Data-Driven Optimal Closures for Mean-Cluster Models: Beyond the Classical Pair Approximation

Avesta Ahmadi

School of Computational Science & Engineering,

McMaster University, Hamilton, Ontario, Canada, L8S 4L8

Jamie M. Foster

School of Mathematics & Physics, University of Portsmouth,

Portsmouth, Hampshire, UK, PO1 2UP

Bartosz Protas*

Department of Mathematics & Statistics,

McMaster University, Hamilton, Ontario, Canada, L8S 4L8

(Dated: March 1, 2022)

Abstract

This study concerns the mean-clustering approach to modelling the evolution of lattice dynamics. Instead of tracking the state of individual lattice sites, this approach describes the time evolution of the concentrations of different cluster types. It leads to an infinite hierarchy of ordinary differential equations which must be closed by truncation using a so-called closure condition. This condition approximates the concentrations of higher-order clusters in terms of the concentrations of lower-order ones. The pair approximation is the most common form of closure. Here, we consider its generalization, termed the "optimal approximation", which we calibrate using a robust data-driven strategy. To fix attention, we focus on a recently proposed structured lattice model for a nickel-based oxide, similar to that used as cathode material in modern commercial Li-ion batteries. The form of the obtained optimal approximation allows us to deduce a simple sparse closure model. In addition to being more accurate than the classical pair approximation, this "sparse approximation" is also physically interpretable which allows us to a posteriori refine the hypotheses underlying construction of this class of closure models. Moreover, the mean-cluster model closed with this sparse approximation is linear and hence analytically solvable such that its parametrization is straightforward. On the other hand, parametrization of the mean-cluster model closed with the pair approximation is shown to lead to an ill-posed inverse problem.

I. INTRODUCTION

Evolution of particles on a structured lattice is typically described by discrete lattice models rather than continuous space models. These models are usually not solvable exactly and have to be studied through computer simulations. One approach to describing the evolution of particles on a structured lattice is to keep track of all interacting particles as is done in various Monte-Carlo techniques such as simulated annealing. However, these methods are costly as they determine the lattice structure which is unnecessary in many applications. What is often sufficient is knowledge of the type and the number of different clusters in the lattice, which can then be used for model fitting purposes along with experimental measurements such as, e.g., Nuclear Magnetic Resonance (NMR) data [1]. Hence, as an alternative to Monte-Carlo methods, one can develop a simplified description of particle

---

* bprotas@mcmaster.ca; https://ms.mcmaster.ca/bprotas/

interactions in terms of evolving probabilities of particle clusters of different types in the form of a dynamical system which is sufficient for many applications. These approaches are referred to as "mean-field clustering methods" and find applications in many areas of science and engineering. The Ising model, as a canonical application of mean-field methods, is a model of ferromagnetism describing the evolution of magnetic moments in a lattice. Both Monte-Carlo methods [2] and mean-field methods [3] have been employed to study this problem. Another example of the application of such models is the contact process which is a stochastic process describing the growth of a population on a structured or unstructured lattice. Cluster approximations are used to find mean-field properties of such systems. Population dynamics in ecology [4, 5] is one example of such processes. Another example is the disease spread in epidemiology that has been widely studied on structured networks [6–12] and complex networks [13, 14]. Failure propagation [15] and emergence of marriage networks [16] are some other examples of contact processes.

The focus of the present study is on cluster-based modelling of systems of interacting particles on two-dimensional (2D) structured lattices. The specific application which motivates the present study is related to prediction of the structure of materials used in Lithium-ion batteries [1]. Using a cluster approximation method, one can construct a hierarchical dynamical system describing the evolution of concentrations of different clusters in the lattice during a real annealing process. In other words, the evolution of concentrations of clusters of size n involves concentrations of clusters of size $(n + 1)$. To solve this system of equations one is required to close it by prescribing the evolution of concentrations of $(n + 1)$ clusters, which in turn will be determined by probabilities of clusters of a still higher order. This process therefore gives rise to an infinite hierarchy of equations which is exact but is intractable both analytically and computationally. Thus, one needs to truncate and close this infinite hierarchy of equations. Various moment closure approximations have been used for this purpose. Ben-Avraham et al. [17] proposed a class of approximations for 1D lattices with extensions developed in higher dimensions, namely, the mean-field and pair approximations. These techniques take into account local interactions between neighbouring elements only and completely neglect interactions between non-nearest neighbours on a lattice. Applications of mean-field and pair approximation methods to various problems in science and engineering can be found in [9, 13, 18, 19] and in [6, 7, 9, 13, 14, 18–20], respectively. Some

extensions of the pair approximation technique are also introduced in [21] where interactions between different elements are considered to be generic functions of distance. In the present study our goal is to develop and validate a general data-driven methodology that will allow us to optimally close (in a mathematically precise sense) the infinite hierarchy of equations. We will refer to this approach as the "optimal approximation". This approach leads to a general simple and mathematically interpretable closure model.

As an emerging application of lattice dynamics, Harris et al. [1] used a simulated annealing approach to investigate the crystalline structure of cathode materials used in state-of-the-art Lithium-ion batteries. More precisely, they focused on layers of NMC (Nickel-Manganese-Cobalt) used in most modern commercial Li-ion batteries. These cathodes are described by the chemical formula $Li(NMC)O_2$, where 2D layers of Lithium, Oxygen and NMC are stacked on top of each other. The capacity enhancement observed in such materials is attributed to changes in the local microscopic structure of the cathode layers [22, 23], however, important aspects of this structure are not yet completely understood. Hence, further refinement of this battery technology requires more information about the arrangement of elements inside these layers. In [1] simulated annealing was used to generate statistical information about arrangements of different species on the lattice in the NMC layer of a cathode, which was very costly and did not scale up to large lattice sizes. The model developed in the present study aims to address this limitation. While the proposed approach is general and can be applied to many lattice systems, to fix attention, we will develop it here for the problem from [1] as an example. Other applications of approaches based on lattice dynamics in physics and chemistry include organic synthesis reactions in the fields of heterogeneous catalysis and materials engineering [24], adsorption models of binary mixtures [25] and microstructure mapping of perovskite materials [26].

In this work, we use the mean-clustering approach to build a hierarchical system of equations for the evolution of concentrations of different clusters inside a structured lattice of the NMC cathode layer. We assume a triangular lattice compatible with the structure of the NMC layer [1]. This spatial structure is important in detecting the rotational symmetries of the system. A dynamical system is constructed to describe reactions between different species which are limited to swaps between nearest-neighbour elements. The underlying principle is that as the "temperature" decreases the lattice converges to a certain equilibrium state

through a series of element swaps, controlled by specific rate constants. Our new approach consists of two distinct steps: first, the truncated hierarchical dynamical system is closed using an optimal approximation whose parameters are inferred from simulated annealing data; it is demonstrated that such an optimal closure is in fact both simpler and more accurate than the nearest-neighbour approximation proposed in [17]. Additionally, robustness of the predictive performance of the obtained model is demonstrated based on problems with different stoichiometries. Second, the reaction rates parameterizing the dynamical system with the three types of closure, i.e., pair approximation, optimal approximation and sparse approximation, are inferred from the simulated annealing data using a Bayesian approach which also allows us to estimate the uncertainty of these reconstructions; this will show that the model with the optimal closure is also less prone to calibration uncertainty than the model closed with the nearest-neighbour approximation.

The paper is organized as follows: further details about our model problem are presented in Section II; then, in Section III we introduce a dynamical system governing the evolution of the concentrations of different clusters and in Section IV we describe the closures we consider which are the pair approximation and the optimal closure; Bayesian approach for estimation of the reaction rates is introduced in Section V; computational results are presented in Section VI together with a justification for a suitably sparse approximation, whereas discussion and conclusions are deferred to Section VII. Some technical material is collected in Appendix A.

II.   MODEL PROBLEM

In this section we provide some details about a lattice evolution problem that will serve as our test case. In [1] Harris et al. used a simulated annealing method to identify an evolving arrangement of particles on the lattice and keep track of their interactions. One material similar to the materials actually used in Li-ion batteries is $Li[Li_{1/3}Mn_{2/3}]O_2$, where 2D sheets of an oxygen layer, transition metal layer and Lithium layer are stacked on top of each other, as shown in Figure 1. Transition metal layer consists of Manganese and Lithium.

In the simulated annealing method, the energy of the system is calculated by considering the local charge neutrality at oxygen sites. Each oxygen element is surrounded by six
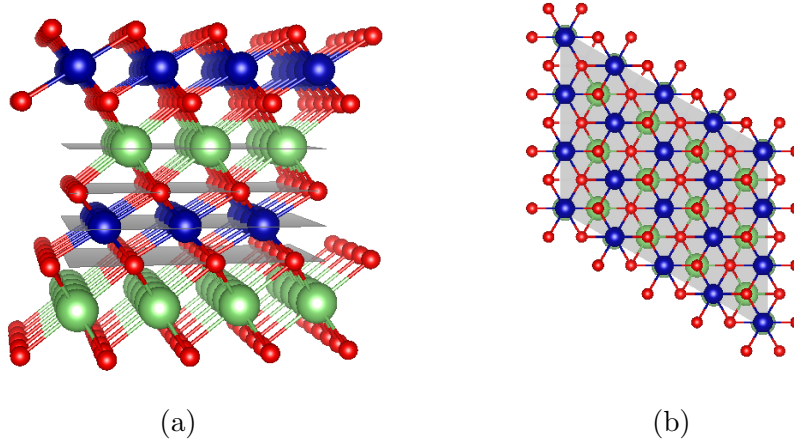
FIG. 1: The $Li[Li_{1/3}Mn_{2/3}]O_2$ lattice considered in [1] and shown here in (a) a 3D view and (b) a 2D view. The red elements are oxygen atoms, green elements are lithium and the blue elements represent the transition metal layer elements which can be either lithium or manganese.

nearest neighbours, cf. Figure 1. The energy of each oxygen site is then determined by considering the charge contributions of the neighbouring sites to its charge balance with the goal of achieving neutrality. The simulated annealing approach attempts to find a 2D lattice configuration minimizing the total energy of the system $E = \sum_i E_i$ corresponding to a specific "temperature", where E_i is the energy over each oxygen site. This is a probabilistic approach to finding global optima in a discrete space which mimics the annealing process applied to actual materials. These materials are annealed at a high temperature, followed by quenching to the desired temperature. Higher energy levels of the system occur at higher temperatures and the evolution of the system from higher temperatures to lower ones is controlled by a probabilistic rule via the Boltzmann distribution. This distribution gives the probability that the system is in a certain state, given the temperature and its energy level. In simulated annealing, state updates occur through random swaps of the elements on the lattice which destroys the ordering of elements and can lower the energy level of the lattice. The acceptance probability of an element swap is

$$P = \begin{cases} 1, & \Delta E \leq 0, \\ \exp\left(\frac{-\Delta E}{T}\right), & \Delta E > 0, \end{cases} \tag{1}$$

where T plays the role of the "temperature". Swaps leading to more probable lower-energy

states are always accepted, however, swaps producing higher-energy states may also be accepted with some probability in order to prevent rapid quenching of the system which might result in convergence to a local minimum. We note that the concept of "temperature" used here is not equivalent to the thermodynamic temperature of the system. The choice of this pseudo-temperature, which controls the annealing protocol of the system, requires knowledge of the change of the system energy resulting from an element swap. In other words, the magnitude of the exponent $\frac{-\Delta E}{T}$ determines the quenching rate of the annealing process such that T needs to be suitably adjusted. This effective parameter T is therefore related to the term $k_B \theta$ in the Boltzmann distribution, where k_B is the Boltzmann constant and θ the thermodynamic temperature of the system. The choice of how the temperature is decreased is in principle arbitrary, however, the equilibrium state must be reached at the end of the annealing process for every arbitrarily chosen temperature profile. The details of this approach can be found in [1].

In the crystal structure of the annealed metal layer of $\text{Li}[\text{Li}_{1/3}\text{Mn}_{2/3}]\text{O}_2$ each triangle consists of two Mn elements and one Li element. In this structure, the energy E_i over each oxygen site becomes zero and the total energy of the system will be zero accordingly, as shown in Figure 3b. In the simulated annealing study of this structure the temperature was reduced in a stepwise manner, cf. Figure 2a, and enough time was allowed for the structure to stabilize at an equilibrium at each intermediate temperature, cf. Figure 2b. The results obtained for the system with $\text{Li}_{1/3}\text{Mn}_{2/3}$ are shown in the form of the final lattice structure in Figure 3. Annealing experiments with the same protocol were also performed for systems with different ratios of Li and Mn in $\text{Li}_x\text{Mn}_{1-x}$ where $x \in \{0.25, 0.30, 0.33, 0.36, 0.42, 0.50, 0.58, 0.64, 0.70, 0.75\}$, but these results are not shown here for brevity. Our goal is to build a model that will accurately predict the evolution of concentrations of different particle clusters present in the lattice without having to solve the entire annealing problem. We note that the elements Mn and Li have charges, respectively, of $(+4)$ and $(+1)$. In simulated annealing these element charges are used to calculate the energy changes ΔE caused by element swaps and to determine the evolution of the lattice. However, the cluster approximation model, cf. Section III, makes no assumptions about the charges of the elements and hence for simplicity the symbols $(+)$ and $(-)$ will hereafter represent the elements Mn and Li, respectively. The concentrations $\widetilde{C}_i$, $i \in \{(++), (+-), (--)\}$ of 2-clusters as functions of time (or temperature),
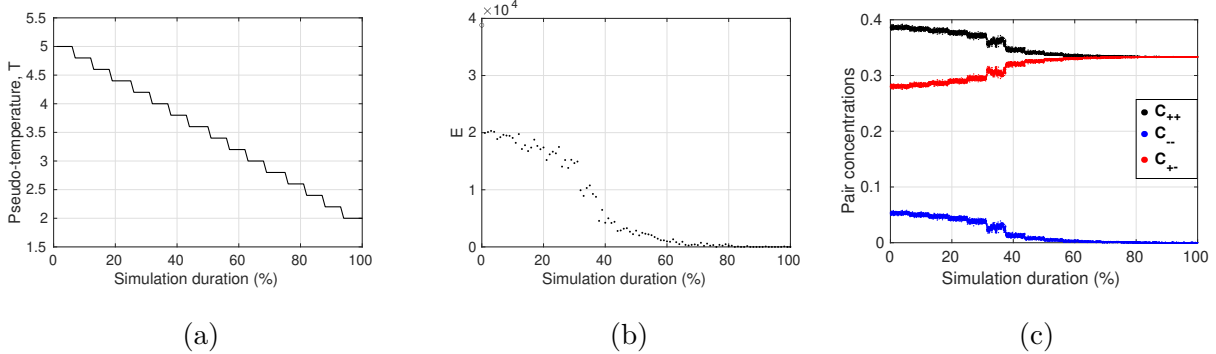
FIG. 2: Dependence of (a) the "temperature" of the system, T, (b) its lattice energy E and (c) concentrations $\widetilde{C}_i$, $i \in \{(++),(+-),(--)\}$, of different 2-clusters on time expressed as a fraction of the entire annealing experiment for the $\mathrm{Li}_{1/3}\mathrm{Mn}_{2/3}$ system.

cf. Figure 2c, will be used as data to construct the optimal closure approximation and to infer the reaction rates in the model. However, while these concentrations will be provided for a single stoichiometry only, the resulting model will be shown to remain accurate for a broad range of stoichiometries. The lattice evolution in this method does not have a natural time scale and for concreteness we will assume that the unit of time is set by an individual iteration of the simulated annealing experiment. Notably, in this model all concentrations are independent of location on the lattice due to spatial homogeneity.

III.   CLUSTER APPROXIMATION

In this section we develop a system of evolution equations for concentrations of clusters in a two-element system with elements denoted $(+)$ (or $\oplus$) and $(-)$ (or $\ominus$). We note that these notations need not correspond to the charge of the elements. In this study, a cluster of size n is referred to as a n-cluster and elements inside the cluster form a closed or an open chain. The concentration of a cluster is defined as the probability of finding that particular cluster among all clusters of the same shape but with different compositions. As an example, the concentration of the 3-cluster shown in Figure 4 is denoted C_{ijk}, where $i,j,k \in \{+,-\}$.

Remark 1. *The normalization condition requires that the sum of the concentrations of all*
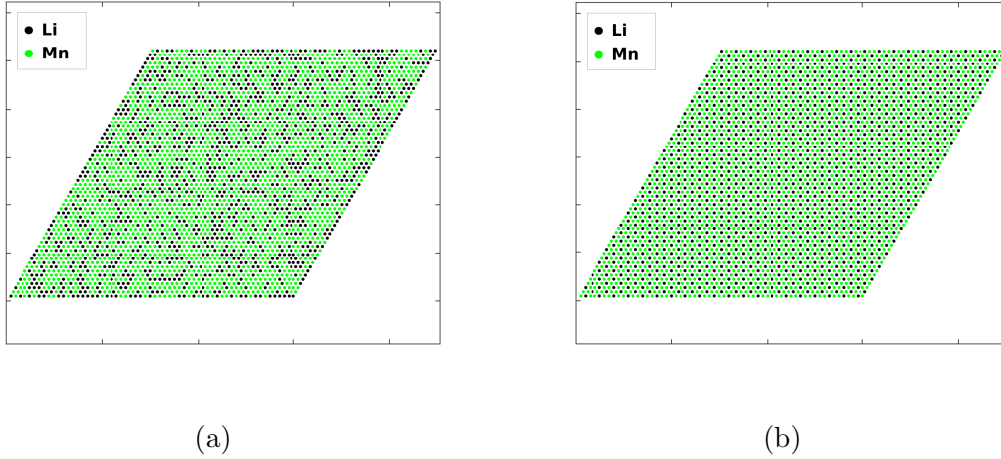
(a)   (b)

FIG. 3: (a) Initial random state and (b) the final ordered state of the lattice for the $Li_{1/3}Mn_{2/3}$ system obtained via simulated annealing [1]. Black and green dots represent Li ions (more generally, negative elements) and Mn ions (more generally, positive elements), respectively.
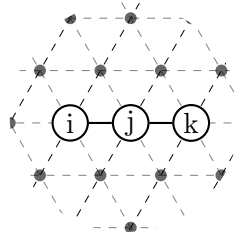

FIG. 4: An example of a linear chain 3-cluster on a 2D lattice.

possible n-clusters with the same geometry must be equal to one [17]:

$$\sum_{S_1,S_2,\ldots,S_n} C_{S_1 S_2 \cdots S_n} = 1, \qquad (2)$$

where the indices $1, 2, 3, \ldots, n$ enumerate different sites within a cluster with two consecutive ones corresponding to nearest neighbours and $S_i \in \{+, -\}$ denotes the state of that specific site. Applying this to 1-clusters and 2-clusters in our model, the following equations are derived from the normalization condition:

$$C_+ + C_- = 1 \qquad (3a)$$

$$C_{++} + C_{--} + C_{+-} + C_{-+} = 1$$
$$\Rightarrow C_{++} + C_{--} + 2C_{+-} = 1. \qquad (3b)$$

The concentrations of the $(+-)$ and $(-+)$ clusters are the same due to the rotational symmetry of the system, as stated in Theorem A.1 in the Appendix.

The aim is to deduce a dynamical system describing the evolution of the probabilities of 2-clusters. There are three different types of 2-clusters found on the lattice, namely, $\oplus\oplus$, $\ominus\ominus$ and $\oplus\ominus$.

A.   Production and Destruction of 2-Clusters

The rate of change of the concentration of specific clusters is determined by the rate at which they are produced and destroyed. Production or destruction of a certain cluster occurs through swaps among nearest-neighbour elements on the lattice. Each swap of nearest-neighbour elements is called here a reaction. The rate equations can then be derived using the window method [17]. In this approach we consider all possible reactions that change the composition of a particular 2-cluster in a certain window containing this cluster, via a swap between one of the elements inside the window and one of its nearest-neighbour elements outside the window. For example, in order to derive the rate equation for the $(\oplus\oplus)$ cluster, in Figure 5 we show all possible reactions that will produce or destroy this cluster via nearest-neighbour element swaps. In each of the reactions, the neighbour element (highlighted in red) will swap with one of the elements of the window (highlighted in blue) to produce a $(\oplus\oplus)$ cluster in the forward reaction. Conversely, reverse reactions destroy the $(\oplus\oplus)$ cluster and produce a $(\oplus\ominus)$ cluster. The rotational symmetry of the lattice allows us to reduce the number of possible reactions to those shown in Figure 5. Moreover, reactions taking place inside a triangular-shaped 3-cluster do not change the total count of 2-clusters inside the triangle and are therefore disregarded. Each reaction has a unique rate constant denoted $k_1, k_2, \ldots$. The rate constants have the units of 1/sec and control the evolution of different clusters participating in a reaction. We note that in deriving the rate equations each reaction is accounted for in proportion to the number of its rotational symmetries on the lattice.

As can be observed in Figure 5, 3-clusters with three types of bonds are involved in the derivation of rate equations. The first type is the linear 3-cluster in which the two bonds are colinear. The second type is the cluster in which there is an obtuse angle of 120 degrees
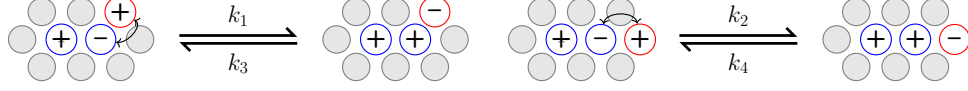
FIG. 5: All unique (up to rotational and translational symmetries) reversible reactions to destroy or produce clusters ($\oplus\oplus$) and ($\oplus\ominus$).
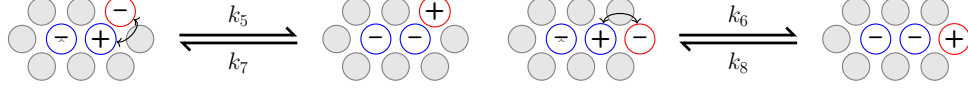

FIG. 6: All unique (up to rotational and translational symmetries) reversible reactions to destroy or produce clusters ($\ominus\ominus$) and ($\oplus\ominus$).

between the bonds due to the triangular shape of the lattice. The third type is the triangular cluster in which the elements form a triangle with 60 degrees between the bonds. We will refer to these as the linear, angled and triangular clusters, respectively. For simplicity, linear clusters will be represented as a combination of elements with a straight line $[\overline{(\bullet\,\bullet\,\bullet)}]$, angled clusters as a combination of elements with a hat sign $[\overset{\frown}{(\bullet\,\bullet\,\bullet)}]$ and triangular clusters as a combination of elements with a triangle $[\overset{\frown}{\overline{(\bullet\,\bullet\,\bullet)}}]$, where $\bullet$ is either $+$ or $-$. The set of all 3-cluster types will be denoted

$$
\begin{aligned}
\Theta = \Big\{ & \overline{+++}, \overline{---}, \overline{++-}, \overline{--+}, \overline{+-+}, \overline{-+-}, \\
& \overset{\frown}{+++}, \overset{\frown}{---}, \overset{\frown}{+-+}, \overset{\frown}{-+-}, \overset{\frown}{++-}, \overset{\frown}{--+}, \\
& \overset{\frown}{\overline{+++}}, \overset{\frown}{\overline{---}}, \overset{\frown}{\overline{++-}}, \overset{\frown}{\overline{--+}} \Big\}.
\end{aligned}
\tag{4}
$$

The rate equations for the ($\ominus\ominus$) and ($\oplus\ominus$) clusters can be derived in a similar way, by considering all possible reactions that produce or destroy these two clusters as shown in Figure 6. We thus obtain the following system of rate equations for the concentrations C_{++}, C_{--} and C_{+-}

$$
\frac{d}{dt}C_{++} = 4k_1 C_{\overset{\frown}{+-+}} + 2k_2 C_{\overline{+-+}} - 4k_3 C_{\overset{\frown}{++-}} - 2k_4 C_{\overline{++-}},
\tag{5a}
$$

$$
\frac{d}{dt}C_{--} = 4k_5 C_{\overset{\frown}{-+-}} + 2k_6 C_{\overline{-+-}} - 4k_7 C_{\overset{\frown}{--+}} - 2k_8 C_{\overline{--+}},
\tag{5b}
$$

$$
\begin{aligned}
\frac{d}{dt}C_{+-} = {} & 2k_3 C_{\overset{\frown}{++-}} + 2k_7 C_{\overset{\frown}{--+}} + k_4 C_{\overline{++-}} + k_8 C_{\overline{--+}} \\
& - 2k_1 C_{\overset{\frown}{+-+}} - 2k_5 C_{\overset{\frown}{-+-}} - k_2 C_{\overline{+-+}} - k_6 C_{\overline{-+-}}.
\end{aligned}
\tag{5c}
$$

An important aspect of system (5) is its hierarchical structure in the sense that the rates of change of concentrations of 2-clusters are given in terms of the concentrations of 3-clusters and if one were to write down equations for their rates of change they would involve concentrations of 4-clusters, etc. Thus, system (5) is not closed and needs to be truncated which we will do so here at the level of 2-clusters. Two strategies for closing the truncated system are discussed in Section IV.

In addition, the normalization condition (3b) can be modified to a dynamic form by taking the derivative with respect to time

$$\frac{d}{dt}C_{++} + \frac{d}{dt}C_{--} + 2\frac{d}{dt}C_{+-} = 0. \tag{6}$$

As can be verified, this equation is satisfied automatically by system (5a)–(5c). Moreover, the rate of the forward reaction will be equal to the rate of corresponding reverse reaction in the chemical equilibrium. As we are interested in the equilibrium state of reactions, the following relations can be written for each pair of forward and reverse reactions in equilibrium

$$k_1 C_{\overparen{+-+}} = k_3 C_{\overparen{++-}} \qquad \Longrightarrow \qquad Q_1 = \frac{k_1}{k_3} = \frac{C_{\overparen{++-}}}{C_{\overparen{+-+}}}, \tag{7a}$$

$$k_2 C_{\overgroup{+-+}} = k_4 C_{\overgroup{++-}} \qquad \Longrightarrow \qquad Q_2 = \frac{k_2}{k_4} = \frac{C_{\overgroup{++-}}}{C_{\overgroup{+-+}}}, \tag{7b}$$

$$k_5 C_{\overparen{-+-}} = k_7 C_{\overparen{--+}} \qquad \Longrightarrow \qquad Q_3 = \frac{k_5}{k_7} = \frac{C_{\overparen{--+}}}{C_{\overparen{-+-}}}, \tag{7c}$$

$$k_6 C_{\overgroup{-+-}} = k_8 C_{\overgroup{--+}} \qquad \Longrightarrow \qquad Q_4 = \frac{k_6}{k_8} = \frac{C_{\overgroup{--+}}}{C_{\overgroup{-+-}}}, \tag{7d}$$

where Q_i, $i = 1, \ldots, 4$, denote the equilibrium constants for each reversible reaction.

IV. CLOSURE APPROXIMATIONS

In this section we discuss two strategies for closing system (5), by which we mean expressing the concentration of 3-clusters on the right-hand side (RHS) of this system in terms of a suitable function of the concentrations of 2-clusters. In other words, the goal is to replace each of the triplet concentrations C_i, $i \in \Theta$, in (5) with suitably chosen functions

$g_i(C_+, C_-, C_{++}, C_{--}, C_{+-})$, such that the closed system will have the form

$$\frac{d}{dt}C_{++} = 4k_1 g_{\overparen{+-+}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) + 2k_2 g_{\overline{+-+}}(C_+, C_-, C_{++}, C_{--}, C_{+-})$$
$$-4k_3 g_{\overparen{++-}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) - 2k_4 g_{\overline{++-}}(C_+, C_-, C_{++}, C_{--}, C_{+-}),$$
$$(8a)$$

$$\frac{d}{dt}C_{--} = 4k_5 g_{\overparen{-+-}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) + 2k_6 g_{\overline{-+-}}(C_+, C_-, C_{++}, C_{--}, C_{+-})$$
$$-4k_7 g_{\overparen{--+}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) - 2k_8 g_{\overline{--+}}(C_+, C_-, C_{++}, C_{--}, C_{+-}),$$
$$(8b)$$

$$\frac{d}{dt}C_{+-} = 2k_3 g_{\overparen{++-}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) + 2k_7 g_{\overparen{--+}}(C_+, C_-, C_{++}, C_{--}, C_{+-})$$
$$+k_4 g_{\overline{++-}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) + k_8 g_{\overline{--+}}(C_+, C_-, C_{++}, C_{--}, C_{+-})$$
$$-2k_1 g_{\overparen{+-+}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) - 2k_5 g_{\overparen{-+-}}(C_+, C_-, C_{++}, C_{--}, C_{+-})$$
$$-k_2 g_{\overline{+-+}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) - k_6 g_{\overline{-+-}}(C_+, C_-, C_{++}, C_{--}, C_{+-}) .$$
$$(8c)$$

 The first approach to finding these functions is the pair approximation based on the classical method introduced in [17] and the second is a new optimal closure approximation. The problem of finding the rate constants $k_1, \ldots, k_8$ in (5) will be addressed in Section V.

A. Pair Approximation

The pair approximation is a classical approach to closing truncated hierarchical dynamical systems. It was first used by Dickman [27] in a surface-reaction model and later by Matsuda et al. [4] for a structured lattice appearing in a population dynamics problem. In our model, we use the pair approximation approach in order to close the dynamical system (5) at the level of 2-clusters. The state of a site is denoted $i, j, k \in \{+, -\}$ for a two-element system. Global concentrations are denoted C_i giving the probability that a randomly chosen site in the lattice is in state $i \in \{+, -\}$. Similarly, C_{ij} is the global concentration of 2-clusters in state ij. In addition, *local* concentrations are denoted $P_{j|i}$ and give the conditional probability that a randomly chosen nearest neighbour of a site in state i is in state j. These local concentrations can be expressed in terms of global concentrations using the

rules governing conditional probabilities as [4, 28]

$$C_{ij} = C_{ji} = C_i P_{j|i} = C_j P_{i|j}, \tag{9a}$$

$$\sum_{i \in \{+,-\}} C_i = 1, \tag{9b}$$

$$\sum_{i \in \{+,-\}} P_{i|j} = 1 \quad for \ any \ j \in \{+, -\}. \tag{9c}$$

Equation (9a) is invariant with respect to the rotational symmetries of the lattice, cf. Appendix A. Also, the global concentration of a triplet in state (ijk) can be derived in a similar approach as Equation (9a),

$$C_{ijk} = C_i P_{j|i} P_{k|ij} = C_{ij} P_{k|ij}. \tag{10}$$

The $P_{k|ij}$ term in this equation involves 3 elements in a triplet. In order to break down the triplet concentration in terms of pair and singlet concentrations, one is required to find an equivalent expression for the $P_{k|ij}$ term. The underlying assumption of the pair approximation method is to neglect the interaction between the non-nearest neighbour elements, i and k in this case, according to Figure 4 [4, 5, 28]. This results in an approximation at the level of 3-clusters expressed in terms of quantities defined at the level of 2-clusters as

$$P_{k|ij} \approx P_{k|j}. \tag{11}$$

A different approach could also be adopted to derive the pair approximation formulation resulting in the same closure model. In this approach, assuming a triplet in state (ijk) on a random lattice (in which all non-nearest-neighbour elements are decoupled), the global concentration of this triplet can be written as

$$C_{ijk} = C_i C_j C_k Q_{ij} Q_{jk} T_{ijk}, \tag{12a}$$

$$Q_{ij} = \frac{C_{ij}}{C_i C_j}, \tag{12b}$$

where C_i, C_j and C_k denote the global concentrations of singlets, Q_{ij} and Q_{ik} are the pair correlations of nearest neighbours and T_{ijk} is the triple correlation of the chain. Note that element i and element k on a random lattice are considered not to be nearest-neighbours. Also, there is no factor Q_{ik} in equation (12a) as the correlation of non-nearest-neighbours is represented by T_{ijk}. According to the underlying assumption of pair approximation, the
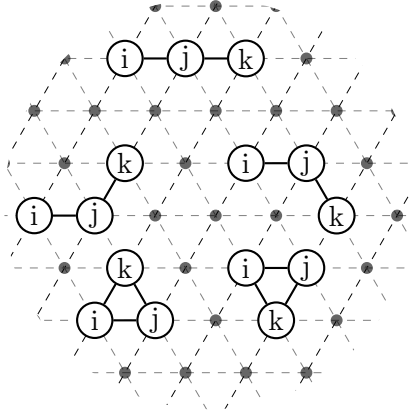
FIG. 7: Schematic of a 2D triangular lattice with chains of 3-clusters with 180-degree bonds, 120-degree bonds, and 60-degree bonds. These cluster types are referred to, respectively, as linear, angled and triangular throughout this text. The clumping intensity of this lattice is equal to the proportion of the triangles over all triplets types, which is equal to $\frac{2}{5}$.

non-nearest-neighbour elements are decoupled. There is no deterministic way of calculating correlations of non-nearest neighbour elements [29] and some additional assumptions have to be made in order to close (5). The standard pair approximation method neglects all triple correlations such that $T_{ijk} = 1$. This is an equivalent approximation to Equation (11).

Each regular lattice can be described by two parameters: the number of neighbours per site (m) and the proportion of triangles to triplets (θ), which determines the clumping intensity of the lattice. A triangular lattice has $m = 6$ neighbours per site and $\theta = \frac{2}{5}$, as shown in Figure 7. Similarly, chain-like triplets in a triangular lattice can be categorized into two groups: linear triplets with 180-degree bonds, and angled triplets with 120-degree bonds. As is evident from Figure 7, the probability of finding a triplet in a closed form, angled form and linear form is equal to $\frac{2}{5}$, $\frac{2}{5}$, and $\frac{1}{5}$, respectively. As the shape of triplets is important in our model, these probabilities have to be taken into account as coefficients when calculating the corresponding concentrations. Morris [30] and Keeling [11] have proposed formulas for approximating the fraction of closed and open chains in a certain state (ijk) on a regular lattice by taking into account the clumping effect of triangles in the lattice. Following these

studies, the concentrations of each type of triplet are approximated as

$$C_{\overline{ijk}} \approx g_{\overline{ijk}} = (1-\theta)\frac{1}{3}\frac{C_{ij}C_{jk}}{C_j}, \tag{13a}$$

$$C_{\widehat{ijk}} \approx g_{\widehat{ijk}} = (1-\theta)\frac{2}{3}\frac{C_{ij}C_{jk}}{C_j}, \tag{13b}$$

$$C_{\widehat{\widehat{ijk}}} \approx g_{\widehat{\widehat{ijk}}} = \theta\frac{C_{ij}C_{jk}C_{ki}}{C_iC_jC_k}, \tag{13c}$$

where $\overline{ijk}$ denotes a linear cluster, $\widehat{ijk}$ denotes an angled triplet with 120-degree bonds and $\widehat{\widehat{ijk}}$ denotes a triangular cluster with 60-degree bonds. The specific forms taken by expressions (13a)–(13b) for different $i,j,k \in \{+,-\}$ are collected in Table I. Applying this pair approximation to close system (5) gives

$$\frac{d}{dt}C_{++} = 4k_1\frac{2}{3}(1-\theta)\frac{C_{+-}^2}{C_-} + 2k_2\frac{1}{3}(1-\theta)\frac{C_{+-}^2}{C_-}$$
$$- 4k_3\frac{2}{3}(1-\theta)\frac{C_{++}C_{+-}}{C_+} - 2k_4\frac{1}{3}(1-\theta)\frac{C_{++}C_{+-}}{C_+}, \tag{14a}$$

$$\frac{d}{dt}C_{--} = 4k_5\frac{2}{3}(1-\theta)\frac{C_{+-}^2}{C_+} + 2k_6\frac{1}{3}(1-\theta)\frac{C_{+-}^2}{C_+}$$
$$- 4k_7\frac{2}{3}(1-\theta)\frac{C_{--}C_{+-}}{C_-} - 2k_8\frac{1}{3}(1-\theta)\frac{C_{--}C_{+-}}{C_-}. \tag{14b}$$

We note that equation (5c) is eliminated from the system of equations as we use the normalization condition $\frac{d}{dt}C_{++} + \frac{d}{dt}C_{--} + 2\frac{d}{dt}C_{+-} = 0$, cf. (6), to close the system.

B. Optimal Approximation

As will be shown in Section VI A, the closure based on the pair approximation introduced above is not very accurate. In order to improve the accuracy of the closure, here we propose a new approach based on nonlinear regression analysis of simulated annealing data. This is a data-driven strategy where an *optimal* form of the closure is obtained by fitting an expression in an assumed well-justified form to the data. The pair approximation scheme attempts to predict the concentrations of the higher-order clusters in terms of concentrations of lower-order ones using expressions with the functional forms given in (13). In the new approach, we close system (5) using relations generalizing the expressions in (13) which depend on a number of adjustable parameters. These parameters, representing the exponents of different concentrations, are then calibrated against the simulated annealing data by solving a suitable

constrained optimization problem. Information about the new more general closure relations and how they compare to the pair approximation for different 3-clusters is collected in Table I where we also group the parameters to be determined in the vector $\mathbf{V}_i$, with $i \in \Theta$ representing different cluster types.

Notably, the new functional forms are generalizations of the expressions used in the pair approximation obtained by allowing for more freedom in how the new expressions for closures depend on the cluster concentrations. The numerators of the new expressions involve concentrations of *all* nearest-neighbour 2-clusters such that the effect of non-nearest-neighbour clusters is still neglected. The denominators, on the other hand, involve the concentrations of singlets present in the triplet which makes the functional form of the new closure different from the pair approximation in some cases. The parameters (exponents) defining the proposed optimal closures in Table I are subject to the following constraints ensuring well-posedness of the resulting system (8)

1. the difference of the sums of the exponents in the numerators and in the denominators is equal to one, i.e., $\sum_j \gamma_j - \sum_j \xi_j = 1$, ensuring that the terms representing the closure have the units of concentration,

2. the exponents in the numerators need to be non-negative, i.e., $\gamma_j \geq 0$, since otherwise the corresponding terms representing the closure model may become unbounded as the concentration approaches zero, causing solutions of the ODE system (8) to blow up,

3. the exponents in the numerators need to be bounded $\gamma_1, \gamma_2 \leq \delta$, where δ is the upper bound on the exponent which needs to be specified, as otherwise the corresponding terms representing the closure model may also become large causing solutions of the ODE system (8) to blow up,

4. while denominators involve concentrations of singlets only, which are time independent, in some cases it is necessary to restrict the corresponding exponents as otherwise the terms representing the closure model will have large prefactors which may also cause the solutions of the ODE system (8) to blow up; hence, we impose $\beta_1 \leq \xi_1, \xi_2 \leq \beta_2$, where β_1 and β_2 are the lower and upper bounds on the exponents to be specified;

Triplet Type i	Pair Approximation $g_i(C_+,\ldots,C_{+-})$	Optimal Approximation $g_i(C_+,\ldots,C_{+-};\mathbf{V}_i)$	Parameters (exponents) $\mathbf{V}_i$
$\overline{+++}$	$\dfrac{1}{5}\dfrac{C_{++}^2}{C_+}$	$\dfrac{1}{5}\dfrac{C_{++}^{\gamma_1}}{C_+^{\xi_1}}$	$\mathbf{V}_{\overline{+++}}=[\gamma_1\ \xi_1]$
$\overline{---}$	$\dfrac{1}{5}\dfrac{C_{--}^2}{C_-}$	$\dfrac{1}{5}\dfrac{C_{--}^{\gamma_1}}{C_-^{\xi_1}}$	$\mathbf{V}_{\overline{---}}=[\gamma_1\ \xi_1]$
$\overline{+-+}$	$\dfrac{1}{5}\dfrac{C_{+-}^2}{C_-}$	$\dfrac{1}{5}\dfrac{C_{+-}^{\gamma_1}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\overline{+-+}}=[\gamma_1\ \xi_1\ \xi_2]$
$\overline{-+-}$	$\dfrac{1}{5}\dfrac{C_{+-}^2}{C_+}$	$\dfrac{1}{5}\dfrac{C_{+-}^{\gamma_1}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\overline{-+-}}=[\gamma_1\ \xi_1\ \xi_2]$
$\overline{++-}$	$\dfrac{1}{5}\dfrac{C_{++}C_{+-}}{C_+}$	$\dfrac{1}{5}\dfrac{C_{++}^{\gamma_1}C_{+-}^{\gamma_2}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\overline{++-}}=[\gamma_1\ \gamma_2\ \xi_1\ \xi_2]$
$\overline{--+}$	$\dfrac{1}{5}\dfrac{C_{--}C_{+-}}{C_-}$	$\dfrac{1}{5}\dfrac{C_{--}^{\gamma_1}C_{+-}^{\gamma_2}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\overline{--+}}=[\gamma_1\ \gamma_2\ \xi_1\ \xi_2]$
$\widehat{+++}$	$\dfrac{2}{5}\dfrac{C_{++}^2}{C_+}$	$\dfrac{2}{5}\dfrac{C_{++}^{\gamma_1}}{C_+^{\xi_1}}$	$\mathbf{V}_{\widehat{+++}}=[\gamma_1\ \xi_1]$
$\widehat{---}$	$\dfrac{2}{5}\dfrac{C_{--}^2}{C_-}$	$\dfrac{2}{5}\dfrac{C_{--}^{\gamma_1}}{C_-^{\xi_1}}$	$\mathbf{V}_{\widehat{---}}=[\gamma_1\ \xi_1]$
$\widehat{+-+}$	$\dfrac{2}{5}\dfrac{C_{+-}^2}{C_-}$	$\dfrac{2}{5}\dfrac{C_{+-}^{\gamma_1}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\widehat{+-+}}=[\gamma_1\ \xi_1\ \xi_2]$
$\widehat{-+-}$	$\dfrac{2}{5}\dfrac{C_{+-}^2}{C_+}$	$\dfrac{2}{5}\dfrac{C_{+-}^{\gamma_1}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\widehat{-+-}}=[\gamma_1\ \xi_1\ \xi_2]$
$\widehat{++-}$	$\dfrac{2}{5}\dfrac{C_{++}C_{+-}}{C_+}$	$\dfrac{2}{5}\dfrac{C_{++}^{\gamma_1}C_{+-}^{\gamma_2}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\widehat{++-}}=[\gamma_1\ \gamma_2\ \xi_1\ \xi_2]$
$\widehat{--+}$	$\dfrac{2}{5}\dfrac{C_{--}C_{+-}}{C_-}$	$\dfrac{2}{5}\dfrac{C_{--}^{\gamma_1}C_{+-}^{\gamma_2}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\widehat{--+}}=[\gamma_1\ \gamma_2\ \xi_1\ \xi_2]$
$\widehat{+++}$	$\dfrac{2}{5}\dfrac{C_{++}^3}{C_+^3}$	$\dfrac{2}{5}\dfrac{C_{++}^{\gamma_1}}{C_+^{\xi_1}}$	$\mathbf{V}_{\widehat{+++}}=[\gamma_1\ \xi_1]$
$\widehat{---}$	$\dfrac{2}{5}\dfrac{C_{--}^3}{C_-^3}$	$\dfrac{2}{5}\dfrac{C_{--}^{\gamma_1}}{C_-^{\xi_1}}$	$\mathbf{V}_{\widehat{---}}=[\gamma_1\ \xi_1]$
$\widehat{++-}$	$\dfrac{2}{5}\dfrac{C_{++}C_{+-}^2}{C_+^2 C_-}$	$\dfrac{2}{5}\dfrac{C_{++}^{\gamma_1}C_{+-}^{\gamma_2}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\widehat{++-}}=[\gamma_1\ \gamma_2\ \xi_1\ \xi_2]$
$\widehat{--+}$	$\dfrac{2}{5}\dfrac{C_{--}C_{+-}^2}{C_+ C_-^2}$	$\dfrac{2}{5}\dfrac{C_{--}^{\gamma_1}C_{+-}^{\gamma_2}}{C_+^{\xi_1}C_-^{\xi_2}}$	$\mathbf{V}_{\widehat{--+}}=[\gamma_1\ \gamma_2\ \xi_1\ \xi_2]$

TABLE I: The functional forms of the closures based on the pair approximation and on the proposed optimal closures for each triplet type. Unknown parameters (exponents) are indicated in the last column.

Optimal parameters $\mathbf{V}_i$ of the closure model are obtained separately for each cluster type i by minimizing the mean-square error between the experimental concentration data $\widetilde{C}_i(t)$

obtained from simulated annealing experiments, and the predictions of the corresponding ansatz function $g_i(\widetilde{C}_+, \widetilde{C}_-, \widetilde{C}_{++}(t), \widetilde{C}_{--}(t), \widetilde{C}_{+-}(t); \mathbf{V}_i)$, cf. Table I, obtained with the parameter vector $\mathbf{V}_i$ over the time window $[0, T]$, where T corresponds to the end of the simulated annealing process. Then, for each $i \in \Theta$, error functional is defined as

$$J_i(\mathbf{V}_i) = \frac{1}{2} \int_0^T \left[g_i(\widetilde{C}_+, \widetilde{C}_-, \widetilde{C}_{++}(t), \widetilde{C}_{--}(t), \widetilde{C}_{+-}(t); \mathbf{V}_i) - \widetilde{C}_i(t) \right]^2 dt \qquad (15)$$

which leads to the following family of constrained optimization problems

$$\min_{\mathbf{V}_i} \; J_i(\mathbf{V}_i),$$

$$\text{subject to:} \quad \begin{cases} 0 \leq \gamma_j \leq \delta, & 1 \leq j \leq \Gamma_i \\ \beta_1 \leq \xi_j \leq \beta_2, & 1 \leq j \leq \Xi_i, \\ \sum_j \gamma_j - \sum_j \xi_j = 1 \end{cases} \qquad (16)$$

for each $i \in \Theta$, where $\Gamma_i, \Xi_i \in \{1, 2\}$ are the numbers of the exponents appearing in the numerator and the denominator for a given cluster type, cf. Table I.

We note that choosing different values of the adjustable parameters δ, β_1 and β_2, which determine how stringent the constraints in the optimization problem (16) are, has the effect of regularizing the solutions of this problem. We will consider the following two cases (when the lower/upper bound is equal to $-\infty/\infty$, this means that effectively there is no bound)

- "soft" regularization with $\beta_1 = -\infty$, $\beta_2 = \infty$, $\delta = 6$, and

- "hard" regularization with $\beta_1 = 0$ and $\beta_2 = \delta = 2$.

In each case optimization problem (16) is solved numerically in MATLAB using the nonlinear programming routine `fmincon`. The optimal closures determined in these two ways are compared to the pair approximation in Section VI A.

V. DETERMINING REACTION RATES VIA BAYESIAN INFERENCE

In order for the truncated model (8) closed with either the pair or optimal approximation to predict the time evolution of 2-cluster concentrations, it must be equipped with correct values of the rate constants $k_1, \ldots, k_8$, cf. Figures 5 and 6. Here we show how these constants can be determined by solving an appropriate inverse problem. It will be demonstrated that

this problem is in fact ill-posed and a suitable solution will be obtained using Bayesian inference which also provides information about the uncertainty of this solution.

We define the error functional as

$$\mathcal{J}(\mathbf{K}) = \frac{1}{2}\int_0^T \left\|\mathbf{C}(t,\mathbf{K}) - \widetilde{\mathbf{C}}(t)\right\|_2^2 dt + \alpha\left\|\mathbf{Q}(\mathbf{K}) - \widetilde{\mathbf{Q}}\right\|_2^2, \tag{17}$$

where $\widetilde{\mathbf{C}}(t) = \left[\widetilde{C}_{++}(t), \widetilde{C}_{--}(t), \widetilde{C}_{+-}(t)\right]$ is the vector of pair concentrations obtained from the simulated annealing experiment, cf. Figure 2c, $\mathbf{K} = [k_1, k_2, \cdots, k_8]$ is the vector of unknown rate constants, and $\mathbf{C}(t,\mathbf{K})$ is the vector of pair concentrations predicted by model (8) equipped with the rate constants $\mathbf{K}$. The second term in (17) is the mean-square error between the equilibrium constants $\mathbf{Q}(\mathbf{K}) = [Q_1, Q_2, Q_3, Q_4]$, cf. relation (7), predicted by model (8) equipped with parameters $\mathbf{K}$ and the equilibrium constant $\widetilde{\mathbf{Q}} = [\widetilde{Q}_1, \widetilde{Q}_2, \widetilde{Q}_3, \widetilde{Q}_4]$ obtained experimentally via simulated annealing. We note that the equilibrium constants in (7) are written in terms of 3-cluster concentrations and one of the closure models (i.e., the pair or the optimal approximation) is used to express the equilibrium constants in terms of 2-cluster concentrations. The parameter α weights the relative importance of matching the equilibrium constants versus matching the time-dependent concentrations in (17).

The optimal reaction rates are then obtained by solving the problem

$$\min_{\mathbf{K}\in\mathbb{R}^8} \mathcal{J}(\mathbf{K}) \tag{18}$$
$$\text{subject to system (8)}$$

separately for the case of the pair and the optimal approximations. We note that the minimization problems (16) and (18) are in fact quite different: in the former the mismatch between the evolution of 3-cluster concentrations is minimized with respect to a suitably-parameterized structure of the closure model, whereas in the latter one seeks to minimize the mismatch between the evolution of 2-clusters in order to find the optimal reaction rates in the closed system (8).

Inverse problems such as (18) are often ill-posed, in the sense that they usually do not admit a unique exact solution, but rather many, typically infinitely many, approximate solutions. This is a result of the presence of multiple local minima, which is a consequence of the non-convexity of the error functional (17), and the fact that these minima are often "shallow"

reflecting weak dependence of the model predictions $\mathbf{C}$ on the parameters $\mathbf{K}$. As will be evident from the results presented in Section VI, it is thus not very useful to solve problem (18) directly using standard methods of numerical optimization [31]. Instead, we will adopt a probabilistic approach based on Bayesian inference where the unknown parameters in the vector $\mathbf{K}$ and the corresponding model predictions $\mathbf{C}$ will be represented in terms of suitable conditional probability densities. This will allow us to systematically assess the relative uncertainty of the many approximate solutions admitted by problem (18). The mathematical foundations of Bayesian inference are reviewed in the monographs [32–34].

In the Bayesian framework the distribution of the model parameters is given by the *posterior* probability distribution $\mathbb{P}\left(\mathbf{K}|\widetilde{\mathbf{C}}\right)$ defined as the probability of obtaining parameters $\mathbf{K}$ given the observed experimental data $\widetilde{\mathbf{C}}$. According to Bayes' rule, we then have

$$\mathbb{P}\left(\mathbf{K}|\widetilde{\mathbf{C}}\right) = \frac{\mathbb{P}\left(\widetilde{\mathbf{C}}|\mathbf{K}\right)\mathbb{P}\left(\mathbf{K}\right)}{\mathbb{P}\left(\widetilde{\mathbf{C}}\right)}, \tag{19}$$

where $\mathbb{P}\left(\widetilde{\mathbf{C}}|\mathbf{K}\right)$ is the *likelihood* function describing the likelihood of obtaining observations $\widetilde{\mathbf{C}}$ given the model parameters $\mathbf{K}$, $\mathbb{P}\left(\mathbf{K}\right)$ is the *prior* probability distribution reflecting some a priori assumptions on the parameters $\mathbf{K}$ (based, e.g., on direct measurements or literature data), whereas $\mathbb{P}\left(\mathbf{C}\right)$ can be viewed as a normalizing factor.

A common approach to choosing the prior distribution $\mathbb{P}\left(\mathbf{K}\right)$ is to use an uniform distribution, leading to the so-called uninformative prior, and this is the approach we adopt here. As regards the likelihood function, it is usually defined as

$$\mathbb{P}\left(\widetilde{\mathbf{C}}|\mathbf{K}\right) \propto e^{-\mathscr{J}(\mathbf{K})}. \tag{20}$$

This definition of the likelihood function arises from the fact that parameter values are considered more likely if they produce model predictions $\mathbf{C}$ closer to the data $\widetilde{\mathbf{C}}$. Moreover, if the error functional is a quadratic function of the model parameters $\mathbf{K}$, then the distribution in (20) is Gaussian.

The main challenge is efficient sampling of the likelihood function $\mathbb{P}\left(\widetilde{\mathbf{C}}|\mathbf{K}\right)$ and this can be performed using a Markov-Chain Monte-Carlo (MCMC) approach. It is a form of a random walk in the parameter space designed to preference the sampling of high-likelihood regions of the space while also exploring other regions. MCMC methods are commonly used

to sample arbitrary distributions known up to a normalizing factor. In particular, these methods are used to sample distributions in high dimensions where exploration of the entire space with classical methods is computationally intractable. MCMC techniques have found applications in many different fields such as electrochemistry [35], medical imaging [36, 37], environmental and geophysical sciences [38, 39] and ecology [40].

In the MCMC algorithm, a kernel $\mathcal{Q}\left(\mathbf{K}^*|\mathbf{K}\right)$ is used to generate a proposal for a move in the parameter space from the current point $\mathbf{K}$ to a new point $\mathbf{K}^*$. This new point is accepted with a probability given by the Hastings ratio; otherwise, it is rejected (the "Metropolis rejection"). In order to preserve the reversibility of the Markov chain, the Hastings ratio for the acceptance probability is defined as

$$\alpha\left(\mathbf{K}^*, \mathbf{K}\right) = \min\left\{1, \frac{\mathbb{P}\left(\mathbf{K}^*|\widetilde{\mathbf{C}}\right)\mathcal{Q}\left(\mathbf{K}|\mathbf{K}^*\right)}{\mathbb{P}\left(\mathbf{K}|\widetilde{\mathbf{C}}\right)\mathcal{Q}\left(\mathbf{K}^*|\mathbf{K}\right)}\right\}. \tag{21}$$

Thus, the Markov chain is reversible with respect to the posterior distribution, meaning that a transition in space is equally probable during forward and backward evolution. This property makes the posterior distribution invariant on the Markov chain. In other words, if given enough iterations, the distribution converges to its equilibrium distribution. The most common choice of the random walk is in the form

$$\mathbf{K}^* = \mathbf{K} + \boldsymbol{\xi} \tag{22}$$

such that $\mathcal{Q}\left(\mathbf{K}^*|\mathbf{K}\right) = \mathcal{Q}\left(\mathbf{K}^* - \mathbf{K}\right) = \mathcal{Q}\left(\boldsymbol{\xi}\right)$, where $\boldsymbol{\xi}$ is an 8-dimensional random variable drawn from a uniform distribution with scale $\boldsymbol{\sigma} \in \mathbb{R}^8$, i.e., $\boldsymbol{\xi} \sim \mathcal{U}\left[-\boldsymbol{\sigma}, \boldsymbol{\sigma}\right]$. Note that the components of the scale $\boldsymbol{\sigma}$ represent intervals defining the uniform distribution. It has been suggested that uniform kernels outperform Gaussian ones in terms of convergence of the MCMC algorithm [41], hence, we adopt the uniform kernel in our study. The choice of symmetric kernels simplifies relation (21) as the factors representing the density in the numerator and denominator cancel. However, the choice of scale for the proposal kernel is nontrivial. Small scales will result in slow convergence to the posterior distribution, whereas large scales will prevent sampling of desirable regions in the parameter space. Moreover, in our model there is no prior information about an appropriate scale for the proposal kernel. In order to tackle this issue, a two-step Delayed-Rejection Metropolis-Hastings (DR-MH) algorithm is used [38, 42, 43]. In this algorithm, the rejection of the first proposed point

at a given iteration of the Markov chain is delayed by proposing a new step in the space based on a different scale. Normally, the scale of the first kernel is chosen to be large in order to explore a wider region of the high-dimensional parameter space and the scale of the second kernel is small to gather more samples from higher-likelihood regions. This approach combines exploration of large regions in a high dimensional space with focus on high-likelihood neighbourhoods. The DR-MH algorithm also ensures the reversibility of the Markov chain, meaning that the direction of time in which the random walk is taking place does not affect the dynamics of the Markov chain. In other words, a random walk in the forward direction of the chain from state n to state $n+1$ is equally probable as the reverse walk from state $n+1$ to state n. This ensures that the chain remains in an equilibrium state as it evolves. This is an important property as the Markov chain is essentially a random walk in the posterior space and reversibility is required to ensure it remains in the same posterior space. The acceptance probability of the delayed proposed point is calculated using relation (24). To initialize the DR-MH algorithm, we require an initial set of model parameters. They could be random, without any prior information about the parameters. As an alternative, we determine this initial point $\boldsymbol{\mathcal{Y}}_1$ by solving problem (18) using a standard numerical optimization method [31] several times with random initial guesses and then taking the best parameter set corresponding to the lowest value of the cost functional $\mathscr{J}(\mathbf{K})$. Algorithm 1 outlines the entire procedure needed to approximate the posterior probability distribution $\mathbb{P}\left(\mathbf{K}|\widetilde{\mathbf{C}}\right)$. Additional details concerning MCMC approaches can be found in monographs [44, 45].

Algorithm 1: Random walk delayed rejection algorithm

Input: M — Number of samples to be drawn from the posterior distribution

$\mathcal{Q}_1\left(\mathbf{K}^*|\mathbf{K}\right)$ — Proposal density of the first trial

$\boldsymbol{\mathcal{Y}}_1$ — initial point for the random walk in the space $\mathbb{R}^8$

$\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2$ — scales defining the random walk

Output: $\mathbb{P}\left(\mathbf{K}|\widetilde{\mathbf{C}}\right)$ — Posterior probability distribution

---

$n \leftarrow 1$

repeat

    Propose a step: $\boldsymbol{\xi} \sim \mathcal{U}\left[-\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1\right]$

    Propose a candidate: $\boldsymbol{\mathcal{Y}}_1 = \mathbf{K}^{n-1} + \boldsymbol{\xi}$

    Accept the proposed step with probability α_1:

$$\alpha_1\left(\boldsymbol{\mathcal{Y}}_1, \mathbf{K}^{n-1}\right) \propto \min\left\{1, \frac{\exp(-\mathscr{J}\left(\boldsymbol{\mathcal{Y}}_1\right))}{\exp(-\mathscr{J}\left(\mathbf{K}^{n-1}\right))}\right\} \qquad (23)$$

    Draw a random number: $r \sim \mathcal{U}\left[0, 1\right]$

    if $\alpha_1\left(\boldsymbol{\mathcal{Y}}_1, \boldsymbol{K}^{n-1}\right) < r$ **then**

        Propose a new step with scale $\boldsymbol{\sigma}_2$: $\boldsymbol{\xi} \sim \mathcal{U}\left[-\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2\right]$

        Propose a new candidate: $\boldsymbol{\mathcal{Y}}_2 = \mathbf{K}^{n-1} + \boldsymbol{\xi}$

        Accept the new proposed point with probability α_2:

$$\alpha_2\left(\mathbf{K}^{n-1}, \boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{Y}}_2\right) = \min\left\{1, \frac{\mathbb{P}\left(\boldsymbol{\mathcal{Y}}_2|\widetilde{\mathbf{C}}\right)\mathcal{Q}_1\left(\boldsymbol{\mathcal{Y}}_2|\boldsymbol{\mathcal{Y}}_1\right)\left[\max\left(0, 1 - \frac{\mathbb{P}(\boldsymbol{\mathcal{Y}}_1|\widetilde{\mathbf{C}})}{\mathbb{P}(\boldsymbol{\mathcal{Y}}_2|\widetilde{\mathbf{C}})}\right)\right]}{\mathbb{P}\left(\mathbf{K}^{n-1}|\widetilde{\mathbf{C}}\right)\mathcal{Q}_1\left(\boldsymbol{\mathcal{Y}}_1|\mathbf{K}^{n-1}\right)\left[\max\left(0, 1 - \frac{\mathbb{P}(\boldsymbol{\mathcal{Y}}_1|\widetilde{\mathbf{C}})}{\mathbb{P}(\mathbf{K}^{n-1}|\widetilde{\mathbf{C}})}\right)\right]}\right\}$$

$$(24)$$

        Draw a random number: $r \sim \mathcal{U}\left[0, 1\right]$

        if $\alpha_2\left(\boldsymbol{K}^{n-1}, \boldsymbol{\mathcal{Y}}_1, \boldsymbol{\mathcal{Y}}_2\right) > r$ **then**

            $\mathbf{K}^n \leftarrow \boldsymbol{\mathcal{Y}}_2$

        $n \leftarrow n + 1$

        else

            discard $\boldsymbol{\mathcal{Y}}_2$

    else

        $\mathbf{K}^n \leftarrow \boldsymbol{\mathcal{Y}}_1$

        $n \leftarrow n + 1$

until M *samples are drawn*;

Construct the posterior probability distribution

In our computations reported in Section VI D we employ an uniform positive prior $\mathbb{P}(\mathbf{K})$, whereas the scales of the first and second trial of the two-step delayed rejection algorithm are defined as the initial guess $\boldsymbol{\mathcal{Y}}_1$ multiplied by a factor of 0.1 and 0.01, respectively. The total number of samples in the Markov chain is $M = 10^5$.

VI. RESULTS

In this section we first determine the optimal structure of the closure models given in Table I by solving optimization problem (16) for each type of 3-cluster in the set Θ, cf. (4), as described in Section IV B. Then, based on these results, in Section VI B we propose a new closure model which we refer to as *Sparse Approximation* (SA) and in Section VI C we assess the predictive capability of the considered models by analyzing how accurately they predict the time evolution of 3-cluster concentrations for a range of different stoichiometries. Finally, we determine the reaction rates in the truncated model (8) closed with the pair approximation, optimal approximation and sparse approximation using Bayesian inference to solve problem (18), as described in Section V.

A. Optimal Closures

Parameters of the closure relations given in Table I are determined separately for each cluster type by solving problem (16) and the obtained results are collected in the form of the values of the exponents in Table II, where, for comparison we also show the exponents corresponding to the pair approximation, cf. Section IV A. We recall that for each 3-cluster type problem (16) is solved with both soft and hard regularization. In Table II, the optimal results are presented for solving problem (16) subject to hard regularization ($\beta_1 = 0$, $\delta = \beta_2 = 2$) by separately fitting the closure models to the data obtained for two systems with $\mathrm{Li}_{1/2}\mathrm{Mn}_{1/2}$ and $\mathrm{Li}_{1/3}\mathrm{Mn}_{2/3}$. The first system is interesting since, as we shall see below, due to the symmetry in the concentrations of Li and Mn, closure models calibrated based on the data from this system are particularly robust with respect to different stoichiometries. The second system is considered in our analysis due to its interesting behaviour at low temperatures where physically relevant crystalline microstructure are obtained, as discussed in Section II. This system is also used as a benchmark in [1]. In Table II we note that most of the exponents

in the optimal closure approximation tend to be different from the corresponding exponents in the pair approximation. Interestingly, we observe that many exponents obtained for the optimal closure by fitting to the data for the system $Li_{1/2}Mn_{1/2}$ are equal to zero or one, opening the possibility of finding a simpler closure model to be investigated in Section VI B.

The accuracy of representing the concentrations of 3-clusters based on 2-cluster concentrations is investigated for different closure approximations in Figure 8 for 4 representative triplet types, namely, $\overline{+++}$, $\overline{-+-}$, $\overline{\smile{---}}$ and $\overline{\frown{---}}$. A significant improvement is evident for most cluster types when the optimal closure is used. This is confirmed in quantitative terms in Figure 9 showing the mean-square error (15) for each 3-cluster type for the pair approximation and the optimal closure fitted to $Li_{1/2}Mn_{1/2}$ and $Li_{1/3}Mn_{2/3}$ systems. For both systems and for almost all 3-cluster types the optimal closure leads to a more accurate description with errors (15) smaller by a few orders of magnitude than when the pair approximation is used. In the next section we will simplify the obtained optimal closure and will propose an interpretation of the resulting structure.

Triplet Type	PA	OA-1/3	OA-1/2	PA	OA-1/3	OA-1/2	PA	OA-1/3	OA-1/2	PA	OA-1/3	OA-1/2
	γ_1			ξ_1								
$\overline{+\,+\,+}$	2	1.12	1.00	1	0.12	0.00	-	-	-	-	-	-
$\overline{-\,-\,-}$	2	1.19	1.00	1	0.19	0.00	-	-	-	-	-	-
$\widehat{+\,+\,+}$	2	1.00	1.00	1	0.00	0.00	-	-	-	-	-	-
$\widehat{-\,-\,-}$	2	1.39	1.00	1	0.39	0.00	-	-	-	-	-	-
$\widehat{+\,+\,+}$	3	2.00	2.00	3	0.99	0.99	-	-	-	-	-	-
$\widehat{-\,-\,-}$	3	1.76	1.18	3	0.76	0.18	-	-	-	-	-	-
	γ_1			ξ_1			ξ_2					
$\overline{+\,-\,+}$	2	1.00	0.99	1	0.00	0.00	0	0.00	0.00	-	-	-
$\overline{-\,+\,-}$	2	2.00	0.99	1	0.99	0.00	0	0.00	0.00	-	-	-
$\widehat{+\,-\,+}$	2	1.00	1.00	1	0.00	0.00	0	0.00	0.00	-	-	-
$\widehat{-\,+\,-}$	2	0.99	1.00	1	0.00	0.00	0	0.00	0.00	-	-	-
	γ_1			γ_2			ξ_1			ξ_2		
$\overline{+\,+\,-}$	1	2.00	0.00	1	0.00	1.00	1	0.00	0.00	0	1.00	0.00
$\overline{-\,-\,+}$	1	0.38	0.00	1	0.62	1.00	1	0.00	0.00	0	0.00	0.00
$\widehat{+\,+\,-}$	1	2.00	0.72	1	0.52	0.28	1	1.52	0.00	0	0.00	0.00
$\widehat{-\,-\,+}$	1	0.66	0.15	1	0.34	0.85	1	0.00	0.00	0	0.00	0.00
$\widehat{+\,+\,-}$	1	2.00	0.00	2	0.00	1.00	2	0.00	0.00	1	0.99	0.00
$\widehat{-\,-\,+}$	1	0.60	0.00	2	0.40	1.00	1	0.00	0.00	2	0.00	0.00

TABLE II: Exponents defining the optimal closure models, cf. Table I, found by solving problem (16) with hard regularization ($\beta_1 = 0$, $\beta_2 = \delta = 2$) based on the data for the system $\mathrm{Li}_{1/3}\mathrm{Mn}_{2/3}$ (OA-1/3) and the system $\mathrm{Li}_{1/2}\mathrm{Mn}_{1/2}$ (OA-1/2) for each 3-cluster type indicated in the first column. For comparison, the exponents characterizing the pair approximation (PA) are also shown. The results are rounded to two decimal places.
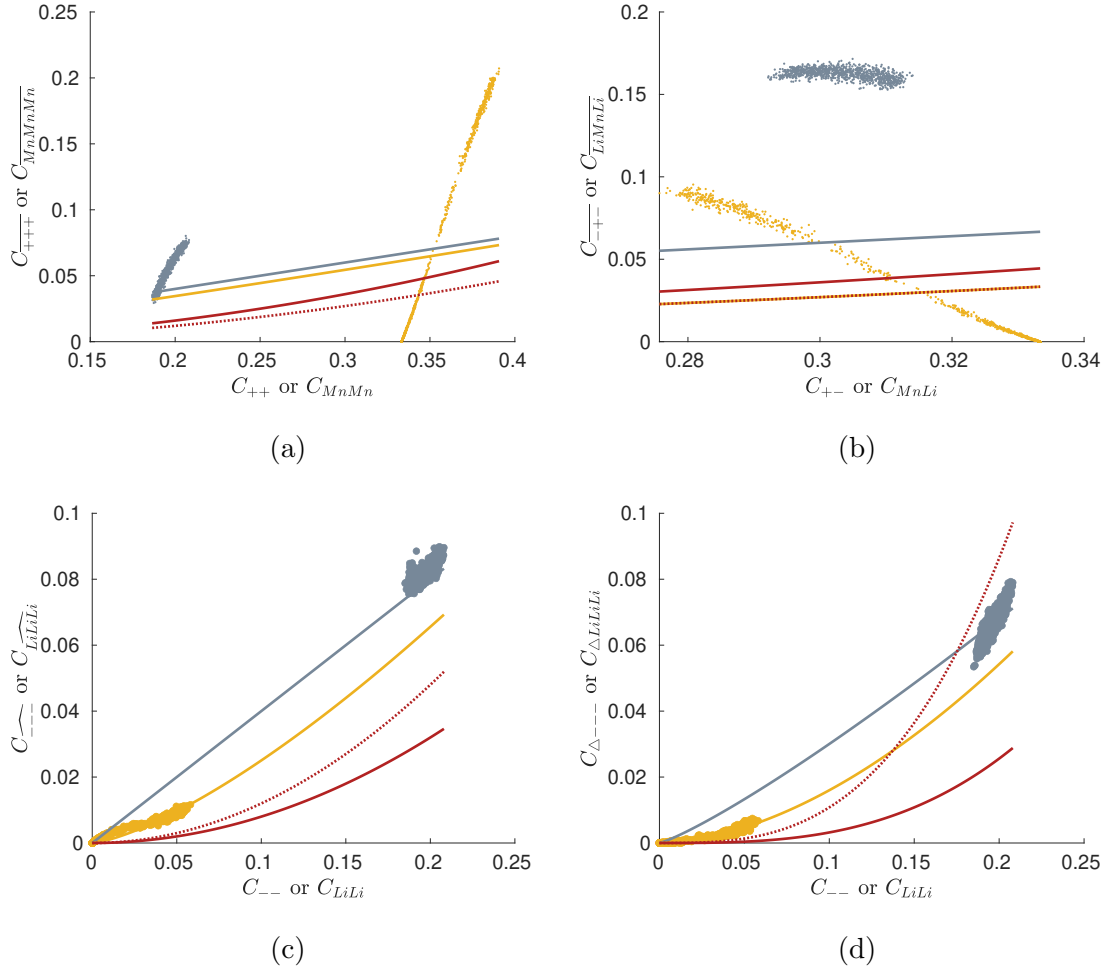
FIG. 8: Experimental triple concentrations for (gray symbols) the system $\text{Li}_{1/2}\text{Mn}_{1/2}$ and (yellow symbols) the system $\text{Li}_{1/3}\text{Mn}_{2/3}$ as functions of the corresponding pair concentrations for the 3-cluster types: $\overline{+++}$ (a), $\overline{-+-}$ (b), $\overset{\frown}{---}$ (c) and $\overset{\triangle}{---}$ (d). The corresponding reconstructions of triple concentrations from lower-order concentrations obtained via the optimal approximation and the pair approximation are shown with the grey lines and red solid lines for the system $\text{Li}_{1/2}\text{Mn}_{1/2}$, and with the yellow and red dotted lines for the system $\text{Li}_{1/3}\text{Mn}_{2/3}$. Note that the yellow and red dotted lines overlap in (b).
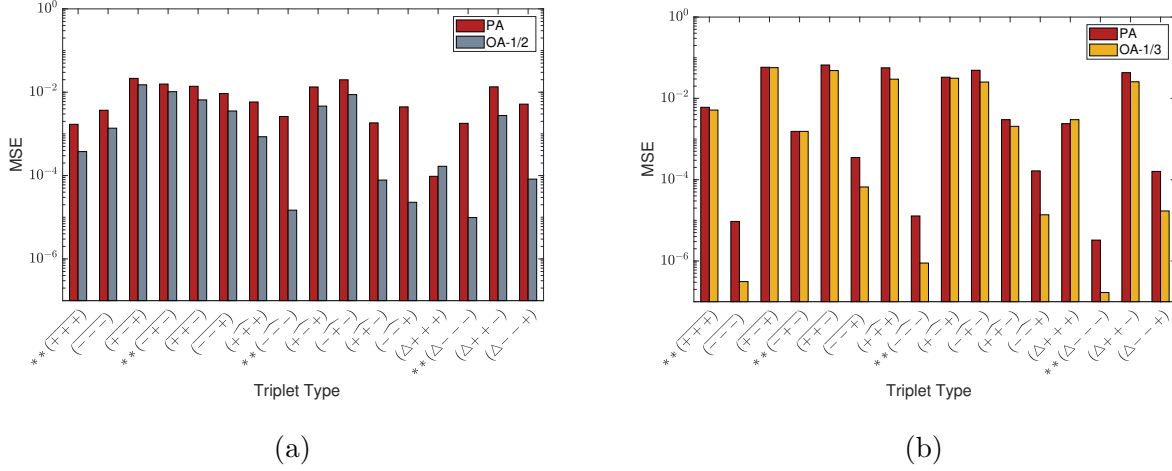
FIG. 9: The mean-square errors (15) for the pair approximation (PA) and optimal approximation subject to hard regularization for (a) the system $Li_{1/2}Mn_{1/2}$ (OA-1/2) and (b) for the system $Li_{1/3}Mn_{2/3}$ (OA-1/3). Predictions of the closure models for the triplet types marked with ($**$) are analyzed in Figure 8.

B. Sparse Approximation and its Interpretation

In this section we investigate the exponents characterizing the optimal closure presented in Table II. As can be observed, many exponents in the optimal closure relations are equal or close to zero and this trend is more pronounced in the optimal closure obtained by fitting the data for the symmetric system $\text{Li}_{1/2}\text{Mn}_{1/2}$ (when an exponent is zero, then the closure relation does not depend on the corresponding 2-cluster concentration). Thus, as is evident from Table III, the resulting structure of the closure is much simpler ("sparser") for the optimal approximation than for the closure obtained based on the pair approximation. More specifically, note that for all triplet types, except for $(\widehat{+\,+\,-})$, $(\widehat{-\,-\,+})$, $(\widehat{+\,+\,+})$ and $(\widehat{-\,-\,-})$, the optimal closure depends on the concentration of one 2-cluster only. In order to make the structure of the closure model more uniform which will facilitate its interpretation, we adjust the expressions which do not follow the pattern. More specifically, in the optimal closure relations for the clusters $(\widehat{-\,-\,+})$ and $(\widehat{-\,-\,-})$ the exponents are rounded up and down to the nearest integer, whereas for $(\widehat{+\,+\,-})$ and $(\widehat{+\,+\,+})$ the change is more significant and involves adjusting the structure of the closure relation. We refer to this simplified closure model as the *Sparse Approximation* (SA) and its functional form is presented in Table III.

We now comment on how to interpret the structure of the sparse approximation. As discussed in Section IV A, the pair approximation model neglects the correlation between non-nearest neighbour elements. This is due to the lack of information about the triple correlation term T_{ijk} in (12a). Considering relations (12) for the sparse approximation, the triplet correlation term is $T_{ijk} = \frac{C_j}{C_{ij}}$ for the linear and angled triplets, and $T_{ijk} = \frac{C_i C_j C_k}{C_{ij} C_{jk}}$ for the triangular triplets. This is contrary to the assumption that $T_{ijk} = 1$ which is central to the pair approximation. With the data in Table III we are now in the position to refine the assumptions underlying these approximations. Referring to relations (12), the concentration of the triplet (C_{ijk}) can be written as the global pair concentration (C_{ij}) times the conditional probability of finding a nearest-neighbour element to the pair in a certain state $(P_{k|ij})$. Considering the linear and angled triplets in the sparse approximation formulation, we obtain

$$C_{ijk} = C_i C_j C_k Q_{ij} Q_{jk} T_{ijk} = C_i C_j C_k \frac{C_{ij}}{C_i C_j} \frac{C_{jk}}{C_j C_k} \frac{C_j}{C_{ij}} = C_{ij} \frac{C_{jk}/C_j}{C_{ij}/C_j} = C_{ij} \frac{P_{k|j}}{P_{i|j}}. \qquad (25)$$

xxx

Triplet Type	Pair Approximation	Optimal Approximation	Sparse Approximation
$\overline{+\,+\,+}$	$\frac{1}{5}\frac{C_{++}^2}{C_+}$	$\frac{1}{5}C_{++}$	$\frac{1}{5}C_{++}$
$\overline{-\,-\,-}$	$\frac{1}{5}\frac{C_{--}^2}{C_-}$	$\frac{1}{5}C_{--}$	$\frac{1}{5}C_{--}$
$\overline{+\,-\,+}$	$\frac{1}{5}\frac{C_{+-}^2}{C_-}$	$\frac{1}{5}C_{+-}$	$\frac{1}{5}C_{+-}$
$\overline{-\,+\,-}$	$\frac{1}{5}\frac{C_{+-}^2}{C_+}$	$\frac{1}{5}C_{+-}$	$\frac{1}{5}C_{+-}$
$\overline{+\,+\,-}$	$\frac{1}{5}\frac{C_{++}C_{+-}}{C_+}$	$\frac{1}{5}C_{+-}$	$\frac{1}{5}C_{+-}$
$\overline{-\,-\,+}$	$\frac{1}{5}\frac{C_{--}C_{+-}}{C_-}$	$\frac{1}{5}C_{+-}$	$\frac{1}{5}C_{+-}$
$\widehat{+\,+\,+}$	$\frac{2}{5}\frac{C_{++}^2}{C_+}$	$\frac{2}{5}C_{++}$	$\frac{2}{5}C_{++}$
$\widehat{-\,-\,-}$	$\frac{2}{5}\frac{C_{--}^2}{C_-}$	$\frac{2}{5}C_{--}$	$\frac{2}{5}C_{--}$
$\widehat{+\,-\,+}$	$\frac{2}{5}\frac{C_{+-}^2}{C_-}$	$\frac{2}{5}C_{+-}$	$\frac{2}{5}C_{+-}$
$\widehat{-\,+\,-}$	$\frac{2}{5}\frac{C_{+-}^2}{C_+}$	$\frac{2}{5}C_{+-}$	$\frac{2}{5}C_{+-}$
$\widehat{+\,+\,-}$	$\frac{2}{5}\frac{C_{++}C_{+-}}{C_+}$	$\left[\frac{2}{5}C_{++}^{0.72}C_{+-}^{0.28}\right]$	$\frac{2}{5}C_{+-}$
$\widehat{-\,-\,+}$	$\frac{2}{5}\frac{C_{--}C_{+-}}{C_-}$	$\left[\frac{2}{5}C_{--}^{0.15}C_{+-}^{0.85}\right]$	$\frac{2}{5}C_{+-}$
$\triangle{+\,+\,+}$	$\frac{2}{5}\frac{C_{++}^3}{C_+^3}$	$\left[\frac{2}{5}\frac{C_{++}^2}{C_+}\right]$	$\frac{2}{5}C_{++}$
$\triangle{-\,-\,-}$	$\frac{2}{5}\frac{C_{--}^3}{C_-^3}$	$\left[\frac{2}{5}\frac{C_{--}^{1.18}}{C_-^{0.18}}\right]$	$\frac{2}{5}C_{--}$
$\triangle{+\,+\,-}$	$\frac{2}{5}\frac{C_{++}C_{+-}^2}{C_+^2 C_-}$	$\frac{2}{5}C_{+-}$	$\frac{2}{5}C_{+-}$
$\triangle{-\,-\,+}$	$\frac{2}{5}\frac{C_{--}C_{+-}^2}{C_+ C_-^2}$	$\frac{2}{5}C_{+-}$	$\frac{2}{5}C_{+-}$

TABLE III: Closure relations for 3-clusters of different types derived based on the pair approximation, the optimal approximation using the data for the system $\mathrm{Li}_{1/2}\mathrm{Mn}_{1/2}$, cf. Table II, and the sparse approximation discussed in Section VI B.

In a similar way one can consider the triangular triplets where

$$C_{ijk} = C_i C_j C_k Q_{ij} Q_{jk} Q_{ik} T_{ijk} = C_{ij}\frac{C_{ik}/C_i}{C_{ij}/C_i} = C_{ij}\frac{P_{k|i}}{P_{j|i}}. \tag{26}$$

We thus deduce

$$P_{k|ij} = \frac{P_{k|j}}{P_{i|j}}, \qquad \text{for linear and angled clusters,} \tag{27a}$$

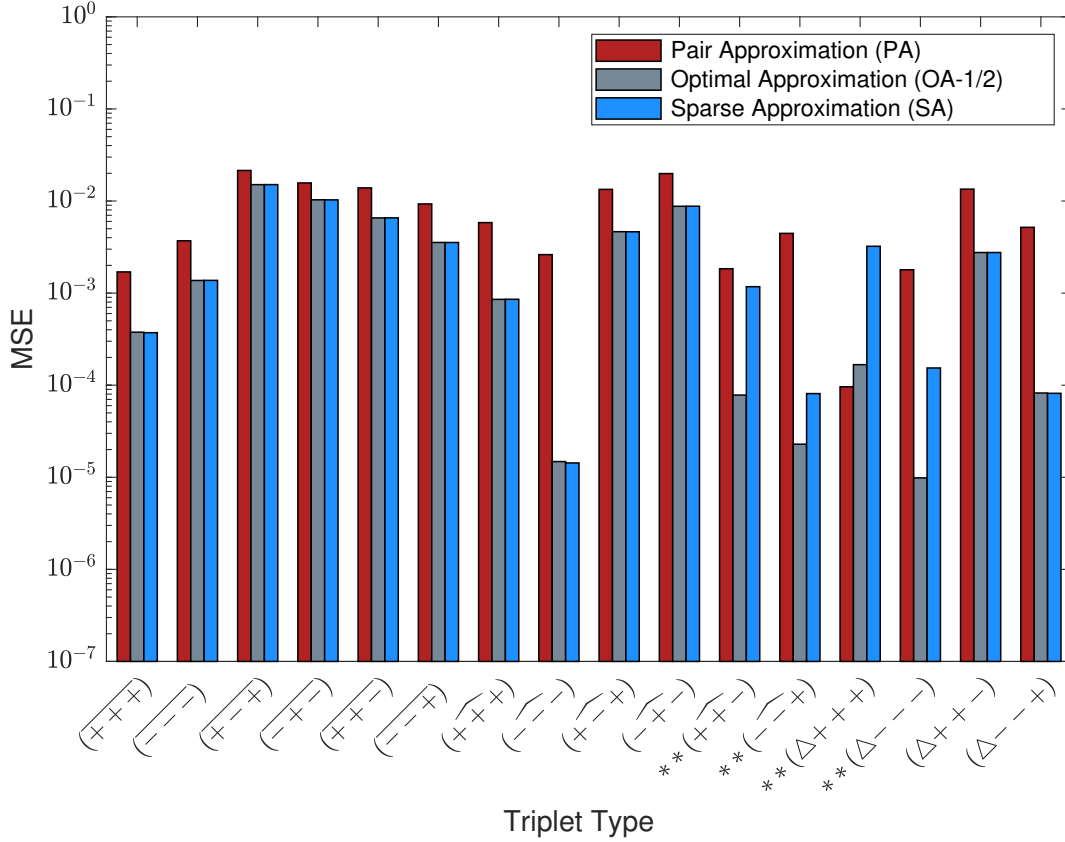$$P_{k|ij} = \frac{P_{k|i}}{P_{j|i}}, \qquad \text{for triangular clusters.} \tag{27b}$$

FIG. 10: The mean-square reconstructions errors (15) for the pair approximation, the optimal approximation constructed subject to hard regularization based on the data for the system $Li_{1/2}Mn_{1/2}$ and for the corresponding sparse approximation for different cluster types, cf. Table III. Note that the results for the last two closures differ only for the clusters marked with ($**$).

These relations break down the probability of a 3-cluster in terms of probabilities of two 2-clusters. They can be regarded as generalizations of the pair approximation model, cf. relation (11), with the inclusion of a term in the denominator. To understand the meaning of this extension of the pair approximation, we refer to relation (11). It is clear that closure is achieved using the pair approximation by assuming that the conditional probability of an element k being a nearest-neighbour of j is equal to the conditional probability of k being a nearest-neighbour of an ij pair. In other words, the pair approximation model assumes that an element j is always a nearest-neighbour of i, and we cannot find an element j which is not a nearest-neighbour of i. However, we know that this simplifying assumption is not

correct in general and there is always a possibility of finding an element j which is not a nearest-neighbour of i. By re-arranging relation (27a) in the form $P_{k|j} = P_{i|j}P_{k|ij}$, it is evident that the SA model assumes that j might not always be a nearest-neighbour of i and accounts for this possibility through the term $P_{i|j}$. A similar interpretation can be adopted for triangular clusters.

The accuracy of the optimal approximation is certainly affected when the exponents in the closure relations for the four triplet types are adjusted as discussed above, cf. Table III. Figure 10 shows the reconstruction errors for triplet concentrations obtained using different closure models for the system $\mathrm{Li}_{1/2}\mathrm{Mn}_{1/2}$. As can be expected, the SA model is less accurate in comparison to the OA model for the triplets $(\widehat{+ + -})$, $(\widehat{- - +})$, $(\widehat{+ + +})$ and $(\widehat{- - -})$. However, the performance of SA model is still better than that of the pair approximation model for the triplets $(\widehat{+ + -})$, $(\widehat{- - +})$ and $(\widehat{- - -})$. To conclude, the adjustments to the OA model sacrifice a degree of the accuracy in reconstructing the triplet concentration for $(\widehat{+ + +})$ while achieving a simpler and interpretable model.

As a result of the simple structure of the SA closure, cf. Table III, system (8) closed with this model becomes linear and hence analytically solvable. It takes the form

$$\frac{d}{dt}C_{++} = 2\alpha_1 C_{+-}, \tag{28a}$$

$$\frac{d}{dt}C_{--} = 2\alpha_2 C_{+-}, \tag{28b}$$

$$\frac{d}{dt}C_{+-} = (-\alpha_1 - \alpha_2)\,C_{+-}, \tag{28c}$$

where the parameters $\alpha_1 = \frac{4}{5}k_1 + \frac{1}{5}k_2 - \frac{4}{5}k_3 - \frac{1}{5}k_4$ and $\alpha_2 = \frac{4}{5}k_5 + \frac{1}{5}k_6 - \frac{4}{5}k_7 - \frac{1}{5}k_8$ are linear combinations of the reaction rates. The solution then is

$$C_{+-}(t) = \mu_1 e^{(-\alpha_1 - \alpha_2)t}, \qquad\qquad \mu_1 = C_{+-_0}, \tag{29a}$$

$$C_{++}(t) = \frac{2\alpha_1 \mu_1}{-\alpha_1 - \alpha_2} e^{(-\alpha_1 - \alpha_2)t} + \mu_2, \qquad \mu_2 = C_{++_0} - \frac{2\alpha_1 \mu_1}{-\alpha_1 - \alpha_2}, \tag{29b}$$

$$C_{--}(t) = \frac{2\alpha_2 \mu_1}{-\alpha_1 - \alpha_2} e^{(-\alpha_1 - \alpha_2)t} + \mu_3, \qquad \mu_3 = C_{--_0} - \frac{2\alpha_2 \mu_1}{-\alpha_1 - \alpha_2}, \tag{29c}$$

where C_{+-_0}, C_{++_0} and C_{--_0} are the initial concentrations of the corresponding 2-clusters. As is evident in (29), the concentrations C_{++} and C_{--} decrease exponentially in time with the decay rate $-(\alpha_1 + \alpha_2)$. These two parameters instead of eight reaction rates k_1 to k_8

are sufficient to describe the evolution of concentrations of different clusters in time. In addition to producing an analytically solvable model, an advantage of the SA closure is that the inverse problem (18) also simplifies and needs to be solved with respect to α_1 and α_2 only which does not require Bayesian inference. Results will be presented in Section VI D.

C. Prediction Capability of the Closure Models

In order to assess the predictive capability of the truncated model closed with the optimal approximation or the sparse approximation, the 3-cluster concentrations are reconstructed as functions of time from 2-cluster concentrations. We are interested in evaluating the prediction accuracy of these models in comparison to the model equipped with the pair approximation. In order to assess the robustness of these predictions, we will do this for stoichiometries other than the one for which the models were calibrated, cf. Sections VI A and VI B. More specifically, while the simulated annealing data for the system with the composition $Li_{1/3}Mn_{2/3}$ was used for calibration, cf. Figure 2, accuracy of the models will be analyzed here for 10 different stoichiometries Li_xMn_{1-x}, $x \in \{0.25, 0.30, 0.33, 0.36, 0.42, 0.50, 0.58, 0.64, 0.70, 0.75\}$. In particular, we are interested in the effect of regularization — soft versus hard with different parameters δ, β_1 and β_2 — in the solution of problem (16).

Robustness of the model performance will be assessed in terms of the mean-square error (15) averaged over all types of 3-clusters, i.e.,

$$\mathcal{E} = \frac{1}{|\Theta|} \sum_{i \in \Theta} J_i, \tag{30}$$

where $|\Theta| = 16$ is the total number of 3-clusters, cf. (4), and the true 3-cluster concentrations $\widetilde{C}_i(t)$ are obtained from simulated annealing experiments performed for each considered stoichiometry. The corresponding 2-cluster concentrations are used to reconstruct the 3-cluster concentrations as a function of time for each triplet type via the optimal and sparse closure approximations. Thus, this diagnostic is designed to asses only the accuracy of the closure relations given in Table III, rather than of the entire truncated model (8).

Error (30) is shown as function of the stoichiometry for the optimal closure obtained for the system $Li_{1/3}Mn_{2/3}$ subject to hard and soft regularization in Figures 11a and 11b, respec-

tively. In addition, in these figures we also show the errors obtained with the model based on the pair approximation. As can be observed, harder regularization results in larger prediction errors for stoichiometries close to $Li_{1/3}Mn_{2/3}$ in comparison to softer regularization strategies. On the other hand, harder regularization reveals better predictive performance for stoichiometries different from $Li_{1/3}Mn_{2/3}$. In other words, less aggressive regularization performs better on stoichiometries close to the stoichiometry for which the calibration of the closure relations from Table I was performed in Section VI A, and the performance gradually degrades as the stoichiometries become more different from $Li_{1/3}Mn_{2/3}$. We thus conclude that there is a trade-off between robustness and accuracy of the closure models, in the sense that models optimized for a particular stoichiometry tend to be less robust when used to describe other stoichiometries.

Finally, robustness of the closures based on the pair approximation, the optimal approximation subject to hard regularization for the system $Li_{1/3}Mn_{2/3}$ and the corresponding sparse approximation is compared for a range of stoichiometries in Figure 11. Note that solving the minimization problem (16) subject to hard regularization produces more versatile closure models that can be applied to a range of stoichiometries without significant loss of accuracy. Hence, the optimal approximation models of interest are achieved by hard regularization in (16). Figure 12 shows the mean error (30) for a range of stoichiometries for the three aforementioned closure models. A significant improvement with respect to the performance of the pair approximation model is achieved by the optimal closure models for all stoichiometries. As can be observed, the SA model performs better than the OA-1/3 model for most of the stoichiometries, except the ones that are close to the system $Li_{1/3}Mn_{2/3}$. This is due to the fact that in the OA-1/3 model the minimization problem (16) is solved for the system $Li_{1/3}Mn_{2/3}$, and hence fits are more accurate in the neighbourhood of this stoichiometry. We conclude by noting that when averaged over all stoichiometries, the performance of the sparse approximation model is improved by 36.13% over the performance of the pair approximation model.
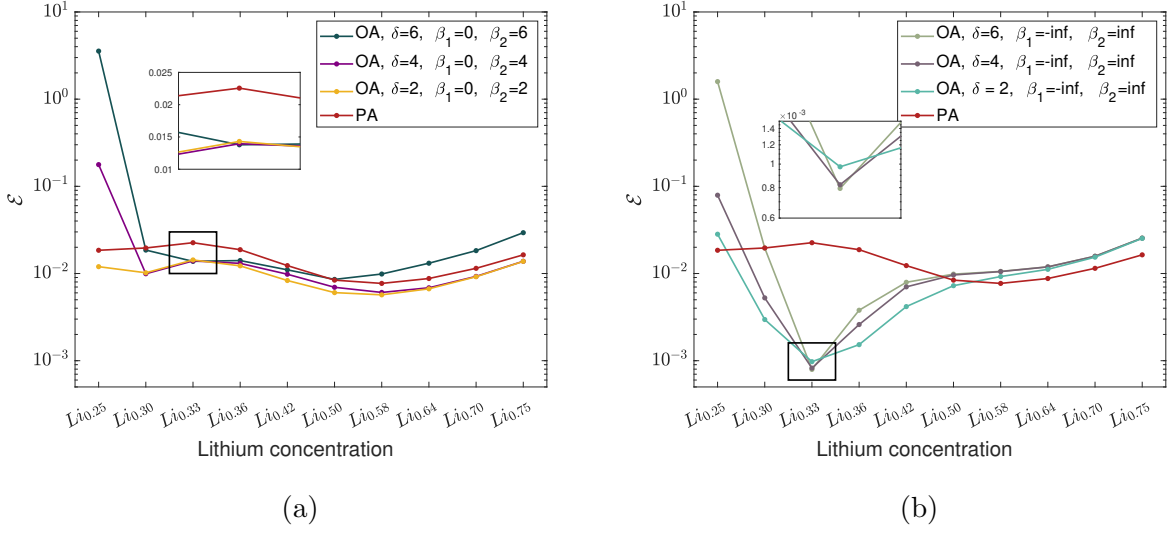
FIG. 11: Dependence of the mean error (30) characterizing the accuracy of the different closure relations on the stoichiometry for (a) hard regularization and (b) soft regularization employed in the solution of optimization problem (16) with parameters indicated in the legend for $Li_{1/3}Mn_{2/3}$ system. "PA" and "OA" refer to, respectively, the pair and the optimal approximation.
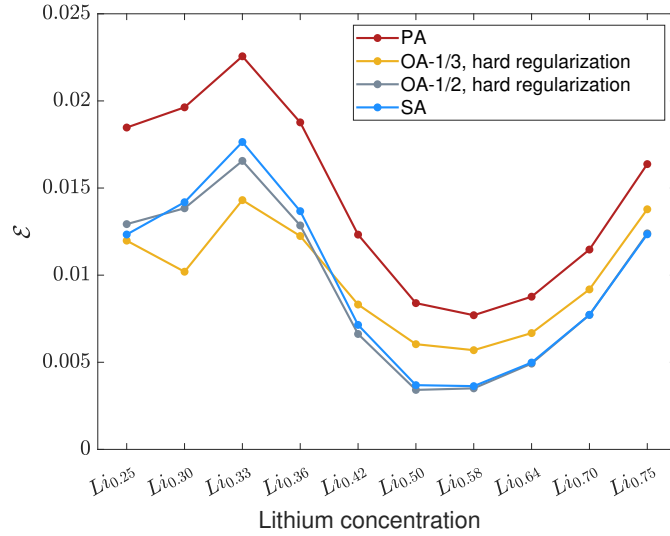

FIG. 12: The mean error (30) characterizing the accuracy of the different closure relations indicated in the legend for a range of different stoichiometries.

D. Inferring Reaction Rates

The reaction rates $k_1, \ldots, k_8$ in system (8) are determined in probabilistic terms using Bayesian inference for the pair approximation and the optimal closure models. On the other hand, for the sparse approximation there are only two unknown parameters (α_1 and α_2) so they can be inferred by solving the problem $\min_{(\alpha_1, \alpha_2) \in \mathbb{R}^2} \mathscr{J}(\alpha_1, \alpha_2)$ where the concentrations in the error functional are evaluated using the closed-form relations (29). Although this minimization problem is not convex, a global minimum can be found using standard optimization methods.

In the problems involving the pair approximation and the optimal closure models some of the reaction rates were found to be essentially equal to zero (or vanishingly small), so here the results are presented for the remaining rates only. In Figures 13a and 13c we visualize the Markov chains obtained with Algorithm 1 for system (8) closed with, respectively, the pair approximation, the optimal approximation with exponents determined subject to hard regularization (OA-1/2), cf. Table II. The Cartesian coordinates of each point in Figures 13a,c represent three of the parameters characterizing an individual Monte-Carlo sample, whereas information about the remaining parameters is encoded in the color of the symbol via the red-green-blue (RGB) mapping, as shown in the color maps in Figures 13b,d. The size of the symbols is proportional to $\mathscr{J}(\mathbf{K})^{-1}$ such that parameter values producing better fits stand out as they are represented with larger symbols. Note that, for clarity, the entire Markov chains are not presented in Figure 13 as the data is filtered based on the value of the cost function (i.e., data points are shown only if $\mathscr{J}(\mathbf{K})$ is smaller than some threshold).

It is evident from Figures 13a,c that in each case parameter values producing good fits form a number of clusters, which reflects the fact that problem (18) indeed admits multiple local minima. It is also interesting to see that good fits are obtained with some of the reaction rates varying by 200% or more which is a manifestation of the ill-posedness of problem (18) when the outputs $\mathbf{C}(\mathbf{K})$ reveal weak dependence on some of the parameters in $\mathbf{K}$. In order to compare the quality of fits obtained with the pair and optimal approximations, in Figures 14a,b we show the histograms of the values of the error functional $\mathscr{J}(\mathbf{K})$ obtained along the Markov chains. Overall, the quality of the fits is comparable in both cases and exhibits significant uncertainty, although poor fits appear more likely when the closure based on the
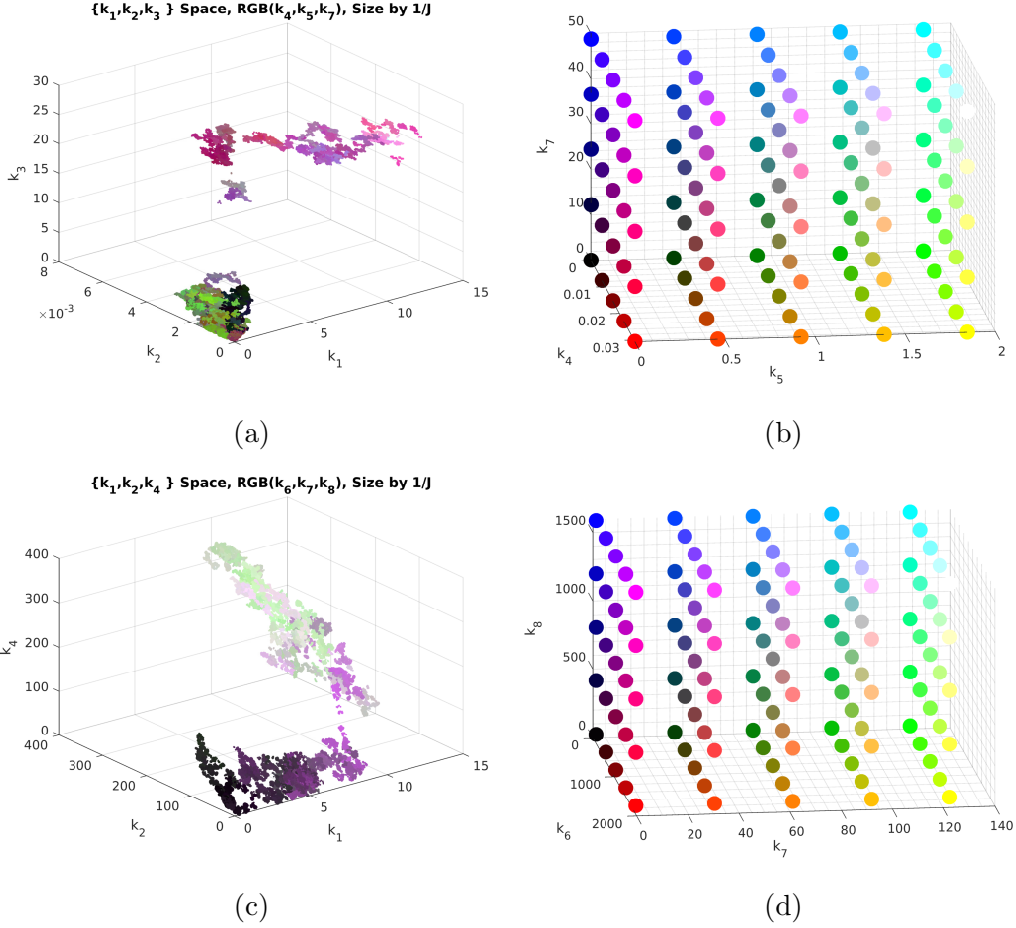
FIG. 13: Posterior probability densities $\mathbb{P}\left(\mathbf{K}|\widetilde{\mathbf{C}}\right)$ obtained using Algorithm 1 for problem (18) with system (8) closed using (a) the pair approximation and (c) the optimal approximation with exponents determined subject to hard regularization (OA-1/2). The parameters k_1, k_2 and k_3 are represented in term of the Cartesian coordinates whereas the remaining three nonzero rate constants are encoded in terms of the color of the symbols via the color maps shown in panel (b) and (d). The size of the symbols in panels (a) and (c) is proportional to $J(\mathbf{K})^{-1}$.

the pair approximation is used. The optimal parameter values for the closure based on the SA model are $(\alpha_1^*, \alpha_2^*) = (-0.083, -0.166)$ and, as we can see in Figures 14a,b, while the accuracy of the fit is lower than in the previous two cases, there is effectively no uncertainty in the determination of the parameters.

Finally, some additional comments are in place as regards the results shown in Figure 13. The parameters k_6 and k_8 are close to zero in the model with the closure based on the pair
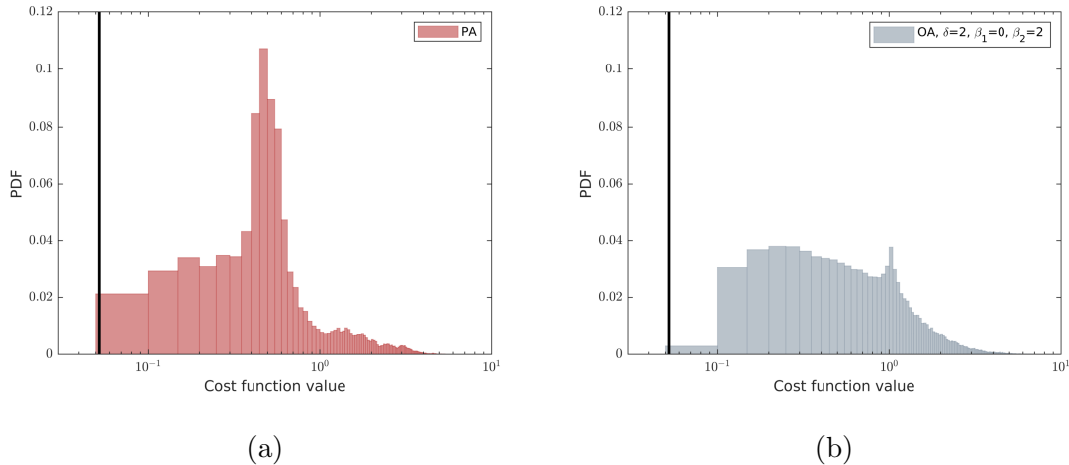
FIG. 14: Histograms of the error functional $\mathcal{J}(\mathbf{K})$ obtained along the Markov chains for problem (18) with system (8) closed using (a) the pair approximation and (b) the optimal approximation with exponents determined subject to hard regularization (OA-1/2). The black vertical lines represent the values of the error functional $\mathcal{J}(\alpha_1^*, \alpha_2^*)$ obtained when the model based on the SA closure is used.

approximation. These two parameters along with k_5 and k_7 contribute to the production and destruction of the $(--)$ cluster, cf. Figure 6. In equilibrium, the concentration of the $(--)$ (or Li-Li) cluster is zero as is evident in Figure 3. Hence, the reaction rates have values needed to annihilate the $(--)$ cluster. The linear triplets $(\overline{--+})$ and $(\overline{-+-})$ do not exist in the equilibrium state and therefore both k_6 and k_8 are very close to zero, cf. Figure 6. On the other hand, the $(\overset{\frown}{-+-})$ cluster does exist in the equilibrium state, and hence the reaction rates k_5 and k_7 are not zero. However, k_7 is bigger than k_5, highlighting the fact that the $(--)$ cluster needs to be destroyed at equilibrium.

The analysis presented above is also true from the point of view of the equilibrium constants in (7). Indeed, the equilibrium constants Q_2 and Q_4 reduce to unity when we use the model closure based on the optimal approximation subject to hard regularization. This forces the parameters k_2 and k_6 to be highly correlated with k_4 and k_8, respectively. Additionally, the equilibrium constants Q_3 and Q_4 control the production and destruction of the $(--)$ cluster, cf. (7). The constant Q_4 reduces to unity and hence does not contribute to destruction of the $(--)$ cluster at equilibrium, which is controlled by the parameters k_5 and k_7. The wide range of obtained values of k_6 and k_8, cf. Figures 13a,c, indicates the low sensitivity of the

model to these parameters, which is in agreement with our analysis.

VII.   SUMMARY & CONCLUSIONS

We have considered a mathematical model for the evolution of different cluster types in a structured lattice. We focused our attention on the structured lattice of a nickel-based oxide similar to those used in Li-ion batteries. That being said, the approach used here is much more broadly applicable. As is usual, the mean-clustering approach gives rise to an infinite hierarchy of ordinary differential equations, where concentrations of clusters of a certain size are described in terms of concentrations of clusters of higher order. This infinite hierarchy must be truncated at an arbitrary level and closed with a suitable closure model (or closure condition) in order to be solvable. This closure requires an approximation of the concentrations of the higher-order clusters in terms of the concentrations of lower-order ones. As a point of departure, we consider the pair approximation which is a classical closure model, and then introduce its generalization referred to as the optimal approximation which is calibrated using a novel data-driven approach.

The optimal approximation can be tuned for different levels of accuracy and robustness by adjusting the degree of regularization employed in the solution of the optimization problem. Our analysis shows that the model subject to soft regularization results in highly accurate approximations for the local stoichiometry but the accuracy deteriorates for other stoichiometries. On the other hand, the model subject to hard regularization has a lower accuracy at the local stoichiometry but is more robust with respect to changes of stoichiometry. The model subject to hard regularization produces more accurate results than the pair approximation for a broad range of stoichiometries. More importantly, the closure model found in this way turns out to have a simple structure with many exponents having nearly integer values. Exploiting this structure, we arrive at the sparse approximation model which is linear and therefore analytically solvable.

In addition to being simpler, the sparse approximation model is also more accurate and robust than the pair approximation, in that it can be applied to a wide range of stoichiometries without a significant loss of accuracy. This model is interpretable as it makes it possible to refine some of the simplifying assumptions at the heart of the pair approximation. One of

these assumptions states that the conditional probability of k being a nearest neighbour of ij in a triplet (ijk) is equal to that of k being a nearest neighbour of j. In other words, it is assumed that every j element in the lattice has a nearest neighbour in state i. The sparse approximation refines this assumption by adding a term that takes into account the conditional probability of j being a nearest neighbour of i. This correction makes the model both simpler and more accurate.

The reaction rates in system (8) closed using one of the closure models are determined by formulating a suitable inverse problem. We solve these problems using a state-of-the-art Bayesian inference approach which also allows us to estimate the uncertainties of the reconstructed parameters. The results obtained show that the inverse problem is in fact ill-posed in the case of the closures based on the pair and optimal approximation, in the sense that the corresponding optimization problems admit multiple local minima. Moreover, these minima tend to be "shallow" reflecting the low sensitivity of the models closed with the pair and optimal approximations to the reaction rates. As a result, the inferred values of these parameters suffer from uncertainties on the order of 200%. In contrast, the model closed using the sparse approximation is well-posed with respect to α_1 and α_2 which are linear combinations of reaction rates. This model is analytically solvable which completely eliminates the uncertainty in the reconstruction of its parameters. Based on these observations, we conclude that the sparse approximation is superior to the pair approximation.

Notably, the mean-cluster modelling approach considered in the present work can be used to describe the evolution of clusters of arbitrary size and type defined on structured lattices various types. The size and shape of the cluster and the structure of the lattice determine the reactions between elements. More complicated lattices and bigger cluster sizes involve more possible nearest-neighbour element swaps, resulting in a larger number of parameters in the model. The sparse approximation methodology could be utilized in a similar way to close the corresponding hierarchical models.

Appendix A: Rotational Symmetry

Theorem A.1. *In a 2D triangular lattice (where each element is surrounded by 6 nearest-neighbours), different spatial orientations of a particular 2-cluster retain the same concentration, i.e., the probability of finding a particular 2-cluster in the lattice is independent of its spatial orientation.*

Proof. Assuming one site in a $\oplus$ state, the concentration of this element can be obtained by summing over concentrations of all 2-clusters in which the second element iterates over the possible elements in the system. Using ($\bullet$) to denote an unspecified state in the lattice, we then have

$$
\begin{aligned}
C\left(\oplus\right) = C\left(\oplus\bullet\right) &= C\left(\oplus\oplus\right) + C\left(\oplus\ominus\right) \\
= C\left(\bullet\oplus\right) &= C\left(\oplus\oplus\right) + C\left(\ominus\oplus\right) \\
= C\left(\oplus^{\bullet}\right) &= C\left(\oplus^{\oplus}\right) + C\left(\oplus^{\ominus}\right) \\
= C\left(^{\bullet}\oplus\right) &= C\left(^{\oplus}\oplus\right) + C\left(^{\ominus}\oplus\right) \\
= C\left(\oplus_{\bullet}\right) &= C\left(\oplus_{\oplus}\right) + C\left(\oplus_{\ominus}\right) \\
= C\left(_{\bullet}\oplus\right) &= C\left(_{\oplus}\oplus\right) + C\left(_{\ominus}\oplus\right) \\
\end{aligned}
\tag{A1}
$$

$$
\implies \quad C\left(\oplus\ominus\right) = C\left(\ominus\oplus\right) = C\left(\oplus^{\ominus}\right) = C\left(^{\ominus}\oplus\right) = C\left(\oplus_{\ominus}\right) = C\left(_{\ominus}\oplus\right).
$$

$\square$

[1] K. J. Harris, J. M. Foster, M. Z. Tessaro, M. Jiang, X. Yang, Y. Wu, B. Protas, and G. R. Goward, Structure Solution of Metal-Oxide Li Battery Cathodes from Simulated Annealing and Lithium NMR Spectroscopy, Chemistry of Materials **29**, 5550 (2017).

[2] M. Hasenbusch, Monte carlo studies of the three-dimensional ising model in equilibrium, International Journal of Modern Physics C **12**, 911 (2001).

[3] J. Strecka and M. Jascur, A brief account of the ising and ising-like models: Mean-field, effective-field and exact results, arXiv preprint arXiv:1511.03031 (2015).

[4] H. Matsuda, N. Ogita, A. Sasaki, and K. Sato, Statistical Mechanics of Population: The Lattice Lotka-Volterra Model, Progress of Theoretical Physics **88**, 1035 (1992).

[5] Y. Harada and Y. Iwasa, Lattice population dynamics for plants with dispersing seeds and Vegetative propagation, Researches on Population Ecology **36**, 237 (1994).

[6] D. H. Silva, F. A. Rodrigues, and S. C. Ferreira, High prevalence regimes in the pair-quenched mean-field theory for the susceptible-infected-susceptible model on networks, Physical Review E **102**, 1 (2020).

[7] M. J. Keeling, The effects of local spatial structure on epidemiological invasions, The Structure and Dynamics of Networks **9781400841**, 480 (2011).

[8] K. T. Eames and M. J. Keeling, Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases, Proceedings of the National Academy of Sciences of the United States of America **99**, 13330 (2002).

[9] T. B. Pedro, W. Figueiredo, and A. L. Ferreira, Mean-field theory for the long-range contact process with diffusion, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics **92**, 1 (2015).

[10] K. Satō, H. Matsuda, and A. Sasaki, Pathogen invasion and host extinction in lattice structured populations, Journal of Mathematical Biology **32**, 251 (1994).

[11] M. J. Keeling, D. A. Rand, and A. J. Morris, Correlation models for childhood epidemics, Proceedings of the Royal Society B: Biological Sciences **264**, 1149 (1997).

[12] C. T. Bauch, The spread of infectious diseases in spatially structured populations: An invasory pair approximation, Mathematical Biosciences **198**, 217 (2005).

[13] M. M. De Oliveira, S. G. Alves, and S. C. Ferreira, Dynamical correlations and pairwise theory for the symbiotic contact process on networks, Physical Review E **100**, 52302 (2019), arXiv:1909.03981.

[14] A. S. Mata and S. C. Ferreira, Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks, Epl **103**, 10.1209/0295-5075/103/48003 (2013), arXiv:1305.5153.

[15] Z. H. Lin, M. Feng, M. Tang, Z. Liu, C. Xu, P. M. Hui, and Y. C. Lai, Non-Markovian recovery makes complex networks more resilient against large-scale failures, Nature Communications **11**, 1 (2020), arXiv:1902.07594.

[16] X. Pei, X. X. Zhan, and Z. Jin, Application of pair approximation method to modeling and analysis of a marriage network, Applied Mathematics and Computation **294**, 280 (2017).

[17] D. Ben-Avraham and J. Köhler, Mean-field (n,m)-cluster approximation for lattice models, Physical Review A **45**, 8358 (1992).

[18] K. E. Sugden and M. R. Evans, A dynamically extending exclusion process, Journal of Statistical Mechanics: Theory and Experiment **2007**, 10.1088/1742-5468/2007/11/P11013 (2007), arXiv:0707.4504.

[19] J. Joo and J. L. Lebowitz, Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercubic lattice, Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics **70**, 9 (2004).

[20] A. S. Mata, R. S. Ferreira, and S. C. Ferreira, Heterogeneous pair-approximation for the contact process on complex networks, New Journal of Physics **16**, 10.1088/1367-2630/16/5/053006 (2014).

[21] J. A. Filipe and M. M. Maule, Analytical methods for predicting the behaviour of population models with general spatial interactions, Mathematical Biosciences **183**, 15 (2003).

[22] D. H. Seo, J. Lee, A. Urban, R. Malik, S. Kang, and G. Ceder, The structural and chemical origin of the oxygen redox activity in layered and cation-disordered Li-excess cathode materials, Nature Chemistry **8**, 692 (2016).

[23] A. K. Shukla, Q. M. Ramasse, C. Ophus, H. Duncan, F. Hage, and G. Chen, Unravelling structural ambiguities in lithium- and manganese-rich transition metal oxides, Nature Communications **6**, 1 (2015).

[24] J. Lisiecki and P. Szabelski, Designing 2D covalent networks with lattice Monte Carlo simu-

lations: precursor self-assembly, Physical Chemistry Chemical Physics **23**, 5780 (2021).

[25] F. O. Sanchez-Varretti, F. M. Bulnes, and A. J. Ramirez-Pastor, Order and disorder in the adsorption model of repulsively interacting binary mixtures on triangular lattices: theory and Monte Carlo simulations, European Physical Journal E **44**, 10.1140/epje/s10189-021-00037-6 (2021).

[26] H.-A. Chen, P.-H. Tang, G.-J. Chen, C.-C. Chang, and C.-W. Pao, Microstructure Maps of Complex Perovskite Materials from Extensive Monte Carlo Sampling Using Machine Learning Enabled Energy Model, The Journal of Physical Chemistry Letters **12**, 3591 (2021).

[27] R. Dickman, Kinetic phase transitions in a surface-reaction model: Mean-field theory, Physical review. A, General physics **34**, 10.1103/physreva.34.4246 (1986).

[28] K. Satō and Y. Iwasa, Pair Approximations for Lattice-based Ecological Models, in *The Geometry of Ecological Interactions* (Cambridge University Press, 2000) pp. 341–358.

[29] M. van Baalen, Pair Approximations for Different Spatial Geometries, in *The Geometry of Ecological Interactions* (Cambridge University Press, 2000) pp. 359–387.

[30] J. Morris, Andrew, *Representing spatial interactions in simple ecological models*, Phd, University of Warwick (1997).

[31] J. Nocedal and S. Wright, *Numerical Optimization* (Springer, 2002).

[32] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM, 2005).

[33] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems* (Springer, 2005).

[34] R. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications* (SIAM, 2013).

[35] A. Sethurajan, S. Krachkovskiy, G. Goward, and B. Protas, Bayesian uncertainty quantification in inverse modeling of electrochemical systems, Journal of Computational Chemistry **40**, 740 (2019), arXiv:1806.00036.

[36] Q. Zhou, T. Yu, X. Zhang, and J. Li, Bayesian inference and uncertainty quantification for medical image reconstruction with poisson data, SIAM Journal on Imaging Sciences **13**, 29 (2019), arXiv:1903.02075.

[37] S. Huo, *Bayesian Modeling of Complex High-Dimensional Data Bayesian Modeling of Complex High-Dimensional Data*, Ph.D. thesis, Virginia Polytechnic Institute and State University (2020).

[38] M. Laine, *Adaptive MCMC methods with applications in environmental and geophysical models*, Ph.D. thesis, Lappeenranta University of technology (2008).

[39] D. D. Lucas, M. Simpson, P. Cameron-Smith, and R. L. Baskett, Bayesian inverse modeling of the atmospheric transport and emissions of a controlled tracer release from a nuclear power plant, Atmospheric Chemistry and Physics **17**, 13521 (2017).

[40] O. Camli and Z. Kalaylioglu, Bayesian predictive model selection in circular random effects models with applications in ecological and environmental studies, Environmental and Ecological Statistics 10.1007/s10651-020-00471-3 (2020).

[41] Y. Thawornwattana, D. Dalquen, and Z. Yang, Designing simple and efficient Markov chain Monte Carlo proposal kernels, Bayesian Analysis **13**, 1033 (2018).

[42] M. Bédard, R. Douc, and E. Moulines, Scaling Analysis of Delayed Rejection MCMC Methods, Methodology and Computing in Applied Probability **16**, 811 (2014).

[43] P. J. Green and A. Mira, Delayed rejection in reversible jump Metropolis–Hastings, Biometrika **88**, 1035 (2001).

[44] J. Kaipo and E. Somersalo, *Applied Mathematical Sciences*, Vol. 160 (Springer Science \& Business Media, 2006).

[45] P. C. Robert and C. George, *Monte Carlo Statistical Methods*, 2nd ed. (Springer, New York, 2004).