] Friedler, F.; Fan, L.T.; Kalotai, L.; Dallos, A.: A combinatorial approach for generating candidate molecules with desired properties based on group contribution, *Comp. Chem. Engng.*, 1998 **22**(6), 809–817 DOI:10.1016/S0098-1354(97)00253-6

[15] Lin, B.; Chavali, S.; Camarda, K.; Miller, D.C.: Computer-aided molecular design using Tabu search, *Comp. Chem. Engng.*, 2005 **29**(2), 337–347 DOI: 10.1016/j.compchemeng.2004.10.008

[16] Soto, A.J.; Cecchini, R.L.; Vazquez, G.E.; Ponzoni, I.: Multi-objective feature selection in QSAR using a machine learning approach, *QSAR & Comb. Sci.*, 2009 **28**(11–12), 1509–1523 DOI: 10.1002/qsar.200960053

[17] Hii, C.E.A.: Evolving toxicity models using multigene symbolic regression and multiple objectives, *Int. J. Mach. Learn. Comp.*, 2011 **1**, 30–35 DOI: 10.7763/IJMLC.2011.V1.5

[18] Manoharan, P.E.A.: Rationalizing fragment-based drug discovery for BACE1: Insights from FB-QSAR, FB-QSSR, multi-objective-QSPR, and MIF studies, *J. Comput. Aided Mol. Des.,* 2010 **24**, 843–864 DOI: 10.1007/s10822-010-9378-9

[19] Herring III, R.H.; Eden, M.R.: Evolutionary algorithm for *de novo* molecular design with multi-dimensional constraints, *Comp. Chem. Engng.*, 2015 **83**, 267–277 DOI: 10.1016/j.compchemeng.2015.06.012

[20] Weber, L.: Evolutionary combinatorial chemistry: application of genetic algorithms, *Drug Discovery Today*, 1998 **3**(8), 379–385 DOI: 10.1016/S1359-6446(98)01219-7

[21] Venkatasubramanian, V.; Sundaram, A.; Chan, K.; Caruthers, J.M.: Computer-aided molecular design using neural networks and genetic algorithms. Genetic algorithms in molecular modeling (Ed.: Devillers, J.: Academic Press, London, UK) 1996 DOI: 10.1016/B978-012213810-2/50012-8

[22] Nicolaou, C.A.; Brown, N.: Multi-objective optimization methods in drug design, *Drug Discovery Today: Technologies*, 2013 **10**(3), e427–e435 DOI: 10.1016/j.ddtec.2013.02.001

[23] Joback, K.G.; Reid, R.C.: Estimation of pure-component properties from group-contributions, *Chem. Engng. Commun.*, 1987 **57**(1–6), 233–243 DOI: 10.1080/00986448708960487

[24] Shin Hyo Bang, S.J.L.; Taeseon Y.: Boiling point estimation program especially for aromatic compounds supplementing the Joback method, *Int. J. Chem. Engng. Appl.*, 2014 **5**(4), 331–334 DOI: 10.7763/IJCEA.2014.V5.404

[25] Deb, K.; Agrawal, S.; Pratap, A.; Meyarivan, T.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, in Parallel problem solving from nature PPSN (Eds.: Schoenauer VI, M.: Deb, K.; Rudolph, G.; Yao, X.; Lutton, E.; Merelo, J. J.; Schwefel, H.-P., Springer, Berlin, Germany) 2000 pp. 849–858 DOI: 10.1007/3-540-45356-3_83

[26] Deb, K.: Multi-objective genetic algorithms: problem difficulties and construction of test problems, *Evolut. Comp.,* 1999 **7**, 205–230 DOI: 10.17877/DE290R-5636

[27] Song, L.: NGPM - A NSGA-II Program in MATLAB, User Manual, 2011 www.mathworks.com/matlabcentral/fileexchange/31166-ngpm-a-nsga-ii-program-in-matlab-v1-4

[28] Odele, O.; Macchietto, S.: Computer aided molecular design: A novel method for optimal solvent selection, *Fluid Phase Equil.*, 1993 **82**, 47–54 DOI: 10.1016/0378-3812(93)87127-M

Degree programmes in English at the Faculty of Engineering

Graduate Programmes

MSc in Environmental Engineering
Duration: 4 semesters
Number of credits: 120
Tuition fee: 4500 USD/semester
Head of the programme: Dr Endre Domokos, Institute of Environmental Engineering
Contact: domokose@uni-pannon.hu

MSc in Environmental Sciences
Duration: 4 semesters
Number of credits: 120
Tuition fee: 4500 USD/semester
Head of the programme: Prof. Judit Padisak, Institute of Environmental Sciences
Contact: padisak@almos.uni-pannon.hu

BSc and MSc in Chemical Engineering
Duration: BSc 7 semesters, MSc 4 semesters
Number of credits: BSc: 210 credits MSc: 120 credit
Tuition fee: 4500 USD/semester
Head of the programme: Dr Sandor Nemeth, Institute of Chemical and Process Engineering
Contact: nemeth@fmt.uni-pannon.hu

Postgraduate Programmes

Hydrocarbon Technology Development Engineer
Duration: 2 semesters
Number of credits: 60
Tuition fee: 2000 USD/semester
Head of the programme: Dr Zoltan Varga, Institute of Chemical and Process Engineering
Contact: vargaz@almos.uni-pannon.hu

Water and Wastewater Treatment System Operation
Duration: 2 semesters
Number of credits: 60
Tuition fee: 2000 USD/semester
Head of the programme: Dr Rita Szakacsne-Foldes, Institute of Environmental Sciences
Contact: foldenyi@almos.uni-pannon.hu

DE GRUYTER OPEN

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
Vol. 44(1) pp. 51–54 (2016)
hjic.mk.uni-pannon.hu
DOI: 10.1515/hjic-2016-0006

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

SELECTIVE HYDROGEN SULPHIDE REMOVAL FROM ACID GAS BY ALKALI CHEMISORPTION IN A JET REACTOR

JANKA BOBEK,* DÓRA RIPPEL-PETHŐ, ÉVA MOLNÁR, AND RÓBERT BOCSI

Department of Chemical Engineering Science, University of Pannonia, Egyetem str. 10, Veszprém, 8200, HUNGARY

Natural gas is a primary energy source that contains a number of light paraffins. It also contains several undesirable components, such as water, ammonia, hydrogen sulphide, etc. In our study, a selective hydrogen sulphide removal process was achieved by alkali chemisorption in a custom-designed jet reactor. Several model gas compositions (CO_2-H_2S-N_2) were evaluated to find parameters that enable H_2S absorption instead of CO_2. The negative effect of the presence of CO_2 in the raw gas on the efficiency of H_2S removal was observed. The beneficial effect of the low residence time (less than 1 s) on the efficiency of H_2S removal was recognized. Optimal operational parameters were defined to reach at least a 50% efficiency of H_2S removal and minimal alkali consumption.

Keywords: acid gas, H_2S selective removal, CO_2, competition with H_2S, chemisorption

1. Introduction

Natural gas is one of our primary energy sources, which contains mainly methane. However, it us comprised of several undesirable components like carbon dioxide (CO_2), hydrogen sulphide (H_2S), ammonia (NH_3), water (H_2O), etc. [1]. *Table 1* shows a typical composition of natural gas [2]; however, the content significantly depends on locality. In most cases, natural gas contains H_2S in various quantities between 10 to 20,000 ppm [1]. The gases with a measurable amount of H_2S are called sour gases. The acid gases are defined as gases containing some acidic component such as CO_2 or H_2S [3].

The H_2S containing hydrocarbon gases causes problems during the delivery, processing, and storage. H_2S is converted into SO_2 during combustion, which poses a health hazard and causes acid rain, smog. In the presence of water, acid components cause corrosion in pipelines and containers. Consequently, H_2S removal from natural gas is absolutely necessary [3].

There are several methods for reducing the H_2S content of natural gas. Membrane techniques also exist, but the adsorption and absorption processes are the most widespread. In the adsorption process, the fixed bed construction is the most common. The adsorber is usually filled with metal ions (iron, copper, zinc, cobalt, etc.) and an impregnated solid host (zeolite, activated-carbon, etc.). The disadvantage of this technique is the huge energy demand of adsorber regeneration. In the absorption process, one of the main points is the high pH value of the medium due to H_2S dissociation. There are numerous solvents for absorbing H_2S, namely alcanol-amines (MEA, DEA, DIPA, TEA, MDEA, etc.), alkali-hydroxides (KOH, NaOH), water, and ammonia. The alcanol-amines and the alkali-hydroxides are the most efficient. The alcanol-amines are widely used in H_2S removal, but their selectivity can be problematic and foaming appears during the process [4]. The use of alkali-hydroxides seems to be the most efficient process. By choosing the correct parameters, such as residence time, pH, solvent concentration, and intake, the procedure can be H_2S selective. In an alkali-hydroxide medium competitive chemisorption takes place between CO_2 and H_2S. Although CO_2 is a stronger acid than H_2S, it is a slower adsorber, thus H_2S absorption can be achieved over a short residence time. Intensive phase connection and fast phase separation afterwards are essential steps to facilitate a H_2S selective process [2]. The spray technique is a widespread method for the intensification of the reaction between the reactants. The pneumatic nozzles act as two-phase sprayers, because the gas at high speed breaks up the liquid into little droplets [5].

Table 1. A typical composition of natural gas [2].

component	concentration (%, m^3/m^3)
methane (CH_4)	97
nitrogen (N_2)	0.936
ethane (C_2H_6)	0.919
carbon dioxide (CO_2)	0.527
propane (C_3H_8)	0.363
butane (C_4H_{10})	0.162
oxygen (O_2)	0-0.800
noble gases (Ar, He, Ne)	trace
other (e.g. H_2S)	0-0.001

*Correspondence: bobekj@almos.uni-pannon.hu

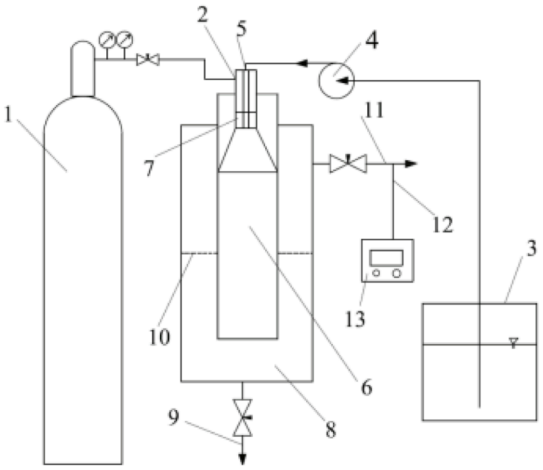

Figure 1. Experimental device equipped with a 1. gas cylinder, 2. gas inlet, 3. alkali vessel, 4. chemical feeder pump, 5. alkali inlet, 6. reactor space, 7. nozzle, 8. separation space, 9. wastewater removal, 10. drop catcher, 11. outlet of purified gas, 12. gas sampling, and 13. gas analyzer.

2. Experimental

The aim of our research is selective hydrogen sulphide removal from model gases that also contain CO_2. Our goal is to achieve the highest H_2S removal efficiency with the lowest alkali specificity as defined by the ratio of NaOH and H_2S expressed in moles. To find the parameters that support H_2S removal several experiments were carried out in a custom-designed jet reactor *(Fig.1)*. Owing to the construction of the reactor, the gas pressure, gas flow, alkali inlet flow, and alkali concentration were variable. All experiments were carried out at 30 bar total pressure. The absorbent was an aqueous NaOH solution of different concentrations, such as 0.5, 1.5, and 2.5% (g/g).

The model gas mixtures *(Table 2)* were produced in an acid-proof gas mixing bridge. For the first set of samples the H_2S content of the model gas mixtures was kept approximately constant; thus, the effect of CO_2 could be studied. For the last three samples, the CO_2 content was kept approximately constant; thus, the sensitivity of the process with regards to the variation of H_2S concentration could be investigated.

3. Results and Analysis

First, the effect of NaOH concentration, NaOH inlet flow, gas flow (residence time), and CO_2 concentration were investigated on the efficiency of H_2S removal.

3.1. Effect of Residence Time

To observe the effect of residence time on the efficiency of H_2S removal, the gas flow rate as a single parameter was varied. By increasing the gas flow rate, the residence time decreased. The gas flow rates were 3.9, 3.2, 2.4, and 1.6 N m^3 h^{-1}, which correspond to residence time rates of 0.05, 0.06, 0.09, and 0.13 s, respectively. *Fig.2* shows the effect of decreasing residence time. At a constant specific alkali value, the efficiency of H_2S removal increased as a result of a decrease in residence time. Furthermore, *Fig.2* also shows that the alkali specificity values decreased by raising the gas flow rate under a constant efficiency of H_2S removal.

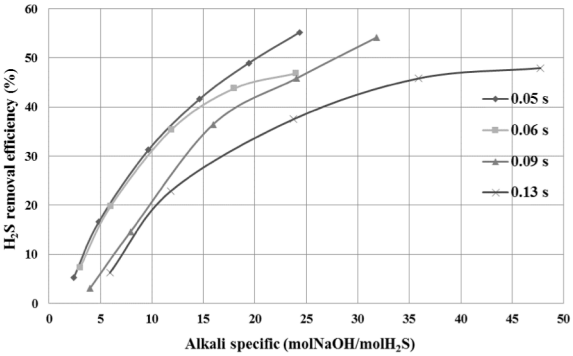

Figure 2. Effect of different residence times on the efficiency of H_2S removal (gas mixture 4, 30 bar, 2.5% (g/g) NaOH).

Table 2. Composition of the tested model gas mixture samples.

samples	CO_2 % (m^3/m^3)	H_2S ppmv	N_2 % (m^3/m^3)
1	0	100	99.999
2	23	90	76.999
3	41	80	58.999
4	60	80	39.999
5	76	85	23.999
6	72	520	27.999

3.2. Effect of NaOH Concentration

The value of alkali specificity depends on the H_2S content of the raw gas, the concentration and the flow rate of the absorbent. By increasing the concentration and the flow rate of the absorbent, the efficiency of H_2S removal is increased. However, the efficiency could not be improved after a point by the absorbent concentration or flow rate, because the efficiency reached a nearly constant value while the alkali specificity continued to increase *(Fig.3)*.

3.3. Effect of CO_2 Concentration

Model gases of different CO_2 concentrations were used to study the effect of CO_2 concentration on the efficiency of H_2S removal. The difference in H_2S concentrations of model gases is a result of non-exact gas mixing, but this does not affect the comparability of the results. *Fig.4* shows that the efficiency of H_2S removal is decreased by increasing CO_2 content. The competition between H_2S and CO_2 in alkali absorbents is documented.

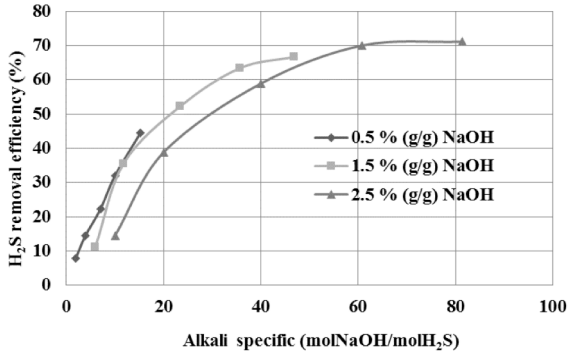

Figure 3. Effect of different NaOH concentrations on the efficiency of H_2S removal (gas mixture 2, 30 bar, 0.2 s residence time).

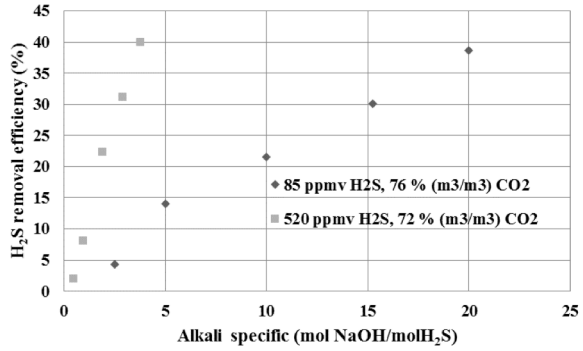

Figure 5. Effect of different H_2S concentrations on the efficiency of H_2S removal (gas mixtures 5-6, 30 bar, 0.09 s residence time, 1.5% (g/g) NaOH).

3.4. Effect of H_2S Concentration

The influence of H_2S concentration on the efficiency of H_2S removal was investigated under a nearly constant CO_2 level (76 and 72% (m^3/m^3)) and greatly differing H_2S (85 and 520 ppmv) containing model gases. When the 85 ppmv H_2S containing gas was compared to the 520 ppmv H_2S sample, the alkali specificity value measured was five times less *(Fig.5)*. On the other hand, *Fig.5* shows that the efficiency of H_2S removal does not depend on the H_2S concentration in this process. The alkali hydroxide absorbent technique shows little sensitivity to the changes in the H_2S content of the inlet gas.

3.5. Optimization of Operational Parameters

Based on the above-mentioned results, our aim was to find the optimal operational parameters for model gases of any composition in order to achieve an H_2S removal efficiency of at least 50%, while applying the minimal amount of alkali specificity. This efficiency of H_2S removal can be achieved by increasing the NaOH concentration. A low alkali specificity value can be achieved by adopting a low residence time.

As shown in *Table 3*, when the CO_2 content is below 50% (m^3/m^3), 1.5% (g/g) NaOH absorbent is enough to achieve an H_2S removal efficiency of 50% in the given type of reactor at a pressure of 30 bar. A gas flow rate of 2.5 $Nm^3\ h^{-1}$ with a 0.08 s residence time is needed.

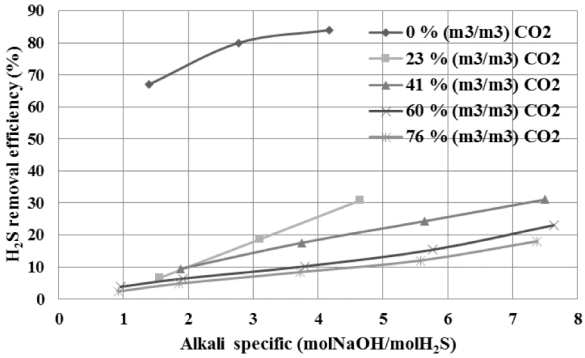

Figure 4. Effect of different CO_2 concentrations on the efficiency of H_2S removal (gas mixtures 0-5, 30 bar, 0.08 s residence time, 0.5% g/g NaOH).

When the CO_2 content is above 50% (m^3/m^3), 2.5% (g/g) NaOH is necessary to achieve a removal efficiency of 50%. The applied gas flow rate needs to be 3.8 $Nm^3\ h^{-1}$ corresponding to 0.05 s residence time in the given type of reactor at a pressure of 30 bar.

4. Discussion

In this study, model gases with different H_2S-CO_2-N_2 contents were investigated in a custom-designed jet reactor. Our aim was to achieve a H_2S removal efficiency of at least 50% with minimal alkali consumption.

The effect of the NaOH, CO_2, and H_2S concentrations, and the residence time on the efficiency of H_2S removal was studied. During our experiments CO_2 absorption was not investigated because the Dräger X-am 7000 analyser we used is only able to measure the CO_2 concentration in percent magnitude.

A positive effect of low residence time on H_2S removal was observed. By increasing the gas flow rate, the efficiency of H_2S removal was increased under constant alkali specificity. If the efficiency of H_2S removal is constant, the alkali specificity can be reduced by decreasing the residence time.

By increasing the NaOH concentration and flow rate, the efficiency of H_2S removal was improved until a point after which it nearly remained constant while the alkali specificity was still rising.

To study the effect of different CO_2 concentrations on the efficiency of H_2S removal, several CO_2 concentrations were investigated under nearly the same H_2S levels. The removal efficiency was reduced radically by increasing the CO_2 concentration.

When comparing the model gases that contain different H_2S concentrations, a reduction in the alkali specificity was observed. The alkali specificity value decreased as the H_2S content increased. The removal efficiency remained constant irrespective of the H_2S concentration of the model gases, which improves the efficiency of the alkali absorption process in terms of selective removal of H_2S.

Table 3. The best operational parameters of the tested model gases at a pressure of 30 bar.

CO_2 content, % (m^3/m^3)	NaOH concentration, % (g/g)	alkali specificity, mol NaOH (mol H_2S)$^{-1}$	H_2S removal efficiency, %	Residence time, s	Gas flow rate, $Nm^3\ h^{-1}$
	0.5	15	44	0.20	1.0
23	1.5	14	51	0.08	2.5
	2.5	19	51	0.10	2.0
	0.5	12	44	0.20	1.0
41	1.5	15	50	0.08	2.5
	2.5	20	50	0.10	2.0
	0.5	6	27	0.06	3.0
60	1.5	24	41	0.09	2.3
	2.5	24	55	0.05	3.8
	0.5	6	20	0.07	3.0
76	1.5	16	47	0.05	3.8
	2.5	22	56	0.05	3.8

We observed that when the CO_2 concentration was less than 50% (m^3/m^3), a 1.5% (g/g) NaOH concentration and 0.08 s residence time is necessary to achieve an H_2S removal efficiency of 50% at a pressure of 30 bar under the given experimental conditions. When the CO_2 concentration was above 50% (m^3/m^3), we found that this is sufficient to provide a NaOH concentration of 2.5% (g/g) over a residence time of 0.05 s at a pressure of 30 bar. Based on our experiments a high efficiency of H_2S selective removal can be achieved by NaOH absorption.

REFERENCES

[1] Balogh, K.: Sedimentology III (Akadémia Kiadó, Budapest, Hungary), 1992 (in Hungarian)

[2] Vágó, Á.; Rippel-Pethő, D.; Horváth, G.; Tóth, I.; Oláh, K.: Removal of hydrogen sulphide from natural gas, a motor vehicle fuel, *Hung. J. Ind. Chem.*, 2011, **39**(2) 283–287

[3] Wu, Y.; Caroll, J.J.; Zhu, W.: Sour gas and related technologies (Scrivener Publishing LLC, Beverly, MA, USA) 2012 pp. xiv–xvii

[4] Kohl, A.L.; Nielsen, R.B.: Gas purification (Gulf Publishing Company, Houston, TX, USA) 1997 pp. 40–466

[5] Tuba, J.: Carburators (Műszaki Könyvkiadó, Budapest, Hungary), 1976 pp. 23–24 (in Hungarian)

DE GRUYTER OPEN

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
Vol. 44(1) pp. 55–70 (2016)
hjic.mk.uni-pannon.hu
DOI: 10.1515/hjic-2016-0007

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

HYDROGEN SULPHIDE CORROSION OF CARBON AND STAINLESS STEEL ALLOYS IMMERSED IN MIXTURES OF RENEWABLE FUEL SOURCES AND TESTED UNDER CO-PROCESSING CONDITIONS

ANDRÁS GERGELY,* [1] ROLAND LOCSKAI,[1] PÉTER SZABÓ,[2] ANTAL KRÓJER,[3] AND TAMÁS KRISTÓF[1]

[1] Department of Physical Chemistry, Institute of Chemistry, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY
[2] Department of General and Inorganic Chemistry, Institute of Chemistry, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY
[3] Department of Inspection and Maintenance, MOL Co., Olajmunkás u. 2, Százhalombatta, 2443, HUNGARY

In accordance with modern regulations and directives, the use of renewable biomass materials as precursors for the production of fuels for transportation purposes is to be strictly followed. Even though, there are problems related to processing, storage and handling in wide range of subsequent uses, since there must be a limit to the ratio of biofuels mixed with mineral raw materials. As a key factor with regards to these biomass sources pose a great risk of causing multiple forms of corrosion both to metallic and non-metallic structural materials. To assess the degree of corrosion risk to a variety of engineering alloys like low-carbon and stainless steels widely used as structural metals, this work is dedicated to investigating corrosion rates of economically reasonable engineering steel alloys in mixtures of raw gas oil and renewable biomass fuel sources under typical co-processing conditions. To model a desulphurising refining process, corrosion tests were carried out with raw mineral gasoline and its mixture with used cooking oil and animal waste lard in relative quantities of 10% (g/g). Co-processing was simulated by batch-reactor laboratory experiments. Experiments were performed at temperatures between 200 and 300 ºC and a pressure in the gas phase of 90 bar containing 2% (m^3/m^3) hydrogen sulphide. The time span of individual tests were varied between 1 and 21 days so that we can conclude about changes in the reaction rates against time exposure of and extrapolate for longer periods of exposure. Initial and integral corrosion rates were defined by a weight loss method on standard size of coupons of all sorts of steel alloys. Corrosion rates of carbon steels indicated a linear increase with temperature and little variation with composition of the biomass fuel sources. Apparent activation energies over the first 24-hour period remained moderate, varying between 35.5 and 50.3 kJ mol^{-1}. Scales developed on carbon steels at higher temperatures were less susceptible to spall and detach. Nonetheless, moderate deceleration of corrosion rates as a function of time are due to the less coherent, frequently spalling and low compactness, higher porosity of the scales evolved at lower and higher temperatures, respectively. On the surface of high alloy steels, sulphide scales of an enhanced barrier nature formed during the induction periods and the layer formation mechanism was found to be assisted by the increasing temperature as initial reaction rates considerably decreased over time. Nevertheless, corrosion-related sulphide conversion of metals and mass loss of the high alloys are strongly affected by the composition of the biomass fuel sources especially animal waste lard. Thermal activation in the first 24 hours decreased from 68.9 to 35.2 kJ mol^{-1}. A greater degree of failure to develop protective sulphide scales was experienced by changing to composition of the biomass fuel sources than the impact of thermal activation between a narrow temperature range at around 100 ºC. In accordance with the literature, high free fatty acid contents lead to high corrosion rates accounted for direct corrosion of high alloy steels and assisted solubilisation of corrosion products. In addition, the pronounced acceleration of sulphide corrosion is connected to the diminishing inhibition effect of the sulphide scales.

Keywords: Hydrogen sulphide corrosion, gas phase corrosion, carbon and high alloy steels, renewable biofuels, used cooking oil, animal waste lard

1. Introduction

In an increasing power-hungry society coupled with the age of unprecedented global warming, the relevance of the substitution of fossil fuels from mineral sources and their mixtures with materials obtained from renewable biomass sources in the highest possible proportions is one of the most urgent priorities for current and future generations to observe. Sustainable biomass resources, vegetable oils, and waste lard are renewable feedstocks for partial substitution of mineral oils in production of biofuels. Although their availability is limited at present, there are clear benefits of biofuel production, namely meeting local consumption demands, reduced emissions of greenhouse gases during transportation over long distances in supply chains. In addition, the lower dependence on fossil fuels unequivocally contributes to increased national security. By 2010, the

*Correspondence: gergelyandras432@gmail.com

European Commission set a goal that EU biofuels must account for 5.75% of transportation fuels. Biodiesel as a mono-alkyl ester of vegetable oils and waste lard [1-3] is distinguished from second-generation type renewable feedstocks of inedible oils and algae. Using vegetable oils, composed of a majority of natural triacylglycerols in similar to waste lard, biodiesel production involves processing by either trans-esterification with methanol or ethanol besides production of glycerol or catalytic hydrocracking. The latter is an economically more feasible option given the absence of investment costs [4, 5] as no dedicated facility is required [6]. Cracking, pyrolysis, catalytic hydrodesulphurisation (HDS), deoxygenation [7], and the saturation of olefins and aromatics favour the production of ultra-low sulphur petrodiesel [8, 9]. The fuel derived from lipid feedstock of biological sources emulating the composition of petro-diesel is renewable diesel. Several crops such as rapeseed, soybean, palm, sunflower, etc. are excellent renewable oil sources recycled for energy storage and production lines, currently maize is the most frequently used feedstock for biofuel production primarily due to its high profitability. However, when the profit margin is a priority, used cooking oil trumps virgin vegetable oils, since it is around three times cheaper. As for the chemical stability aspects, fossil fuels containing methyl esters of fatty acids are less oxidation-resistant and more hygroscopic, so their long-term storage is not recommended. To process biomass in conventional oil refineries, vegetable oil and vacuum gas oil mixtures are preferred for quality diesel production [10, 11] as hydrotreating the latter along with heavy gas oil results in normal alkanes of C15-C18 [12]. Moderate reaction temperatures yield diesel, while higher temperatures lead to gasoline production. Technological parameters of hydroprocessing under mild conditions [13] generally mean temperatures of 320–430 ºC and pressures of 3.5–5.5 MPa with a relative quantity of the biomass source of ~5.5%. High isoparaffin content is obtained by catalytic hydrogenation of triglycerides stemming from waste lard [14]. Nevertheless, there are vegetable oils (tall oil) that have high fatty acid contents [15-19] similar to waste lard. If they are not processed properly, they will be corrosive in nature in subsequent applications. Their conversion to liquid hydrocarbons by catalytic hydrocracking seems a viable option. Nevertheless, aside from indispensable processes relating to biomass-based fuel production, all sorts of sources including biofuels are to be refined to decrease heteroatom content, such as sulphur, nitrogen and phosphorous, which might be potential sources of pollutants. To preserve safety and profitability one should consider possible corrosion risks of co-processing of renewable fuel sources to metallic structures from which equipment of existing petroleum refineries has been built and used for decades.

Industrial methods and technological means to protect oil production equipment were reviewed in terms of corrosion protection [20]. In relation to highly corrosive effects caused by waste lard, it depends on the quantity and composition of free fatty acids as a way of dissolution of metal-acid complexes coupled with reduction of a depolarisator. Some fatty acids, such as stearic, palmitic, myristic and lauric acid, are even used for heat storage systems and their corrosivity is enhanced by thermal cycling [21]. Acceptable resistance of the 304L stainless steel (1.4307) and an aluminium alloy with passive oxide layers on their surfaces was pointed out. The kinetics of naphthenic acid and sulphide corrosion and their interactions with Q235 and 5Cr1/2Mo steel alloys were investigated at 270 ºC in oil with naphthenic acid and sulphur contents of up to 1.0% (g/g) as synthetic refining media [22]. In the liquid phase with naphthenic acids at 230 ºC, the low alloy and carbon steels indicate nearly the same naphthenic corrosion rates but above 230 ºC the former gave higher rates. This is attributed to its higher activation energy (63 kJ mol^{-1}) compared to that of carbon steel (54 kJ mol^{-1}) leading to greater acceleration with increasing temperature. In dimethyl disulphide containing oil, sulphide corrosion follows parabolic kinetics. Film growth was faster on carbon steel than on low alloy steel. In oil containing naphthenic acids and dimethyl disulphide, low alloy steel showed lower corrosion rates than carbon steel; thus, sulphide corrosion proved to be rate-governed by pseudo-passive films. Different resistance of the alloys was connected to the chromium sulphide (Cr_5S_8) in the pseudo-passive film formed on low alloy steel, in contrast with pyrrhotite (Fe_7S_8) and troilite (FeS) films obtained on carbon steel. At oil refineries, especially in the HDS streamline part, the rate of metal loss either in the form of general or local corrosion is one of the highest. Thus, to achieve an optimal balance between production, inspection and maintenance careful planning is required.

Aside from works on corrosion risks of biofuels at low temperatures without the presence of hydrogen sulphide rather for the automotive than refining industry, to our best knowledge, there is a rather limited number of studies in the literature to date, giving insights into either the corrosion rates of engineering alloys or the integrity risk assessment of plant equipment regarding the HDS co-processing of renewable fuel sources such as used cooking oil and waste lard in low proportion mixtures with raw gas oils in refining plants originally designed to withstand mineral oils. Therefore, the aim of this work is to investigate corrosion rates of inexpensive engineering alloys, two types of carbon steels (1.0425 and St35.8) and a low alloy steel sample (1.7335), which are all basic materials of heat-resistant tube and pipe systems. In addition, for the purpose of making a comparison, as structural materials of absorbers, strippers, and overlays in reactors, three high alloy steel samples (1.4541, 1.4571, 1.4401) were selected for hydrogen sulphide corrosion, tested at a pressure of 90 bar and temperatures between 200 and 300 ºC. Under the aforementioned parameters, hydrogen sulphide and naphthenic acid corrosion of steel alloys are investigated focusing on the dependence of corrosion rates on temperature and time.

Table 1. Chemical compositions of autoclave experiment-tested steel alloys in weight percent (Fe in balance).

Alloys	C	Cu	Si	P	S	Mn	Cr	Ni	Mo	N	Ti
1.0425	≤0.20	≤0.30	≤0.40	≤0.025	≤0.015	≤0.80-1.40	≤0.30	≤0.30	≤0.08	≤0.012	≤0.03
St35.8			0.10-0.35	≤0.04	≤0.04	0.40-0.80					
1.7335	≤0.08-0.18	≤0.30	≤0.35	≤0.025	≤0.01	0.4-1.0	0.70-1.15		0.5		0.012
1.4541	≤0.06		≤0.75	≤0.04	≤0.015	≤2.0	17.0-19.0	9.0-11.0	-		5*C-0.70
1.4571	≤0.08		≤1.0	≤0.045	≤0.015	≤2.0	16.5-18.5	10.5-13.5	2.5		5*C≤0.70
1.4401	≤0.07		≤1.0	≤0.045	≤0.03	≤2.0	16.5-18.5	10.0-13.0	2.5	≤0.1	

2. Experimental

In processing technology before the hydrotreating reactor is heated to ~420 ºC, generally the high proportions of unsaturated and saturated free fatty acids cause the main corrosion problem, after that hydrogen sulphide is usually responsible for most of the corrosion risks. Therefore, experiments were planned to satisfy the goals of investigating the corrosion rates of engineering steel alloys representative of materials in HDS plants caused by the evolution of hydrogen sulphide during processes, combined with the increased risks exerted by the mixture of used cooking oils and animal fats that should all be co-processed in the same plant altogether under the same conditions.

Batch type experiments were carried out in two reactors with a wall thickness of 25 mm (manufactured by Lampart Vegyipari Gépgyár Zrt) having internal sizes of 2 and 4 litres. The temperature was measured inside the reactors through a stainless steel tube casing leading from the head into the core of the inner volume, with platinum100 resistive thermometers and adjusted with HAGA KD481DD controllers within ±6 ºC after auto-tuning refining with liquid- and gas-filled reactors at pressure targets. Heating of the reactors was performed with 6 resistance-heating wires, which were evenly distributed around the encasings wrapped in a 4 cm thick thermo-insulation layer of glass wool lined inside steel scaffolds. The apparent effective performance of the heating systems with power consumptions of 3.7 and 6 kW at 3*380 V and drew currents of 4.5 and 7.2 A to the small and large reactors, respectively. Heat ramp and cooling down phases of batch-type experiments were carried out as rapidly as possible especially in the cases of the 24- and 72-hour tests. Fast temperature ramp ups were achieved by providing the highest laboratory temperature and the close location of the nearby working reactor to utilise its dissipated heat (initial pressure was corrected for fill-ups). During cooling down phases at the end of corrosion tests, reactor encasings were removed from scaffolds. The inside pressure was slowly reduced by letting the gas slowly out of the reactor through pressure valves so that gas phase reactions were frozen within the shortest possible period. The time when the temperature decreased to 200 ºC and/or the gas pressure dropped to at least 70% of its initial target was regarded as the end time of reactor experiments. Short stainless steel tubes (outer- and inner-diameters of 6 and 4 mm) were employed for fill-ups and long ones for the evacuation of reactor gases. For safety reasons, used reactor gases still containing about 2% (m^3/m^3) hydrogen sulphide were led over several steps through Erlenmeyer flasks containing an aqueous solution of sodium bicarbonate with a moderately high pH to work as alkaline absorbers.

Coupons of steel alloys of a standard size (32 mm outer- and 10.5 mm inner-diameters with a thickness of 2 mm) were used for the experiments. All coupons were grated and polished with silicon carbide emery papers (#400 and #600) soaked in water to remove rust and stain spots. Then they were ultrasonicated twice for 5 minutes in isopropanol and acetone. The cleaning process obeyed and adapted principles of the ASTM G1 standard ('Standard practice for preparing, cleaning, and evaluating corrosion test specimens'). After drying, mass of the coupons was measured with an analytical balance. Compositions of the investigated steel alloys are summarised in *Table 1*. With proper engravings, coupons were assembled on glass rods, each with annular glass cylinders (8 mm long) for proper spacing. Then glass rods with steel alloys were inserted into large glass tubes holding the media of raw or refined gas oil and its mixture with used cooking oil (UCO in 10% (g/g)) or waste lard (WL in 10% (g/g)). The composition of gas oil and renewable sources, supplied by the Department of Analytics and the Institute of Biotechnology, and Bay Zoltán Non-profit Ltd. for Applied Research, are summarised in *Table 2*. Animal waste lard was one of the most corrosive types available in the region owing to its high free fatty acid content. All materials were supplied by MOL Co., Hungary. Glass tubes accommodated five coupons (parallel samples) four times (according to the four different chemical compositions of steel alloys). When integral corrosion rates were defined for a time period by weight loss measurements, the standard deviation was also assessed by the 5 parallel specimens. Error bars in figures based on gathered data are directly related to the standard deviation of population samples of the experiments obtained by five coupons in one autoclave test. The level of the fluids was set well above the upper coupons. To minimize loss of the fluids owing to evaporation especially around 300 ºC despite the high pressure, inner and outer glass caps were employed for each glass tube. Glass caps with an appropriate figure did not hinder efficient gas diffusion into sample holder glass tubes, ensuring good access to the fluids and steel coupons.

Table 2. Basic characteristics of the renewable biomass fuel sources and distillation data of raw gas oil.

Fuel source	Used cooking oil	Animal waste lard	Distillation data of raw gas oil Boiling point of fractions	Proportion of fractions
Organic acids (C1-C4, mg kg^{-1})			136.4	Initial
Citric acid	1.90	-	187.2	5
Lactic acid	32.69	-	197.3	10
Formic acid	8.46	-	214.1	20
Acetic acid	64.43	94.13	227.6	30
Isobutyric acid	0.00	30.83	241.8	40
Butyric acid	27.69	175.58	258.2	50
Organic acids (C8-C20, mg kg^{-1})			274.2	60
Caprylic acid	1	1.5	291.5	70
Capric acid	-	1	310.9	80
Lauric acid	2	1	331.9	90
Myristic acid	3	11	347.3	95
Palmitoleic acid	2.5	18	Recovery	96.7
Heptadecanoic acid	-	3.5	Residue	1.5
γ-linoleic acid	0.6	5	Loss	1.8
Arachidic acid	-	3.5	-	-
Gondoic acid	-	7.5	-	-
Nervonic acid	-	1	-	-
Palmitic acid	100	165	-	-
Stearic acid	25	95	-	-
Oleic acid	285	310	-	-
Linoleic acid	190	65	-	-
Extractable petroleum Hydrocarbons (C10-C40, mg kg^{-1})				
C36	1.15	0.30	-	-
C37	0.10	-	-	-
C38	0.15	0.24	-	-
C39	0.12	-	-	-
Iodine number (10^{-2} mg g^{-1})	96	65	-	-
Total acid number (KOH mg g^{-1})	6	74	-	-
Alcohol components (mg kg^{-1})				
Glycerol	-	7014.05	-	-
Methanol	-	100.33	-	-
Ethanol	11.12	6.35	-	-
Water content (mg kg^{-1})	611	1171	-	-
Metals & other elements as impurities				
Al	0.02	0.016	-	-
Fe	0.001	0.004	-	-
Na	0.005	0.740	-	-
P	-	0.021	-	-
S	-	0.038	-	-
K	-	0.021	-	-
Ca	-	0.036	-	-
Amino acids (in % g/g)	--	Less than 0.3	-	-

For the experiments, bottled hydrogen gas (2.5ip, with a purity of 99.5%) with an initial pressure of 150 bar (provided by Messer Hungarogáz Kft.) was used as received to fill up reactors at around room temperature (or the temperature when the experiment started). Initial fill-up pressures were calculated according to the pressure of hydrogen refining at the target temperature according to the p-V-T database [23]. There were pressure increments almost in all experiments, which were caused by the varying volatility of raw and refining gas oil phases. Hence real pressure of reactor gas phases was recorded during the experiments. The mixtures of gas oil and used cooking oil or animal fats were not so volatile as pure gas oil so pressure increases were less pronounced with those materials. At the time of filling up, the gas pressure inside the reactors was measured both by a reductor (200/200, supplied by Messer Hungarogáz Kft.) and pressure gauges (MERR-KFM PRLT S91, pressure measuring was calibrated before and after experiments) nested in reactor heads.

For practical and safety reasons, hydrogen sulphide of stoichiometric quantity was in situ produced from elemental sulphur. The quantity of hydrogen sulphide was calculated to be ~2 percent by volume in all autoclave experiments at the target pressure, but the overall pressure was greater because of the changing volatility of the organic fluid phases. Thus, the pressure in autoclave experiments was higher than the hydrogen

sulphide contents would have been lower with a maximum of 15%, but stoichiometric quantities were initially set with an excess of 10% in the experiments so the concentration of 2% (m^3/m^3) was considered to be assured. Before the series of autoclave experiments, blind tests were performed to validate the proportion of hydrogen sulphide of the gas phase (2 volume percent) after an induction time of one hour at a temperature of 250 ºC and at least 40 bar of pressure. The reactor gas was diluted by 100 times with a nitrogen stream then the mixture was entrained into a Draeger X-am 5600 H_2S detector to validate the hydrogen sulphide contents of the reactors.

After continuous decompression at the end of the autoclave experiments, reactor heads were removed then glass tube holders (made of stainless steel) with the fluid and steel samples were gathered. Depending on the kind of alloy, the heavy oil and greasy residue of steel coupons were removed with paper towels then purification was performed three times for 10 minutes each using ultrasonication to eliminate traces of organic phases. Sulphide corrosion scale was removed by careful polishing, while continuously monitoring the surface for intact metallic steel. The use of Clark's solution was attempted to use but it proved to be inefficient in cleaning metal sulphides first rather than attacking some of the remaining metallic parts of steel coupons near the surface. The masses of the coupons were measured on the same analytical balance as before the tests then by relying on the weight loss method corrosion rates were assessed for the measured periods. In almost all experiments, high mass changes were obtained with low relative variance of data. Corrosion tested coupons were not investigated for internal penetration forms of corrosion products deep inside the bulk lattice and the corrosion of grain boundaries, only external corrosion scale products were analysed.

Although the temperature highly impacts the efficiency of hydroprocessing of biomass sources and composition of diesel fuels [24] at the proposed target temperature of 420 ºC and pressure of 100 bar to reproduce original conditions, severe limitations were faced because of the rapid coke formation from the mixture of raw gas oil with used cooking oil and animal fat, yield complete conversion within less than 24 hours. This made it impossible to carry out experiments at such a high temperature. Therefore, the temperature rage had to be compromised and balanced with the dwelling time of the samples. Then, all experiments could be carried out without any noticeable loss of the organic fluid phase caused by evaporation or cock deposition.

2.1. X-ray diffraction

After reactor testing, spalled layers were crushed. The powdered and low adhering sulphide scales were scrubbed, gathered, and purified twice over 10 minutes each using ultrasonication in isopropanol.

XRD analysis was performed using a Philips PW 3710 type powder diffractometer equipped with a PW 3020 vertical goniometer and curved graphite diffracted beam monochromator. The radiation applied was CuKα from a broad-focus Cu tube, operating at 50 kV and 40 mA. Samples were measured in a continuous scan mode with a 0.02° s^{-1} scanning rate. Data collections and evaluations were carried out with X'Pert Data Collector and X'Pert High Score Plus software.

2.2. Scanning Electron Microscopy and Energy Dispersive Spectroscopy

At the end of reactor testing, steel coupons containing multiple layers of low-adhering sulphide scales and powders gathered from removed surface scales were purified twice over 10 minutes each using ultrasonication in isopropanol. The morphology of corroded steel coupons was measured with a scanning electron microscope (SEM, Philips XL31 ESEM) in the secondary electron mode with an accelerating voltage of 20 kV and recording images produced by back-scattered electrons, as well. For rough elemental analysis, energy-dispersive X-ray spectroscopy (EDS) was also carried out.

2.3. Micro-Raman Spectroscopy

After gathering the metal coupons from fuel mixtures and removing the traces of organic phases twice of over 5 minutes each using ultrasonication both in isopropanol and acetone, sulphide scales were probed five times at different spots on the surface. Raman measurements were performed at atmospheric pressure and gas composition. A Bruker SENTERRA Raman microscope was used to analyse samples. The source was a green semiconductor laser with a diameter of 532 nm and a maximum power of 10 mW. For the microscope, a 10x objective was used. Raman signals were collected with a thermoelectrically-cooled charge-coupled device detector and recorded for typically 15 scans. The typical integration time to record Raman spectra was 15 s on average with a spectral resolution of 4 cm^{-1}.

3. Results and Analysis

3.1. Reactor Experiments

Temperature dependence of the corrosion rates of steel alloys immersed in raw gas oil and its mixtures with renewable biomass sources are presented in *Fig.1*. Over a narrow temperature range of 100 ºC, the 1.0425 carbon steel indicated saturation of sulphidation rates at 275 ºC in all organic fluids. Corrosion rates in raw gas oil at 300 ºC remained similarly high but in used cooking oil and waste lard mixtures decreased beyond a random scatter of data. The St35.8 carbon steel indicated a similar trend. In case of the 1.7335 carbon steel, the highest corrosion rates were obtained with almost linear relationship between the reaction rate and temperature. Reaction rates over the entire temperature range were the least affected by the types of organic fluids. As expected, the 1.4541 high alloy steel gave the

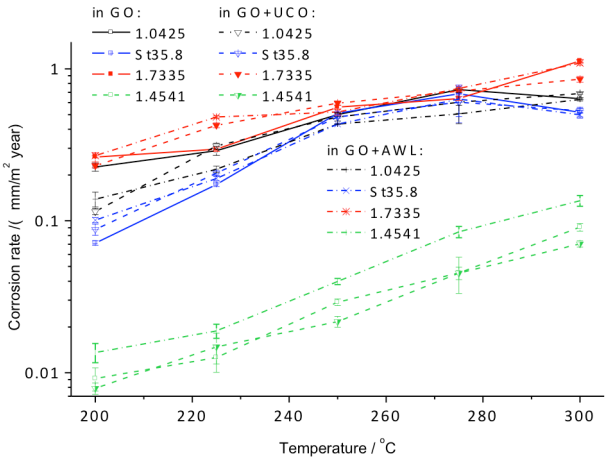

Figure 1. Integral corrosion rates over 504 hours of test period (as a function of temperature) of steel alloys immersed in raw gas oil (GO) and its mixtures with used cooking oil (UCO) and animal waste lard (AWL) at an overall pressure of 90 bar in the presence of 2% (m^3/m^3) hydrogen sulphide, as well as the integral corrosion rates for test periods of 24 hours for the alloys of (b) 1.0425, (c) St35.8, (d) 1.7335, and (e) 1.4541.

Table 3. Temperature dependence of logarithmic corrosion rates of steel alloys (in mol m^2 s^{-1}) expressed in activation energies ($\Delta E_a^{\#}$ in kJ mol^{-1}) over the first 24 hours tested at a pressure of 90 bar in the presence of 2% (m^3/m^3) hydrogen sulphide.

Steel alloys in organic fluids	raw GO $\Delta E_a^{\#}$	GO-UCO $\Delta E_a^{\#}$	GO-WL $\Delta E_a^{\#}$
1.0425	50.3	45.8	52.0
St35.8	45.4	53.3	63.1
1.7335	35.5	31.0	45.6
1.4541	68.9	67.8	35.2

Table 4. Temperature dependence of logarithmic integral corrosion rates (in mm m^{-2} $year^{-1}$) expressed in activation energies ($\Delta E_a^{\#}$ in kJ mol^{-1}) of steel alloys over 21 days. All other parameters are the same as detailed in *Table 2*.

Steel alloys in organic fluids	raw GO $\Delta E_a^{\#}$	GO-UCO $\Delta E_a^{\#}$	GO-WL $\Delta E_a^{\#}$
1.0425	30.5	41.1	36.7
St35.8	78.8	46.5	43.1
1.7335	30.3	30.3	33.3
1.4541	30.3	41.6	53.3

lowest sulphidation rates in all experiments. The higher corrosion resistance of this alloy is well reflected by the almost levelled rates up to 225 ºC, although its markedly different behaviour can be observed by the steeply increasing sulphidation rates above 250 ºC. This temperature range is associated with the evolution of adherent sulphide scale in appreciable amounts. The special behaviour of this alloy is also underscored by the high sensitivity to composition of the organic phase, i.e. waste lard, showing various corrosion rates even around 200 ºC which remained decisive over the entire temperature range. Corrosion rates were heavily dependent on fluid phase composition with biomass sources mixed in 10% (g/g) than on variation of the temperature up to 250 ºC. Generally, sulphidation rates of metals should be similarly more temperature dependent than hydrotreating and desulphurising conversion of the hydrocarbons [25]. In addition, according to the literature, corrosion rates of the 18/8 stainless steels are 10–50 µm $year^{-1}$ between 250 and 300 ºC in the presence of hydrogen sulphide at 2% (m^3/m^3). In our case, carbon steels indicated much greater sulphidation rates than the advisable as maximum of 0.25 mm $year^{-1}$, which should be acceptable up to 390 ºC under a pressure of 34 atm with 10% (m^3/m^3) of hydrogen sulphide [26].

Logarithmic corrosion rates as a function of reciprocal absolute temperature were used to evaluate thermodynamic activation. As results over a short timescale characterise thermodynamics of the sulphide corroding alloys immersed in various organic fluid phases, long-term results are much more dependent on inherent kinetic controls such as barrier effect of the scales. Therefore, reaction rates were both evaluated over the first 24 h and the entire 504 h periods. Estimated activation energies are assumed to be closely related to the overall free energy change of the reactions, as presented in *Tables 3 and 4*. Over a short time period, the highest activation energy of the 1.4541 was ascertained to be in raw gas oil and its mixture with used cooking oil. Nevertheless, about half of the energy was required for the activation in the waste lard mixture of gas oil. This tendency was found to be reversed over long periods when activation was the lowest in gas oil and the highest in the waste lard mixture. This result might only be interpreted as acid corrosion of the 1.4541 taking place over short time periods and sulphidic corrosion occurring over longer periods in parallel with the hydrodesulphurisation gas oil. Adsorption of the fluid phase components and dissolution of the ionized metals are probably inhibited by the firm and adherent sulphide scales. The 1.0425 and St35.8 carbon steels indicated lower activation energies in the short term almost independent of the composition of the organic fluid phase, although the St35.8 might yield some increase with the presence of biomass sources. Over longer periods, except for the St35.8 in gas oil, lower activation with regards to sulphide corrosion manifested, which leads to the recognition that corrosion of these alloys is primarily governed by the hydrogen sulphide. Furthermore, if activation of the interfacial reactions changes it almost certainly decreases over time, which is anticipated in many metal sulphide forming reactions. The 1.7335 low alloy experienced high reaction rates featuring low thermal activation irrespective of the nature of the organic phase and time exposure. Similarly to carbon steels, activation tends to decrease over time, suggesting a catalysed reaction by the scales of precipitating corrosion products, which were porous, susceptible to spalling and semiconductors. This behaviour reveals an insignificant difference between the rates of sulphidic and acidic corrosion, whereas an inefficient barrier effect of the sulphide scale. In short term experiments,

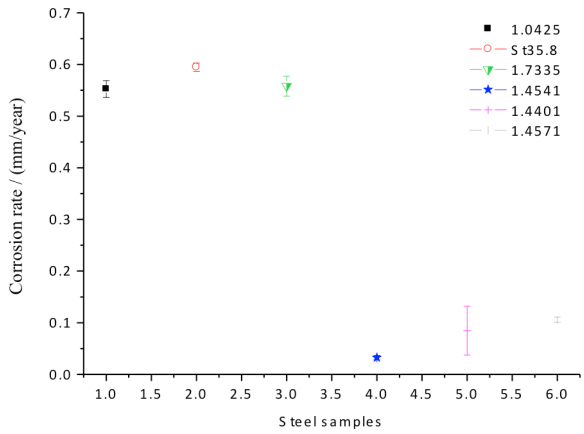

Figure 2. Corrosion rate of steel alloys immersed in refined gas oil tested in a blind experiment for 504 hours at a pressure of 90 bar and temperature of 250 ºC in the presence of 2% (m^3/m^3) hydrogen sulphide.

higher activation energy of the 1.4541 for sulphide corrosion is due to the formation of metal sulphides with the chromium and nickel alloying elements. In contrast, over long time exposures the lower activation energy barrier is partly related to the increased thermodynamic preference in complete balance of the reaction between hydrogen sulphide and iron [27]. Furthermore, the activation of surface processes such as adsorption and dissociation is more attributed to a less preferable free energy balance increasing with the organic materials of used cooking oil and animal fat having a greater absorption abilities. In this regard, the 1.4541 high alloy steel proved to be more sensitive exhibiting more than 2 orders of magnitude changes depending on composition of the organic phase, rather than that of carbon steel alloys indicating changes within a magnitude of 3 times or less. Consequently, the reaction mechanism leading to sulphide corrosion should partly be impacted by adsorption and accelerated by saturated free fatty acids composing a considerable proportion of waste lard. Moreover, the increasing temperature is known to play a vital role in governing the mechanism of naphthenic acid corrosion either by chemical or electrochemical mechanisms [28]. The latter is more dependent on thermal activation as a greater rate of dissociation of the weak acids occurs at elevated temperatures [29]. Electrochemical acceleration may lead to local corrosion and pit formation, whereas chemical naphthenic acid corrosion at lower temperatures results in general processes. Pearlite as an interpenetrating bicrystal phase composed of ferrite and cementite in plain carbon steels is a relatively good cathodic area due to the effective depolarisation for hydrogen ion discharge [23]. In our case, selective corrosion like pitting on specimens of the steel alloys was not observed as a function of temperature between 200 and 300 ºC except for the 1.0425 carbon steel exhibiting pitting corrosion of metal lattice regardless of the organic fluid phase. Process of the latter must be governed by the hydrogen sulphide which with its increasing concentration after a threshold

Table 5. General corrosion rates of steel alloys (in mm $year^{-1}$) immersed in raw and refined gas oils, tested at 250ºC over 21 days under a pressure of 90 bar in the presence of 2% (m^3/m^3) hydrogen sulphide.

Steel alloys	Raw gas oil	Refined gas oil
1.0425	0.10 – 0.50	0.55
St35.8	0.10 – 0.50	0.60
1.7335	0.30 – 1.00	0.55
1.4541	0.01 – 0.10	0.05
1.4401	-	0.10
1.4571	-	0.10

enhances the rate of local process around metal crystal boundaries and sulphide inclusions.

As a blind experiment to probe the effects of an HDS plant using a batch reactor experiment, alloys were tested under the same conditions as for temperature dependence at 250 ºC and the results are presented in *Fig.2* and *Table 5*. The corrosion rates of carbon steels (1.0425, St35.8 and 1.7335) were greater with low variances of data compared to the ones measured for the 1.4541, 1.4571 and 1.4401 high alloy steels that exhibited reaction rates an order of magnitude lower with high variances, except for the 1.4541. Thus, the firm corrosion resistivity of the 1.4541 alloy was observed, which is primarily connected to its high chromium [30] and moderate nickel contents and the absence of molybdenum. Nickel and molybdenum are of high and moderate concentrations in tested steel alloys causing somewhat decreasing sulphation resistance [31, 32]. Furthermore, from the results of an independent experiment, a less favourable activation of metal sulphide formation is suspected due to the inefficient adsorption and dissociation of hydrogen sulphide from the gas phase besides the absence of dissolved small molecule acids in the fluid phase. Thus, corrosion rates of steel alloys might be the least affected by the adsorption of species from the organic phase only the gaseous phase, such as hydrogen and rather the hydrogen sulphide. The difference between the medium sulphidation resistivity of the 1.4571 and 1.4401 high alloy steels at low temperatures can be explained. On the other hand, by comparing the results of the blind experiment with that obtained in raw gas oil, the corrosion rates were roughly the same or somewhat greater, since the refining of the raw gas oil and the desulphurisation of the raw gas oil content should not have a notable effect on the resistivity of investigated alloys and their observed corrosion rates at least up to 250 ºC.

The timespan, over which parabolic corrosion rates appear after a short term linear reaction rate decrease in the induction period is related to development of thick sulphide scales. Therefore, the time dependence of integral corrosion rates is presented in *Figs.3a-3d*. The 1.0425 and St35.8 carbon steels indicated parabolic or even more intense decay of corrosion rates over time periods above 250 ºC in all organic fluids but especially in waste lard mixture (*Figs.3a* and *3b*). The only exception was the St35.8 at 275 ºC that exhibited a less pronounced decay of sulphidation rates over time. This

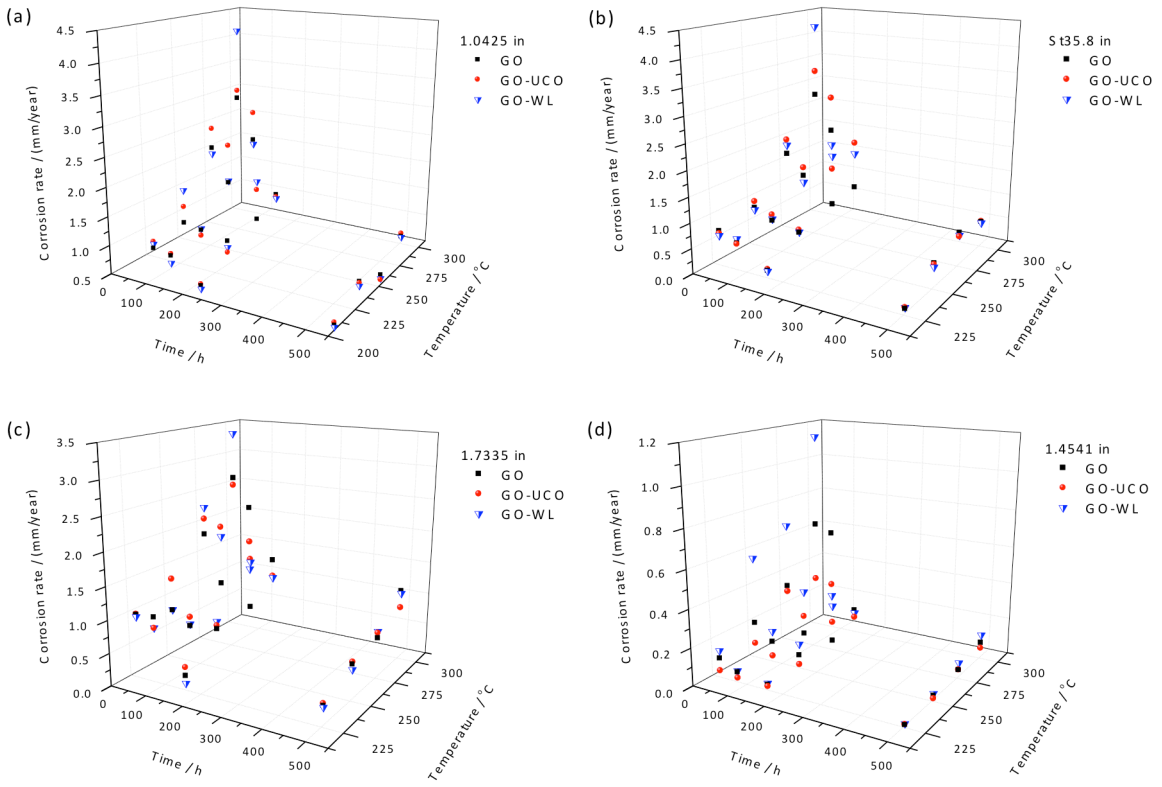

Figure 3. Integral sulphide corrosion rates of steel alloys over time at various temperatures: (a) 1.0425, (b) St35.8(A38), (c) 1.7335, (d) 1.4541, parameters as defined in *Fig.1*.

may be attributed to the partially incorrect experimental data owing to the lack of a large number of samples and parallel experiments to provide a more reliable prediction of the most probable corrosion rates with the least scattering of the mean values as possible. Below 250 ºC, a less remarkable decay of reaction rates was ascertained for both metals. The 1.7335 low and 1.4541 high alloys exhibited moderately intense parabolic

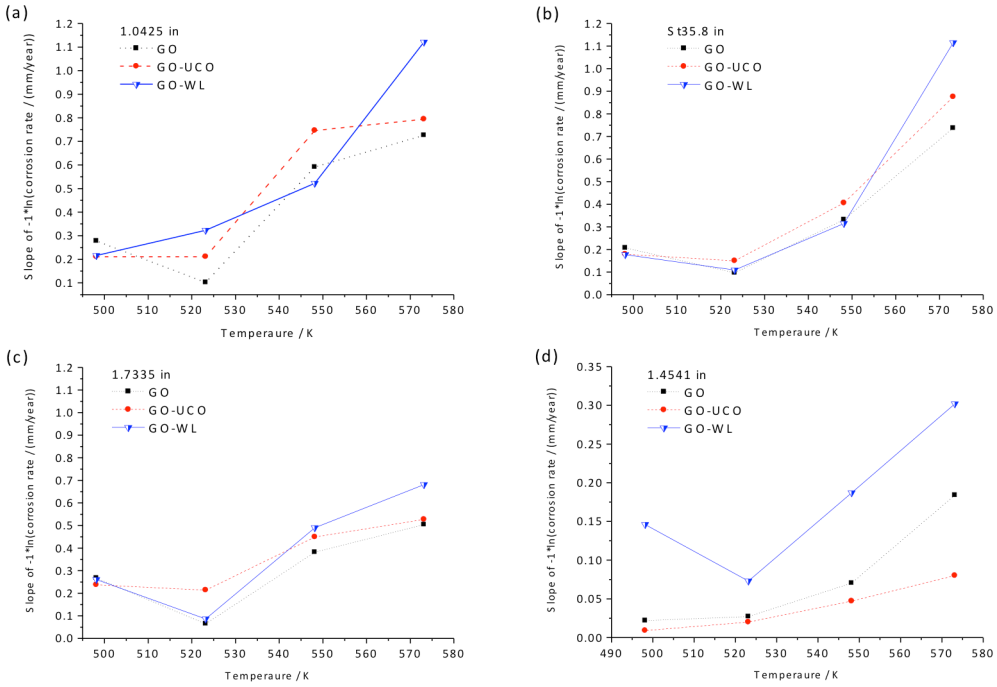

Figure 4. Gradients (in terms of absolute value) of logarithms of integral corrosion rates over 21 days in raw gas oil (GO) and its mixtures with used cooking oil (UCO) and waste lard (WL) as a function of temperature: (a) 1.0425, (b) St35.8, (c) 1.7335, (d) 1.4541, parameters as defined in *Fig.1*.

decreases of reaction rates over time. Nevertheless, the gradients of decay of the reaction rates were lower (greater in absolute value) at higher temperatures than quadratic rate decreases at and under 250 ºC as a function of time (*Figs.3c and 3*).

The 1.7335 yielded gave almost constant corrosion rates at 250 ºC between 3 and 21 days of exposure (*Fig.3c*). Its induction period seemed to be restrained to the first 24 h term in the mixture of used cooking oil. The 1.4541 steel experienced the lowest corrosion rates in the used cooking oil mixture over the parameter space (*Fig.3d*), which exhibited a rather linear dependence as a function of temperature. Overall, all steel alloys indicated the most severe corrosion-type integrity loss in the waste lard mixture of gas oil.

The absolute values of the gradients obtained by linear fitting to the logarithms of the integral corrosion rates measured over 21 days is presented in *Figs.4a-4d*. Sulphidation rates of the 1.0425 and St35.8 carbon steels exhibited the highest temperature dependence in the waste lard mixture (a 5 times increase), whereas the lower gradients probably indicate that a less sensitive thermal activation occurs in raw gas oil and the used cooking oil mixture (enhanced by around 3 times, *Figs.4a* and *4b*). As for the 1.7335 low and the 1.4541 high alloy steels, the low and the lowest gradients with restrained enhancement of around two times were obtained as a function of temperature (*Figs.4c* and *4d*). This signifies a less pivotal role of thermal activation in the interfacial sulphidic corrosion of the latter two alloys but in all cases waste lard was the fluid phase that was proven to be the most susceptible to open thermally-allowed reaction mechanism leading in some cases to the highest corrosion rates by far. Among all steels, only the 1.4541 exhibited low thermal rate dependence in used cooking oil compared to the results obtained in other phases. As a general impression, either the most intense dependence of thermal activation on sulphidation or the highest magnitudes of reaction rates occurred in the waste lard mixture of gas oil. Thus, not only sulphidation takes place by the hydrogen sulphide, but also acidic corrosion occurs and competes with the former at a rate of the latter becoming higher by increasing temperatures. This finding may also point out that the contributing role of free fatty acids in parts of the corrosion processes could be adsorption of mercapto compounds assisted by possible chelating effect and thermally favoured dissociation of the anchoring functional groups. The greater oxidation ability of the organic compounds than hydrogen sulphide, and their more positive redox potentials are due to their stronger electron withdrawing ability boosted by conjugated electron systems of various extents. The desorption of corrosion products is promoted by the proper dissolution of metal ions with lipophilic complexes, and diffusion through porous, cracked or spalled scales, allowing a less hindered access of the steel surface to the species taking part in subsequent processes. In the case of the exclusive hydrogen sulphide corrosion, effective desorption and diffusion from the surface can only be realised by high porosity, cracking or spalling,

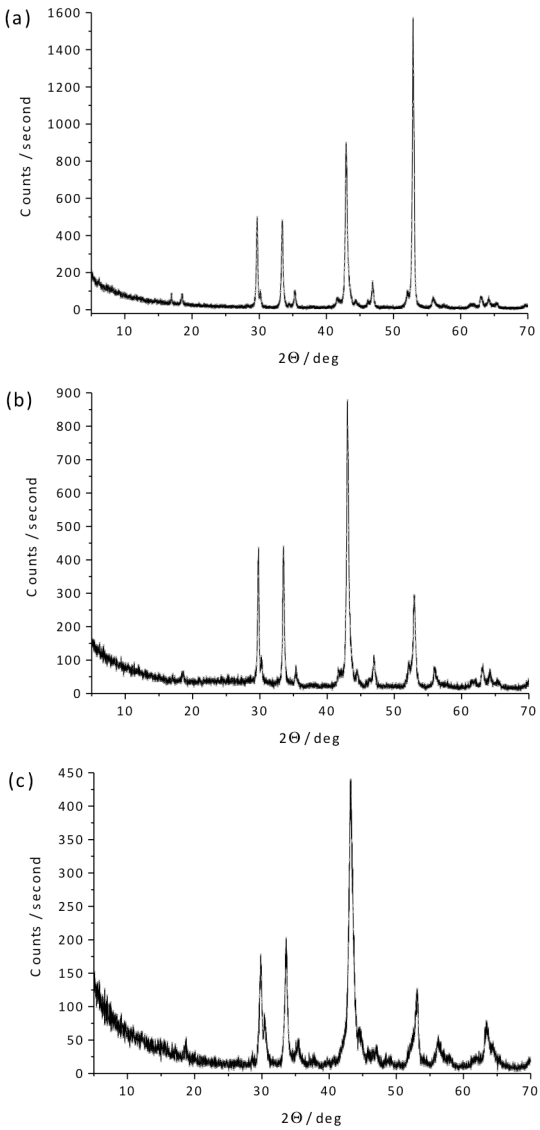

Figure 5. XRD patterns of the (a) 1.0425 carbon steel immersed in raw gas oil and (b) refined gas oil tested at 250 ºC, (c) 1.7335 low alloy steel immersed in a gas oil mixture with waste lard (10% g/g) tested at 300 ºC, all for 21 days at a pressure of 90 bar.

and failure of the scales as observed with many of the carbon and low alloy steels but it should not happen to high alloy steels developing quasi-passivating, coherent and adherent sulphide scales. In the latter case, fluid phases with improved wettability based on the closer surface energy of the solid phases and tension of the fluid phases compared to that of the raw gas oil and steel combination is beneficial as less hindered interfacial reactions result.

3.2. X-ray Diffraction

In *Fig.5a*, diffraction pattern of troilite, a 1:1 stoichiometric ratio of iron sulphide (FeS) was identified on the whole in the powder of the spalled scales of the 1.0425 steel tested in raw gas oil at 250 ºC. Low crystallinity metallic iron Fe(0) with high magnetic susceptibility formed the majority of the rest phase.

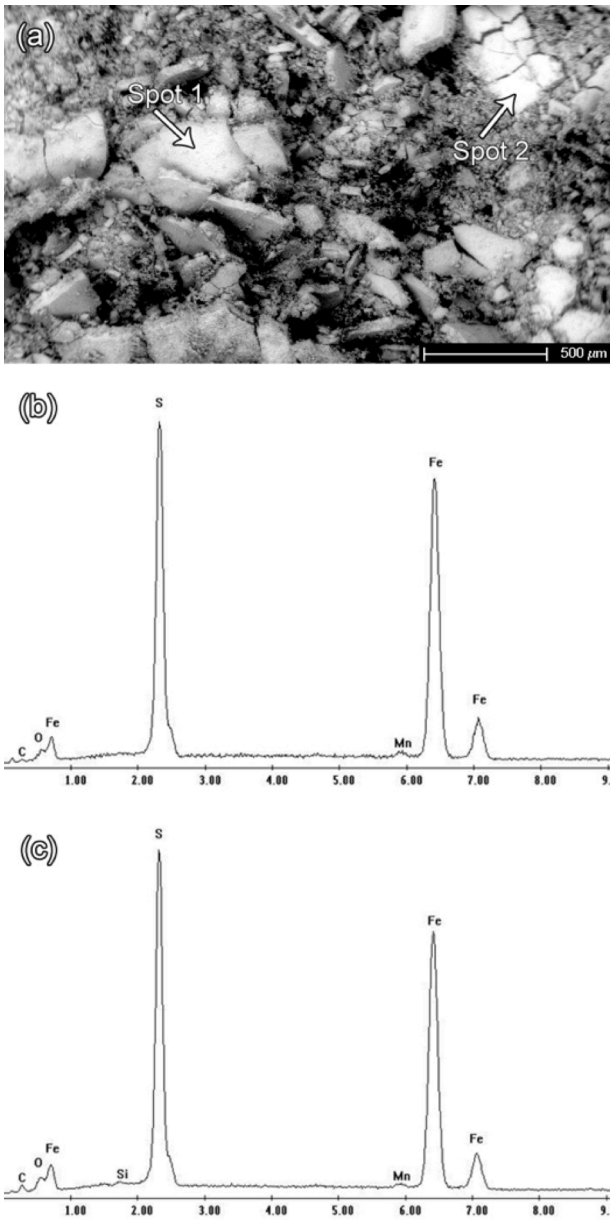

Figure 6. SEM image of (a) the powdered form of sulphide scale removed from the surface of the 1.0425 carbon steel coupons immersed in raw gas oil and tested at a temperature of 250 ºC and pressure of 90 bar; (b and c) the elemental composition of the scale removed from surface of the coupons obtained by EDS analysis of Spots 1 and 2 marked by arrows, respectively.

Table 6. *SEM*-EDS composition (in atomic %) of the corrosion product of sulphide corrosion scales removed from the surface of carbon steel coupons (1.0425) immersed in raw gas oil tested at a temperature of 250 °C and pressure of 90 bar.

Elements	Spot 1	Spot 2
C (K)	0.00	6.95
O (K)	6.02	7.42
Si (K)	0.00	0.38
S (K)	36.58	35.77
Mn (K)	1.06	1.02
Fe (K)	56.33	48.00

Because typical high intensity reflections at 2θ values of 12º, 15º, and 19º [33] and 40º, 50º, and 59º [34] were not measured, the low intensity reflections gained at 2θ values of 17.5º and 30º might not be connected to mackinawite. The absence of its phase in metal sulphides in the long-term tested samples is attributed to its formation during the early reaction phase and rapid transformation to more stable pyrrhotite, the sub-phase of troilite. The reflections together at 2θ values of 30º and 53º may not be related to greigite as its stability region is above the H^+/H_2 equilibrium line. Troilite is more stable in weakly acidic aqueous solutions under reducing conditions, i.e. below the H^+/H_2 equilibrium line [35]. The intense reflection at 2θ value of 44º cannot unanimously be attributed to Pyrrhotite [36], as it undoubtedly overlaps with one of the main reflections of polycrystalline iron [37]. In *Fig.7b*, a different diffraction pattern of the corrosion scale can be identified by the varied intensity ratios of reflections at 43º and 53º. This is due to a composition change of the scale leading to metallic iron Fe^0 and crystalline troilite formation (FeS) in molar ratio of about a 2:1. Thus, a considerable amount of metallic iron segregated over the course of high rates of sulphide corrosion, but its amorphous phase did not result in its conversion to sulphide by either hydrogen sulphide or naphthenic acids in refined gas oil subjected to testing over 21 days at 250 ºC and 90 bar. As expected, long-term medium temperature exposure resulted in high sulphide conversion of metallic iron and the formation of iron sulphide crystals in raw gas oil of a high sulphur content. On the other hand, roughly the same reaction rates resulted in the evolution of a similar amount of sulphide scales in refined gas oil, which proved to be more brittle. Thus, high proportions of the scale detached and rapid disintegration of the metallic phase occurred, resulting in corrosion products with far lower ratio of the elements than stoichiometry. The lower sulphur content of the refined (hydrodesulphurised) gas oil is accountable for this finding because of the lower plasticity of the more compact sulphide scale. An XRD investigation of the sulphide scale formed on the 1.7335 low alloy steel coupons yielded a similar diffraction pattern (*Fig.5c*) to that of the 1.0425 carbon steel (*Fig.5b*) tested under the same conditions. Nevertheless, in this case the massive amount of highly porous corrosion scale formed with a low spalling ability (high coherence and adherence) should mean less frequent occasions of the surface to renew by spalling or detachment resulting in further hydrogen sulphide attack but lower diffusion hindrance to accessing the metallic surface by sulphide species.

3.3. SEM and EDX Analyses

Morphology of the sulphide scales gathered from the bottom of the glass tubes and removed from the surface of the 1.0425 carbon steel specimens is presented in

Fig.6a. The multiple layers of sulphide scale were moderately easy to detach, brittle, and highly porous. Spalled and detached scales were of a lateral size of about 500 µm and a thickness of 40 µm. The spots of 1 and 2 (arrowed in the image) were analysed to identify the elemental composition of the scale and the SEM-EDS results (*Figs.6b* and *6c*) is presented in *Table 6*. The molar ratios of the iron sulphide (surmised to be 1:1 stoichiometric ratio of FeS) and metallic iron (Fe^0) were 3:1 and 1.8:1 at Spots 1 and 2, respectively. This is in acceptable agreement with and supplementary to XRD results.

3.4. Micro-Raman Spectroscopy

The Raman spectra of the sulphide scales gathered from the surface of the steel coupons (tested at 250 ºC over various time exposures) are presented in *Fig.7*. Similar spectra were obtained for the St35.8 carbon (b) and 1.7335 low alloy (c) steels. The 1.4541 high alloy steel (d) indicated a strong feature in the wavenumber region of 550-800 cm^{-1}. In the low wavenumber range, a similar Raman spectrum was detected to that of the 1.0425 carbon steel (a). Because Raman spectra depend strongly on the intensity and wavelength of the incident beam, based on the weak feature of the spectra around the main Raman characteristic wavenumbers, it is thought that small proportion of troilite in the crystalline phase [38] should probably have formed after the chemical conversion of the initial metal sulphide on the St35.8 and 1.7335 (lines at 289 and 290 cm^{-1}) over a short time period. Metal sulphides, detected in a somewhat greater proportion of crystallinity (yielding a relatively low amount of scale) formed on the slower corroding 1.4541, resulting in to the more pronounced features in the wavenumbers between 550 and 800 cm^{-1}. Corrosion scales of St35.8, 1.7335, and 1.4541 alloys yielded intense broad and weak features at 1292 and 218 cm^{-1} as a result of the troilite phase. These characteristic wavenumbers were only found in the spectra in accordance with the literature [40]. This reveals that the low crystallinity of corrosion scales evolved within 24 hours. Other Raman features of troilite, a weak intensity peak at 456 cm^{-1} and the sharp ones at 600, 398, and 283 cm^{-1} [41], and 400 and 295 cm^{-1} [38] were not found. On the other hand, iron sulphide in a form close to mackinawite [42] might keep evolving in the early stages on 1.0425, St35.8 and 1.4541 as wavenumbers of moderate intensity at 321, 317, and 315 cm^{-1}, respectively, were detected. They usually appear as intense overlapping peaks between 310 and 320 cm^{-1} depending on the degree of crystallinity [43]. Thus, this seems to concur with greater reaction rates leading to a lower proportion of crystallinity in the product as the rate of nucleation overtakes to a level of far beyond orientation and growth. Insignificant Raman intensities in the regions of 280-360 cm^{-1} (at 206, 257, and 296 cm^{-1} and at 256, 312, and 355 cm^{-1}) are generally connected to mackinawite ($Fe(II)S_{1-x}$ and $Fe(II\text{-}III)S_{1-x}$) and iron-rich mackinawite, and greigite, respectively [44]. In the

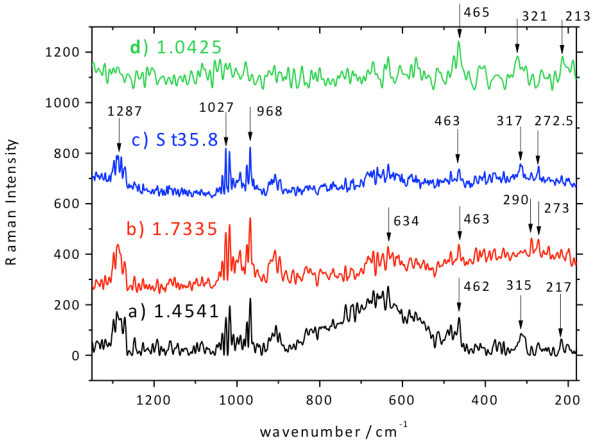

Figure 7. Micro-Raman survey spectra of steel alloys tested under specified conditions as stated: a) 1.0425 (250 ºC/20 bar, 8 h), b) St.35.8 (250 ºC/50 bar, 24 h), c) 1.7335 (250 ºC/20 bar, 8 h) and d) 1.4541 (250 ºC/50 bar, 24 h).

1.4541, the very low Raman intensity at 240 cm^{-1} is probably due to the presence of chromium-rich troilite whereas the low intensity features at 255 (290) and 365 cm^{-1} are partly connected to daubrelite ($FeCr_2S_4$ and $FeCr_2S_3$). Nonetheless, these phases should have been present in negligible quantity, because none of the supporting XRD results could confirm the presence of such crystals, besides theoretical thermodynamic calculations of stable regions of iron sulphides (see discussion). In all cases, the low crystallinity of the corrosion products is attributed to the short time-scales and low temperatures of the experiments. It should also be noted that amorphous iron sulphide scales, which evolved in the induction periods were far less coherent, more crystalline, and more susceptible to spalling and detaching than the ones formed in slower reactions. Thus to explain this phenomenon, lower mechanical stress might have accumulated over time in scales generated by slower and decreasing rates of reactions. As a result, the induction time of sulphide scale formation should be very low at a pressure as low as 20 bar but crystallisation must occur over longer periods. Therefore, time constants in terms of spalling, and the failure of the scales to detach might be related to relaxation of accumulated stresses, probably at a rate faster than the crystallisation of iron sulphides. For troilite, it must be longer than 8 hours, but much less than 21 days at 250 ºC and a pressure of 90 bar. In fact [45], due to kinetic reasons in connection with the highest reaction rate of scale growing without diffusion-type mass transport control [46], it has been stated that [47] generally mackinawite is expected to form rapidly first almost regardless of reaction conditions. For thermodynamic reasons, if metastable cubic FeS formed, it should have also transformed into pyrrhotite and troilite over the weeks the experiment was undertaken at a moderate temperature and pressure both in the presence [48] and absence of chloride [49]. Phase transformation of metal sulphides to Troilite takes place in days or even hours as was found in our case. The evolution of pyrrhotite might occur over a couple of

days so its detection in samples tested over at least 3 days can be anticipated. Not exactly in line with the location of the pyrrhotite-troilite transition boundary [50, 51], in our case elevated temperatures favoured the formation of troilite rather than pyrrhotite under less than 2 bar of partial pressure of the hydrogen sulphide.

4. Discussion

The hypothetical growth in thickness of the sulphide scales over selected time periods based on the mass changes of the coupons is given in *Figs.8a-8d*. The 1.0425 at 225 ºC exhibited thicker scale development in biomass mixtures with lower reductions in rate over time than in neat gas oil (*Fig.8a*). At 250 ºC, all corrosion rates exhibit proportional growth as time of exposure increases. Such linear rates of sulphidation started to decelerate as a function of time above 250 ºC. The lowest variation of corrosion rates over time manifested in the used cooking oil mixture. The St35.8 exhibited similar sulphidation resistance behaviour up to 275 ºC but at 300 ºC a higher drop in corrosion rates was observed in mixed organic environment of the biomass and a smaller reduction in neat gas oil over time (*Fig.8b*). The 1.7335 experienced increasing corrosion rates in all fluids as a function of temperature, without any notable decrease related to barrier effects over longer time periods (*Fig.9c*). Depending on the type of the organic phase, moderate drops and increases in the corrosion rates over intermittent periods were observed at 74, 168, and 168 h for 1.0425, St35.8 and 1.7335 steels, respectively. According to our experience, the growth rate in the thickness of stable scales on carbon steels never exceeded 10 µm over the first 3 days. Up to 250 ºC, the 1.0425 exhibited a higher and the St35.8 a somewhat lower susceptibility of the sulphide layers to spalling and detachment. The 1.7335 manifested the development of more coherent scales that were not susceptible to spalling or failure over time especially at and above 275 ºC. A less frequent variation in time of reaction rates along with the observed quantity of detached scales in the form of a sediment at the bottom of the glass tubes led to the conclusion that the about low compactness, and high porosity of the sulphide scales are the primary reasons for the long-lasting high sulphidation rate of this alloy. Non-protective scales exhibit linear growth and a mass change over time similar to that of the 1.733. This is typical for porous or cracked films, which facilitate the less hindered transport of reactants and access to the metallic phase at faster rates than at which the chemical reaction proceeds. As far as the 1.0425 and St35.8 carbon steels are concerned, due to the high barrier nature of the sulphide scales at elevated temperatures, variations in the time needed to renew the surface almost to the metallic phase were necessary to maintain the high reaction rates over long periods of time. Scale failures should have been caused by the accumulation of high internal stresses along with high growth rates taking place on polycrystalline surfaces. Such failures do not occur within scales at the scale/metal interface since it is cohesive in nature due to the high electrostatic forces that exceed the yield strength of the scales [52].

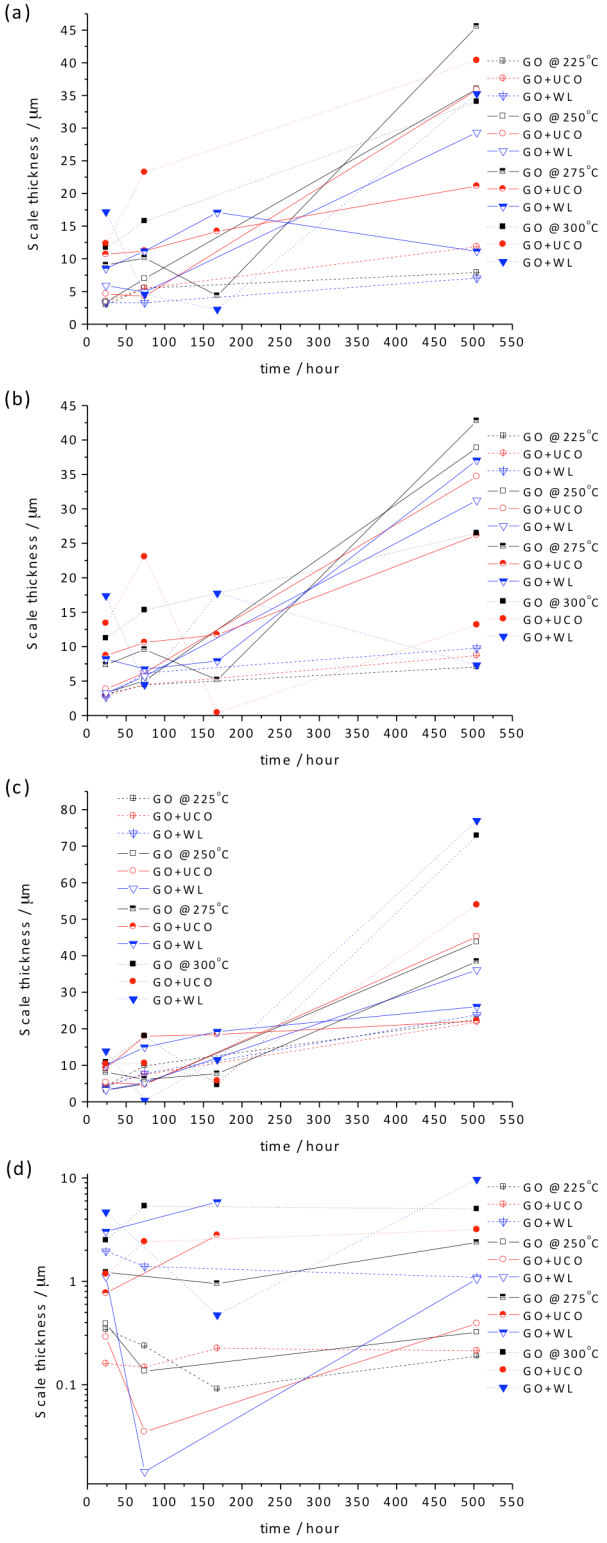

Figure 8. Hypothetical differential metal-sulphide scale thicknesses developed over time periods based on the mass change of the coupons of steel alloys immersed in refined gas oil (GO) and its mixtures with used cooking oil (UCO) or waste lard (WL) tested at a pressure of 90 bar in the presence of 2% (m^3/m^3) hydrogen sulphide.

Thus, hypothetically speaking, further scale growth as a solid-state reaction must trigger the breakdown of scales [53]. This is basically in line with the literature [54] as the corrosion resistances of carbon and low alloy steels are approximately equal. The low chromium alloys of up to 5% (g/g) warrant little or no advantages over carbon steels in terms of resistance to reactions with hydrogen sulphide. Nonetheless, the 18Cr-8Ni stainless steels feature acceptable resistances over a wide range of pressures, temperatures, and hydrogen sulphide concentrations, except under severe conditions [55].

In other aspects, the protection potential of sulphide scales for carbon steels is usually questionable when well adsorbing species or compounds are in fluid media. After the first phase of mackinawite growth, corrosion rates can accelerate because the tarnish is not a uniform film, but rather composed of discrete crystallites that are highly permeable [56]. If the electrochemical mechanism of sulphide corrosion occurs and cathodic depolarisation is efficient, pyrrhotite formation becomes anodically controlled by cation diffusion that takes place through sulphides, which are in part limited by maximum reaction rates. Thus, as thermodynamic simulations delineated the region of stability of troilite [57], in our experiment the electrode potential of the steel coupons should have been between -0.4 and -0.65 V *versus* standard hydrogen electrode (SHE), which is a narrow stability region in neutral media. At 250 ºC, the 1.4541 showed slight variations in sulphidation over time regardless of the fluid phase (*Fig.8d*). The scale structure was less compact and easily removable from the surface. At 250 ºC, the highest decrease in reaction rate was obverserved over moderate periods of time that could only be accounted for by the greater initial sulphidation rates, increased compactness, and packing of the corrosion film, resulting in the metallic surface becoming more rapidly blocked, thus preventing subsequent reactions leading to less hindered ionic diffusion. As the temperature increased to 275 ºC and above, time variations in the reaction rates became less remarkable as a function of time with similar changes in all organic fluids. This may suggested as the factors of utmost importace that determines the type of sulphide scales that could grow on the specimens, which slows down further metal sulphide forming reactions. As recognised first by Wagner [58], some metals develop oxide scales exhibiting excellent protection characteristics as a result of the decreasing diffusion rate of ions through compact scales. Such quasi-passivating behaviour was found in the sulphide corrosion of the 1.4541 under the experimental conditions, as mass changes obeyed the parabolic law as a function of time. This is in accordance with the nature of chromium, which is considered to be present in relatively high amounts as the most significant alloying element that increases sulphidation resistance. In some instances, the logarithmic rate of sulphide conversion by mass change was also observed in the initial phases of the reactions. The induction periods in the formation of metal sulphides were intense probably because of the lack of mass transport hindrances at the interfaces which evidently evolved to various extents in later phases of the experiments as thermal activation between 200 and 300 ºC is inefficient to maintain high ion fluxes through the lattice of scales of moderate thicknesses and compactnesses. Thermodynamic calculations based on Gibbs free energy changes of the reactions of sulphide corrosion processes do not justify remarkable changes in reaction rates as a function of temperature. Therefore, kinetic factors are likely to play a pivotal role in the enhancement of reaction rates in all cases. As a comparison of the protection performances provided by sulphide scales on different sorts of steel alloys, corrosion layers on the 1.4541 and the St35.8 evolved better barrier characteristics at higher temperatures than under 250 ºC, which is generally in agreement with the terms of scaling tendency [59]. More coherent and adherent scales of metal sulphides formed at greater rates as corrosion products were not as solubile in GO and UCO compared to the WL mixture. So precipitates shrank and became more compact, which is partly is accounted to their mediocre thermal activations. In waste lard mixtures, corrosion through both chemical and electrochemical mechanisms assisting each other resulted in a pronounced efficiency of thermal activation was experienced along with the impact of acids relegating protective function of the chromium-, nickel-, and molybdenum-alloying elements. Aside from the 1.7335 exhibiting good mechanical protection characteristics independent of temperature, the 1.0425 yielded rapidly increasing sulphidic corrosion with temperature. The former is attributed to high porosity of the sulphide film and the latter is explained by detachment of the layers thereof regularity depends on the corrosion rate and surface structure, orientation of the metallic crystallites, affecting the rate of stress accumulation in the corrosion products. The pronounced sensitivity of corrosion resistance in the 1.4541 regarding composition of the organic fluid phase resulted in high reaction rates especially in waste lard mixtures not only in the short run, but over longer periods, suggests the existence of an altered microstructure, lower compactness and increased solubility of metals and sulphide scales in acid containing fluid than in neat gas oils. This is in agreement with the literature [60] as naphthenic acids are most accountable for the highly corrosive nature oils [61-64]. The high corrosion rate of the St35.8 at low temperatures probably is a result of diffusion along short-circuit paths at grain boundaries or dislocations rather than of lattice diffusion, as thermal activation is high compared to over long periods when kinetic control of barrier effects besides the seldom failure of scales. Slow growing scales proved to be more coherent and protective, exerting better mass transport control on during subsequent reactions. The Gibbs free energy change of the metal sulphide reactions is inferior to diffusion activation in the solid state. The growth mechanism of metal sulphide scales takes place via outward diffusion of cations which leads to lower rate of stress accumulation in the layer. In carbon steel

samples, spalling and scale failure were experienced over the first 24 hours as one of the four types of stress-relief mechanisms. In the case of thin scales, maximum-stress affects metal/scale interface. The imperfections of the flat surface and the geometry of the coupons is also partially connected to the detachment of sulphide scales [65] especially at the inner and outer edges as flat interfaces are less susceptible to spalling compared to the ones with small radii, high curvatures [66]. Low film ductility at low temperatures and high corrosion rates might inevitably lead to spalling, undermining and detachment.

5. Conclusion

The corrosion rates of carbon steels yielded an almost linear dependence as a function of temperature between 200 and 300 ºC regardless of the biomass fuel sources. Although carbon steels developed less spalling and detaching sulphide scales with increasing temperature, the porous layers did not contribute to the higher protection of the metallic phase as a result of their enhanced barrier natures. In contrast, on the surface of high alloy steels sulphide scales with excellent protection characteristics evolved during the induction period and the layer formation is assisted by temperature.

Integrity loss of the high alloys can severely be affected by biomass fuel sources such as animal waste lard when they are used in high proportions along with raw gas oil. Free fatty acids play probably a dominant role in the corrosion mechanism.

The higher corrosion rates observed than reported in the literature lead us to conclude that the proportion of the biomass sources used, especially those the ones having high free fatty acid contents, should preferably be kept low to prevent rapid losses in integrity of steel alloys, ensuring long and safe life cycles of the equipment used in HDS refining plants.

Acknowledgement

The authors are indebted to Ferenc Molnár, Corrosion Research Group, Department of Physical Chemistry, Institute of Chemistry, University of Pannonia, for his invaluable discussions and advice on planning experiments and evaluating results. Bence Nagy, MOL Co., is gratefully acknowledged for his mediation and help in the procurement of precursors for the experiments. The authors are indebted to Róbert Bocsi, Department of Chemical Processing, Institute of Chemical and Process Engineering, University of Pannonia, for his instruction, laboratory work and help in detecting and measuring the concentrations of hydrogen sulphide during reactor experiments. The authors are grateful to Csaba Tóth, Department of MOL Hydrocarbon and Coal Processing, Institute of Chemical and Process Engineering, University of Pannonia, for his variety of contributions to the experimental work, and to Margit Balázs for her analytical results of the biomass fuel sources.

REFERENCES

[1] Knothe, G.: Biodiesel and renewable diesel: A comparison, *Prog. Energy Combus. Sci.*, 2010 **36**(3), 364–373 DOI: 10.1016/j.pecs.2009.11.004

[2] ASTM D6751-15ce1, Standard Specification for Biodiesel Fuel Blend Stock (B100) for Middle Distillate Fuels (ASTM International, West Conshohocken, PA) 2015 DOI: 10.1520/D6751-15CE01

[3] Standard EN 14214:2008, Automotive fuels - fatty acid methyl esters (FAME) for diesel engines - requirements and test methods (European Committee for Standardization, Brussels, Belgium) 2008

[4] Lapuerta, M.; Villajos, M.; Agudelo, J.R.; Boehman, A.L.: Key properties and blending strategies of hydrotreated vegetable oil as biofuel for diesel engines, *Fuel Process Technol.*, 2011 **92**(12), 2406–2411 DOI: 10.1016/j.fuproc.2011.09.003

[5] Bezergianni, S.; Kalogianni, A.: Hydrocracking of used cooking oil for biofuels production, *Bioresource Technol.*, 2009 **100**(17), 3927–3932 DOI: 10.1016/j.biortech.2009.03.039

[6] Lappas, A.A.; Bezergianni, S.; Vasalos, I.A.: Production of biofuels via co-processing in conventional refining processes, *Catal. Today*, 2009 **145**(1-2), 55–62 DOI: 10.1016/j.cattod.2008.07.001

[7] Choudhary, T.V.; Phillips, C.B.: Renewable fuels via catalytic hydrodeoxygenation, *Appl. Catal. A: Gen.*, 2011 **397**(1-2), 1–12 DOI: 10.1016/j.apcata.2011.02.025

[8] ASTM D7467-15ce1, Standard Specification for Diesel Fuel Oil, Biodiesel Blend (B6 to B20), (ASTM International, West Conshohocken, PA) 2015 DOI: 10.1520/D7467-15CE01

[9] Standard EN 590:2004. Automotive fuels - diesel - requirements and test methods (European Committee for Standardization, Brussels, Belgium) 2009

[10] Bezergianni, S.; Kalogianni, A.; Vasalos, I.A.: Hydrocracking of vacuum gas oil - vegetable oil mixtures for biofuels production, *Bioresource Technol.*, 2009 **100**(12), 3036–3042 DOI: 10.1016/j.biortech.2009.01.018

[11] Huber, G.W.; O'Connor, P.; Corma, A.: Processing biomass in conventional oil refineries: Production of high quality diesel by hydrotreating vegetable oils in heavy vacuum oil mixtures, *Appl. Catal. A: Gen.*, 2007 **329**, 120–129 DOI: 10.1016/j.apcata.2007.07.002

[12] Farrauto, R.J.; Bartholomew, C.H.: Fundamentals of industrial catalytic processes (Blackie Academic and Professional, Chapman & Hall, London, UK) 1997

[13] Mikulec, J.; Cvengros, J.; Joríkov, L.; Banica, M.; Kleinová, A.: Second generation diesel fuel from renewable sources, *J. Cleaner Prod.*, 2010 **18**(9), 917–926 DOI: 10.1016/j.jclepro.2010.01.018

[14] Hancsók, J.; Baladincz, P.; Kasza, T.; Kovács, S.; Tóth, Cs.; Varga, Z.: Biogas oil production from waste lard, *J. Biomed. Biotechnol.*, 2011 ID 384184, p. 9 DOI: 10.1155/2011/384184

[15] Craig, W.K.; Soveran, D.W.: Production of hydrocarbons with a relatively high cetane rating (US Patent 4,992,605) 1991
[16] Stumborg, M.; Wong, A.; Hogan, E.: Hydroprocessed vegetable oils for diesel fuel improvements, *Bioresource Technol*., 1996 **56**(1), 13-18 DOI: 10.1016/0960-8524(95)00181-6
[17] Monnier, J.; Tourigny, G.; Soveran, D.W.; Wong, A.; Hogan, E.N.; Stumborg, M.: Conversion of biomass feedstock to diesel fuel additive assigned to natural resources (US Patent 5,705,722) 1998
[18] Klass, D.L.: Biomass for renewable energy, fuels and chemicals (in Cleveland, C.J.: Encyclopedia of energy, 1st ed., Elsevier Science Inc., Oxford, UK) 2004 pp. 1-20
[19] Knothe, G.; van Gerpen, J.; Krahl, J.: The Biodiesel Handbook (AOCS Press, Champaign, IL, USA) 2005
[20] Topolnitskij, P.: Corrosion protection of oil production and refinery equipment, *Chem. Chem. Technol*., 2007 **1**(1), 45-54
[21] Sari, A.; Kaygusuz, K.: Some fatty acids used for latent heat storage: thermal stability and corrosion of metals with respect to thermal cycling, *Renewable Energy*, 2003 **28**(6), 939–948 DOI: 10.1016/S0960-1481(02)00110-6
[22] Qu, D.R.; Zheng, Y.G.; Jing, H.M.; Yao, Z.M.; Ke, W.: High temperature naphthenic acid corrosion and sulphidic corrosion of Q235 and 5Cr1/2Mo steels in synthetic refining media, *Corr. Sci*., 2006 **48**(8), 1960–1985 DOI: 10.1016/j.corsci.2005.08.016
[23] NIST Standard Reference Data: Thermophysical properties of fluid systems, webbook.nist.gov/chemistry/fluid/ (last accessed 21-08-2016)
[24] Bezergianni, S.; Dimitriadis, A.: Temperature effect on co-hydroprocessing of heavy gas oil waste cooking oil mixtures for hybrid diesel production, *Fuel*, 2013 **103**, 579–584 DOI: 10.1016/j.fuel.2012.08.006
[25] Sedriks, A.J.: The corrosion monograph series: Corrosion of stainless steels (2nd Ed., Verlag John Wiley & Sons, Chichester, UK) 1979
[26] John, R.C.: Compilation and use of corrosion data for alloys in various high-temperature gases, *Nace Int*., 1999 Corrosion Paper no. 99073
[27] Barin, I.I.: Thermochemical data of pure substances (VCH Verlag, Weinheim, Germany) 2008 DOI: 10.1002/9783527619825
[28] Slavcheva, E.; Shone, B.; Turnbull, A.: Review of naphthenic acid corrosion in oil refining, *Br. Corr. J*., 1999 **34**(2), 125–131 DOI: 10.1179/000705999101500761
[29] Wu, X.Q.; Jing, H.M.; Zheng, Y.G.; Yao, Z.M.; Ke, W.: Resistance of Mo-bearing stainless steels and Mo-bearing stainless-steel coating to naphthenic acid corrosion and erosion–corrosion *Corr. Sci*, 2004 **46**(4), 1013–1032 DOI: 10.1016/S0010-938X(03)00192-6
[30] Coutsouradis, D.; Davin, A.: The sulphidation of cobalt and cobalt alloys by sulphur vapour and hydrogen sulphide (in Ed.: Foroulis, Z.A.; High temperature metallic corrosion of sulfur and its compounds, Electrochemical Society, New York, USA) 1970 pp. 235-254
[31] Davin, A.; Coutsouradis, D.: What are the effects of alloying elements simply or in combination on hot corrosion in high temperature corrosion of aerospace alloys, *AGARD Conf. Proc*., 1973 **120**, 221-234
[32] Davin, A.; Coutsouradis, D.; Habraken, L.: Influence of alloying elements on the hot-corrosion resistance of Co-Cr alloys, *Mat. Corr*., 1971 **22**(6), 517-527 DOI: 10.1002/maco.19710220609
[33] Lennie, A.R.; Redfern, S.A.T.; Schofield, P.F.; Vaughan, D.J.: Synthesis and Rietveld crystal structure refinement of mackinawite, tetragonal FeS, *Mineral. Mag*., 1995 **59**(4), 677-683 DOI: 10.1180/minmag.1995.059.397.10
[34] Jeong, H.Y.; Lee, J.H.; Hayes, K.F.: Characterization of synthetic nanocrystalline mackinawite: crystal structure, particle size, and specific surface area, *Geochim. Cosmochim. Acta*, 2008 **72**(2), 493–505 DOI: 10.1016/j.gca.2007.11.008
[35] Schoonen, M.A.A.; Barnes, H.L.: Reactions forming pyrite and marcasite from solution: I. Nucleation of FeS_2 below 100°C, *Geochim. Cosmochim. Acta*, 1991 **55**(6), 1495-1504 DOI: 10.1016/0016-7037(91)90122-L
[36] Csákberényi-Malasics, D.; Rodriguez-Blanco, J.D.; Kovács Kis, V.; Rečnik, B.; Aleksander L.G.; Pósfai, M.: Structural properties and transformations of precipitated FeS, *Chem. Geol*., 2012 **294–295**, 249–258 DOI: 10.1016/j.chemgeo.2011.12.009
[37] Saw, C.K.; Lian, T.; Day, S.D.; Farmer, J.C.: X-ray diffraction techniques for structural determination of amorphous materials, Report No, UCRL-TR-225388, *INIS Repository*, 2006 **38**(24), 1-30 DOI: 10.2172/900132
[38] Wang, H.; Salveson, I.: A review on the mineral chemistry of the non-stoichiometric iron sulphide, $Fe_{(1-x)}S$ ($0 \leq x \leq 0.125$): polymorphs, phase relations and transitions, electronic and magnetic structures, *Phase Transit*., 2005 **78**, 547-567 DOI: 10.1080/01411590500185542
[39] Keller-Besrest, F.; Collin, G.: Structural aspects of the α transition in stoichiometric FeS: identification of the high-temperature phase, *J. Solid State Chem*., 1990 **84**, 194-210 DOI: 10.1016/0022-4596(90)90319-S
[40] Ma, C.; Beckett, J.R.; Rossman, G.R.: Buseckite, (Fe,Zn,Mn)S, a new mineral from the Zakłodzie meteorite, *Am. Mineral*., 2012 **97**, 1226–1233 DOI: DOI: 10.2138/am.2012.4110
[41] Xie, X.; Chen, M.: Suizhou Meteorite: Mineralogy and shock metamorphism (Springer geochemistry/mineralogy, Guangdong Science and Technology Press Co., Ltd., Guangzhou, China) 2016 pp. 64-85

[42] Lennie, A.R.; Redfern, S.A.T.; Schofield, P.F.; Vaughan, D.J.: Synthesis and Rietveld crystal structure refinement of mackinawite, tetragonal FeS, *Mineral. Mag.*, 1995 **59**, 677-683

[43] Rémazeilles, C.; Saheb, M.; Neff, D.; Guilminot, E.; Tran, K.; Bourdoiseau, J.-A.; Sabot, R.; Jeannin, M.; Matthiesen, H.; Dillmann, P.; Refait, P.: Micro-Raman spectra of FeS phases, *J. Raman Spectr.*, 2010 **41**(11), 1425-1433 DOI: 10.1002/jrs.2717

[44] Colomban, P.: Potential and drawbacks of Raman (micro)spectrometry for the understanding of iron and steel corrosion (in Ed.: Chiaberge, M.; New trends and developments in automotive system engineering, InTech Europe, Rijeka, Croatia) 2011 Chapter 28, pp. 567-584 DOI: 10.5772/13436

[45] Rickard, D.; Luther, G.W.: Chemistry of iron sulfides, *Chem. Rev.*, 2007 **107**(2), 514-562 DOI: 10.1021/cr0503658

[46] Ren, C.; Li, D.; Bai, Z.; Li, T.: Corrosion behavior of oil tube steel in stimulant solution with hydrogen sulfide and carbon dioxide, *Mater. Chem. Phys.*, 2005 **93**(2-3), 305-309 DOI: 10.1016/j.matchemphys.2005.03.010

[47] Smith, S.N.: Current understanding of corrosion mechanisms due to H_2S in oil and gas production environments, *Nace Int.*, 2015 Corrosion Paper no. 5485

[48] Choi, H.J.; Al-Bannai, N.S.; Al-Beheiri, F.I.; Warnken, D.K.: Field corrosion assessment of l80 carbon steel downhole production tubing in Khuff gas wells, *Nace Int.*, 2006 Corrosion Paper no. 6653

[49] Dehkordi, E.H.; Tavakoli, A.R.: The effect of time on the formation and growth of passive layer on carbon steel A516 GR.70 in sour water, *Int. J. Iron Steel Soc. Iran*, 2007 **4**(1-2), 28-33

[50] Shoesmith, D.W.; Taylor, P.; Bailey, M.G.; Owen, D.G.: The formation of ferrous monosulfide polymorphs during the corrosion of iron by aqueous hydrogen sulfide at 21 ºC, *J. Electrochem. Soc.*, 1980 **127**(5), 1007-1015 DOI: 10.1149/1.2129808

[51] Wikjord, A.G.; Rummery, T.E.; Doern, F.E.; Owen, D.G.: Corrosion and deposition during the exposure of carbon steel to hydrogen sulphide-water solutions, *Corr. Sci.*, 1980 **20**(5), 651-671 DOI: 10.1016/0010-938X(80)90101-8

[52] Stringer, J.: Stress generation and adhesion in growing oxide scales, *Mat. Corr.*, 1972 **23**(9), 747-755 DOI 10.1002/maco.v23:9

[53] Singer, M.; Brown, B.; Camacho, A.; Nesic, S.: Combined effect of CO_2, H_2S and acetic acid on bottom of the line corrosion, *Nace Int.*, 2007, Paper no. 07661

[54] Sorell, G.; Hoyt, W.B.: Collection and correlation of high temperature hydrogen sulfide data, *Corr.*, 1956 **12**(5), 33-54 DOI: 10.5006/0010-9312-12.5.33

[55] Backensto, E.B.; Drew, R.D.; Stapleford, C.C.: High temperature hydrogen sulfide corrosion, *Corr.*, 1956 **12**(1), 6-16 DOI: 10.5006/0010-9312-12.1.22

[56] Meyer, F.H.; Riggs, O.L.; McGlasson, R.L.; Sudbury, J.D.: Corrosion products of mild steel in hydrogen sulfide environments, *Corr.*, 1958 **14**(2), 69–75 DOI: 10.5006/0010-9312-14.2.69

[57] Anderko, A.; Shuler, P.J.: A computational approach to predicting the formation of iron sulfide species using stability diagrams, *Comp. Geosci.*, 1997 **23**(6), 647–658 DOI: 10.1016/S0098-3004(97)00038-1

[58] Wagner, C.: Theory of the tarnishing process, *Z Physik Chem., Abteil B: Chem. Element. Proz., Aufbau Mater.,* 1933 **B21**, 25-29 (in German)

[59] Gulbrandsen, E.: Acetic acid and carbon dioxide corrosion of carbon steel covered with iron carbonate, *Nace Int.*, 2007, Paper no.. 07322

[60] Silva, C.C.; Farias, J.P.; de Sant'Ana, H.B.: Evaluation of AISI 316L stainless steel welded plates in heavy petroleum environment, *Mat. Design*, 2009 **30**(5), 1581–1587 DOI: 10.1016/j.matdes.2008.07.050

[61] Turnbull, A.; Slavcheva, E.; Shone, B.: Factors controlling naphthenic acid corrosion, *Corr.*, 1998 **54**(11), 922-930 DOI: 10.5006/1.3284811

[62] Tebbal, S.; Kane, R.D.: Assessment of crude oil corrosivity, *Nace Int.*, 1998, Corrosion Paper no. 98578

[63] Babaian-Kibala, E.; Nugent, M.J.: Naphthenic acid corrosion literature survey, *NACE Int.*, 1999, Corrosion Paper no. 99378

[64] Kane, R.D.; Cayard, M.S.: A comprehensive study on naphthenic acid corrosion, *NACE Int.*, 2002, Corrosion Paper no. 02555

[65] Hancock, P.; Hurst, R.C.: The mechanical properties and breakdown of surface oxide films at elevated temperatures (in Eds: Staehle, R.W.; Fontana, M.G., Advances in Corrosion Science and Technology, Plenum, New York, USA) 1974 Vol. 4, pp. 1-84

[66] Evans, A.G.; Hutchinson, J.W.: On the mechanics of delamination and spalling in compressed films, *Int. J. Solids Structures*, 1984 **20**(5), 455-466 DOI: 10.1016/0020-7683(84)90012-X

Guide for Authors

1. Electronic copies of manuscripts should be uploaded to the Author's Pages at **hjic.mk.uni-pannon.hu** or emailed directly to **hjic@almos.uni-pannon.hu**.
   - Please visit the Journal's website for downloading a **Word or Latex template** and follow closely the suggested layout for the text, figures, tables, and references. Manuscripts without the recommended structure and formatting for publication will be returned without a review.
   - The use of template is highly desired in order to avoid imposing any excessive article processing charges (APCs), page fees, or article submission charges.
2. In a brief letter attached to each manuscript, ***authors must declare that their work is original*** and has not previously been published elsewhere. The editorial board requests an originality evaluation score to be below 10% according to CrossCheck/Ithenticate excluding quotes, references, and three-word expressions. If originality of a given manuscript is between 10-20% the authors are kindly requested to rewrite the overlapping sections. Nothing above 20% similarity will be accepted for publication and will not be sent out for review.
3. For the purposes of correspondence, the authors are asked to give their current ***address***, ***telephone***, ***professional website (if applicable), and regularly accessed E-mail address***.
4. ***Authors are requested to adhere to the rules*** given below and follow the examples:
   - The layout of manuscripts should be A4 page size (21 cm × 29.7 cm) with double columns, single-spaced (approximately 45 characters per line, 55 lines per page), and numbered pages. ***The text of the papers should be concise, and not exceed twelve pages***; even with Tables and Figures the manuscript should not exceed a total of twelve pages. These limits may be lifted for reviews and invited articles for themed issues.
   - The first page should give the title, the full name (no abbreviations please) of the author(s) in the sequence to be published, the name and address of the institution where the work was completed, and a brief summary of the article (maximum 15 lines). The title of the paper should not exceed 15 words. At least three keywords must be specified, but not more than seven.
   - The corresponding author or multiple authors should be clearly marked.
   - The article should include the following sections: Introduction, Experimental or Theoretical Methods, Results and Analysis, Discussion, and Conclusion. A Conclusion section is mandatory with at least a paragraph presenting the main outcome(s) of the given study, not just repeating the abstract or summarizing the study.
   - ***Tables and Figures must be inserted into the text*** at their closest position to the first mention, flushed with the top margin as much as possible. Drawings must be of high contrast and have continuous curves. Please attach the source of the figure/scheme graphics if available in a resolution of at least 300 dpi. Characters and labels should be sufficiently large to allow for reproduction. All figures must contain decimal points and not commas.
   - Note that the electronic deposition of the published version of the manuscript will contain colored graphics; however, the printed version will only be reproduced in black and white unless otherwise requested for a fee.
   - The Tables and Figures should not exceed one A/4 size page, (***maximum width of Figures: portrait 8 cm; landscape 16.5 cm***). If diagrams are presented, only the explanation and dimensions of the abscissa, ordinate and marking numbers are required, further explanatory texts can be given in the caption. Please remove any excess white space around the figures.
   - Equations should always be stand-alone, i.e. occupy an extra line and marked with Arabic numerals in parentheses on the right-hand margin. Efforts avoiding equations in two columns are appreciated.
   - The summary of the symbols used must be included at the end of the manuscripts after the Conclusion section under the heading: ‘‘SYMBOLS”. Symbols and abbreviations that represent variables, constants, quantities, properties etc., must be defined in the text at their first occurrence.
   - ***Only SI (System International d'Unites) units are to be used.*** If data with non-SI units are reported, they should be in parentheses following the corresponding data in SI units.
5. References should be numbered throughout the text and written in square brackets and not in superscript position. The references are also to be listed at the end of the paper. *Abbreviations of journal titles should conform to international standards, as accepted by Chemical Abstracts*. The style and punctuation of the references should conform to the ACS Style Guide. For a quick reference please consult the following examples:
   a) Journal articles: Debye, P.; Hückel, E.: The theory of electrolytes I. Lowering of freezing point and related phenomena, *Physik. Z.* 1923 **24**(1), 185–206
   b) Books: Verwey, E.J.W.; Overbeek, J.T.G.: Stability of Lyophobic Colloids (Elsevier, Amsterdam) 1948
   c) Book Chapters: Gunn, A.M.; Winnard, D.A.; Hunt, D.T.E.: Trace metal speciation in sediment and soils, in *Metal speciation: Theory, analysis and application* (Eds.: Kramer, J.R.; Allen, H.E.: Lewis, Boca Raton, FL, USA) 1988
   d) Patents: U.S. Pat. 3,425,442 (1984)
   e) Published lectures, conference proceedings: Hih, S.; Hu, C.; Leech, W.J.: A Discrete Method to Heat Conduction Problems, *Proc. 5th Int. Heat Transfer Conf., Tokyo, Japan*, 1975 pp. 2-4. Citations of conference presentations are strongly discouraged and should be avoided unless the talk was made available electronically.
   f) Papers that are unpublished, but have been submitted to a journal may be cited with the journal's name, followed by “submitted for publication” or “in press”. This will be accepted if the author uploads the submitted manuscript as ‘editorial’ material. No reference to “unpublished work” or “personal communication” will be accepted.
   g) All references must be provided with a digitial object identifier (DOI or arXiv). Please provide the editorial office with a rationale for any citations without DOI or arXiv numbers and explain why these references are necessary.
6. Authors receive galley proofs, which they are to be corrected and returned as soon as possible, but no later than 1 week after receipt. No new material may be inserted into the text at the time of proofreading.
7. ***Authors who are less experienced with written English are urged to have their manuscripts checked by scientists working in the Authors' respective field and proficient in English***.

Editorial Board of the
Hungarian Journal of Industry and Chemistry