

# Laws of Large Numbers and Langevin Approximations for Stochastic Neural Field Equations

Martin G. Riedler · Evelyn Buckwar

Received: 4 July 2012 / Accepted: 14 January 2013 / Published online: 23 January 2013

© 2013 M.G. Riedler, E. Buckwar; licensee Springer. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract** In this study, we consider limit theorems for microscopic stochastic models of neural fields. We show that the Wilson–Cowan equation can be obtained as the limit in uniform convergence on compacts in probability for a sequence of microscopic models when the number of neuron populations distributed in space and the number of neurons per population tend to infinity. This result also allows to obtain limits for qualitatively different stochastic convergence concepts, e.g., convergence in the mean. Further, we present a central limit theorem for the martingale part of the microscopic models which, suitably re-scaled, converges to a centred Gaussian process with independent increments. These two results provide the basis for presenting the neural field Langevin equation, a stochastic differential equation taking values in a Hilbert space, which is the infinite-dimensional analogue of the chemical Langevin equation in the present setting. On a technical level, we apply recently developed law of large numbers and central limit theorems for piecewise deterministic processes taking values in Hilbert spaces to a master equation formulation of stochastic neuronal network models. These theorems are valid for processes taking values in Hilbert spaces, and by this are able to incorporate spatial structures of the underlying model.

**Keywords** Stochastic neural field equation · Wilson–Cowan model · Piecewise deterministic Markov process · Stochastic processes in infinite dimensions · Law of large numbers · Martingale central limit theorem · Chemical Langevin equation

**Mathematics Subject Classification (2000)** 60F05 · 60J25 · 60J75 · 92C20

---

M.G. Riedler (✉) · E. Buckwar  
Institute for Stochastics, Johannes Kepler University, Linz, Austria  
e-mail: [martin.riedler@jku.at](mailto:martin.riedler@jku.at)

E. Buckwar  
e-mail: [evelyn.buckwar@jku.at](mailto:evelyn.buckwar@jku.at)

## 1 Introduction

The present study is concerned with the derivation and justification of neural field equations from finite size stochastic particle models, i.e., stochastic models for the behaviour of individual neurons distributed in finitely many populations, in terms of mathematically precise probabilistic limit theorems. We illustrate this approach with the example of the Wilson–Cowan equation

$$\tau \dot{v}(t, x) = -v(t, x) + f \left( \int_D w(x, y) v(t, y) dy + I(t, x) \right). \quad (1.1)$$

We focus on the following two aspects:

- (A) Often one wants to study deterministic equations such as Eq. (1.1) in order to obtain results on the ‘behaviour in the mean’ of an intrinsically stochastic system. Thus, we first discuss limit theorems of the law of large numbers type for the limit of infinitely many particles. These theorems connect the trajectories of the stochastic particle models to the deterministic solution of mean field equations, and hence provide a justification studying Eq. (1.1) in order to infer on the behaviour of the stochastic system.
- (B) Secondly, we aim to characterise the internal noise structure of the complex discrete stochastic models as in the limit of large numbers of neurons the noise is expected to be close to a simpler stochastic process. Ultimately, this yields a stochastic neural field model in terms of a stochastic evolution equation conceptually analogous to the *Chemical Langevin Equation*. The Chemical Langevin Equation is widely used in the study of chemical reactions networks for which the stochastic effects cannot be neglected but a numerical or analytical study of the exact discrete model is not possible due to its inherent complexity.

In this study, we understand as a *microscopic model* a description as a stochastic process, usually a Markov chain model, also called a *master equation formulation* (cf. [3, 5, 8, 9, 22] containing various master equation formulations of neural dynamics). In contrast, a *macroscopic model* is a deterministic evolution equation such as (1.1). Deterministic mean field equations have been used widely and for a long time to model and analyse large scale behaviour of the brain. In their original deterministic form, they are successfully used to model geometric visual hallucinations, orientation tuning in the visual cortex and wave propagation in cortical slices to mention only a few applications. We refer to [7] for a recent review and an extensive list of references. The derivation of these equations is based on a number of arguments from statistical physics and for a long time a justification from microscopic models has not been available. The interest in deriving mean field equations from stochastic microscopic model has been revived recently as it contains the possibility to derive deterministic ‘corrections’ to the mean field equations, also called second-order approximations. These corrections might account for the inherent stochasticity, and thus incorporate so called finite size effects. This has been achieved by either applying a path-integral approach to the master equation [8, 9] or by a van Kampen system-size expansion of the master equation [5]. In more detail, the author in the latter reference proposes a particular master equation for a finite number of neuron populations and derives the

Wilson–Cowan equation as the first-order approximation to the mean via employing the van Kampen system size expansion and then taking the continuum limit for a continuum of populations. In keeping also the second-order terms, a ‘stochastic’ version of the mean field equation is also presented in the sense of coupling the first moment equation to an equation for the second moments.

However, the van Kampen system size expansion does not give a precise mathematical connection, as it neither quantifies the type of convergence (quality of the limit), states conditions when the convergence is valid nor does it allow to characterise the speed of convergence. Furthermore, particular care has to be taken in systems possessing multiple fixed points of the macroscopic equation, and we refer to [5] for a discussion of this aspect in the neural field setting. The limited applicability of the van Kampen system size expansion was already well known to Sect. 10 in van Kampen [33]. In parallel to the work of van Kampen, T. Kurtz derived precise limit theorems connecting sequences of continuous time Markov chains to solutions of systems of ordinary differential equations; see the seminal studies [19, 20] or the monograph [15]. Limit theorems of that type are usually called the *fluid limit*, *thermodynamic limit*, or *hydrodynamic limit*; for a review, see, e.g., [13].

As is thoroughly discussed in [5] establishing the connection between master equation models and mean field equations involves two limit procedures. First, a limit which takes the number of particles, in this case neurons per considered population, to infinity (thermodynamic limit), and a second which gives the mean field by taking the number of populations to infinity (continuum limit). In this ‘double limit’, the theorems by Kurtz describe the connection of taking the number of neurons per population to infinity yielding a system of ordinary differential equation, one for each population. Then the extension from finite to infinite dimensional state space is obtained by a continuum limit. This procedure corresponds to the approach in [5]. Thus, taking the double limit step by step raises the question what happens if we first take the spatial limit and then the fluid limit, thus reversing the order of the limit procedures, or in the case of taking the limits simultaneously. Recently, in an extension to the work of Kurtz, one of the present authors and co-authors established limit theorems that achieve this double limit [27], thus being able to connect directly finite population master equation formulations to spatio-temporal limit systems, e.g., partial differential equation or integro-differential equations such as the Wilson–Cowan equation (1.1). In a general framework, these limit theorems were derived for Piecewise Deterministic Markov Processes on Hilbert spaces, which in addition to the jump evolution also allow for a coupled deterministic continuous evolution. This generality was motivated by applications to neuron membrane models consisting of microscopic models of the ion channels coupled to a deterministic equation for the transmembrane potential. We find that this generality is also advantageous for the present situation of a pure jump model as it allows to include time-dependent inputs. In this study, we employ these theorems to achieve the aims (A) and (B) focussing on the example of the deterministic limit given by the Wilson–Cowan equation (1.1).

Finally, we state what this study does *not* contain, which in particular distinguishes the present study from [5, 8, 9] beyond mathematical technique. Presently, the aim is not to derive moment equations, i.e., a deterministic set of equations that approximate the moments of the Markovian particle model, but rather processes (deterministic or

stochastic) to which a sequence of microscopic models converges under suitable conditions in a probabilistic way. This means that a microscopic model, which is close to the limit—presently corresponding to a large number of neurons in a large number of populations—can be assumed to be close to the limiting processes in structure and pathwise dynamics as indicated by the quality of the stochastic limit. Hence, the present work is conceptually—though neither in technique nor results—close to [30] wherein using a propagation to chaos approach in the vicinity of neural field equations the author also derives in a mathematically precise way a limiting process to finite particle models. However, it is an obvious consequence that the convergence of the models necessarily implies a close resemblance of their moment equations. This provides the connection to [5, 8, 9], which we briefly comment on in Appendix B.

As a guide, we close this introduction with an outline of the subsequent sections and some general remarks on the notation employed in this study. In Sects. 1.1 to 1.3, we first discuss the two types of mean field models in more detail, on the one hand, the Wilson–Cowan equation as the macroscopic limit and, on the other hand, a master equation formulation of a stochastic neural field. The main results of the paper are found in Sect. 2. There we set up the sequence of microscopic models and state conditions for convergence. Limit theorems of the law of large numbers type are presented in Theorem 2.1 and Theorem 2.2 in Sect. 2.1. The first is a classical weak law of large numbers providing uniform convergence on compacts in probability and the second convergence in the mean uniformly over the whole positive time axis. Next, a central limit theorem for the martingale part of the microscopic models is presented in Sect. 2.2 characterising the internal fluctuations of the model to be of a diffusive nature in the limit. This part of the study is concluded in Sect. 2.3 by presenting the Langevin approximations that arise as a result of the preceding limit theorems. The proofs of the theorems in Sect. 2 are deferred to Sect. 4. The study is concluded in Sect. 3 with a discussion of the implications of the presented results and an extension of these limit theorems to different master equation formulations or mean field equations.

*Notations and Conventions* Throughout the study, we denote by  $L^p(D)$ ,  $1 \leq p \leq \infty$ , the Lebesgue spaces of real functions on a domain  $D \subset \mathbb{R}^d$ ,  $d \geq 1$ . Physically reasonable choices are  $d \in \{1, 2, 3\}$ , however, for the mathematical theory presented the spatial dimension can be arbitrary. In the present study, spatial domains  $D$  are always bounded with a sufficiently smooth boundary, where the minimal assumption is a strong local Lipschitz condition; see [2]. For bounded domains  $D$ , this condition simply means that for every point on the boundary its neighbourhood on the boundary is the graph of a Lipschitz continuous function. Furthermore, for  $\alpha \in \mathbb{N}$  we denote by  $H^\alpha(D)$  the Sobolev spaces, i.e., subspaces of  $L^2(D)$ , with the corresponding Sobolev norm. For  $\alpha \in \mathbb{R}_+ \setminus \mathbb{N}$  we denote by  $H^\alpha(D)$  the interpolating Besov spaces. In this study,  $H^{-\alpha}(D)$  is the dual space of  $H^\alpha(D)$ , which is in contrast to the widespread notation to denote by  $H^{-\alpha}(D)$ ,  $\alpha \geq 0$ , the dual space of  $H_0^\alpha(D)$ . As usual, we have  $H^0(D) = L^2(D) = H^{-0}(D)$ . We thus obtain a continuous scale of Hilbert spaces  $H^\alpha(D)$ ,  $\alpha \in \mathbb{R}$ , which satisfy that  $H^{\alpha_1}(D)$  is continuously embed-

ded<sup>1</sup> in  $H^{\alpha_2}(D)$  for all  $\alpha_1 < \alpha_2$ . Next, a pairing  $(\cdot, \cdot)_{H^\alpha}$  denotes the inner product of the Hilbert space  $H^\alpha(D)$  and pairings in angle brackets  $\langle \cdot, \cdot \rangle_{H^\alpha}$  denote the duality pairing for the Hilbert space  $H^\alpha(D)$ . That is, for  $\psi \in H^\alpha(D)$  and  $\phi \in H^{-\alpha}(D)$  the expression  $\langle \phi, \psi \rangle_{H^\alpha}$  denotes the application of the real, linear functional  $\phi$  to  $\psi$ . Furthermore, the spaces  $H^\alpha(D)$ ,  $L^2(D)$  and  $H^{-\alpha}(D)$  form an evolution triplet, i.e., the embeddings are dense and the application of linear functionals and the inner product in  $L^2(D)$  satisfy the relation

$$\langle \phi, \psi \rangle_{H^\alpha} = (\phi, \psi)_{L^2} \quad \forall \phi \in L^2(D), \psi \in H^\alpha(D). \tag{1.2}$$

Norms in Hilbert spaces are denoted by  $\|\cdot\|_{H^\alpha}$ ,  $\|\cdot\|_0$  is used to denote the supremum norm of real functions, i.e., for  $f : \mathbb{R} \rightarrow \mathbb{R}$  we have  $\|f\|_0 = \sup_{z \in \mathbb{R}} |f(z)|$ , and  $|\cdot|$  denotes either the absolute value for scalars or the Lebesgue measure for measurable subsets of Euclidean space. Finally, we use  $\mathbb{N}_0$  to denote the set of integers including zero.

### 1.1 The Macroscopic Limit

Neural field equations are usually classified into two types: *rate-based* and *activity-based* models. The prototype of the former is the Wilson–Cowan equation; see Eq. (1.1), which we also restate below, and the Amari equation, see Eq. (3.7) in Sect. 3, is the prototype of the latter. Besides being of a different structure, due to their derivation, the variable they describe has a completely different interpretation. In rate-based models, the variable describes the average rate of activity at a certain location and time, roughly corresponding to the fraction of active neurons at a certain infinitesimal area. In activity-based models, the macroscopic variable is an average electrical potential produced by neurons at a certain location. For a concise physical derivation that leads to these models, we refer to [5]. In the following, we consider rate-based equations, in particular, the classical Wilson–Cowan equation, to discuss the type of limit theorems we are able to obtain. We remark that the results are essentially analogous for activity based models.

Thus, the macroscopic model of interest is given by the equation

$$\tau \dot{v}(t, x) = -v(t, x) + f\left(\int_D w(x, y)v(t, y) dy + I(t, x)\right), \tag{1.3}$$

where  $\tau > 0$  is a decay time constant,  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  is a gain (or response) function that relates inputs that a neuron receives to activity. In (1.3), the value  $f(z)$  can be interpreted as the fraction of neurons that receive at least threshold input. Furthermore,  $w(x, y)$  is a weight function, which states the connectivity strength of a neuron located at  $y$  to a neuron located at  $x$ , and finally,  $I(t, x)$  is an external input, which is received by a neuron at  $x$  at time  $t$ . For the weight function  $w : D \times D \rightarrow \mathbb{R}$  and the external input  $I$ , we assume that  $w \in L^2(D \times D)$  and  $I \in C(\mathbb{R}_+, L^2(D))$ . As for

---

<sup>1</sup>A normed space  $X$  is continuously embedded in another normed space  $Y$ , in symbols  $X \hookrightarrow Y$ , if  $X \subset Y$  and there exists a constant  $K < \infty$  such that  $\|u\|_Y \leq K \|u\|_X$  for all  $u \in X$ .

the gain function  $f$ , we assume in this study that  $f$  is non-negative, satisfies a global Lipschitz condition with constant  $L > 0$ , i.e.,

$$|f(a) - f(b)| \leq L|a - b| \quad \forall a, b \in \mathbb{R}, \quad (1.4)$$

and it is bounded. From an interpretive point-of-view, it is reasonable and consistent to stipulate that  $f$  is bounded by one—being a fraction—as well as being monotone. The latter property corresponds to the fact that higher input results in higher activity. In specific models,  $f$  is often chosen to be a sigmoidal function, e.g.,  $f(z) = (1 + e^{-(\beta_1 z + \beta_2)})^{-1}$  in [6] or  $f(z) = (\tanh(\beta_1 z + \beta_2) + 1)/2$  in [3], which both satisfy  $f \in [0, 1]$ . Moreover, the most common choices of  $f$  are even infinitely often differentiable with bounded derivatives, which already implies the Lipschitz condition (1.4).

The Wilson–Cowan equation (1.3) is well-posed in the strong sense as an integral equation in  $L^2(D)$  under the above conditions. That is, Eq. (1.3) possesses a unique, continuously differentiable global solution  $v$  to every initial condition  $v(0) = v_0 \in L^2(D)$ , i.e.,  $v \in C^1([0, T], L^2(D))$  for all  $T > 0$ , which depends continuously on the initial condition. Furthermore, if the initial condition satisfies  $v_0(x) \in [0, \|f\|_0]$  almost everywhere in  $D$ , then it holds for all  $t > 0$  that  $v(t, x) \in (0, \|f\|_0)$  for almost all  $x \in D$ . For a brief derivation of these results, we refer to Appendix A where we also state a result about higher spatial regularity of the solution: Let  $\alpha \in \mathbb{N}$  be such that  $\alpha > d/2$ . If now  $v_0 \in H^\alpha(D)$  and if  $f$  is at least  $\alpha$ -times differentiable with bounded derivatives and the weights and the input function satisfy  $w \in H^\alpha(D \times D)$  and  $I \in C(\mathbb{R}_+, H^\alpha(D))$ , then the equation is well-posed in  $H^\alpha(D)$ , i.e., for all  $T > 0$  in  $v \in C^1([0, T], H^\alpha(D))$ . In particular, this implies that the solution  $v$  is jointly continuous on  $\mathbb{R}_+ \times D$ .

## 1.2 Master Equation Formulations of Neural Network Models

For the microscopic model, we concentrate on a variation of the model considered in [5, 6], which is already an improvement on a model introduced in [11]. We extend the model including variations among neuron populations and foremost time-dependent inputs. We chose this model over the master equation formulations in [8, 9] as it provides a more direct connection of the microscopic and macroscopic models; see also the discussion in Sect. 3. We describe the main ingredients of the model beginning with the simpler, time-independent model as prevalent in the literature. Subsequently, in Sect. 1.3 the final, time-dependent model is defined.

We denote by  $P$  the number of neuron populations in the model. Further, we assume that the  $k$ th neuron population consists of identical neurons which can either be in one of two possible states, *active*, i.e., emitting action potentials, and *inactive*, i.e., quiescent or not emitting action potentials. Transitions between states occur instantaneously and at random times. For all  $k = 1, \dots, P$ , the random variables  $\Theta_t^k$  denote the number of active neurons at time  $t$ . An integer  $l(k)$  is used to characterise the population size. This number  $l(k)$  can be interpreted as the number of neurons in the  $k$ th population, at least for sufficiently large values. However, this is not accurate in the literal sense as it is possible with positive probability for populations to contain more than  $l(k)$  active neurons. Nevertheless, a posteriori the interpretation can be sal-

vaged from the obtained limit theorems.<sup>2</sup> It is a corollary of these that the probability of more than  $l(k)$  neurons being active for some time becomes arbitrarily small for large enough  $l(k)$ . Hence, for physiological reasonable neuron numbers the probability in these models of observing ‘non-physiological’ trajectories in the interpretation becomes ever smaller.

Proceeding with notation,  $\Theta_t = (\Theta_t^1, \dots, \Theta_t^P)$  is a (unbounded) piecewise constant stochastic process taking values in  $\mathbb{N}_0^P$ . The stochastic transitions from inactive to active states and vice versa for a neuron in population  $k$  are governed by a constant inactivation rate  $\tau^{-1} > 0$ —uniformly for all populations—and inputs from other neurons depending on the current network state. This non-negative activation rate is given by  $\tau^{-1}l(k)\bar{f}_k(\theta)$  for  $\theta \in \mathbb{N}_0^P$ . For the definition of  $\bar{f}_k$ , we consider weights  $\bar{W}_{kj}$ ,  $k, j = 1, \dots, P$ , which weigh the input one neuron in population  $k$  receives from a neuron in population  $j$ . Then the activation rate of a neuron in population  $k$  is proportional to

$$\bar{f}_k(\theta) = f\left(\sum_{j=1}^P \bar{W}_{kj}\theta^j\right) \tag{1.5}$$

for a non-negative function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , which obviously corresponds to the gain function  $f$  in the Wilson–Cowan equation (1.3). We remark that here  $f$  is *not* the rate of activation of one neuron. In this model, the activation rate of a population is not proportional to the number of inactive neurons but it is proportional to  $l(k)$ , which stands for the total number of neurons in the population. In [5], this rate is thus interpreted as the rate with which a neuron *becomes or remains* active.

It follows that the process  $(\Theta_t)_{t \geq 0}$  is a continuous-time Markov chain which is usually defined via the following master equation, where  $e_k$  denotes the  $k$ th basis vector of  $\mathbb{R}^P$ ,

$$\begin{aligned} \frac{d\mathbb{P}[\theta, t]}{dt} &= \frac{1}{\tau} \sum_{k=1}^P (l(k)\bar{f}_k(\theta - e_k)\mathbb{P}[\theta - e_k, t] \\ &\quad - (\theta^k + l(k)\bar{f}_k(\theta))\mathbb{P}[\theta, t] + (\theta^k + 1)\mathbb{P}[\theta + e_k, t]) \end{aligned} \tag{1.6}$$

which is endowed with the boundary conditions  $\mathbb{P}[\theta, t] = 0$  if  $\theta \notin \mathbb{N}_0^P$ . In (1.6), the variable  $\mathbb{P}[\theta, t]$  denotes the probability that the process  $\Theta_t$  is in state  $\theta$  at time  $t$ . Finally, the definition is completed with stating an initial law  $\mathcal{L}$ , the distribution of  $\Theta_0$ , i.e., providing an initial value for the ODE system (1.6).

Another definition of a continuous-time Markov chain is via its generator; see, e.g., [15]. Although the master equation is widely used in the physics and chemical reactions literature the mathematically more appropriate object for the study of a Markov process is its generator and the master equation is an object derived from

---

<sup>2</sup>The derivation of limit theorems for bounded populations sizes, where  $l(k)$  actually *is* the number of neurons per population, is much more delicate than the subsequent presentation as the transition rate functions become discontinuous. Although this would be a desirable result, we have not yet been able to prove such a theorem, though it is clear that the Wilson–Cowan equation would be the only possible limit. See also a discussion of this aspect in Sect. 3.2.

the generator, see Sect. V in [33]. The generator of a Markov process is an operator defined on the space of real functions over the state space of the process. For the above model defined by the master equation (1.6), the generator is given by

$$\mathcal{A}g(\theta) = \lambda(\theta) \int_{\mathbb{N}_0^P} (g(\xi) - g(\theta))\mu(\theta, d\xi) \tag{1.7}$$

for all suitable  $g : \mathbb{N}_0^P \rightarrow \mathbb{R}$ . For details, we refer to [15]. Here,  $\lambda$  is the total instantaneous jump rate, given by

$$\lambda(\theta) := \frac{1}{\tau} \sum_{k=1}^P (\theta^k + l(k)\bar{f}_k(\theta)), \tag{1.8}$$

and defines the distribution of the waiting time until the next jump, i.e.,

$$\mathbb{P}[\Theta_{t+s} = \Theta_t \ \forall s \in [0, \Delta t] | \Theta_t = \theta] = e^{-\lambda(\theta)\Delta t}.$$

Further, the measure  $\mu$  in (1.7) is a Markov kernel on the state space of the process defining the conditional distribution of the post-jump value, i.e.,

$$\mathbb{P}[\Theta_t \in A | \Theta_t \neq \Theta_{t-}] = \mu(\Theta_{t-}, A) \tag{1.9}$$

for all sets  $A \subseteq \mathbb{N}_0^P$ . In the present case for each  $\theta$ , the measure  $\mu$  is given by the discrete distribution

$$\begin{aligned} \mu(\theta, \{\theta - e_k\}) &= \frac{1}{\tau} \frac{\theta^k}{\lambda(\theta)}, \\ \mu(\theta, \{\theta + e_k\}) &= \frac{1}{\tau} \frac{l(k)\bar{f}_k(\theta)}{\lambda(\theta)} \quad \forall k = 1, \dots, P. \end{aligned} \tag{1.10}$$

The importance of the generator lies in the fact that it fully characterises a Markov process and that convergence of Markov processes is strongly connected to the convergence of their generators; see [15].

### 1.3 Including External Time-Dependent Input

Until now, the microscopic model does not incorporate any time-dependent input into the system. In analogy to the macroscopic equation (1.3), this input enters into the model inside the active rate function  $\bar{f}_k$ . Thus, let  $\bar{I}_k(t)$  denote the external input into a neuron in population  $k$  at time  $t$ , then the time-dependent activation rate is given by

$$\bar{f}_k(\theta, t) = f\left(\sum_{j=1}^P \bar{W}_{kj}\theta^j + \bar{I}_k(t)\right). \tag{1.11}$$

The most important qualitative difference when substituting (1.5) by (1.11) is that the corresponding Markov process is no longer homogeneous. In particular, the waiting



time distributions in between jumps are no longer exponential, but satisfy

$$\mathbb{P}[\Theta_{t+s} = \Theta_t \ \forall s \in [0, \Delta t] | \Theta_t = \theta] = e^{-\int_0^{\Delta t} \lambda(\theta, s) ds}.$$

Hence, the resulting process is an inhomogeneous continuous-time Markov chain; see, e.g., Sect. 2 in [36]. It is straightforward to write down the corresponding master equation analogously to (1.6) yielding a system of non-autonomous ordinary differential equations, cf. the master equation formulation in [8]. Similarly, there exists the notion of a time-dependent generator for inhomogeneous Markov processes, cf. Sect. 4.7 in [15]. Employing a standard trick, that is, suitably extending the state space of the process, we can transform a inhomogeneous to a homogeneous Markov process [15, 28]. That is, the space-time process  $Y_t := (\Theta_t, t)$  is again a homogeneous Markov process. The initial law of the associated space-time process is  $\mathcal{L} \times \delta_0$  on  $\mathbb{N}^P \times \mathbb{R}_+$ . We emphasise that definitions of the space-time process and its initial law imply that the time-component starts at 0 a.s. and, moreover, moves continuously and deterministically. That is, the trajectories satisfy in between jumps the differential equation

$$\begin{pmatrix} \dot{\theta} \\ \dot{t} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

where the jump intensity  $\lambda$  is given by the sum of all individual time-dependent rates analogously to (1.8). Finally, the post jump value is given by a Markov kernel  $\mu((\theta, t), \cdot) \times \delta_t$  as there clearly do not occur jumps in the progression of time and  $\mu$  is the obvious time-dependent modification of (1.10).

It thus follows, that the space-time process  $(\Theta_t, t)_{t \geq 0}$  is a homogeneous Piecewise Deterministic Markov Process (PDMP); see, e.g., [14, 16, 26]. This connection is particularly important as we apply in the course of the present study limit theorems developed for this type of processes; see [27]. Finally, for the space-time process  $(\Theta_t, t)_{t \geq 0}$ , we obtain for suitable functions  $g : \mathbb{N}_0^P \times \mathbb{R}_+ \rightarrow \mathbb{R}$  the generator

$$Ag(\theta, t) = \partial_t g(\theta, t) + \lambda(\theta, t) \int_{\mathbb{N}_0^P} (g(\xi, t) - g(\theta, t)) \mu((\theta, t), d\xi). \tag{1.12}$$

## 2 A Precise Formulation of the Limit Theorems

In this section, we present the precise formulations of the limit theorems. To this end, we first define a suitable sequence of microscopic models, which gives the connection between the defining objects of the Wilson–Cowan equation (1.3) and the microscopic models discussed in Sect. 1.2. Thus,  $(Y_t^n)_{t \geq 0} = (\Theta_t^n, t)_{t \geq 0}$ ,  $n \in \mathbb{N}$ , denotes a sequence of microscopic PDMP neural field models of the type as defined in Sect. 1.3. Each process  $(Y_t^n)_{t \geq 0}$  is defined on a filtered probability space  $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n)_{t \geq 0}, \mathbb{P}^n)$ , which satisfies the usual conditions. Hence, the defining objects for the jump models are now dependent on an additional index  $n$ . That is  $P(n)$  denotes the number of neuron populations in the  $n$ th model,  $l(k, n)$  is the number of neurons in the  $k$ th population of the  $n$ th model and analogously we use the notations

$\overline{W}_{kj}^n$  and  $\overline{I}_{k,n}$  and  $\overline{f}_{k,n}$ . However, we note from the beginning that the decay rate  $\tau^{-1}$  is independent of  $n$  and  $\tau$  is the time constant in the Wilson–Cowan equation (1.3). In the following paragraphs, we discuss the connection of the defining components of this sequence of microscopic models to the components of the macroscopic limit.

*Connection to the Spatial Domain  $D$*  A key step of connecting the microscopic models to the solution of Eq. (1.3) is that we need to put the individual neuron populations into relation to the spatial domain  $D$  the solution of (1.3) lives on. To this end, we assume that each population is located within a sub-domain of  $D$  and that the sub-domains of the individual populations are non-overlapping. Hence, for each  $n \in \mathbb{N}$ , we obtain a collection  $\mathcal{D}_n$  of  $P(n)$  non-overlapping sub-sets of  $D$  denoted by  $D_{1,n}, \dots, D_{P(n),n}$ . We assume that each subdomain is measurable and convex. The convexity of the sub-domains is a technical condition that allows us to apply Poincaré’s inequality, cf. (4.1). We do not think that this condition is too restrictive as most reasonable partition domains, e.g., cubes, triangles, are convex. Furthermore, for all reasonable domains  $D$ , e.g., all Jordan measurable domains, a sequence of convex partitions can be found such that additionally the conditions imposed in the limit theorems below are also satisfied. One may think of obtaining the collection  $\mathcal{D}_n$  by partitioning the domain into  $P(n)$  convex sub-domains  $D_{1,n}, \dots, D_{P(n),n}$  and confining each neuron population to one sub-domain. However, it is not required that the union of the sets in  $\mathcal{D}_n$  amounts to the full domain  $D$  nor that the partitions consists of refinements. Necessary conditions on the limiting behaviour of the sub-domains are very strongly connected to the convergence of initial conditions of the models, which is a condition in the limit theorems; see below. For the sake of terminological simplicity, we refer to  $\mathcal{D}_n$  simply as the partitions.

We now define some notation for parameters characterising the partitions  $\mathcal{D}_n$ : the minimum and maximum Lebesgue measure, i.e., length, area, or volume depending on the spatial dimension, is denoted by

$$v_-(n) := \min_{k=1, \dots, P(n)} |D_{k,n}|, \quad v_+(n) := \max_{k=1, \dots, P(n)} |D_{k,n}|, \quad (2.1)$$

and the maximum diameter of the partition is denoted by

$$\delta_+(n) := \max_{1, \dots, P(n)} \text{diam}(D_{k,n}), \quad (2.2)$$

where the diameter of a set  $D_{k,n}$  is defined as  $\text{diam}(D_{k,n}) := \sup_{x,y \in D_{k,n}} |x - y|$ . In the special case of domains obtained by unions of cubes with edge length  $n^{-1}$ , it obviously holds that  $v_{\pm}(n) = n^{-d}$  and  $\delta_+(n) = \sqrt{d}n^{-1}$ . It is a necessary condition in all the subsequent limit theorems that  $\lim_{n \rightarrow \infty} \delta_+(n) = 0$ . This condition implies on the one hand that  $\lim_{n \rightarrow \infty} v_+(n) = 0$  as the Lebesgue measure of a set is bounded in terms of its diameter, and on the other hand—at least in all but degenerate cases due to the necessary convergence of initial conditions that  $\lim_{n \rightarrow \infty} P(n) = \infty$ . That is, in order to obtain a limit the sequence of partitions usually consists of ever finer sets and the number of populations diverges. Finally, each domain  $D_{k,n}$  of the partition  $\mathcal{D}_n$  contains one neuron population ‘consisting’ of  $l(k, n) \in \mathbb{N}$  neurons. Then

we denote by  $\ell_{\pm}(n)$  the maximum and minimum number of neurons in populations corresponding to the  $n$ th model, i.e.,

$$\ell_{-}(n) := \min_{k=1, \dots, P(n)} l(k, n), \quad \ell_{+}(n) := \max_{k=1, \dots, P(n)} l(k, n). \tag{2.3}$$

*Connection to the Weight Function  $w$*  We assume that there exists a function  $w : D \times D \rightarrow \mathbb{R}$  such that the connection to the discrete weights is given by

$$\overline{W}_{kj}^n := \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \left( \int_{D_{j,n}} w(x, y) dy \right) dx, \tag{2.4}$$

where  $w$  is the same function as in the Wilson–Cowan equation (1.3). For the definition of activation rate at time  $t$ , we thus obtain

$$\overline{f}_{k,n}(\theta^n, t) := f \left( \sum_{j=1}^P \overline{W}_{kj}^n \frac{\theta^{j,n}}{l(j, n)} + \overline{I}_{k,n}(t) \right). \tag{2.5}$$

As already highlighted by Bressloff [5], the transition rates are not uniquely defined by the requirement that a possible limit to the microscopic models is given by the Wilson–Cowan equation (1.3). If in (2.5), the definition of the transition rates is changed to

$$\overline{f}_{k,n}(\theta^n, t) := f^n \left( \sum_{j=1}^P \overline{W}_{kj}^n \frac{\theta^{j,n}}{l(j, n)} + \overline{I}_{k,n}(t) \right),$$

where  $f^n, n \in \mathbb{N}$ , is a sequence of functions converging uniformly to  $f$ , then all limit theorems remain valid. The proof can be carried out as presented adding and subtracting the appropriate term where the additional difference term vanishes due to  $\sup_{x \in \mathbb{R}} |f^n(x) - f(x)| \rightarrow 0$  for  $n \rightarrow \infty$ . Hence, any microscopic model with gain rates  $\overline{f}_{k,n}$  of such a form reduces to the same Wilson–Cowan equation in the limit. Clearly, the same applies analogously to the decay rate  $\tau$ , the weights  $w$ , and the input  $I$ .

*Connection to the Input Current  $I$*  The external input which is applied to neurons in a certain population is obtained by spatially averaging a space-time input over the sub-domain that population is located in, i.e.,

$$\overline{I}_{k,n}(t) := \frac{1}{|D_{k,n}|} \int_{D_{k,n}} I(t, x) dx. \tag{2.6}$$

This completes the definition of the Markov jump processes  $(\Theta_t^n, t)_{t \geq 0}$ . For the sake of completeness, we repeat the definition of the total jump rate

$$\lambda^n(\theta^n, t) := \frac{1}{\tau} \sum_{k=1}^P (\theta^{k,n} + l(k, n) \overline{f}_{k,n}(\theta^n, t))$$

and the transition measure  $\mu^n$  is defined by

$$\begin{aligned} \mu^n((\theta^n, t), \{\theta^n - e_k\}) &:= \frac{1}{\tau} \frac{\theta^{k,n}}{\lambda^n(\theta^n, t)}, \\ \mu^n((\theta^n, t), \{\theta^n + e_k\}) &:= \frac{1}{\tau} \frac{l(k, n) \overline{f}_{k,n}(\theta^n, t)}{\lambda^n(\theta^n, t)} \end{aligned}$$

for all  $k = 1, \dots, P(n)$ .

*Connection to the Solution  $v$*  As functions of time, the paths of the PDMP  $(\Theta_t^n, t)_{t \geq 0}$  and the solution  $v$  live on different state spaces. The former takes values in  $\mathbb{N}_0^P \times \mathbb{R}_+$  and the latter in  $L^2(D)$ . Thus, in order to compare these two, we have to introduce a mapping that maps the stochastic process onto  $L^2(D)$ . In [27], the authors called such a mapping a *coordinate function*, which is also the terminology used in [13]. In fact, the limit theorems we subsequently present actually are for the processes we obtain from the composition of the coordinate functions with the PDMPs. Here, it is important to note that for each  $n \in \mathbb{N}$  the coordinate functions may—and usually do—differ, however, they project the process into the common space  $L^2(D)$ . For the mean field models, we define the coordinate functions for all  $n \in \mathbb{N}$  by

$$v^n : \mathbb{N}_0^P \rightarrow L^2(D) : \theta^n \mapsto \sum_{k=1}^P \frac{\theta^{k,n}}{l(k, n)} \mathbb{I}_{D_{k,n}}. \tag{2.7}$$

Clearly, each  $v^n$  is a measurable map into  $L^2(D)$ . For the composition of  $v^n$  with the stochastic process  $(\Theta_t^n, t)_{t \geq 0}$ , we also use the abbreviation  $v_t^n := v^n(\Theta_t^n)$ , and hence the resulting stochastic process  $(v_t^n)_{t \geq 0}$  is an adapted càdlàg process taking values in  $L^2(D)$ . This process thus states the activity at a location  $x \in D$  as the fraction of active neurons in the population, which is located around this location.

*Connection of the Initial Conditions* One condition in the subsequent limit theorems is the convergence of initial conditions in probability, i.e., the assumption that

$$\lim_{n \rightarrow \infty} \mathbb{P}^n \left[ \|v^n(\Theta_0^n) - v_0\|_{L^2(D)} > \epsilon \right] = 0 \quad \forall \epsilon > 0. \tag{2.8}$$

It is easy to see that such a sequence of initial conditions  $\Theta_0^n, n \in \mathbb{N}$ , can be found for any deterministic initial condition  $v_0$  under some reasonable conditions on the domain  $D$  and the sequence of partitions  $\mathcal{D}_n$ . Hence, the assumption (2.8) can always be satisfied. For example, we may define such a sequence of initial conditions by

$$\Theta_0^{k,n} = \operatorname{argmin}_{i=1, \dots, l(k,n)} \left| \frac{i}{l(k, n)} - \frac{1}{|D_{k,n}|} \int_{D_{k,n}} v(0, x) \, dx \right|.$$

Next, assuming that partitions fill the whole domain  $D$  for  $n \rightarrow \infty$ , i.e.,  $\lim_{n \rightarrow \infty} |D \setminus \bigcup_{k=1}^{P(n)} D_{k,n}| = 0$ , and that the maximal diameter of the sets decreases to zero, i.e.,  $\lim_{n \rightarrow \infty} \delta_+(n) = 0$ , it is easy to see using the Poincaré inequality (4.1) that the

above definition of the initial condition implies that  $\|v_0^n - v(0)\|_{L^2(D)} \rightarrow 0$  and  $\sup_{n \in \mathbb{N}} \|v_0^n\|_{L^2(D)}^{2r} < \infty$  for all  $r \geq 1$ . Then (2.8) holds trivially as the initial condition is deterministic and converges. A simple non-degenerate sequence of initial conditions is obtained by choosing random initial conditions with the above value as their mean and sufficiently fast decreasing fluctuations. Furthermore, a sequence of partitions, which satisfy the above conditions also exists for a large class of reasonable domains  $D$ . Assume that  $D$  is Jordan measurable, i.e., a bounded domain such that the boundary is a Lebesgue null set, and let  $\mathcal{C}_n$  be the smallest grid of cubes with edge length  $1/n$  covering  $D$ . We define  $\mathcal{D}_n$  to be the set of all cubes, which are fully in  $D$ . As  $D$  is Jordan measurable, these partitions fill up  $D$  from inside and  $\delta_+(n) \rightarrow 0$ . For a more detailed discussion of these aspects, we refer to [26].

In the remainder of this section, we now collect the main results of this article. We start with the law of large numbers, which establishes the connection to the deterministic mean field equation, and then proceed to central limit theorems which provide the basis for a Langevin approximation. The proofs of the results are deferred to Sect. 4.

### 2.1 A Law of Large Numbers

The first law of large numbers takes the following form. Note that the assumptions imply that the number of neuron populations diverges.

**Theorem 2.1** (Law of large numbers) *Let  $w \in L^2(D) \times L^2(D)$  and  $I \in L^2_{\text{loc}}(\mathbb{R}_+, H^1(D))$ . Assume that the sequence of initial conditions converges to  $v(0)$  in probability in the space  $L^2(D)$ , i.e., (2.8) holds, that  $\mathbb{E}^n \Theta_0^{k,n} \leq l(k, n)$ , and that*

$$\lim_{n \rightarrow \infty} \delta_+(n) = 0, \quad \lim_{n \rightarrow \infty} \ell_-(n) = \infty \tag{2.9}$$

*holds. Then it follows that the sequence of  $L^2(D)$ -valued jump-process  $(v_t^n)_{t \geq 0}$  converges uniformly on compact time intervals in probability to the solution  $v$  of the Wilson–Cowan equation (1.3), i.e., for all  $T, \epsilon > 0$  it holds that*

$$\lim_{n \rightarrow \infty} \mathbb{P}^n \left[ \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2(D)} > \epsilon \right] = 0. \tag{2.10}$$

*Moreover, if for  $r \geq 1$  the initial conditions satisfy in addition  $\sup_{n \in \mathbb{N}} \mathbb{E}^n \|v_0^n\|_{L^2(D)}^{2r} < \infty$ , then convergence in the  $r$ th mean holds, i.e., for all  $T > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{E}^n \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2(D)}^r = 0. \tag{2.11}$$

**Remark 2.1** The norm of the uniform convergence  $\sup_{t \in [0, T]} \|\cdot\|_{L^2(D)}$ , which we used in Theorem 2.1 is a very strong norm on the space of  $L^2(D)$ -valued càdlàg functions on  $[0, T]$ . Hence, due to continuous embeddings, the result immediately extends to weaker norms, e.g., the norms  $L^p((0, T), L^2(D))$  for all  $1 \leq p \leq \infty$ . Also, for the state space, weaker spatial norms can be chosen, e.g.,  $L^p(D)$  with  $1 \leq p \leq 2$  or any norm on the duals  $H^{-\alpha}(D)$  of Sobolev spaces with  $\alpha > 0$ . If weaker norms for the

state space are considered, it is possible to relax the conditions of Theorem 2.1 by sharpening some estimates in the proof of the theorem. The results in the following corollary cover the whole range of  $\alpha \geq 0$  and splits it into sections with weakening conditions. In particular note that after passing to weaker norms, the convergence does not necessitate that the neuron numbers per population diverge. However, regarding the divergence of the neuron populations, this condition  $(\delta_+(n) \rightarrow 0)$  cannot be relaxed.

**Corollary 2.1** *Let  $\alpha \geq 0$  and set*

$$q := \begin{cases} \frac{2d}{d+2\alpha} & \text{if } 0 \leq \alpha < d/2, \\ 1- & \text{if } \alpha = d/2, \\ 1 & \text{if } d/2 < \alpha < \infty. \end{cases} \tag{2.12}$$

*Further, assume that  $w \in L^q(D) \times L^2(D)$  and  $I \in L^2_{loc}(\mathbb{R}_+, H^1(D))$  and that the sequence of initial conditions converges to  $v(0)$  in probability in the space  $H^{-\alpha}(D)$ , that  $\lim_{n \rightarrow \infty} \delta_+(n) = 0$  and*

$$\left. \begin{aligned} \lim_{n \rightarrow \infty} \frac{v_+(n)^{2\alpha/d}}{\ell_-(n)} = 0 & \quad \text{if } 0 \leq \alpha < d/2, \\ \lim_{n \rightarrow \infty} \frac{v_+(n)^{1-}}{\ell_-(n)} = 0 & \quad \text{if } \alpha = d/2, \\ \lim_{n \rightarrow \infty} \frac{v_+(n)}{\ell_-(n)} = 0 & \quad \text{if } d/2 < \alpha < \infty, \end{aligned} \right\} \tag{2.13}$$

*where  $1-$  denotes an arbitrary positive number strictly smaller than 1. Then it holds for all  $T, \epsilon > 0$  that*

$$\lim_{n \rightarrow \infty} \mathbb{P}^n \left[ \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{H^{-\alpha}(D)} > \epsilon \right] = 0$$

*and for  $r \geq 1$ , if the additional boundedness assumptions of Theorem 2.1 are satisfied, that for all  $T > 0$*

$$\lim_{n \rightarrow \infty} \mathbb{E}^n \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{H^{-\alpha}(D)}^r = 0.$$

**Remark 2.2** We believe that fruitful and illustrative comparisons of these convergence results and their conditions to the results in Kotelenez [17, 18], and particularly, Blount [4] can be made. Here, we just mention that the latter author conjectured the conditions (2.13) to be optimal for the convergence, but was not able to prove this result in his model of chemical reactions with diffusions for the region  $\alpha \in (0, d/2]$ . For our model, we could achieve these rates.

### 2.1.1 Infinite-Time Convergence

In the law of large numbers, Theorem 2.1, and its Corollary 2.1 we have presented results of convergence over finite time intervals. Employing a different technique, we are also able to derive a convergence result over the whole positive time axis

motivated by a similar result in [32]. The proof of the following theorem is deferred to Sect. 4.3. Restricted to finite time intervals, the subsequent result is strictly weaker than Theorem 2.1. However, the result is important when one wants to analyse the mean long time behaviour of the stochastic model via a bifurcation analysis of the deterministic limit as (2.14) suggests that  $\mathbb{E}^n v_t^n$  is close to  $v(t)$  for all times  $t \geq 0$  for sufficiently large  $n$ .

**Theorem 2.2** *Let  $\alpha \geq 0$  and assume that the conditions of Corollary 2.1 are satisfied. We further assume that the current input function  $I \in L^2_{\text{loc}}(\mathbb{R}_+, H^1(D))$  satisfies  $\|\nabla_x I\|_{L^\infty(\mathbb{R}_+, L^2(D))} < \infty$ , i.e., it is square integrable in  $H^1(D)$  over bounded intervals, and possesses first spatial derivatives bounded for almost all  $t \geq 0$  in  $L^2(D)$ . Then it holds that*

$$\lim_{n \rightarrow \infty} \sup_{t \geq 0} \mathbb{E}^n \|v_t^n - v(t)\|_{H^{-\alpha}(D)} = 0. \tag{2.14}$$

### 2.2 A Martingale Central Limit Theorem

In this section, we present a central limit theorem for a sequence of martingales associated with the jump processes  $v^n$ . A brief, heuristic discussion of the method of proof for the law of large numbers explains the importance of these martingales and motivates their study. In the proof of the law of large numbers, the central argument relies on the fact that the process  $(v_t^n)_{t \geq 0}$  satisfies the decomposition

$$v_t^n = v_0^n + \int_0^t \lambda(\Theta_s^n, s) \int_{\mathbb{N}_0^p} (v^n(\xi) - v^n(\Theta_s^n)) \mu^n((\Theta_s^n, s), d\xi) ds + M_t^n. \tag{2.15}$$

Here, the process  $(M_t^n)_{t \geq 0}$  is a Hilbert space-valued, square-integrable, càdlàg martingale using (2.15) as its definition. We have used this representation of the process  $v^n$  in the proof of Theorem 2.2; see Sect. 4.3. We note that the Bochner integral in (2.15) is a.s. well defined due to bounded second moments of the integrand; see (4.7) in the proof of Theorem 2.1. Now an heuristic argument to obtain the convergence to the solution of the Wilson–Cowan equation is the following: The initial conditions converge, the martingale term  $M^n$  converges to zero and the integral term in the right-hand side of (2.15) converges to the right-hand side in the Wilson–Cowan equation (1.3). Hence, the ‘solution’  $v^n$  of (2.15) converges to the solution  $v$  of the Wilson–Cowan equation (1.3). Now interpreting Eq. (2.15) as a stochastic evolution equation, which is driven by the martingale  $(M_t^n)_{t \geq 0}$  sheds light on the importance of the study of this term. Because, from this point of view, the martingale part in the decomposition (2.15) contains all the stochasticity inherent in the system. Then the idea for deriving a Langevin or linear noise approximation is to find a stochastic non-trivial limit (in distribution) for the sequence of martingales and substituting heuristically this limiting martingale into the stochastic evolution equation. Then it is expected that this new and much less complex process behaves similarly to the process  $(v_t^n)_{t \geq 0}$  for sufficiently large  $n$ . Deriving a suitable limit for  $(M_t^n)_{t \geq 0}$  is what we set to do next. The result can be found in Theorem 2.3 below and takes the form of a central limit theorem.

First of all, what has been said so far implies the necessity of re-scaling the martingale with a diverging sequence in order to obtain a non-trivial limit. The conditions in the law of large numbers imply in particular that the martingale converges uniformly in the mean square to zero, i.e.,

$$\lim_{n \rightarrow \infty} \mathbb{E}^n \sup_{t \in [0, T]} \|M_t^n\|_{L^2(D)} = 0,$$

which in turn implies convergence in probability and convergence in distribution to the zero limit.

Furthermore, in contrast to Euclidean spaces norms on infinite-dimensional spaces are usually not equivalent. In Corollary 2.1, we exploited this fact as it allowed us to obtain convergence results under less restrictive conditions by changing to strictly weaker norms. In the formulation and proof of central limit theorems, the change to weaker norms even becomes an essential ingredient. It is often observed in the literature, see, e.g., [4, 17, 18] that central limit theorems cannot be proven in the strongest norm for which the law of large numbers holds, e.g.,  $L^2(D)$  in the present setting, but only in a strictly weaker norm. Here, this norm is the norm in the dual of an appropriate Sobolev space. Hence, from now on, we consider for all  $n \in \mathbb{N}$  the processes  $(v_t^n)_{t \geq 0}$  and the martingales  $(M_t^n)_{t \geq 0}$  as taking values in the space  $H^{-\alpha}(D)$  for an  $\alpha > d$ , where  $d$  is the dimension of the spatial domain  $D$ , using the embedding of  $L^2(D)$  into  $H^{-\alpha}(D)$ . The technical significance of the restriction  $\alpha > d$  is that these are the indices such that there exists an embedding  $H^\alpha(D)$  into a  $H^{\alpha_1}(D)$  with  $d/2 < \alpha_1 < \alpha$ , which is of Hilbert–Schmidt type<sup>3</sup> due to Maurin’s theorem and  $H^{\alpha_1}(D)$  is embedded into  $C(\overline{D})$  due to the Sobolev embedding theorem. These two properties are essential for the proof of the central limit theorem and their occurrence will be made clear subsequently.

The limit we propose for the re-scaled martingale sequence is a *centred diffusion process* in  $H^{-\alpha}(D)$ , that is, a centred continuous Gaussian stochastic process  $(X_t)_{t \geq 0}$  taking values in  $H^{-\alpha}(D)$  with independent increments and given covariance  $C(t)$ ,  $t \geq 0$ ; see, e.g., [12, 25] for a discussion of Gaussian processes in Hilbert spaces. Such a process is uniquely defined by its covariance operator and conversely, each family of linear, bounded operators  $C(t) : H^\alpha(D) \rightarrow H^{-\alpha}(D)$ ,  $t \geq 0$ , uniquely defines a diffusion process<sup>4</sup> if

<sup>3</sup>A continuous embedding of two Hilbert spaces  $X \hookrightarrow Y$  is of Hilbert–Schmidt type if for every orthonormal basis  $\varphi_j$ ,  $j \in \mathbb{N}$ , of  $X$  it holds that  $\sum_{j=1}^\infty \|\varphi_j\|_Y^2 < \infty$ . Then, more precisely, Maurin’s theorem states that for non-negative integers  $m, k$ , the embedding of  $H^{m+k}(D)$  into  $H^m(D)$  is of Hilbert–Schmidt type for  $k > d/2$ ; see [2]. The result was generalised to fractional order Sobolev spaces in [35]: Let  $D$  be a bounded, strong local Lipschitz domain in  $\mathbb{R}^d$  and  $0 \leq \alpha_1 < \alpha_2$  are real numbers. Then it holds that the embedding of  $H^{\alpha_2+d/2}(D)$  into  $H^{\alpha_1}(D)$  is of Hilbert–Schmidt type.

<sup>4</sup>Usually, the covariance operator for a Hilbert space-valued process is an operator mapping from the state space into the state space and not into the dual, i.e., in the present situation mapping  $H^{-\alpha}(D)$  into itself. Due to the canonical embedding of Hilbert spaces into their dual and the Riesz representation, however, we can effortlessly change from the usual definition to ours and vice versa. Moreover, the symmetry condition thus implies due to the Hellinger–Toeplitz theorem that the operator is self-adjoint, and hence of trace class if and only if (2.16) is satisfied. The choice of the presentation here is due to the fact that it is simpler to evaluate the duality pairing on  $H^{-\alpha}(D)$  than the inner product thereon, as the former usually is just the inner product in  $L^2(D)$ .



(i) each  $C(t)$  is *symmetric* and *positive*, i.e.,

$$\langle C(t)\phi, \psi \rangle_{H^\alpha(D)} = \langle C(t)\psi, \phi \rangle_{H^\alpha(D)} \quad \text{and} \quad \langle C(t)\phi, \phi \rangle_{H^\alpha(D)} \geq 0,$$

(ii) each  $C(t)$  is of *trace class*, i.e., for one (and thus every) orthonormal basis  $\varphi_j$ ,  $j \in \mathbb{N}$ , in  $H^\alpha(D)$  it holds that

$$\sum_{j=1}^{\infty} \langle C(t)\varphi_j, \varphi_j \rangle_{H^\alpha(D)} < \infty, \tag{2.16}$$

(iii) and the family  $C(t)$ ,  $t \geq 0$ , is *continuously increasing* in  $t$  in the sense that the map  $t \mapsto \langle C(t)\phi, \psi \rangle_{H^\alpha(D)}$  is continuous and increasing for all  $\phi, \psi \in H^\alpha(D)$ .

We next define the process, which will be the limit identified in the martingale central limit theorem via its covariance. In order to define the operator  $C$ , we first define a family of linear operators  $G(v(t), t)$  mapping from  $H^\alpha(D)$  into the dual space  $H^{-\alpha}(D)$  via the bilinear form

$$\begin{aligned} &\langle G(v(t), t)\phi, \psi \rangle_{H^\alpha(D)} \\ &= \int_D \phi(x) \left( \frac{1}{\tau} v(t, x) \right. \\ &\quad \left. + \frac{1}{\tau} f \left( \int_D w(x, y) v(t, y) dy + I(t, x) \right) \right) \psi(x) dx. \end{aligned} \tag{2.17}$$

It is obvious that this bilinear form is symmetric and positive and, as  $v(t)$  is continuous in  $t$ , it holds that the map  $t \mapsto \langle G(v(t), t)\phi, \psi \rangle_{H^\alpha(D)}$  is continuous for all  $\phi, \psi \in H^\alpha(D)$ . Furthermore, it is easy to see that the operator is bounded, i.e.,

$$\|G(v(t), t)\|_{L(H^\alpha(D), H^{-\alpha}(D))} = \sup_{\|\phi\|_{H^\alpha(D)}=1} \sup_{\|\psi\|_{H^\alpha(D)}=1} |\langle G(v(t), t)\phi, \psi \rangle_{H^\alpha(D)}| < \infty,$$

as the solution of the Wilson–Cowan equation  $v$  and the gain function  $f$  are pointwise bounded. Hence, due to the Cauchy–Schwarz inequality, the norm  $|\langle G(v(t), t)\phi, \psi \rangle_{H^\alpha(D)}|$  is proportional to the product  $\|\phi\|_{L^2(D)}\|\psi\|_{L^2(D)}$  and for any  $\alpha \geq 0$  the Sobolev embedding theorem gives now a uniform bound in terms of the norm of  $\phi, \psi$  in  $H^\alpha(D)$ . As a final property, we show that these operators are of trace-class if  $\alpha > d/2$ . Thus, let  $(\varphi_j)_{j \in \mathbb{N}}$  be an orthonormal basis in  $H^\alpha(D)$ , then the Cauchy–Schwarz inequality yields

$$|\langle G(v(t), t)\varphi_j, \varphi_j \rangle_{H^\alpha(D)}| \leq \frac{1}{\tau} (1 + \|f\|_0) |D| \|\varphi_j\|_{L^2(D)}^2.$$

Summing these inequalities for all  $j \in \mathbb{N}$ , we find that the resulting right-hand side is finite as due to Maurin’s theorem the embedding of  $H^\alpha(D)$  into  $L^2(D)$  is of Hilbert–Schmidt type. Moreover, their trace is even bounded independently of  $t$ .

Now, it holds that the map  $t \mapsto G(v(t), t)$  is continuous taking values in the Banach space of trace class operators, hence we define trace class operators  $C(t)$  from

$H^\alpha(D)$  into  $H^{-\alpha}(D)$  via the Bochner integral for all  $t \geq 0$

$$C(t) := \int_0^t G(v(s), s) \, ds. \tag{2.18}$$

Clearly, the resulting bilinear form  $\langle C(t)\cdot, \cdot \rangle_{H^\alpha(D)}$  inherits the properties of the bilinear form (2.17). Moreover, due to the positivity of the integrands, it follows that  $\langle C(t)\phi, \phi \rangle_{H^\alpha(D)}$  is increasing in  $t$  for all  $\phi \in H^\alpha(D)$ . Hence, the family of operators  $C(t)$ ,  $t \geq 0$ , satisfies the above conditions (i)–(iii), and thus uniquely defines an  $H^{-\alpha}(D)$ -valued diffusion process.

We are now able to state the martingale central limit theorem. The proof of the theorem is deferred to Sect. 4.4.

**Theorem 2.3** (Martingale central limit theorem) *Let  $\alpha > d$  and assume that the conditions of Theorem 2.1 are satisfied. In particular, convergence in the mean holds, i.e., (2.11) holds for  $r = 1$ . Additionally, we assume it holds that*

$$\lim_{n \rightarrow \infty} \frac{v_-(n)}{v_+(n)} \frac{\ell_-(n)}{\ell_+(n)} = 1. \tag{2.19}$$

Then it follows that the sequence of re-scaled  $H^{-\alpha}(D)$ -valued martingales

$$\left( \sqrt{\frac{\ell_-(n)}{v_+(n)}} M_t^n \right)_{t \geq 0}$$

converges weakly on the space of  $H^{-\alpha}(D)$ -valued càdlàg function to the  $H^{-\alpha}(D)$ -valued diffusion process defined by the covariance operator  $C(t)$  given by (2.18).

*Remark 2.3* In connection with the results of Theorem 2.3, two questions may arise. First, in what sense is there uniqueness of the re-scaling sequence, and hence of the limiting diffusion? That is, does a different scaling also produce a (non-trivial) limit, or, rephrased, is the proposed scaling the correct one to look at? Secondly, the theorem deals with the norms for the range of  $\alpha > d$  in the Hilbert scale, what can be said about convergence in the stronger norms corresponding to the range of  $\alpha \in [0, d]$ ? Does there exist a limit? We conclude this section addressing these two issues.

Regarding the first question, it is immediately obvious that the re-scaling sequence  $\frac{\ell_-(n)}{v_+(n)}$ , which we denote by  $\rho_n$  in the following, is not a unique sequence yielding a non-trivial limit. Re-scaling the martingales  $M^n$  by any sequence of the form  $\sqrt{c\rho_n}$  yields a convergent martingale sequence. However, the limiting diffusion differs only in a covariance operator, which is also re-scaled by  $c$ , and hence the limit is essentially the same process with either ‘stretched’ or ‘shrunked’ variability. However, the asymptotic behaviour of the re-scaling sequences, which allow for a non-trivial weak limit is unique. In general, by considering different re-scaling sequences  $\rho_n^*$ , we obtain three possibilities for the convergence of the sequence  $\sqrt{\rho_n^*} M^n$ . If  $\rho_n^*$  is of the same speed of convergence as  $\rho_n$ , i.e., for  $\rho_n^* = \mathcal{O}(\rho_n)$ , the thus re-scaled sequence converges again to a diffusion process for which the covariance operator is

proportional to (2.18). This is then just a re-scaling by a sequence (asymptotically) proportional to  $\rho_n$  as discussed above. Secondly, if the convergence is slower, i.e.,  $\rho_n^* = o(\rho_n)$ , then the same methods as in the law of large numbers show that the sequence converges to zero uniformly on compacts in probability, hence also convergence in distribution to the degenerate zero process follows. Thus, one only obtains the trivial limit. Finally, if we rescale by a sequence that diverges faster, i.e.,  $\rho_n = o(\rho_n^*)$ , we can show that there does not exist a limit. This follows from general necessary conditions for the preservation of weak limits under transformation, which presuppose that  $\sqrt{\rho_n^*/\rho_n}M$  has to converge in distribution in order for  $\sqrt{\rho_n^*}M_n$  possessing a limit in distribution; see Theorem 2 in [29]. As the sequence  $\rho_n^*/\rho_n$  diverges, this is clearly not possible to hold.

Unfortunately, an answer to the second question is not possible in this clarity, when considering non-trivial limits. Essentially, we can only say that the currently used methods do not allow for any conclusion on convergence. The limitations are the following: The central problem is that for the parameter range  $\alpha \in [0, d]$  the current method does not provide tightness of the re-scaled martingale sequence, hence we cannot infer that the sequence possesses a convergent subsequence. However, if tightness can be established in a different way then for the range  $\alpha \in (\max\{1, d/2\}, d]$ , the limit has to be the diffusion process defined by the operator (2.18) as follows from the characterisation of any limit in the proof of the theorem. Here, the lower bound of  $\max\{1, d/2\}$  results, on the one hand, from our estimation technique, which necessitates  $\alpha \geq 1$ , and on the other hand, from the definition of the limiting diffusion. Recall that the covariance operator is only of trace class for  $\alpha > d/2$ . Hence, for  $\alpha \in [0, d/2]$ , we can no longer infer that the limiting diffusion even exists.

### 2.3 The Mean-Field Langevin Equation

An important property of the limiting diffusion in view toward analytic and numerical studies is that it can be represented by a stochastic integral with respect to a cylindrical or  $Q$ -Wiener process. For a general discussion of infinite-dimensional stochastic integrals, we refer to [12]. First, let  $(W_t)_{t \geq 0}$  be a cylindrical Wiener process on  $H^{-\alpha}(D)$  with covariance operator being the identity. Then  $G(v(t), t) \circ \iota^{-1}$  is a trace class operator on  $H^{-\alpha}(D)$  for suitable values of  $\alpha$ . Here,  $\iota^{-1} : H^{-\alpha}(D) \rightarrow H^\alpha(D)$  is the Riesz representation, i.e., the usual identification of a Hilbert space with its dual. The operator  $G(v(t), t) \circ \iota^{-1}$  possesses a unique square-root we denote by  $\sqrt{G(v(t), t) \circ \iota^{-1}}$ , which is a Hilbert–Schmidt operator on  $H^{-\alpha}(D)$ . It follows that the stochastic integral process

$$Z_t := \int_0^t \sqrt{G(v(s), s) \circ \iota^{-1}} dW_s \tag{2.20}$$

is a diffusion process in  $H^{-\alpha}(D)$  with covariance operator  $C(t)$ . That is,  $(Z_t)_{t \geq 0}$  is a version of the limiting diffusion in Theorem 2.3. Now, formally substituting for the limits in (2.15) yields the *linear noise approximation*

$$U_t = v_0 + \int_0^t \tau^{-1}(U_s + F(U_s, s)) ds + \epsilon_n \int_0^t \sqrt{G(v(s), s) \circ \iota^{-1}} dW_s,$$

or in differential notation

$$dU_t = \tau^{-1}(U_t + F(U_t, t)) dt + \epsilon_n \sqrt{G(v(t), t) \circ \iota^{-1}} dW_t, \quad U_0 = v_0, \quad (2.21)$$

where  $\epsilon_n = \sqrt{v_+(n)/\ell_-(n)}$  is small for large  $n$ . Here, we have used the operator notation

$$F : H^{-\alpha}(D) \times \mathbb{R}_+ \rightarrow H^{-\alpha}(D) : F(g, t)(x) = f(\langle g, w(x, \cdot) \rangle_{H^\alpha(D)} + I(t, x)).$$

Equation (2.21) is an infinite-dimensional stochastic differential equation with additive (linear) noise. Here, additive means that the coefficient in the diffusion term does not depend on the solution  $U_t$ . A second formal substitution yields the *Langevin approximation*. Here, the dependence of the diffusion coefficient on the deterministic limit  $v$  is formally substituted by a dependence on the solution. That is, we obtain a stochastic partial differential equation with multiplicative noise given by

$$V_t = V_0 + \int_0^t \tau^{-1}(V_s + F(V_s, s)) ds + \epsilon_n \int_0^t \sqrt{G(V_s, s) \circ \iota^{-1}} dW_s,$$

or in differential notation

$$dV_t = \tau^{-1}(V_t + F(V_t, t)) dt + \epsilon_n \sqrt{G(V_t, t) \circ \iota^{-1}} dW_t. \quad (2.22)$$

Note that the derivation of the above equations was only formal, hence we have to address the existence and uniqueness of solutions and the proper setting for these equations. This is left for future work. It is an ongoing discussion and probably undecidable as lacking a criterion of approximation quality which—if any at all—is the correct diffusion approximation to use. First of all note that for both versions the noise term vanishes for  $n \rightarrow \infty$ , and thus both have the Wilson–Cowan equation as their limit. And also, neither of them approximates even the first moment of the microscopic models exactly. This means that for neither we have that the mean solves the Wilson–Cowan equation, which would be only the case if  $f$  were linear. However, they are close to the mean of the discrete process. We discuss this aspect in Appendix B.

Furthermore, we already observe in the central limit theorem, and thus also in the linear noise and Langevin approximation that the covariance (2.18) or the drift and the structure of the diffusion terms in (2.21) and (2.22), respectively, are independent of objects resulting from the microscopic models. They are defined purely in terms of the macroscopic limit. This observation supports the conjecture that these approximations are independent from possible different microscopic models converging to the same deterministic limit. Analogous statements hold also for derivations from the van Kampen system size expansion [5] and in related limit theorems for reaction diffusion models [4, 17, 18]. The only object reminiscent of the microscopic models in the continuous approximations is the re-scaling sequence  $\epsilon_n$ . However, the re-scaling is proportional to the square root of  $\ell_-(n)/v_+(n)$ , i.e., the number of neurons per area divided by the size of the area, which is just the local density of particles. Therefore, in the approximations, the noise scales inversely to the square root of neuron density in this model, which interpreted in this way can also be considered a macroscopic fixed parameter and chosen independently of the approximating sequence.

*Remark 2.4* The stochastic partial differential equations (2.21) and (2.22), which we proposed as the linear noise or Langevin approximation, respectively, are not necessarily unique as the representation of the limiting diffusion as a stochastic integral process (2.20) may not be unique. It will be subject for further research efforts to analyse the practical implications and usability of this Langevin approximation. Let  $Q$  be a trace class operator,  $(W_t^Q)_{t \geq 0}$  be a  $Q$ -Wiener process and let  $B(v(t), t)$  be operators such that  $B(v(t), t) \circ Q \circ B(v(t), t)^* = G(v(t), t) \circ \iota^{-1}$ , where  $*$  denotes the adjoint operator. Then also the stochastic integral process

$$Z_t^Q := \int_0^t B(v(s), s) dW_s^Q$$

is a version of the limiting diffusion in (2.3) and the corresponding linear noise and Langevin approximations are given by

$$dU_t^Q = \tau^{-1}(U_t^Q + F(U_t^Q, t)) dt + \epsilon_n B(v(t), t) dW_t^Q$$

and

$$dV_t^Q = \tau^{-1}(V_t^Q + F(V_t^Q, t)) dt + \epsilon_n B(V_t^Q, t) dW_t^Q.$$

We conclude this section by presenting one particular choice of a diffusion coefficient and a Wiener process. We take  $(W_t^Q)_{t \geq 0}$  to be a cylindrical Wiener process on  $L^2(D)$  with covariance  $Q = \text{Id}_{L^2}$ . Then we can choose  $B(t) = j \circ (\cdot \sqrt{g(t)}) \in L(L^2(D), H^{-\alpha}(D))$ , where  $j$  is the embedding operator  $L^2(D) \hookrightarrow H^{-\alpha}(D)$  in the sense of (1.2) and  $(\cdot \sqrt{g(t)}) \in L(L^2(D), L^2(D))$  denotes a pointwise product of a function in  $L^2(D)$ , i.e.,

$$(\phi \cdot \sqrt{g(t)})(x) = \phi(x) \left( \tau^{-1} v(t, x) + \tau^{-1} f \left( \int_D w(x, y) v(t, y) dy + I(t, x) \right) \right)^{1/2}.$$

We first investigate the operator  $G(v(t), t) \circ \iota^{-1}$  and write it in more detail as the following composition of operators:

$$G(v(t), t) \circ \iota^{-1} = j \circ (\cdot g(t)) \circ k \circ \iota^{-1},$$

where  $k$  is the embedding operator  $H^\alpha(D) \hookrightarrow L^2(D)$ . Next, the Hilbert adjoint  $B^* \in L(H^{-\alpha}, L^2)$  is given by  $B^* = (\cdot \sqrt{g}) \circ k \circ \iota^{-1}$ , which is easy to verify. Hence, the stochastic integral of  $B(t)$  with respect to  $W^Q$  is again a version of the limiting martingale as

$$\begin{aligned} B(t) \circ Q \circ B^*(t) &= j \circ (\cdot \sqrt{g(t)}) \circ \text{Id}_{L^2} \circ (\cdot \sqrt{g(t)}) \circ k \circ \iota^{-1} \\ &= j \circ (\cdot g(t)) \circ k \circ \iota^{-1} = G(v(t), t) \circ \iota^{-1}. \end{aligned}$$

### 3 Discussion and Extensions

In this article, we have presented limit theorems that connect finite, discrete microscopic models of neural activity to the Wilson–Cowan neural field equation. The

results state qualitative connections between the models formulated as precise probabilistic convergence concepts. Thus, the results strengthen the connection derived in a heuristic way from the van Kampen system size expansion.

A general limitation of mathematically precise approaches to approximations, cf. also the propagation to chaos limit theorems in [30], is that the microscopic models are usually defined via the limit. In other words, the limit has to be known a priori, and we look for models which converge to this limit. Thus, in contrast to the van Kampen system size expansion, the presented results are not a step-by-step modelling procedure in the sense that, via a constructive limiting procedure, a microscopic model yields a deterministic or stochastic approximation. Hence, it might be objected that the presented method can only be used a posteriori in order to justify a macroscopic model from a constructed microscopic model and that somehow one has to ‘guess’ the correct limit in advance. Several remarks can be made to answer this objection.

First, this observation is certainly true, but not necessarily a drawback. On the contrary, when both microscopic and macroscopic models are available, then it is rather important to know how these are connected and qualitatively and quantitatively characterise this connection. Concerning neural field models, this precise connection was simply not available so far for the well-established Wilson–Cowan model. Furthermore, when starting from a stochastic microscopic description working through proving the conditions for convergence for given microscopic models, one obtains very strong hints on the structure of a possible deterministic limit. Therefore, our results can also ease the procedure of ‘guessing the correct limit’.

Secondly, often a phenomenological, deterministic model, which is an approximation to an inherently probabilistic process is derived from ad-hoc heuristic arguments. Given that the model has proved useful, one often aims to derive a justification from first principles and/or a stochastic version, which keeps the features of the deterministic model, but also accounts for the formerly neglected fluctuations. A standard, though somewhat simple approach to obtain stochastic versions consists of adding (small) noise to the deterministic equations. This article, provides a second approach which consists of finding microscopic models, which converge to the deterministic limit to obtain a stochastic correction via a central limit argument.

Thirdly and finally, the method also provides an argument for new equations, i.e., the Langevin and linear noise approximations, which can be used to study the stochastic fluctuations in the model. Furthermore, in contrast to previous studies, we do not provide deterministic moment equations but stochastic processes, which can be, e.g., via Monte Carlo simulations, studied concerning a large number of pathwise properties and dynamics beyond first and second moments.

We now conclude this article commenting on the feasibility of our approach connecting microscopic Markov models to deterministic macroscopic equations when dealing with different master equation formulations that appear in the literature. Additionally, the following discussions also relate the model (1.6) considered in this article to other master equation formulations. We conjecture that the analogous results as presented for the Wilson–Cowan equation (1.3) in Sect. 2 also hold for these variations of the master equations. This should be possible to achieve by an adaptation of the methods of proof presented although we have not performed the computations in detail.

### 3.1 A Variation of the Master Equation Formulation

A first variation of the discrete model we discussed in Sect. 1.2 was considered in the articles [8, 9] and a version restricted to a bounded state space also appears in [31]. This model consists of the master equation stated below in (3.2), which closely resembles (1.6). In the earlier reference [8], the model was introduced with a different interpretation called the *effective spike model*. We briefly explain this interpretation before presenting the master equation. Instead of interpreting  $P$  as the number of neuron populations, in this model,  $P$  denotes the number of different neurons in the network located within a spatial domain  $D$ . Then  $\Theta_t^k$ , the state of the  $k$ th neuron, counts the number of ‘effective’ spikes this neuron has emitted in the past up till time  $t$ . Effective spikes are those spikes that still influence the dynamics of the system, e.g., via a post-synaptic potential. Then state transitions adding/subtracting one effective spike for the  $k$ th neuron are governed by a firing rate function  $\tilde{f}_k$ , which depends on the input into neuron  $k$ , and a decay rate  $\tau^{-1}$ . The constant decay rate indicates that emitted spikes are effective for a time interval of length  $\tau$  and the gain function is defined—neglecting external input—by

$$\tilde{f}_k(\theta) = f^* \left( \sum_{j=1}^P \tilde{W}_{kj} \theta^j \right),$$

where  $f^*$  is a certain non-negative, real function. It is stated clearly in [9] that the function  $f^*$  is *not* equal to the gain function  $f$  in the proposed limiting Wilson–Cowan equation (1.3), but rather connected to  $f$  such that

$$\mathbb{E} f^* \left( \sum_{j=1}^P \tilde{W}_{kj} \Theta_t^j \right) = f \left( \sum_{j=1}^P \tilde{W}_{kj} \mathbb{E} \Theta_t^j \right) + \text{higher order terms.} \tag{3.1}$$

The authors in [9] state that for any function  $f$  such a function  $f^*$  can be found. Then the process  $\Theta_t = (\Theta_t^1, \dots, \Theta_t^P)$  is a jump Markov process given by the master equation

$$\begin{aligned} \frac{d\mathbb{P}[\theta, t]}{dt} = & \sum_{k=1}^P \left[ \tilde{f}_k(\theta - e_k) \mathbb{P}[\theta - e_k, t] \right. \\ & \left. - \left( \frac{1}{\tau} \theta^k + \tilde{f}_k(\theta) \right) \mathbb{P}[\theta, t] + \frac{1}{\tau} (\theta^k + 1) \mathbb{P}[\theta + e_k, t] \right] \end{aligned} \tag{3.2}$$

with boundary conditions  $\mathbb{P}[\theta, t] = 0$  if  $\theta \notin \mathbb{N}_0^P$  as stated in [9]. The advantage of the effective spike model interpretation over the interpretation as neurons per population is that the unbounded state space of the model is justified. In principle, there can be an arbitrary number of spikes emitted in the past still active. However, a disadvantage of the master equation (3.2) is that for taking the limit it lacks a parameter corresponding to the system size providing a natural small parameter in the van Kampen system size expansion. This explains the shift in the interpretation of the master equation in the

study [9] following [8], and subsequently in [5] to the interpretation we presented in Sect. 1.2, which provides the system-size parameters  $l(k)$ .

On the level of Markov jump processes, the master equation (3.2) obviously describes dynamics similar to the master equation (1.6) only replacing the activation rate  $\tau^{-1}l(k)\bar{f}_k(\theta)$  in (1.6) by  $\tilde{f}_k(\theta)$  which is independent of the parameter  $l(k)$ . Thus, the model (3.2) can be understood as resulting from (1.6) after a limit procedure taking  $l(k) \rightarrow \infty$  has been applied and the firing rate functions are connected via the formal limit  $\lim_{l(k) \rightarrow \infty} l(k)\bar{f}_k(\theta) = \tilde{f}_k(\theta)$ . A qualitative interpretation of this limit procedure connecting the two types of models is given in [8]. This observation motivated the model in [5] stepping back one limit procedure, and thus providing the correct framework for the derivation of limit theorems.

It would be an interesting addition to the limit theorems in Theorem 2.1 to derive a law of large numbers for the models (3.2) with stochastic mean activity  $\nu^n$  as defined in (2.7) and suitable chosen weights  $\tilde{W}_{kj}$ . Clearly, the macroscopic limit should be given by the Wilson–Cowan equation (1.3). We conjecture that the appropriate condition for the function  $f^*$  in the present setting—including time dependent inputs—is

$$\begin{aligned} \mathbb{E} \left[ l(k, n)^{-1} f^* \left( \sum_{j=1}^P \tilde{W}_{kj}^n \Theta_j + \tilde{I}_{k,n}(t) \right) \right] \\ = f \left( \sum_{j=1}^P \bar{W}_{kj}^n \frac{\mathbb{E} \Theta_j}{l(j, n)} + \bar{I}_{k,n}(t) \right) + \text{h.o.t.}, \end{aligned} \tag{3.3}$$

such that the higher order terms are uniformly bounded and vanish in the limit  $n \rightarrow \infty$ , and where the weights  $\bar{W}_{kj}^n$  and inputs  $\bar{I}_{k,n}(t)$  are defined as in (2.4) and (2.6). Property (3.3) closely resembles condition (3.1) and trivially holds for linear  $f$  with  $f^* = f$ .

### 3.2 Bounded State Space Master Equations

We have already stated when introducing the microscopic model in Sect. 1.2 that the interpretation of the parameter  $l(k)$  as the number of neurons in the  $k$ th population is not literally correct. The state space of the process is unbounded, hence arbitrarily many neurons can be active, and thus each population contains arbitrarily many neurons. In order to overcome this interpretation problem, it was supposed to consider the master equation only on a bounded state space. That is, the  $k$ th population consists of  $l(k)$  neurons, and  $0 \leq \Theta_t^k \leq l(k)$  almost surely. Such master equations are simply obtained by setting the transition rates for transition of  $\theta^k$  from  $l(k) \rightarrow l(k) + 1$  to zero.

A first master equation of this form was considered in [22], which in present notation, takes the form

$$\begin{aligned} \frac{d\mathbb{P}[\theta, t]}{dt} = \frac{1}{\tau} \sum_{k=1}^P [ (l(k) - \theta^k + 1) \bar{f}_k(\theta - e_k) \mathbb{P}[\theta^k - e_k, t] \\ - (\theta^k + (l(k) - \theta^k) \bar{f}_k(\theta)) \mathbb{P}[\theta, t] + (\theta^k + 1) \mathbb{P}[\theta + e_k, t] ]. \end{aligned} \tag{3.4}$$



Versions of such a master equation for, e.g., one population only or coupled inhibitory and excitatory populations were considered in [3, 22], and a van Kampen systems size expansion was carried out. Here, the bound in the state space provides a natural parameter for the re-scaling, thus a small parameter for the expansion. The setup of this problem resembles closely the structure of excitable membranes for which limits have been obtained with the present technique by one of the present author and co-workers in [27]. Therefore, we conjecture that our limit theorems also apply to this setting with minor adaptations with essentially the same conditions and results as in Sect. 2. However, the macroscopic limit, which will be obtained does not conform with the Wilson–Cowan equation but will be given by

$$\tau \dot{v}(t, x) = -v(t, x) + (1 - v(t, x)) f \left( \int_D w(x, y) v(t, y) dy + I(t, x) \right). \quad (3.5)$$

Next, we return to the master equation (1.6) as discussed in this article in Sect. 1.2 and the comment we made regarding bounded state spaces the footnote on page 7. In our primary reference for this model [5], actually a bounded state space version of the master equation was considered where the activation rate for the event  $\theta^k \rightarrow \theta^k + 1$  is

$$l(k) \bar{f}_k(\theta, t) \mathbb{I}_{[\theta_k < l(k)]}, \quad (3.6)$$

replacing  $l(k) \bar{f}_k(\theta, t)$  in (1.6). The van Kampen system size expansion was then applied to this bounded state space master equation, tacitly neglecting possible difficulties, which might arise due to the discontinuity of (3.6) considered as a function on  $\mathbb{R}^P$ . However, for the present, mathematically precise limit convergence results considering bounded state space as originally suggested in [5] are problematic. The discontinuous activation rate (3.6) causes the machinery developed in [27], which depends on Lipschitz-type estimates to break down. However, we strongly expect that also in this case the law of large numbers with the deterministic limit given by the Wilson–Cowan equation (1.3) holds. Furthermore, also the Langevin approximations should agree with the equations discussed in Sect. 2.3. However, we have not yet been able to prove such a theorem. We further conjecture that the results in this article can be used to prove the convergence for the bounded state space model by a domination argument. Heuristically, it seems clear that a bounded process should be dominated by a process that possesses the same dynamics inside the state space of the bounded process, but can stray out from that bounded domain. Hence, as the limit of the potentially larger process lies within the domain where the two processes agree also the dominated process should converge to the same limit. Mathematically, this line of argument relies on non-trivial estimates between occupation measures of high-dimensional Markov processes. This is work in progress.

### 3.3 Activity Based Neural Field Model

Finally, we return also to a difference in neural field theory mentioned in the beginning. In contrast to rate-based neural field models of the Wilson–Cowan type (1.1), there exists a second essential class of neural field models, so-called activity based

models, the prototype of which is the *Amari equation*

$$\tau \dot{v}(t, x) = -v(t, x) + \int_D w(x, y) f(v(t, y)) \, dy + I(t, x). \tag{3.7}$$

We conjecture that also for this type of equations a phenomenological microscopic model can be constructed with a suitable adaptation of the activation rates and that limit theorems analogous to the results in Sect. 2.1 hold. Then also a Langevin equation for this model can be obtained and used for further analysis.

### 4 Proofs of the Main Results

In this section, we present the proofs of the limit theorems. For the convenience of the reader, as it is important tool in the subsequent proofs, we first state the Poincaré inequality. Let  $D \subset \mathbb{R}^d$  be a convex domain, then it holds for any function  $\phi \in H^1(D)$  that

$$\|\bar{\phi}_D - \phi\|_{L^2(D)} \leq \frac{\text{diam}(D)}{\pi} \|\nabla \phi\|_{L^2(D)}, \tag{4.1}$$

where  $\bar{\phi}_D$  is the mean value of the function  $\phi$  on the domain  $D$ , i.e.,

$$\bar{\phi}_D = \frac{1}{|D|} \int_D \phi(x) \, dx. \tag{4.2}$$

Moreover, the constant in the right-hand side of (4.1) is the optimal constant depending only on the diameter of the domain  $D$ , cf. [1, 23]. Whenever we omit to denote the spatial domain for definition of norms or inner products in  $L^2(D)$  or Sobolev spaces  $H^\alpha(D)$ , then it is to be interpreted as the norm over the whole domain  $D$ . If the norm is taken only over a subset  $D_{k,n}$ , then this is always indicated unexceptionally.

For the benefit of the reader, we next repeat the limiting equation

$$\tau \dot{v}(t, x) = -v(t, x) + f\left(\int_D w(x, y) v(t, y) \, dy + I(t, x)\right). \tag{4.3}$$

We denote by  $F$  the Nemytzkii operator on  $L^2(D)$  defined by

$$F(g, t)(x) = f\left(\int_D w(x, y) g(y) \, dy + I(t, x)\right) \quad \forall g \in L^2(D), \tag{4.4}$$

and for all  $\theta \in \mathbb{N}_0^P$  we define a discrete version of the Nemytzkii operator via

$$\begin{aligned} -\frac{1}{\tau} v^n(\theta) + \frac{1}{\tau} \bar{F}^n(v^n(\theta), t) &= \lambda^n(\theta, t) \int_{\mathbb{N}_0^P} (v^n(\xi) - v^n(\theta)) \mu^n((\theta, t), d\xi) \\ &= \frac{1}{\tau} \sum_{k=1}^P \frac{1}{l(k, n)} (-\theta^k + l(k, n) \bar{f}_{k,n}(\theta, t)) \mathbb{1}_{D_{k,n}} \\ &= -\frac{1}{\tau} v^n(\theta) + \frac{1}{\tau} \sum_{k=1}^P \bar{f}_{k,n}(\theta, t) \mathbb{1}_{D_{k,n}}. \end{aligned} \tag{4.5}$$

Note that  $\tau^{-1}(\phi, v^n(\theta))_{L^2} + \tau^{-1}(\phi, \overline{F}^n(v^n(\theta), t))_{L^2}$  for  $\phi \in L^2(D)$  corresponds to the generator of  $(\Theta_t^n, t)_{t \geq 0}$  applied to the function  $(\theta, t) \mapsto (\phi, v^n(\theta))_{L^2}$ .

Finally, another useful property is that the means of the process' components are bounded. For each  $k, n$  it holds that

$$\begin{aligned} \mathbb{E}\Theta_t^{k,n} &= \mathbb{E}\Theta_0^{k,n} + \frac{1}{\tau} \int_0^t l(k, n) \mathbb{E}\overline{f}_{k,n}(Y_s^n) - \mathbb{E}\Theta_s^{k,n} \, ds \\ &\leq \mathbb{E}\Theta_0^{k,n} + \frac{1}{\tau} \int_0^t l(k, n) \|f\|_0 - \mathbb{E}\Theta_s^{k,n} \, ds, \end{aligned}$$

see also (B.1). Therefore, it holds that  $\mathbb{E}\Theta_t^{k,n} \leq m_t^{k,n}$ , where  $m_t^{k,n}$  solves the deterministic initial value problem

$$\dot{m}_t^{k,n} = -\frac{1}{\tau} m_t^{k,n} + \frac{1}{\tau} l(k, n) \|f\|_0, \quad m_0^{k,n} = \mathbb{E}\Theta_0^{k,n},$$

i.e.,

$$\begin{aligned} m_t^{k,n} &= e^{-t/\tau} (m_{k,n}^0 - l(k, n) \|f\|_0) + l(k, n) \|f\|_0 \\ &\leq l(k, n) (1 + \|f\|_0) \quad \forall t \geq 0. \end{aligned} \tag{4.6}$$

Here, we also used the assumption  $\mathbb{E}^n \Theta_0^{k,n} \leq l(k, n)$  on the initial condition.

### 4.1 Proof of Theorem 2.1 (Law of Large Numbers)

In order to prove the law of large numbers, Theorem 2.1, we apply the law of large numbers for Hilbert space valued PDMPs, see Theorem 4.1 in [27], to the sequence of homogeneous PDMPs  $(Y_t^n)_{t \geq 0} = (\Theta_t^n, t)_{t \geq 0}$ . For the application of this theorem, recall that the first, piecewise constant, vector-valued component of this process counts the number of active neurons in each sub-population and the second, deterministic component states time. The process  $(Y_t^n)_{t \geq 0}$  is the usual 'space-time process', i.e., homogeneous Markov process which is obtained via a state-space extension to obtain a homogeneous Markov process from the inhomogeneous process  $(\Theta_t^n)_{t \geq 0}$ . The continuous component satisfies the simple ODE  $\dot{i} = 1, i(0) = 0$ , and thus the full process is a PDMP. In the terminology of [27], the sequence of coordinate functions on the different state spaces of the PDMPs  $(Y_t^n)_{t \geq 0}$  into a common Hilbert space is given by the maps  $v^n$  (2.7) with the common Hilbert space  $L^2(D)$ . Thus, in order to infer convergence in probability (2.10) from Theorem 4.1 in [27], it is sufficient to validate the following conditions:

(LLN1) For fixed  $T > 0$ , it holds that

$$\lim_{n \rightarrow \infty} \mathbb{E}^n \int_0^T \lambda^n(Y_t^n) \int_{\mathbb{N}^p} \|v^n(\xi) - v^n(\Theta_t^n)\|_{L^2}^2 \mu^n(Y_t^n, d\xi) \, dt = 0. \tag{4.7}$$

(LLN2) The Nemytzkii operator  $F$  satisfies a Lipschitz condition in  $L^2(D)$  uniformly with respect to  $t, t \geq 0$ , i.e., there exists a constant  $L_0 > 0$  such

that

$$\|F(g_1, t) - F(g_2, t)\|_{L^2} \leq L_0 \|g_1 - g_2\|_{L^2} \quad \forall t \geq 0, g_1, g_2 \in L^2(D). \tag{4.8}$$

(LLN3) For fixed  $T > 0$ , it holds that

$$\lim_{n \rightarrow \infty} \mathbb{E}^n \int_0^T \|\overline{F}^n(v_t^n, t) - F(v_t^n, t)\|_{L^2} dt = 0. \tag{4.9}$$

Note that the final condition of Theorem 4.1 in [27], i.e., the convergence of the initial conditions, is satisfied by assumption. For a discussion of these conditions, we refer to [27] and proceed to their derivation for the present model in the subsequent parts (a) to (c).

(a) In order to prove condition (4.7), we write the integral with respect to the discrete probability measure  $\mu^n$  as a sum. This yields

$$\begin{aligned} & \mathbb{E}^n \lambda(Y_t^n) \int_{\mathbb{N}^P} \|v^n(\xi) - v^n(\Theta_t^n)\|_{L^2}^2 \mu^n(Y_t^n, d\xi) \\ &= \frac{1}{\tau} \sum_{k=1}^P \mathbb{E}^n \frac{1}{l(k, n)^2} (\Theta_t^{k, n} + l(k, n) \overline{f}_{k, n}(Y_t^n)) |D_{k, n}| \\ &\leq \frac{1}{\tau} \frac{1 + 2\|f\|_0}{\ell_-(n)} |D|, \end{aligned} \tag{4.10}$$

where we have used the upper bound (4.6) on the expectation  $\mathbb{E}^n \Theta_t^{k, n}$  and the assumption on the initial conditions. Next, integrating over  $[0, T]$  and employing the assumption  $\lim_{n \rightarrow \infty} \ell_-(n) = \infty$  in (2.9) establishes condition (4.7).

(b) The Lipschitz condition (4.8) of the Nemytzkii operators is a straightforward consequence of the Lipschitz continuity (1.4) of the gain function  $f$  as

$$\begin{aligned} & \|F(g_1, t) - F(g_2, t)\|_{L^2}^2 \\ &= \int_D \left| f \left( \int_D w(x, y) g_1(y) dy + I(x, t) \right) \right. \\ &\quad \left. - f \left( \int_D w(x, y) g_2(y) dy + I(x, t) \right) \right|^2 dx \\ &\leq L^2 \int_D \left| \int_D w(x, y) (g_1(y) - g_2(y)) dy \right|^2 dx \\ &\leq L^2 \int_D \|w(x, \cdot)\|_{L^2}^2 \|g_1 - g_2\|_{L^2}^2 dx \\ &= L^2 \|w\|_{L^2 \times L^2}^2 \|g_1 - g_2\|_{L^2}^2. \end{aligned}$$

Therefore, (4.8) holds with Lipschitz constant  $L_0 := L \|w\|_{L^2 \times L^2}$ .

(c) Finally, we prove the convergence of the generators (4.9). To this end, we employ the characterisation of the norm in  $L^2(D)$  by  $\|\eta\|_{L^2} = \sup_{\|\phi\|_{L^2}=1} |(\phi, \eta)_{L^2}|$  for all  $\eta \in L^2(D)$ , and thus consider first the scalar product of elements  $\phi \in L^2(D)$  with  $\|\phi\|_{L^2} = 1$  and the difference inside the norm in (4.9). On the one hand, we obtain using definition (4.5) that

$$(\phi, \bar{F}^n(v_t^n, t))_{L^2} = \left( \phi, \sum_{k=1}^P \bar{f}_{k,n}(Y_t^n) \mathbb{I}_{D_{k,n}} \right)_{L^2}. \tag{4.11}$$

Next, we apply the Nemytzkii operator  $F$  defined in (4.4) to  $v^n(t)$  and take the inner product of the result with respect to  $\phi$  to obtain on the other hand

$$(\phi, F(v_t^n, t))_{L^2} = \left( \phi, f \left( \sum_{k=1}^P \frac{\Theta_t^{k,n}}{l(k, n)} \int_{D_{k,n}} w(\cdot, y) dy + I(t, \cdot) \right) \right)_{L^2}. \tag{4.12}$$

Subtracting (4.12) from (4.11), we obtain the integrated difference

$$\begin{aligned} & (\phi, \bar{F}^n(v_t^n, t))_{L^2} - (\phi, F(v_t^n, t))_{L^2} \\ &= \int_D \phi(x) \left[ \sum_{k=1}^P \bar{f}_{k,n}(Y_t^n) \mathbb{I}_{D_{k,n}} - f \left( \sum_{j=1}^P \frac{\Theta_j^n(t)}{l(j, n)} \int_{D_{j,n}} w(x, y) dy + I(t, x) \right) \right] dx \\ &= \sum_{k=1}^P \int_{D_{k,n}} \phi(x) \left[ f \left( \sum_{j=1}^P \bar{W}_{kj}^n \frac{\Theta_t^{j,n}}{l(j, n)} + \bar{I}_{k,n}(t) \right) - f \left( \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(j, n)} \int_{D_{j,n}} w(x, y) dy + I(t, x) \right) \right] dx. \end{aligned}$$

We proceed to estimate the norm of the term in the right-hand side. We use the Lipschitz condition (1.4) on  $f$ , the triangle inequality, and finally the Cauchy–Schwarz inequality on the resulting second term to obtain the estimate

$$\begin{aligned} & |(\phi, \bar{F}^n(v_t^n, t))_{L^2} - (\phi, F(v_t^n, t))_{L^2}| \\ &\leq L \sum_{k=1}^P \int_{D_{k,n}} |\phi(x)| \\ &\quad \times \left| \sum_{j=1}^P \frac{\Theta_j^n(t)}{l(j, n)} \left( \bar{W}_{kj}^n - \int_{D_{j,n}} w(x, y) dy \right) + \bar{I}_{k,n}(t) - I(t, x) \right| dx \end{aligned}$$

$$\begin{aligned} &\leq L \underbrace{\sum_{k=1}^P \int_{D_{k,n}} |\phi(x)| \left| \sum_{j=1}^P \frac{\Theta_j^n(t)}{l(j,n)} \left( \overline{W}_{kj}^n - \int_{D_{j,n}} w(x,y) dy \right) \right|}_{(*)} dx \\ &\quad + L \underbrace{\sum_{k=1}^P \|\phi\|_{L^2(D_{k,n})} \|\overline{I}_{k,n}(t) - I(t)\|_{L^2(D_{k,n})}}_{(**)}. \end{aligned}$$

Here, the term in the right-hand side marked (\*\*) is further estimated using the Cauchy–Schwarz inequality and the Poincaré inequality (4.1), which yields

$$(**) \leq \frac{\delta_+(n)}{\pi} \left( \sum_{k=1}^P \|\nabla I(t)\|_{L^2(D_{k,n})}^2 \right)^{1/2} = \frac{\delta_+(n)}{\pi} \|\nabla_x I(t)\|_{L^2}. \tag{4.13}$$

We now consider the term marked (\*). Inserting the definition of  $\overline{W}_{kj}^n$  given in (2.4), the reordering of the summations and changing the order of integration yields

$$\begin{aligned} (*) &= \sum_{k=1}^P \int_{D_{k,n}} |\phi(x)| \\ &\quad \times \left| \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \left( \int_{D_{j,n}} \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} w(z,y) dz \right) - w(x,y) dy \right) \right| dx \\ &\leq \sum_{k=1}^P \int_{D_{k,n}} |\phi(x)| \\ &\quad \times \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \int_{D_{j,n}} \left| \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} w(z,y) dz \right) - w(x,y) \right| dy dx \\ &= \sum_{k=1}^P \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \\ &\quad \times \int_{D_{j,n}} \left[ \int_{D_{k,n}} |\phi(x)| \left| \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} w(z,y) dz \right) - w(x,y) \right| dx \right] dy. \end{aligned}$$

We next apply the Cauchy–Schwarz inequality to the integral inside the square brackets in the last term. Thus, we obtain the estimate

$$\begin{aligned} (*) &\leq \sum_{j=1}^P \int_{D_{j,n}} \sum_{k=1}^P \|\phi\|_{L^2(D_{k,n})} \\ &\quad \times \left[ \int_{D_{k,n}} \left| \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} w(z,y) dz \right) - w(x,y) \right|^2 dx \right]^{1/2} dy. \end{aligned}$$

Now the Poincaré inequality (4.1) is applied to the innermost integral inside the square brackets, which yields

$$(*) \leq \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \int_{D_{j,n}} \sum_{k=1}^P \|\phi\|_{L^2(D_{k,n})} \frac{\text{diam}(D_{k,n})}{\pi} \|\nabla_x w(\cdot, y)\|_{L^2(D_{k,n})} dy.$$

Finally, using once more the Cauchy–Schwarz inequality on the innermost summation we obtain

$$(*) \leq \frac{\delta_+(n)}{\pi} \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \int_{D_{j,n}} \|\nabla_x w(\cdot, y)\|_{L^2} dy. \tag{4.14}$$

Now, a combination of the estimates (4.13) and (4.14) on the terms (\*) and (\*\*) yields

$$\begin{aligned} & |(\phi, \bar{F}^n(v_t^n, t))_{L^2} - (\phi, F(v^n(t), t))_{L^2}| \\ & \leq \delta_+(n) \frac{L}{\pi} \left( \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \int_{D_{j,n}} \|\nabla_x w(\cdot, y)\|_{L^2} dy + \|\nabla_x I(t)\|_{L^2} \right). \end{aligned}$$

Here, the right-hand side is independent of  $\phi$ , hence taking the supremum over all  $\phi$  with  $\|\phi\|_{L^2} = 1$  yields

$$\begin{aligned} & \|\bar{F}^n(v_t^n, t) - F(v^n(t), t)\|_{L^2} \\ & \leq \delta_+(n) \frac{L}{\pi} \left( \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \int_{D_{j,n}} \|\nabla_x w(\cdot, y)\|_{L^2} dy + \|\nabla_x I(t)\|_{L^2} \right). \end{aligned}$$

Finally, integrating over  $(0, T)$  and taking the expectation on both sides results in

$$\begin{aligned} & \mathbb{E}^n \int_0^T \|\bar{F}^n(v_t^n, t) - F(v^n(t), t)\|_{L^2} dt \\ & \leq \delta_+(n) \frac{L}{\pi} (\sqrt{|D|}T(1 + \|f\|_0) \|\nabla_x w\|_{L^2 \times L^2} + \|\nabla_x I\|_{L^1((0,T),L^2)}). \end{aligned} \tag{4.15}$$

Here, we have used (4.6) and a combination of the Cauchy–Schwarz and Poincaré inequality (4.1) in order to estimate

$$\mathbb{E}^n \sum_{j=1}^P \frac{\Theta_t^{j,n}}{l(k,n)} \int_{D_{j,n}} \|\nabla_x w(\cdot, y)\|_{L^2} dy \leq \delta_+(n) \frac{\sqrt{|D|}(1 + \|f\|_0)}{\pi} \|\nabla_x w\|_{L^2 \times L^2}.$$

The upper bound in (4.15) is of order  $\mathcal{O}(\delta_+(n))$  and, therefore, converges to zero for  $n \rightarrow \infty$  due to assumption (2.9). Hence, condition (4.9) is satisfied. The proof of the convergence in probability (2.10) is completed.

It is now easy to extend this result to the convergence in the  $r$ th mean. First of all, the convergence in probability (2.10) implies for all  $r \geq 1$  the convergence in probability of the random variables  $\sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2}^r$  to zero. As convergence in the mean of real valued random variables is equivalent to convergence in probability and uniform integrability it remains to prove the latter for the families  $\sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2}^r, n \in \mathbb{N}$ .

We first consider the case  $r = 1$ , and establish a uniform bound on the second moments  $\mathbb{E}^n \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2}^2$ . Then the de la Vallée–Poussin theorem, cf. App., Proposition 2.2 in [15], implies that the random variables  $\sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2}, n \in \mathbb{N}$ , are uniformly integrable.

Without loss of generality, we can assume that there exist<sup>5</sup> Poisson processes  $(N_t^{k,n})_{t \geq 0}$  with rates  $\Lambda_{k,n} = l(k, n)(1 + \|f\|_0)/\tau$ , which dominate  $(\Theta_t^{k,n} - \Theta_0^{k,n})_{t \geq 0}$  pathwise. Then we obtain almost surely

$$\|v_t^n\|_{L^2}^2 \leq 2\|v_0^n\|_{L^2}^2 + 2 \sum_{k=1}^P \frac{(\Theta_t^{k,n} - \Theta_0^{k,n})^2}{l(k, n)^2} |D_{k,n}| \leq 2\|v_0^n\|_{L^2}^2 + 2 \sum_{k=1}^P \frac{(N_T^{k,n})^2}{l(k, n)^2} |D_{k,n}|.$$

Here, the right-hand side is independent of  $t \leq T$ , and thus we obtain

$$\mathbb{E}^n \sup_{t \in [0, T]} \|v_t^n\|_{L^2}^2 \leq 2\mathbb{E}^n \|v_0^n\|_{L^2}^2 + 2 \sum_{k=1}^P \frac{\mathbb{E}^n (N_T^{k,n})^2}{l(k, n)^2} |D_{k,n}| \leq 2\mathbb{E}^n \|v_0^n\|_{L^2}^2 + C_T,$$

where we have used that  $N_T^{k,n}$  is Poisson distributed with rate  $T\Lambda_{k,n}$ , and thus  $\mathbb{E}^n (N_T^{k,n})^2 = T\Lambda_{k,n} + T^2\Lambda_{k,n}^2$ . Here,  $C_T$  is some finite constant which depends on  $T$  and the overall parameters of the model, i.e.,  $\tau, f, D$ , but is independent of  $k$  and  $n$ . Using this upper bound, the triangle inequality yields the estimate

$$\mathbb{E}^n \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2}^2 \leq 2C_T (\mathbb{E}^n \|v_0^n\|_{L^2}^2 + \|v\|_{C([0, T], L^2)}^2 + 1).$$

Therefore, using the assumption  $\sup_{n \in \mathbb{N}} \mathbb{E}^n \|v_0^n\|_{L^2}^2 < \infty$  it holds that

$$\sup_{n \in \mathbb{N}} \mathbb{E}^n \sup_{t \in [0, T]} \|v_t^n - v(t)\|_{L^2}^2 < \infty.$$

The general case for  $r > 1$  works analogously. Note that the  $r$ th moment of the Poisson distribution is proportional to the  $r$ th power of its rate. Hence, just as in the case of  $r = 1$ , the term

$$\mathbb{E}^n \left( \sum_{k=1}^P \frac{(N_T^{k,n})^2}{l(k, n)^2} |D_{k,n}| \right)^r$$

<sup>5</sup>The Poisson process jumps at a faster rate than the components of the Markov chain regardless of the time and the state these are in. Furthermore, all jumps are upward. Hence, using a coupling argument as discussed in the proof of Theorem 4.3.5 in [16], we find that there exists a probability space supporting two processes with distributions equivalent to the Poisson process and the Markov chain component such that the Poisson process dominates the second process for all paths. Clearly, all moments dominate and these inequalities are then valid for any probability spaces supporting these processes.



can thus be bounded from above by some constant  $C_T$  independent of  $k$  and  $n$ . The proof of Theorem 2.1 is completed.

#### 4.2 Proof of Corollary 2.1 (Corollary to the Law of Large Numbers)

For  $\alpha = 0$ , the statement of the corollary coincides with the statement of Theorem 2.1, hence we consider  $\alpha > 0$ . As in the proof of Theorem 2.1, we apply Theorem 4.1 in [27] to the PDMPs  $(Y_t^n)_{t \geq 0}$ , however, this time for the functions  $v^n$  understood as taking values in the Hilbert space  $H^{-\alpha}(D)$  instead of  $L^2(D)$ . Thus, we have to validate again conditions (LLN1)–(LLN3) wherein the norm in  $L^2(D)$  is always replaced by the norm in  $H^{-\alpha}(D)$ . The essential argument is sharpening the estimates in part (a) of the proof of Theorem 2.1 using optimal Sobolev embedding theorems such that the conditions (2.13) imply (LLN1). This we present in part (a) of the proof below. The Lipschitz condition (LLN2) of the Nemytzkii operator  $F$  in the spaces  $H^{-\alpha}$  is established in part (b). Finally, as the condition  $\delta_+(n) \rightarrow \infty$  remains as in Theorem 2.1, the condition (LLN3) follows immediately from the proof of Theorem 2.1 due to the continuous embedding of  $L^2(D)$  into  $H^{-\alpha}(D)$ .

(a) In the case  $\alpha = 0$ , i.e.,  $H^\alpha(D) = L^2(D)$ , we used in (4.10) that  $\|\mathbb{I}_{D_{k,n}}\|_{L^2}^2 = |D_{k,n}|$ . For general  $\alpha > 0$ , we use the representation

$$\|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}} = \sup_{\|\phi\|_{H^\alpha}} |(\phi, \mathbb{I}_{D_{k,n}})_{L^2}|.$$

In order to estimate the terms inside the supremum in the right-hand side, we use Hölder’s inequality and the Sobolev embedding theorem, i.e.,  $H^\alpha(D) \hookrightarrow L^\infty(D)$  for  $\alpha > d/2$  and  $H^\alpha(D) \hookrightarrow L^r(D)$  with  $r = d/(d/2 - \alpha)$  for  $0 < \alpha < d/2$ , see Theorem 7.34 and Corollary 7.17 in [2]. Thus, we obtain

$$\|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}} \leq \begin{cases} K_{d/(d/2-\alpha)} \|\mathbb{I}_{D_{k,n}}\|_{L^{2d/(d+2\alpha)}} & \text{if } 0 < \alpha < d/2, \\ K_\infty \|\mathbb{I}_{D_{k,n}}\|_{L^1} & \text{if } d/2 < \alpha, \end{cases}$$

where the constants  $K$  are the constants arising from the continuous embeddings of the Sobolev spaces into the Lebesgue spaces. Evaluating the norms in the right-hand side, and further estimating using the maximal Lebesgue measure of the elements of the partition yields

$$\|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}}^2 \leq \begin{cases} K_{d/(d/2-\alpha)}^2 |D_{k,n}| v_+(n)^{2\alpha/d} & \text{if } 0 < \alpha < d/2, \\ K_\infty^2 |D_{k,n}| v_+(n) & \text{if } d/2 < \alpha. \end{cases}$$

Note that the upper bounds are consistent with the condition in Theorem 2.1 for  $\alpha = 0$ . Finally, as  $H^{d/2}(D) \hookrightarrow H^{(d/2-\epsilon)}(D)$  for all small  $\epsilon$ , the result for  $\alpha = d/2$  follows from the result above as

$$\|\mathbb{I}_{D_{k,n}}\|_{H^{-d/2}} \leq \sup_{\|\phi\|_{H^{2\alpha}}} (\|\phi\|_{L^{d/\epsilon}} \|\mathbb{I}_{D_{k,n}}\|_{L^{d/(d-\epsilon)}}) \leq C \|\mathbb{I}_{D_{k,n}}\|_{L^{d/(d-\epsilon)}}$$

where  $C$  is the constant resulting from the continuous embedding of  $H^{d/2}(D)$  into  $H^{d/2-\epsilon}(D)$ . Thus, we obtain for all  $\epsilon > 0$  the estimate

$$\|\mathbb{I}_{D_{k,n}}\|_{H^{-d/2}}^2 \leq C^2 |D_{k,n}| v_+(n)^{(d-2\epsilon)/d}.$$

- (b) Next, we have to establish that the Nemytzkii operator  $F$  on  $L^2(D)$  is also Lipschitz continuous with respect to the norms  $\|\cdot\|_{H^{-\alpha}}$ ,  $\alpha \geq 0$ , i.e., for all  $\alpha \geq 0$  there exists a constant  $L_{-\alpha}$  such that

$$\|F(g_1, t) - F(g_2, t)\|_{H^{-\alpha}} \leq L_{-\alpha} \|g_1 - g_2\|_{H^{-\alpha}} \quad \forall t \geq 0, g_1, g_2 \in L^2(D). \tag{4.16}$$

We obtain due to the Lipschitz continuity of  $f$ , which implies absolute continuity of  $f$ , that

$$\left| \int_D \phi(x) (F(g_1, t)(x) - F(g_2, t)(x)) \, dx \right| = \left| \int_D \phi(x) \int_{z_1(t,x)}^{z_2(t,x)} f'(z) \, dz \, dx \right|,$$

where

$$z_1(t, x) = \int_D w(x, y) g_1(y) \, dy + I(t, x),$$

$$z_2(t, x) = \int_D w(x, y) g_2(y) \, dy + I(t, x).$$

Applying Hölder’s inequality and the essential boundedness of the derivative  $f'$ , we obtain the estimate

$$\begin{aligned} & \left| \int_D \phi(x) (F(g_1, t)(x) - F(g_2, t)(x)) \, dx \right| \\ & \leq \|\phi\|_{L^p} \left( \int_D \left| \int_{z_1(t,x)}^{z_2(t,x)} f'(z) \, dz \right|^q \, dx \right)^{1/q} \\ & \leq \|\phi\|_{L^p} \left( \int_D \|f'\|_{L^\infty} |z_1(t, x) - z_2(t, x)|^q \, dx \right)^{1/q} \\ & = \|\phi\|_{L^p} \|f'\|_{L^\infty} \left( \int_D \left| \int_D w(x, y) (g_1(y) - g_2(y)) \, dy \right|^q \, dx \right)^{1/q}. \end{aligned}$$

Next, as by assumption  $w(x, \cdot) \in H^\alpha(D)$ , we obtain

$$\begin{aligned} & \left( \int_D \left| \int_D w(x, y) (g_1(y) - g_2(y)) \, dy \right|^q \, dx \right)^{1/q} \\ & = \left( \int_D \|w(x, \cdot)\|_{H^\alpha}^q |(w(x, \cdot) / \|w(x, \cdot)\|_{H^\alpha}, g_1 - g_2)_{H^\alpha}|^q \, dx \right)^{1/q} \\ & \leq \|w\|_{L^q \times H^\alpha} \|g_1 - g_2\|_{H^{-\alpha}}. \end{aligned}$$

Overall, this yields the estimate

$$|\langle \phi, F(g_1, t) - F(g_2, t) \rangle_{H^{-\alpha}}| \leq \|\phi\|_{L^p} \|f'\|_{L^\infty} \|w\|_{L^q \times H^\alpha} \|g_1 - g_2\|_{H^{-\alpha}}.$$

Hence, taking the supremum on both sides of this inequality over all  $\|\phi\|_{H^\alpha} = 1$ , we obtain the Lipschitz condition (4.16) with  $L_{-\alpha} := LK_\alpha \|w\|_{L^q \times H^\alpha}$ , where  $K_\alpha$  is the constant resulting from the continuous embedding of  $H^\alpha(D)$  into  $L^p(D)$  and the Lipschitz constant  $L$  of  $f$  satisfies  $L \geq \|f'\|_{L^\infty}$ .

### 4.3 Proof of Theorem 2.2 (Infinite Time Convergence)

- (a) We first present an alternative representation for the jump processes  $(\Theta_t^n)_{t \geq 0}$  and the solution  $v$  of the Wilson–Cowan equation (1.3). Using the generator of the PDMP  $(\Theta_t^n, t)_{t \geq 0}$ , we obtain that the components  $\Theta^{k,n}$  satisfy

$$\begin{aligned} \Theta_t^{k,n} &= \Theta_0^{k,n} + \int_0^t \lambda^n(\Theta_s^n, s) \int_{\mathbb{N}^p} (\xi^k - \Theta_s^{k,n}) \mu^n(\Theta_s^n, s; d\xi) ds + M_t^{k,n} \\ &= \Theta_0^{k,n} + \int_0^t \left( -\frac{1}{\tau} \Theta_s^{k,n} + \frac{1}{\tau} l(k, n) \bar{f}_{k,n}(\Theta_s^n, s) \right) ds + M_t^{k,n}, \end{aligned} \tag{4.17}$$

where  $(M_t^{k,n})_{t \geq 0}$  is a square-integrable càdlàg martingale given by

$$M_t^{k,n} := \Theta_t^{k,n} - \Theta_0^{k,n} - \int_0^t \lambda^n(\Theta_s^n, s) \int_{\mathbb{N}^p} (\xi^k - \Theta_s^{k,n}) \mu^n(\Theta_s^n, s; d\xi) ds. \tag{4.18}$$

As the jump process is regular, this martingale is almost surely of finite variation and it could also be written in terms of a stochastic integral with respect to the associated martingale measure of the PDMP [16]. Next, interpreting  $\Theta^{k,n}$  as the solution of the stochastic evolution equation (4.17) driven by the martingale  $M^{k,n}$ , it follows from the variation of constants formula that it satisfies

$$\begin{aligned} \Theta_t^{k,n} &= e^{-t/\tau} \Theta_0^{k,n} + \frac{1}{\tau} l(k, n) \int_0^t e^{-(t-s)/\tau} \bar{f}_{k,n}(\Theta_s^n, s) ds \\ &\quad + \int_0^t e^{-(t-s)/\tau} dM_s^{k,n}. \end{aligned} \tag{4.19}$$

This formula can also be easily verified path-by-path by inserting (4.19) into (4.17) and using integration by parts. Note that here the stochastic integral with respect to the martingale is just a Riemann–Stieltjes integral as the martingale is of finite variation. For the sake of completeness, we briefly sketch the arguments. Thus, inserting (4.19) into (4.17) yields

$$\begin{aligned} \Theta_t^{k,n} &= \underbrace{\Theta_0^{k,n} - \frac{1}{\tau} \int_0^t e^{-s/\tau} \Theta_0^{k,n} ds}_{(*)} \\ &\quad - \frac{1}{\tau} l(k, n) \end{aligned}$$

$$\begin{aligned} & \times \underbrace{\left( \frac{1}{\tau} \int_0^t \int_0^s e^{-(s-r)/\tau} \bar{f}_{k,n}(\Theta_r^n, r) \, dr \, ds - \int_0^t \bar{f}_{k,n}(\Theta_s^n, s) \, ds \right)}_{(**)} \\ & - \underbrace{\frac{1}{\tau} \int_0^t \int_0^s e^{-(s-r)/\tau} \, dM_r^{k,n} \, ds + M_t^{k,n}}_{(***)}. \end{aligned}$$

Considering the three terms marked (\*)–(\*\*\*) separately, we show that this right-hand side equals (4.19). For the first term (\*) simply evaluating the integral yields

$$\Theta_0^{k,n} - \frac{1}{\tau} \int_0^t e^{-s/\tau} \Theta_0^{k,n} \, ds = \Theta_0^{k,n} - \frac{1}{\tau} (\alpha^{-1} e^{-t/\tau} - \tau) \Theta_0^{k,n} = e^{-t/\tau} \Theta_0^{k,n},$$

which gives the first term in the right-hand side of (4.19). Next, we simplify the term (\*\*) employing integration by parts to the first term in (\*\*), which yields

$$\begin{aligned} & \frac{1}{\tau} \int_0^t \int_0^s e^{-(s-r)/\tau} \bar{f}_{k,n}(\Theta_r^n, r) \, dr \, ds \\ & = \frac{1}{\tau} \int_0^t e^{(t-s)/\tau} \int_0^s e^{-(t-r)/\tau} \bar{f}_{k,n}(\Theta_r^n, r) \, dr \, ds \\ & = \frac{1}{\tau} \left( -\tau e^{(t-s)/\tau} \int_0^s e^{-(t-r)/\tau} \bar{f}_{k,n}(\Theta_r^n, r) \, dr \right) \Big|_0^t \\ & \quad - \frac{1}{\tau} \int_0^t (-\tau) e^{(t-s)\tau} e^{-(t-s)/\tau} \bar{f}_{k,n}(\Theta_s^n, s) \, ds \\ & = - \int_0^t e^{-(t-s)/\tau} \bar{f}_{k,n}(\Theta_s^n, s) \, ds + \int_0^t \bar{f}_{k,n}(\Theta_s^n, s) \, ds. \end{aligned}$$

Thus, we obtain subtracting from this right-hand side the second term in (\*\*) that

$$(**) = - \int_0^t e^{-(t-s)/\tau} \bar{f}_{k,n}(\Theta_s^n, s) \, ds.$$

This term is just the second term in the right-hand side of (4.19). It remains to consider the term marked (\*\*\*). We have already stated that the stochastic integral with respect to the martingale (4.18) is defined path-by-path as a Riemann–Stieltjes integral, and thus satisfies

$$\begin{aligned} & - \frac{1}{\tau} \int_0^s e^{-(s-r)/\tau} \, dM_s^{k,n} \\ & = - \frac{1}{\tau} \sum_{\tau_j^n \leq s} e^{-(s-\tau_j^n)/\tau} (\Theta_{\tau_j^n}^{k,n} - \Theta_{\tau_{j-}^n}^{k,n}) \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{\tau} \int_0^s e^{-(s-r)/\tau} \lambda^n(\Theta_r^n, r) \\
 & \times \int_{\mathbb{N}_0^P} (\xi^k - \Theta_r^{k,n}) \mu^n((\Theta_r^n, r), d\xi) ds, \tag{4.20}
 \end{aligned}$$

where  $\tau_j^n$  denotes the  $j$ th jump time of the  $n$ th PDMP. Integrating the sum in this right-hand side over  $(0, t)$  yields

$$\begin{aligned}
 & - \frac{1}{\tau} \int_0^t \sum_{\tau_j^n \leq s} e^{-(s-\tau_j^n)/\tau} (\Theta_{\tau_j^n}^{k,n} - \Theta_{\tau_j^-}^{k,n}) ds \\
 & = \sum_{\tau_j^n \leq t} e^{-(t-\tau_j^n)/\tau} (\Theta_{\tau_j^n}^{k,n} - \Theta_{\tau_j^-}^{k,n}) - \sum_{\tau_j^n \leq t} (\Theta_{\tau_j^n}^{k,n} - \Theta_{\tau_j^-}^{k,n}) \\
 & = \sum_{\tau_j^n \leq t} e^{-(t-\tau_j^n)/\tau} (\Theta_{\tau_j^n}^{k,n} - \Theta_{\tau_j^-}^{k,n}) - (\Theta_t^{k,n} - \Theta_0^{k,n}).
 \end{aligned}$$

Next, we apply integration by parts to the integral over  $(0, t)$  of the second term above analogously to the application to term (\*\*), and obtain

$$\begin{aligned}
 & \frac{1}{\tau} \int_0^t \int_0^s e^{-\alpha(s-r)} \lambda^n(\Theta_r^n, r) \int_{\mathbb{N}_0^P} (\xi^k - \Theta_r^{k,n}) \mu^n((\Theta_r^n, r), d\xi) dr ds \\
 & = - \int_0^t e^{-(t-s)/\tau} \lambda^n(\Theta_s^n, s) \int_{\mathbb{N}_0^P} (\xi^k - \Theta_s^{k,n}) \mu^n((\Theta_s^n, s), d\xi) ds \\
 & + \int_0^t \lambda^n(\Theta_s^n, s) \int_{\mathbb{N}_0^P} (\xi^k - \Theta_s^{k,n}) \mu^n((\Theta_s^n, s), d\xi) ds.
 \end{aligned}$$

Hence, overall these considerations show that

$$(***) = \int_0^t e^{-(t-s)/\tau} dM_s^{k,n}$$

and we obtain the final, third term in the right-hand side of (4.19). This completes the proof that (4.19) solves Eq. (4.17).

Further, we obtain from the variation of constants formula for  $\Theta_t^{k,n}$  also a representation for the stochastic mean activity  $v^n$  by inserting (4.19) into its definition (2.7). This gives

$$\begin{aligned}
 v_t^n & = e^{-t/\tau} v_0^n + \frac{1}{\tau} \sum_{k=1}^P \int_0^t e^{-(t-s)/\tau} \bar{f}_{k,n}(\Theta_s^n, s) ds \mathbb{I}_{D_{k,n}} \\
 & + \sum_{k=1}^P \frac{1}{l(k,n)} \int_0^t e^{-(t-s)/\tau} dM_s^{k,n} \mathbb{I}_{D_{k,n}}
 \end{aligned}$$

$$\begin{aligned}
 &= e^{-t/\tau} v_0^n + \frac{1}{\tau} \int_0^t e^{-(t-s)/\tau} \overline{F}^n(v_s^n, s) \, ds \\
 &\quad + \sum_{k=1}^P \frac{1}{l(k, n)} \int_0^t e^{-(t-s)/\tau} \, dM_s^{k,n} \mathbb{I}_{D_{k,n}}. \tag{4.21}
 \end{aligned}$$

Finally, in order to compare stochastic and deterministic solutions we use that the solution of the Wilson–Cowan equation can also be given via the variation of constants formula, i.e., it holds that for all  $t \geq 0$

$$v(t) = e^{-t/\tau} v(0) + \frac{1}{\tau} \int_0^t e^{-(t-s)/\tau} F(v(s), s) \, ds. \tag{4.22}$$

Thus, subtracting (4.22) from (4.21), and taking the expectation of the norm in  $H^{-\alpha}(D)$  yields the estimate

$$\begin{aligned}
 &\mathbb{E}^n \left\| v(t) - v_t^n \right\|_{H^{-\alpha}} \\
 &= e^{-t/\tau} \mathbb{E}^n \left\| v(0) - v_0^n \right\|_{H^{-\alpha}} \\
 &\quad + \frac{1}{\tau} \int_0^t e^{-(t-s)/\tau} \mathbb{E}^n \left\| F(v(s), s) - \overline{F}^n(v_s^n, s) \right\|_{H^{-\alpha}} \\
 &\quad + \mathbb{E}^n \left\| \sum_{k=1}^P \frac{1}{l(k, n)} \int_0^t e^{-(t-s)/\tau} \, dM_s^{k,n} \mathbb{I}_{D_{k,n}} \right\|_{H^{-\alpha}}. \tag{4.23}
 \end{aligned}$$

We deal with the terms in the right-hand side of (4.23) separately in the following such that we can apply Gronwall’s inequality. Note that the term containing the initial condition vanishes due to the assumptions of the theorem. We start with the stochastic integrals in the subsequent part (b) of the proof.

- (b) As due to Jensen’s inequality  $\mathbb{E}|Y| \leq \sqrt{\mathbb{E}|Y|^2}$ , it makes sense to calculate the second moment of the stochastic integral in the right-hand side. For the norm in  $H^{-\alpha}(D)$ , we use  $\|\phi\|_{H^{-\alpha}}^2 = (\phi, \phi)_{H^{-\alpha}}$ , and thus obtain using the linearity of the inner product

$$\begin{aligned}
 &\left\| \sum_{k=1}^P \frac{1}{l(k, n)} \underbrace{\int_0^t e^{-(t-s)/\tau} \, dM_s^{k,n} \mathbb{I}_{D_{k,n}}}_{=: \beta_t^{k,n}} \right\|_{H^{-\alpha}}^2 \\
 &= \sum_{k=1}^P \frac{|\beta_{k,n}|^2}{l(k, n)^2} \|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}}^2 + \sum_{\substack{k,j=1 \\ k \neq j}}^P \frac{\beta_{k,n} \beta_{j,n}}{l(k, n)l(j, n)} (\mathbb{I}_{D_{k,n}}, \mathbb{I}_{D_{j,n}})_{H^{-\alpha}}.
 \end{aligned}$$

We next consider the individual expectations of the random terms  $|\beta_{k,n}|^2$  and  $\beta_{k,n} \beta_{j,n}$  in the right-hand side. We have already stated that the stochastic inte-

gral with respect to the martingale (4.18) is defined path-by-path as a Riemann–Stieltjes integral (see (4.20)) and, moreover, (4.20) implies that the stochastic convolution integral can be written as a stochastic integral with respect to the fundamental martingale measure  $M^n$  associated with the PDMP  $(\Theta_t^n, t)_{t \geq 0}$  (see [16]), i.e.,

$$\int_0^t e^{-(t-s)/\tau} dM_s^{k,n} = \int_{[0,t] \times \mathbb{N}_0^p} e^{-(t-s)/\tau} (\xi^k - \Theta_{s-}^{k,n}) M^n(ds, d\xi)$$

with predictable integrand

$$(\xi, s, \omega) \mapsto e^{-(t-s)/\tau} (\xi^k - \Theta_{s-}^{k,n}(\omega)).$$

Then we obtain due to the Itô-isometry following from Proposition 4.6.2 in [16] using (4.6) that

$$\begin{aligned} & \mathbb{E}^n \left| \int_0^t e^{-(t-s)/\tau} dM_s^{k,n} \right|^2 \\ &= \mathbb{E}^n \int_0^t \lambda^n(Y_s^n) \int_{\mathbb{N}_0^p} e^{-2(t-s)/\tau} (\xi^k - \Theta_s^{k,n})^2 \mu^n(Y_s^n, d\xi) ds \\ &\leq \mathbb{E}^n \int_0^t e^{-2(t-s)/\tau} \left( \frac{1}{\tau} \Theta_s^{k,n} + \frac{1}{\tau} l(k, n) \bar{f}_{k,n}(Y_s^n) \right) ds \\ &\leq \frac{1}{\tau} (\mathbb{E}^n \Theta_0^{k,n} + 2l(k, n) \|f\|_0) \int_0^t e^{-2(t-s)/\tau} ds. \end{aligned}$$

It remains to consider the product  $\beta_t^{k,n} \beta_t^{j,n}$  for which we obtain due to the integration by parts formula

$$\beta_t^{k,n} \beta_t^{j,n} = \int_0^t \beta_{s-}^{k,n} d\beta_s^{j,n} + \int_0^t \beta_{s-}^{j,n} d\beta_s^{k,n} + [\beta^{k,n}, \beta^{j,n}]_t, \tag{4.24}$$

where the square brackets denote the quadratic variation process. The expectation of each of the terms in the right-hand side vanishes: The first two are stochastic integrals with respect to martingales, hence martingales themselves which are identical to zero at the origin. Furthermore, as both martingales are càdlàg with paths of finite variation on compacts, hence quadratic pure jump martingales, we obtain for the quadratic variation process

$$[\beta^{k,n}, \beta^{j,n}]_t = \sum_{s \leq t} (\beta_s^{k,n} - \beta_{s-}^{k,n})(\beta_s^{j,n} - \beta_{s-}^{j,n}).$$

However, as all jump times of the two martingales a.s. differ it follows that  $[\beta^{k,n}, \beta^{j,n}]_t = 0$ .

Thus, overall we have established that

$$\begin{aligned} & \mathbb{E}^n \left\| \sum_{k=1}^P \frac{1}{l(k, n)} \int_0^t e^{-(t-s)/\tau} dM_s^{k,n} \mathbb{I}_{D_{k,n}} \right\|_{H^{-\alpha}}^2 \\ &= \frac{1}{2} \sum_{k=1}^P \frac{1 + 2\|f\|_0}{l(k, n)} \|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}}^2, \end{aligned} \tag{4.25}$$

where  $1/2$  is an upper bound for  $\frac{1}{\tau} \int_0^t e^{-2(t-s)/\tau} ds$  independent of  $t$ . Estimating the norm  $\|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}}^2$  just as in the proof of Corollary 2.1, we finally obtain that

$$\mathbb{E}^n \left\| \sum_{k=1}^P \frac{1}{l(k, n)} \int_0^t e^{-(t-s)/\tau} dM_s^{k,n} \mathbb{I}_{D_{k,n}} \right\|_{H^{-\alpha}} \leq \frac{1}{4} \left( (1 + 2\|f\|_0) |D| \frac{v_+(n)^r}{\ell_-(n)} \right)^{1/2},$$

with  $r = 2\alpha/d$  for  $0 \leq \alpha \leq d/2$ ,  $r = 1 - \epsilon$  for  $\alpha = d/2$  and  $r = 1$  for  $\alpha > d/2$ .

(c) We next estimate the term

$$\int_0^t e^{-(t-s)/\tau} \mathbb{E}^n \|F(v(s), s) - \bar{F}^n(v_s^n, s)\|_{H^{-\alpha}}$$

in (4.23). From part (b) of the proof of Theorem 2.1 in Sect. 4.1, it follows that

$$\begin{aligned} & \mathbb{E}^n \|F(v_t^n, t) - \bar{F}^n(v_t^n, t)\|_{H^{-\alpha}} \\ & \leq \delta_+(n) \frac{K_{-\alpha} L}{\pi} (\sqrt{|D|} (1 + \|f\|_0) \|\nabla_x w\|_{L^2 \times L^2} + \|\nabla_x I(t)\|_{L^2}), \end{aligned} \tag{4.26}$$

where  $F$  is the Nemytzkii operator defined in (4.4) and  $K_{-\alpha}$  is a constant resulting from the continuous embedding of  $L^2(D)$  into  $H^{-\alpha}(D)$ . Here, the right-hand side can be further estimated independently of  $t \geq 0$  using the assumption that  $\|\nabla_x I(t)\|_{L^2}$  is uniformly bounded in  $t \geq 0$ . Furthermore, we have shown in Sect. 4.2 in the proof of Corollary 2.1, that under the appropriate assumptions the Nemytzkii operator  $F$  is Lipschitz continuous on  $H^{-\alpha}(D)$ ,  $\alpha \geq 0$ , with Lipschitz constant  $L_{-\alpha} > 0$  independent of  $t \geq 0$ , i.e.,

$$\|F(g_1, t) - F(g_2, t)\|_{H^{-\alpha}} \leq L_{-\alpha} \|g_1 - g_2\|_{H^{-\alpha}} \quad \forall g_1, g_2 \in L^2(D). \tag{4.27}$$

A combination of the triangle inequality and the estimates (4.26) and (4.27) yields

$$\begin{aligned} & \int_0^t e^{-(t-s)/\tau} \mathbb{E}^n \|F(v(s), s) - \bar{F}^n(v_s^n, s)\|_{H^{-\alpha}} \\ & \leq L_{-\alpha} \int_0^t e^{-(t-s/\tau)} \mathbb{E}^n \|v(s) - v_s^n\|_{H^{-\alpha}} ds + \mathcal{O}(\delta_+(n)). \end{aligned}$$



Overall, it thus follows from (4.23) that

$$\begin{aligned} \mathbb{E}^n \left\| v(t) - v_t^n \right\|_{H^{-\alpha}} &\leq \mathbb{E}^n \left\| v(0) - v_0^n \right\|_{H^{-\alpha}} \\ &\quad + \frac{L-\alpha}{\tau} \int_0^t e^{-(t-s)/\tau} \mathbb{E}^n \left\| v(s) - v_s^n \right\|_{H^{-\alpha}} \, ds \\ &\quad + \mathcal{O} \left( \delta_+(n) + \sqrt{\frac{v_+(n)^r}{\ell_-(n)}} \right). \end{aligned}$$

Then an application of Gronwall’s inequality yields

$$\begin{aligned} \mathbb{E}^n \left\| v(t) - v_t^n \right\|_{H^{-\alpha}} &\leq \left( \mathbb{E}^n \left\| v(0) - v_0^n \right\|_{H^{-\alpha}} + \mathcal{O} \left( \delta_+(n) + \sqrt{\frac{v_+(n)^r}{\ell_-(n)}} \right) \right) \\ &\quad \times \exp \left( \frac{L-\alpha}{\tau} \int_0^t e^{-(t-s)/\tau} \, ds \right) \\ &\leq \left( \mathbb{E}^n \left\| v(0) - v_0^n \right\|_{H^{-\alpha}} + \mathcal{O} \left( \delta_+(n) + \sqrt{\frac{v_+(n)^r}{\ell_-(n)}} \right) \right) e^{L-\alpha}. \end{aligned}$$

By assumptions of the theorem, the term in the right-hand side converges to zero for  $n \rightarrow \infty$ . As this convergence is uniform in  $t$ , it holds that

$$\lim_{n \rightarrow \infty} \sup_{t \geq 0} \mathbb{E}^n \left\| v(t) - v_t^n \right\|_{H^{-\alpha}} = 0. \tag{4.28}$$

#### 4.4 Proof of Theorem 2.3 (Martingale Central Limit Theorem)

In order to prove the martingale central limit theorem, we employ the general martingale central limit theorem (Theorem 5.1 in [27]) for the Hilbert space  $H^{-\alpha}(D)$ , i.e., the dual of the Sobolev space  $H^\alpha(D)$ , for  $\alpha > d$ . To apply this theorem, it suffices to prove the following conditions. Subsequently we use  $\rho_n = \sqrt{\ell_-(n)/v_+(n)}$  to denote the re-scaling sequence and use the notation

$$\langle G^n(t)\phi, \phi \rangle_{H^\alpha} = \lambda(Y_t^n) \int_{\mathbb{N}^p} \langle v^n(\xi) - v^n(\Theta_t^n), \phi \rangle_{H^\alpha}^2 \mu^n(Y_t^n, d\xi) \tag{4.29}$$

which corresponds to the quadratic variation of the martingales  $(M_t^n)_{t \geq 0}$ , see [27] for a discussion.

(CLT1) For all  $t > 0$ , it holds that

$$\begin{aligned} \sup_{n \in \mathbb{N}} \rho_n \mathbb{E}^n \int_0^t \left[ \lambda^n(Y_s^n) \int_{\mathbb{N}^p} \left\| v^n(\xi) - v^n(\Theta_s^n) \right\|_{H^{-\alpha}}^2 \mu^n(Y_s^n, d\xi) \, ds \right] \\ < \infty, \end{aligned} \tag{4.30}$$

and there exists an orthonormal basis  $(\varphi_j)_{j \in \mathbb{N}}$  of  $H^\alpha(D)$  such that for all  $j \in \mathbb{N}$  and  $t \geq 0$

$$\rho_n \mathbb{E}^n \langle G^n(Y_t^n) \varphi_j, \varphi_j \rangle_{H^\alpha} \leq \gamma_j C, \tag{4.31}$$

where the constants  $\gamma_j > 0$  are independent of  $n$  and  $t$ , satisfy  $\sum_{j \geq 1} \gamma_j < \infty$ , and the constant  $C > 0$  is independent of  $n$  and  $k$  but may depend on  $t$ .

(CLT2) The jump heights of the re-scaled martingales are almost surely uniformly bounded, i.e., there exists a constant  $\beta < \infty$  such that it holds almost surely for all  $n \in \mathbb{N}$  that

$$\sup_{t \geq 0} \sqrt{\rho_n} \|v^n(\Theta_t^n) - v^n(\Theta_{t-}^n)\|_{H^{-\alpha}} < \beta. \tag{4.32}$$

Further, for all  $\phi \in H^\alpha(D)$  and all  $t > 0$  it holds that

$$\lim_{n \rightarrow \infty} \int_0^t \mathbb{E}^n |\langle G(v(s))\phi, \phi \rangle_{H^\alpha} - \rho_n \langle G^n(Y_s^n)\phi, \phi \rangle_{H^\alpha}| ds = 0. \tag{4.33}$$

On a technical level, we note that the condition (CLT1) guarantees tightness of the sequence of re-scaled martingales  $(\sqrt{\rho_n} M_t^n)_{t \geq 0}$  in the Skorokhod space of càdlàg functions in  $H^{-\alpha}(D)$ . This property is equivalent to relative compactness in the topology of weak convergence of measures, and thus implies the existence of a convergent subsequence. The conditions (CLT2) are then sufficient to establish that any limit possesses the form of a diffusion process defined by the covariance operator  $C$  given in (2.18). In particular, condition (4.33) precisely gives the convergence of the quadratic variations and is thus the central condition. In the subsequent two parts of the proof, we show that they are satisfied: In part (a), we prove conditions (4.30) and (4.31), and part (b) establishes (4.32) and (4.33).

(a) We first prove conditions (4.30) and (4.31). Here, we also observe the significance of the choice of the norm in  $H^{-\alpha}(D)$  with  $\alpha > d$  for establishing the convergence, which is essentially that it guarantees the existence a Sobolev space  $H^{\alpha_1}(D)$  with continuous embeddings  $H^\alpha(D) \hookrightarrow H^{\alpha_1}(D) \hookrightarrow C(\bar{D})$ , where the first is of Hilbert–Schmidt type. For subsequent use, we recall the estimates

$$\|\mathbb{I}_{D_{k,n}}\|_{H^{-\alpha}}^2 \leq K_\alpha^2 |D_{k,n}|^2$$

with a suitable constant  $K_\alpha > 0$ , which we have already established in the proof of Corollary 2.1 due to the Hölder inequality and the Sobolev embedding theorem. Therefore, we obtain for the term inside the expectation in (4.30) the estimate

$$\begin{aligned} & \lambda^n(Y_s^n) \int_{\mathbb{N}_0^p} \|v^n(\xi) - v^n(\Theta_s^n)\|_{H^{-\alpha}}^2 \mu^n(Y_s^n, d\xi) \\ & \leq \frac{1}{\tau} K_\alpha^2 \sum_{k=1}^p \frac{|D_{k,n}|^2}{l(k,n)^2} (\Theta_s^{k,n} + l(k,n) \bar{f}_{k,n}(Y_s^n)). \end{aligned}$$

Next taking the expectation, using the bound (4.6) on  $\mathbb{E}^n \Theta_s^{k,n}$  and integrating over  $[0, t]$  we obtain the estimate

$$\begin{aligned} & \int_0^t \mathbb{E}^n \left[ \lambda^n(Y_s^n) \int_{\mathbb{N}_0^P} \|v^n(\xi) - v^n(\Theta_s^n)\|_{H^{-\alpha}}^2 \mu^n(Y_s^n, d\xi) \right] ds \\ & \leq \frac{t}{\tau} K_\alpha^2 (1 + 2\|f\|_0) \frac{v_+(n)}{\ell_-(n)}. \end{aligned}$$

Multiplying both sides with  $\rho_n = \ell_-(n)/v_+(n)$ , we find that condition (4.30) is satisfied.

We proceed to condition (4.31) and first of all expand the integrand to obtain

$$\begin{aligned} \langle G^n(Y_s^n) \varphi_j, \varphi_j \rangle_{H^\alpha} &= \lambda^n(Y_s^n) \int_{\mathbb{N}_0^P} \langle v^n(\xi) - v^n(\Theta_s^n), \varphi_j \rangle_{H^{-\alpha}}^2 \mu^n(Y_s^n, d\xi) \\ &= \frac{1}{\tau} \sum_{k=1}^P \frac{1}{l(k, n)^2} (\Theta_s^{k,n} + l(k, n) \bar{f}_{k,n}(Y_s^n)) \langle \mathbb{I}_{D_{k,n}}, \varphi_j \rangle_{H^{\alpha_1}}^2. \end{aligned}$$

We next estimate the term  $\langle \mathbb{I}_{D_{k,n}}, \varphi \rangle_{H^{\alpha_1}}^2$ . Here, we use the fact that for a function in  $L^2(D)$  its application as an element of the dual  $H^{-\alpha}(D)$  as well as  $H^{-\alpha_1}(D)$  for any  $\alpha_1$  with  $0 \leq \alpha_1 < \alpha$  coincide. We choose  $\alpha_1$  such that  $d/2 < \alpha_1 < \alpha - d/2$  and obtain

$$\langle \mathbb{I}_{D_{k,n}}, \varphi \rangle_{H^{\alpha_1}}^2 \leq \| \mathbb{I}_{D_{k,n}} \|_{H^{-\alpha_1}}^2 \| \varphi_j \|_{H^{\alpha_1}}^2 \leq K_{\alpha_1}^2 |D_{k,n}|^2 \| \varphi_j \|_{H^{\alpha_1}}^2,$$

where  $K_{\alpha_1}$  is the constant resulting from the Sobolev embedding theorem. Next, taking the expectation, estimating the expectation terms as before and multiplying by  $\rho_n$  yields

$$\rho_n \mathbb{E}^n \langle G^n(Y_t^n) \varphi_j, \varphi_j \rangle_{H^\alpha} \leq \frac{1}{\tau} K_{\alpha_1}^2 (1 + \|f\|_0) \| \varphi_j \|_{H^{\alpha_1}}^2.$$

We chose the constants in (4.31) as  $C := K_{\alpha_1}^2 (1 + \|f\|_0)/\tau$  and  $\gamma_j := \| \varphi_j \|_{H^{\alpha_1}}^2$ . Finally, as due to Maurin’s theorem the embedding of the space  $H^\alpha(D)$  into  $H^{\alpha_1}(D)$  is of Hilbert–Schmidt type, cf. footnote 3 on page 16, it holds that  $\sum_{j \geq 1} \| \varphi_j \|_{H^{\alpha_1}}^2 < \infty$ . Condition (4.31) is satisfied.

- (b) The estimates in part (a) further show that the jump sizes are almost surely uniformly bounded as

$$\sup_{t \geq 0} \sqrt{\rho_n} \| v^n(\Theta_t^n) - v^n(\Theta_{t-}^n) \|_{H^{-\alpha}} \leq K_\alpha \sqrt{\frac{v_+(n)}{\ell_-(n)}}.$$

Here, the upperbound in the right-hand side converges to zero for  $n \rightarrow \infty$ , and thus the left-hand side is bounded over all  $n \in \mathbb{N}$ . Therefore, condition (4.32) holds and we are left to prove the convergence of the quadratic variation (4.33).

For the jump process, the quadratic variation satisfies

$$\begin{aligned} \langle G^n(t)\phi, \phi \rangle_{H^\alpha} &= \lambda(Y_t^n) \int_{\mathbb{N}_0^P} \langle v^n(\xi) - v^n(\Theta_t^n), \phi \rangle_{H^\alpha}^2 \mu^n(Y_t^n, d\xi) \\ &= \frac{1}{\tau} \sum_{k=1}^P \frac{1}{l(k, n)^2} (\Theta_t^{k, n} + l(k, n) \bar{f}_{k, n}(Y_t^n)) \langle \mathbb{I}_{D_{k, n}}, \phi \rangle_{H^\alpha}^2. \end{aligned}$$

The quadratic variation of the limiting diffusion is given by

$$\begin{aligned} &\langle G(v(t), t)\phi, \phi \rangle_{H^\alpha} \\ &= \int_D \phi(x)^2 \left( \frac{1}{\tau} v(t, x) + \frac{1}{\tau} f \left( \int_D w(x, y) v(t, y) dy + I(t, x) \right) \right) dx. \end{aligned}$$

Here, the necessary estimates are split into several parts which are separately considered in the following. Afterwards, the estimates are combined to infer the convergence (4.33). In the following, we use again  $F$  as the Nemytzkii operator defined in (4.4). Hence, for the difference of the quadratic variations, we obtain the estimate

$$\begin{aligned} &\mathbb{E}^n \left| \langle G(v(t), t)\phi, \phi \rangle_{H^\alpha} - \rho_n \langle G^n(t)\phi, \phi \rangle_{H^\alpha} \right| \\ &= \frac{1}{\tau} \mathbb{E}^n \left| \int_D \phi(x)^2 v(t, x) + \phi(x)^2 F(v(t), t)(x) dx \right. \\ &\quad \left. - \sum_{k=1}^P \frac{\rho_n}{l(k, n)^2} (\Theta_t^{k, n} + l(k, n) \bar{f}_{k, n}(Y_t^n)) \langle \mathbb{I}_{D_{k, n}}, \phi \rangle_{H^\alpha}^2 \right| \\ &\leq \frac{1}{\tau} \mathbb{E}^n \left| \int_D \underbrace{\phi(x)^2 v(t, x)}_{(i)} + \underbrace{\phi(x)^2 F(v(t), t)(x)}_{(ii)} dx \right. \\ &\quad \left. - \int_D \underbrace{\phi(x)^2 v^n(\Theta_t^n)(x)}_{(i)} + \underbrace{\phi(x)^2 F(v^n(\Theta_t^n), t)(x)}_{(ii)} dx \right| \\ &\quad + \frac{1}{\tau} \mathbb{E}^n \left| \int_D \underbrace{\phi(x)^2 v^n(\Theta_t^n)(x)}_{(iii)} + \underbrace{\phi(x)^2 F(v^n(\Theta_t^n), t)(x)}_{(iv)} dx \right. \\ &\quad \left. - \sum_{k=1}^P \frac{\rho_n}{l(k, n)} \underbrace{\left( \frac{\Theta_t^{k, n}}{l(k, n)} \langle \mathbb{I}_{D_{k, n}}, \phi \rangle_{H^\alpha}^2 \right)}_{(iii)} \right. \\ &\quad \left. + \underbrace{\frac{l(k, n)}{l(k, n)} \bar{f}_{k, n}(Y_t^n) \langle \mathbb{I}_{D_{k, n}}, \phi \rangle_{H^\alpha}^2}_{(iv)} \right|. \tag{4.34} \end{aligned}$$

Using the triangle inequality once again for each of the two differences grouping the terms marked (i)–(iv) we obtain four terms, which we subsequently estimate separately. Finally, in part (v), we combine the four estimates.

- (i) The first term is the simplest to estimate. Using the Cauchy–Schwarz inequality, we obtain

$$\mathbb{E}^n \left| \int_D \phi^2(x)(v(t, x) - v_t^n(x)) \, dx \right| \leq \|\phi\|_{L^4}^2 \mathbb{E}^n \|v(t) - v_t^n\|_{L^2}. \quad (4.35)$$

- (ii) We next consider the difference arising from the terms marked (ii) and obtain using the Lipschitz condition (1.4) on  $f$  and the Cauchy–Schwarz inequality twice

$$\begin{aligned} & \mathbb{E}^n \left| \int_D \phi(x)^2 (F(v(t), t)(x) - F(v_t^n, t)(x)) \, dx \right| \\ & \leq L \mathbb{E}^n \int_D |\phi(x)|^2 \left| \int_D w(x, y)(v(t, y) - v_t^n(y)) \, dy \right| dx \\ & \leq L \mathbb{E}^n \int_D |\phi(x)|^2 \|w(x, \cdot)\|_{L^2} \|v(t) - v_t^n\|_{L^2} \, dx \\ & \leq L \|\phi\|_{L^4}^2 \|w\|_{L^2 \times L^2} \mathbb{E}^n \|v(t) - v_t^n\|_{L^2}. \end{aligned} \quad (4.36)$$

- (iii) In order to estimate the next term, we use the bound (4.6) on  $\mathbb{E}^n \Theta_t^{k,n}$ , and thus obtain

$$\begin{aligned} & \mathbb{E}^n \sum_{k=1}^P \frac{\Theta_t^{k,n}}{l(k, n)} \left| \int_{D_{k,n}} \phi(x)^2 \, dx - \frac{\rho_n}{l(k, n)} \left( \int_{D_{k,n}} \phi(x) \, dx \right)^2 \right| \\ & \leq (1 + \|f\|_0) \\ & \quad \times \sum_{k=1}^P |D_{k,n}| \left| \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \phi(x)^2 \, dx - \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \phi(x) \, dx \right)^2 \right| \\ & \quad + (1 + \|f\|_0) \sum_{k=1}^P |D_{k,n}| \left| 1 - \frac{\rho_n |D_{k,n}|^2}{l(k, n) |D_{k,n}|} \right| \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \phi(x) \, dx \right)^2 \\ & \leq (1 + \|f\|_0) \sum_{k=1}^P \int_{D_{k,n}} \left( \phi(x) - \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \phi(y) \, dy \right)^2 \, dx \\ & \quad + (1 + \|f\|_0) \left| 1 - \frac{v_-(n) \ell_-(n)}{v_+(n) \ell_+(n)} \right| \sum_{k=1}^P |D_{k,n}| \left( \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \phi(x) \, dx \right)^2. \end{aligned}$$

Then the estimate is completed applying the Poincaré inequality (4.1) to the first term, that is, estimating

$$\sum_{k=1}^P \int_{D_{k,n}} \left( \phi(x) - \frac{1}{|D_{k,n}|} \int_{D_{k,n}} \phi(y) \, dy \right)^2 \, dx \leq \frac{\text{diam}(D_{k,n})^2}{\pi^2} \|\nabla \phi\|_{L^2}^2,$$

and the observation that the second term is proportional to  $\|\bar{\phi}^n\|_{L^2}^2$ , which is the piecewise constant approximation to  $\phi$  based on the partition  $\mathcal{D}_n$ ; see (4.2). Therefore, we overall obtain an upper bound for the difference constituted by the terms (iii) in (4.34) by

$$\begin{aligned} & \mathbb{E}^n \left| \int_D \phi(x)^2 v_t^n(x) \, dx - \rho_n \sum_{k=1}^P \frac{\Theta_t^{k,n}}{l(k,n)^2} \langle \mathbb{I}_{D_{k,n}}, \phi \rangle_{H^\alpha}^2 \right| \\ & \leq \delta_+(n)^2 \frac{1 + \|f\|_0}{\pi^2} \|\phi\|_{H^1}^2 + (1 + \|f\|_0) R(n). \end{aligned} \tag{4.37}$$

In the last term,

$$R(n) := \left| 1 - \frac{v_-(n)}{v_+(n)} \frac{\ell_-(n)}{\ell_+(n)} \right| \|\bar{\phi}^n\|_{L^2}^2$$

converges to zero for  $n \rightarrow \infty$  by assumption (2.19) and as the sequence  $\|\bar{\phi}^n\|_{L^2}$  is bounded as it converges to  $\|\phi\|_{L^2}$  for  $n \rightarrow \infty$ .

(iv) Finally, we consider the difference

$$\begin{aligned} & \mathbb{E}^n \left| \int_D \phi(x)^2 F(v_t^n, t)(x) \, dx - \rho_n \sum_{k=1}^P \frac{l(k,n)}{l(k,n)^2} \bar{f}_{k,n}(Y_t^n) \langle \mathbb{I}_{D_{k,n}}, \phi \rangle_{H^\alpha}^2 \right| \\ & \leq \mathbb{E}^n \sum_{k=1}^P \left| \int_{D_{k,n}} \phi(x)^2 F(v_t^n, t)(x) \, dx - \frac{\rho_n}{l(k,n)} \bar{f}_{k,n}(Y_t^n) \langle \phi, \mathbb{I}_{D_{k,n}} \rangle_{H^\alpha}^2 \right|. \end{aligned}$$

We continue estimating the difference in each summand in the final right-hand side and obtain using the triangle inequality for the term inside the expectation

$$\begin{aligned} & \mathbb{E}^n \sum_{k=1}^P \left| \int_{D_{k,n}} \phi(x)^2 F(v_t^n, t)(x) \, dx - \frac{\rho_n}{l(k,n)} \bar{f}_{k,n}(Y_t^n) \langle \mathbb{I}_{D_{k,n}}, \phi \rangle_{H^\alpha}^2 \right| \\ & \leq \underbrace{\mathbb{E}^n \sum_{k=1}^P \left| \int_{D_{k,n}} \phi(x)^2 (F(v_t^n, t)(x) - \bar{f}_{k,n}(Y_t^n)) \, dx \right|}_{(*)} \\ & \quad + \underbrace{\mathbb{E}^n \sum_{k=1}^P \left| \bar{f}_{k,n}(Y_t^n) \left( \int_{D_{k,n}} \phi(x)^2 \, dx - \frac{\rho_n}{l(k,n)} \langle \mathbb{I}_{D_{k,n}}, \phi \rangle_{H^\alpha}^2 \right) \right|}_{(**)}. \end{aligned}$$

We start with the first term and observe that it possesses the same structure as the term estimated in part (c) of the proof of Theorem 2.1 with the only difference that here the function  $\phi$  in the integrand is squared. Therefore, we obtain the estimate, cf. (4.15),

$$(*) \leq \delta_+(n) \frac{L \|\phi\|_{L^4}^2}{\pi} (\sqrt{|D|} (1 + \|f\|_0) \|\nabla_x w\|_{L^2 \times L^2} + \|\nabla_x I(t)\|_{L^2}).$$

Next, we estimate the second term. Note that  $\bar{f}_{k,n}$  is bounded by  $\|f\|_0$ , and thus the remaining term is just as in part (iii) of the proof. Hence, we obtain the estimate, cf. (4.37),

$$(**) \leq \delta_+(n)^2 \frac{\|f\|_0}{\pi^2} \|\phi\|_{H^1}^2 + \|f\|_0 \left| 1 - \frac{v_-(n) \ell_-(n)}{v_+(n) \ell_+(n)} \right| \|\bar{\phi}^n\|_{L^2}^2.$$

Therefore, we overall obtain an upper bound for the difference generated by the terms (iv) by

$$\begin{aligned} & \mathbb{E}^n \left| \int_D \phi(x)^2 F(v_t^n, t)(x) \, dx - \rho_n \sum_{k=1}^P \frac{1}{l(k, n)} \bar{f}_{k,n}(Y_t^n) \langle \mathbb{I}_{D_{k,n}}, \phi \rangle_{H^\alpha}^2 \right| \\ & \leq \delta_+(n) \frac{L \|\phi\|_{L^4}^2}{\pi} (\sqrt{|D|} (1 + \|f\|_0) \|\nabla_x w\|_{L^2 \times L^2} + \|\nabla_x I(t)\|_{L^2}) \\ & \quad + \delta_+(n)^2 \frac{\|f\|_0}{\pi^2} \|\phi\|_{H^1}^2 + \|f\|_0 R(n), \end{aligned} \tag{4.38}$$

where the term  $R(n)$  is as in (4.37).

(v) To complete the proof, we combine the estimates (4.35)–(4.38) to obtain

$$\begin{aligned} & \mathbb{E}^n \left| \langle G(v(t), t)\phi, \phi \rangle_{H^\alpha} - \rho_n \langle G^n(t)\phi, \phi \rangle_{H^\alpha} \right| \\ & \leq (1 + L \|w\|_{L^2 \times L^2}) \|\phi\|_{L^4}^2 \mathbb{E}^n \|v(t) - v_t^n\|_{L^2} + (1 + 2\|f\|_0) R(n) \\ & \quad + \delta_+(n) \frac{L \|\phi\|_{L^4}^2}{\pi} (\sqrt{|D|} (1 + \|f\|_0) \|\nabla_x w\|_{L^2 \times L^2} + \|\nabla_x I(t)\|_{L^2}) \\ & \quad + \delta_+(n)^2 \frac{1 + 2\|f\|_0}{\pi^2} \|\phi\|_{H^1}^2. \end{aligned}$$

Integrating over  $(0, T)$ , we obtain with a suitable constant  $C_\phi > 0$  independent of  $n$  and  $T$  the estimate

$$\begin{aligned} & \int_0^T \mathbb{E}^n \left| \langle G(v(t), t)\phi, \phi \rangle_{H^\alpha} - \rho_n \langle G^n(t)\phi, \phi \rangle_{H^\alpha} \right| dt \\ & \leq C_\phi (\mathbb{E}^n \|v(t) - v_t^n\|_{L^1((0,T),L^2)} + TR(n)) \\ & \quad + T\delta(n) (1 + \|\nabla_x I(t)\|_{L^1((0,T),L^2)}) + \delta(n)^2. \end{aligned}$$

The constant  $C_\phi$  depends on the norm of  $\phi$  in the spaces  $H^1(D)$  and  $L^4(D)$  where the latter can be estimated in terms of the norm in the Sobolev space  $H^\alpha(D)$  due to the embedding  $H^\alpha(D) \hookrightarrow L^4(D)$ , i.e.,  $C_\phi$  is finite and depends only on  $\phi \in H^\alpha(D)$ . Finally, each term in the right-hand side converges to zero for  $n \rightarrow \infty$ , and hence condition (4.33) follows. The proof of Theorem 2.3 is completed.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

MR performed the analysis and wrote and rewrote versions of the manuscript. EB contributed mathematical discussions and to the writing of the manuscript. All authors read and approved the final manuscript.

**Acknowledgements** The authors thank J. Touboul for directing our attention also toward the infinite-time convergence in Theorem 2.2. We also thank two anonymous reviewers for their helpful comments.

### Appendix A: Well-Posedness of the Wilson–Cowan Equation

This section provides a concise exposition, based on classical existence theory, of the well-posedness of the Wilson–Cowan equation (1.3), and the boundedness and regularity results for its solution as referred to in Sect. 1.1. We understand Eq. (1.3) as an  $L^2(D)$ -valued integral equation, i.e.,

$$v(t) = v_0 + \frac{1}{\tau} \int_0^t (-v(s) + F(v(s), s)) \, ds, \quad t \geq 0, v_0 \in L^2(D), \tag{A.1}$$

where the integral is a Bochner integral and  $F$  is the Nemytzkii operator acting on  $L^2(D)$  defined by

$$F(g, t)(x) = f\left(\int_D w(x, y)g(y) \, dy + I(t, x)\right) \quad \forall g \in L^2(D).$$

As in Sect. 1.1, we assume that  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  is Lipschitz continuous,  $w \in L^2(D \times D)$  and  $I \in C(\mathbb{R}_+, L^2(D))$ , which implies that  $F$  is continuous in  $t$ . Furthermore, it was shown in Sect. 4.1 that under these assumptions  $F(g, t)$  is Lipschitz continuous in the argument  $g$  with Lipschitz constant independent of  $t \geq 0$ . Thus, the integrand in (A.1) is Lipschitz continuous with respect to the  $L^2(D)$ -valued argument for all  $t \geq 0$  and, moreover, uniformly continuous in  $g$  with respect to  $t$ . It follows that the integrand in (A.1), that is, the map  $(g, t) \rightarrow -g + F(g, t)$ , is jointly continuous on  $\mathbb{R}_+ \times L^2(D)$ . Then Theorem 5.1.1 in [10] implies that there exists a unique, strongly continuous, global solution to (A.1) for every initial condition  $v_0 \in L^2(D)$ . By definition, this solution is absolutely continuous, and as  $F$  is jointly continuous, the derivative of the solution is continuous and exists everywhere. Thus, we conclude that there exists a unique continuously differentiable solution, i.e.,  $v \in C^1(\mathbb{R}_+, L^2(D))$ .



Next, we recall an ‘explicit’ representation of the solution is the variation of constants formula (4.22), which we already stated in Sect. 4.3. We have that the solution of the Wilson–Cowan equation satisfies the integral equation

$$v(t) = v_0 + \int_0^t A v(s) + F(v(s), t) / \tau \, ds,$$

where  $A$  is the linear operator in  $L^2(D)$  mapping  $g$  to  $-g/\tau$ . Thus, the solution  $v$  satisfies

$$v(t) = e^{tA} v_0 + \frac{1}{\tau} \int_0^t e^{(t-s)A} F(s, v(s)) \, ds \quad \forall t \geq 0.$$

In the present setting, the application of the linear operator  $e^{tA}$  corresponds to the scalar multiplication with  $e^{-t/\tau}$  as  $A = -\frac{1}{\tau} \text{Id}_{L^2}$ , and thus

$$v(t) = e^{-t/\tau} v_0 + \frac{1}{\tau} \int_0^t e^{-(t-s)/\tau} F(s, v(s)) \, ds \quad \forall t \geq 0.$$

We next discuss the results stated in Sect. 1.1 on the higher spatial regularity of solutions to (A.1). Then a pointwise bound on  $v(t) \in L^2(D)$ , i.e., a constant  $C$  such that  $|v(t, x)| \leq C$  for almost all  $x \in D$  and all  $t \geq 0$ , are then easily obtained by an approximation argument, that is, approximating the less regular solution by solutions of higher regularity. It is possible to prove the pointwise bounds directly; see e.g., [26] for such an argumentation in a similar setting. However, it is easier and more illustrative to use available results for solutions of higher spatial regularity, which are usually arising as the deterministic solution of (A.1) one is interested in. For example, the authors in [34] argue that from an application point of view it is reasonable to consider at least continuous solutions. In particular, the authors in [34] present a detailed existence and uniqueness result for the activity based Amari mean field equation and state that an analogous result hold for the Wilson–Cowan equation (A.1) for spatial dimensions  $d \leq 3$ , which covers all physical relevant domains. Concerning the spatial regularity, they consider the space  $H^\alpha(D)$ , where  $\alpha$  is set to be the smallest integer such that  $\alpha > d/2$ . The significance of the choice of  $\alpha > d/2$  is—as so often in this study—that this implies the embedding of the space  $H^\alpha(D)$  into  $C(\overline{D})$ . Furthermore, we then even obtain that  $C([0, T], H^\alpha(D)) \subset C([0, T] \times \overline{D})$ , i.e., the solution  $v$  is jointly continuous.

Therefore, we have the subsequent theorem, which is sufficient for the set-up in this study. However, we note that existence and uniqueness of solutions of the Amari equation were considered under less strict regularity assumptions on the coefficients in [24], and we conjecture that these are also valid for the Wilson–Cowan equation.

**Theorem A.1** (Sect. 2 in [34]) *The domain  $D$  is bounded and satisfies the strong local Lipschitz property. We assume that  $w \in H^\alpha(D \times D)$ , that  $f \in C^\alpha(D)$  with all derivatives bounded, and that  $I \in C(\mathbb{R}_+, H^\alpha(D))$ . Then there exists a unique global solution  $v \in C([0, T], H^\alpha(D))$  for every  $T > 0$  and every initial condition  $v_0 \in H^\alpha(D)$  to (A.1), which depends continuously on the initial condition and is continuously differentiable. Moreover, the solution is globally bounded in  $H^\alpha(D)$  if the externally applied current  $I$  is globally bounded.*

*Remark A.1* In the work [34], the authors assume for the domain only the cone property, which is implied by the strong local Lipschitz property; see p. 84 in [2]. The latter is the necessary boundary regularity for the present study, cf. footnote 3 on page 16. Furthermore, in [34], it is also assumed that the gain function  $f$  is infinitely often differentiable with bounded derivatives, but it is surely sufficient for  $f$  being  $\alpha$ -times continuously differentiable.

Finally, it remains to show the pointwise bound  $v(t, x) \in (0, \|f\|_0)$  if the initial condition satisfies  $v_0(x) \in [0, \|f\|_0]$  proposed in Sect. 1.1. Under Theorem A.1 the solution  $v(t, x)$  to (A.1) is jointly continuous and, therefore, the Wilson–Cowan equation holds pointwise in  $x$  everywhere and for all  $t \geq 0$ . Furthermore,  $t \mapsto v(t, x)$  is continuously differentiable for every fixed  $x \in D$ , and it is immediate that the bounds are satisfied due to the fact that the derivative of the solution approaching 0 or  $\|f\|_0$  becomes positive or negative, respectively. Now, using an approximation result of smooth solutions converging to the  $L^2(D)$  solution, we obtain that even in this less regular case the pointwise bounds hold almost everywhere.

### Appendix B: Comparisons of Moment Equations

In this section, we discuss the moment equations for the  $L^2(D)$ -valued jump Markov processes  $v_t^n = v^n(\Theta_t^n)$ . These can be derived from the corresponding moment equations of the jump Markov process  $(\Theta_t^n)_{t \geq 0}$  taking values in  $\mathbb{N}^P$ . This process is analogous in structure to the usual model used in chemical reaction kinetics, cf., e.g., [21]. Thus, we can use the formulae derived in this reference to obtain, e.g., for the mean the system of differential equations

$$\frac{d}{dt} \mathbb{E}^n \Theta_t^n = -\frac{1}{\tau} \mathbb{E}^n \Theta_t^n + \frac{1}{\tau} \sum_{k=1}^P l(k, n) \mathbb{E}^n f \left( \sum_{j=1}^P \overline{W}_{kj}^n \Theta_t^{k,n} \right) e_k. \tag{B.1}$$

Furthermore, it is straightforward to state a system for the second moments, however, we are not so much interested in the moments of the Markov chain model, but those of the  $L^2(D)$ -valued processes  $(v_t^n)_{t \geq 0}$ , which we can compare to the Langevin approximation. As  $v^n$  is a linear mapping from  $\mathbb{R}^P$  into  $L^2(D)$ , it holds that  $v^n(\mathbb{E}^n \Theta_t^n) = \mathbb{E}^n v^n(\Theta_t^n)$  and  $v^n(\frac{d}{dt} \mathbb{E}^n \Theta_t^n) = \frac{d}{dt} \mathbb{E}^n v^n(\Theta_t^n)$ , and thus

$$\frac{d}{dt} \mathbb{E}^n v_t^n = -\frac{1}{\tau} \mathbb{E}^n v_t^n + \frac{1}{\tau} \mathbb{E}^n \overline{F}^n(v_t^n, t). \tag{B.2}$$

For the second moments of the  $L^2(D)$ -valued process, we obtain for all  $\phi \in L^2(D)$

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}^n (\phi, v_t^n)_{L^2}^2 \\ &= \mathbb{E}^n \left( \sum_{k=1}^P \frac{\Theta_t^{k,n}}{l(k, n)} \int_{D_k} \phi(x) dx \right)^2 \end{aligned}$$

$$\begin{aligned}
 &= 2 \frac{1}{\tau} \mathbb{E}^n [(\phi, v_t^n)_{L^2} (\phi, -v_t^n + \bar{F}^n(v_t^n, t))_{L^2}] \\
 &\quad + \frac{1}{\tau} \mathbb{E}^n \left[ \sum_{k=1}^P \frac{1}{l(k, n)^2} (\Theta_t^{k, n} + l(k, n) \bar{f}_{k, n}(v_t^n, t)) \left( \int_{D_{k, n}} \phi(s) dx \right)^2 \right] \\
 &= \frac{2}{\tau} \mathbb{E}^n [(\phi, v_t^n)_{L^2} (\phi, -v_t^n + \bar{F}^n(v_t^n, t))_{L^2}] \\
 &\quad + \mathbb{E}^n (G^n(\Theta_t^n, t) \phi, \phi)_{L^2}, \tag{B.3}
 \end{aligned}$$

where the bilinear form  $(G^n(\Theta_t^n, t) \phi, \phi)_{L^2}$  is as defined in (4.29).

Next, we state the moment equations for the stochastic partial differential equations. We assume that the Langevin approximation (2.22) possesses a (strong) solution in an appropriate Hilbert space  $H$  and employ the Itô-formula Sect. 4.5 in [12] which yields for all  $\phi \in H^*$

$$\langle \phi, V_t \rangle_H = \langle \phi, V_0 \rangle_H + \epsilon_n \int_0^t \langle \phi, \sqrt{G(V_s, s)} dW_s \rangle_H + \int_0^t \left\langle \phi, \frac{1}{\tau} V_s + \frac{1}{\tau} F(V_s, s) \right\rangle_H ds$$

and

$$\begin{aligned}
 \langle \phi, V_t \rangle_H^2 &= \langle \phi, V_0 \rangle_H^2 + \epsilon_n \int_0^t \langle 2 \langle \phi, V_s \rangle_H \phi, \sqrt{G(V_s, s)} dW_s \rangle_H \\
 &\quad + 2 \int_0^t \langle \phi, V_s \rangle_H \left\langle \phi, -\frac{1}{\tau} V_s + \frac{1}{\tau} F(V_s, s) \right\rangle_H ds \\
 &\quad + \epsilon_n^2 \int_0^t \langle \phi, G(V_s, s) \phi \rangle_H ds.
 \end{aligned}$$

Next, we take the expectation both sides of these identities and differentiate with respect to  $t$  resulting for the first moment in the differential equations

$$\frac{d}{dt} \mathbb{E} \langle \phi, V_t \rangle_H = \left\langle \phi, \mathbb{E} \left[ -\frac{1}{\tau} V_t + \frac{1}{\tau} F(V_t, t) \right] \right\rangle_H$$

which is equivalent to the abstract evolution equation in  $H$  given by

$$\frac{d}{dt} \mathbb{E} V_t = -\frac{1}{\tau} \mathbb{E} V_t + \frac{1}{\tau} \mathbb{E} F(V_t, t). \tag{B.4}$$

And for the second moment, we obtain the differential equation

$$\frac{d}{dt} \mathbb{E} \langle \phi, V_t \rangle_H^2 = \frac{2}{\tau} \mathbb{E} [ \langle \phi, V_t \rangle_H \langle \phi, -V_t + F(V_t, t) \rangle_H ] + \epsilon_n^2 \mathbb{E} [ \langle G(V_t, t) \phi, \phi \rangle_H ]. \tag{B.5}$$

Further, the linear noise approximation (2.21) satisfies the equations

$$\frac{d}{dt} \mathbb{E} U_t = -\frac{1}{\tau} \mathbb{E} U_t + \frac{1}{\tau} \mathbb{E} F(U_t, t) \tag{B.6}$$

and

$$\frac{d}{dt} \mathbb{E} \langle \phi, U_t \rangle_H^2 = \frac{2}{\tau} \mathbb{E} [ \langle \phi, U_t \rangle_H \langle \phi, -U_t + F(U_t, t) \rangle_H ] + \epsilon_n^2 \langle G(v(t), t) \phi, \phi \rangle_H. \tag{B.7}$$

Finally, we note that exactly the same moment equations hold for the variants of the linear noise and Langevin approximation using a  $Q$ -Wiener process and an appropriate diffusion coefficient, cf. Remark 2.4.

A comparison of the moment equations (B.1), (B.4), (B.6) for the mean and (B.3), (B.5), (B.7) for the second moments show that they are similar in structure, but do not coincide. This is analogous to the properties of the moment equations in finite dimension and as in finite dimensions there is one exception, which is the case of first order transitions: If  $F$  were affine<sup>6</sup> in  $v$ , i.e.,  $F(v, t) = f_1(t) \cdot v + f_2(t)$ , then we obtain that the first moment equations (B.4) and (B.6) of the Langevin and linear noise approximation, respectively, reduce to the Wilson–Cowan equation with  $v(t) = \mathbb{E}V_t = \mathbb{E}U_t$ . Furthermore, if  $F$  is affine, this implies that also  $G$  is affine in  $V_t$ , and thus

$$\langle \phi, G(V_t, t) \phi \rangle_H = \frac{1}{\tau} \langle \phi, V_t \cdot \phi \rangle_H + \langle \phi, f_1(t) \cdot V_t \cdot \phi \rangle_H + \langle \phi, f_2(t) \cdot \phi \rangle_H. \tag{B.8}$$

Taking the expectation on both sides and assuming interchangeability of the expectation with the application of all the linear forms (think of the duality pairing as the inner product in  $L^2(D)$ ), we obtain

$$\begin{aligned} \mathbb{E} \langle G(V_t, t) \phi, \phi \rangle_H &= \frac{1}{\tau} \langle \phi, \mathbb{E}[V_t] \cdot \phi \rangle_H + \langle \phi, f_1(t) \cdot \mathbb{E}[V_t] \cdot \phi \rangle_H + \langle \phi, f_2(t) \cdot \phi \rangle_H \\ &= \langle G(\mathbb{E}[V_t], t) \phi, \phi \rangle_H. \end{aligned} \tag{B.9}$$

As  $\mathbb{E}V_t = \mathbb{E}U_t = v(t)$ , we obtain that the second moment equation for the Langevin approximation and the linear noise approximation coincide. Moreover, they are closed (for each  $\phi$ ), i.e., the system depends only on  $\mathbb{E}V_t$  and  $\mathbb{E} \langle \phi, V_t \rangle_H^2$ . Again, this corresponds to the well-known case from finite-dimensional chemical reaction kinetics.

Finally, if  $F$  is affine also the connection of the moment equations for the resulting Markov chain models is interesting. On the one hand, the equation for the mean coincides with the Wilson–Cowan equation where the gain function in its right-hand side is given by  $\bar{F}^n$ . As  $\bar{F}^n$  is essentially a piecewise constant approximation to  $F$ , the resulting equations for the mean correspond to a spatial discretisation of the Wilson–Cowan equation, cf. the continuum limit in the derivation of the mean field equation in [5].

---

<sup>6</sup>In the case of an affine function  $F(v, t) = f_1(t) \cdot v + f_2(t)$ , the mapping  $f_1$  is a linear form on  $H$ , which is interchangeable with the expectation operator. For example, in the simplest case, the application of  $f_1$  to  $v \in H$  is just a multiplication by a scalar.

## References

- Acosta G, Durán RG: **An optimal Poincaré inequality in  $L_1$  for convex domains.** *Proc Am Math Soc* 2004, **132**(1):195-202.
- Adams RA, Fournier JFJ: *Sobolev Spaces*. 2nd edition. San Diego: Academic Press; 2003.
- Benayoun M, Cowan JD, van Drongelen W, Wallace E: **Avalanches in a stochastic model of spiking neurons.** *PLoS Comput Biol* 2010, **6**(7):e1000846.
- Blount D: **Comparison of stochastic and deterministic models of a linear chemical reaction with diffusion.** *Ann Probab* 1991, **19**(4):1440-1462.
- Bressloff PC: **Stochastic neural field theory and the system-size expansion.** *SIAM J Appl Math* 2009, **70**:1488-1521.
- Bressloff PC: **Metastable states and quasicycles in a stochastic Wilson–Cowan model of neuronal population dynamics.** *Phys Rev E* 2010, **82**:051903.
- Bressloff PC: **Spatiotemporal dynamics of continuum neural fields.** *J Phys A, Math Theor* 2012, **45**:033001.
- Buice MA, Cowan JD: **Field-theoretic approach to fluctuation effects in neural networks.** *Phys Rev E* 2007, **75**:051919.
- Buice MA, Cowan JD, Chow CC: **Systematic fluctuation expansion for neural network activity equations.** *Neural Comput* 2010, **22**(2):377-426.
- Corduneanu C: *Integral Equations and Applications*. Cambridge: Cambridge University Press; 1991.
- Cowan JD: **Stochastic neurodynamics.** In *Advances in Neural Information Processing Systems*. Edited by Touretzky DS, Lippmann RP, Moody JE. San Mateo: Morgan Kaufman; 1991:62-69.
- Da Prato G, Zabczyk J: *Stochastic Equations in Infinite Dimensions*. Cambridge: Cambridge University Press; 1992.
- Darling RWR, Norris JR: **Differential equation approximation for Markov chains.** *Probab Surv* 2007, **5**:37-79.
- Davis MHA: *Markov Models and Optimisation*. London: Chapman & Hall; 1993.
- Ethier SN, Kurtz TG: *Markov Processes: Characterization and Convergence*. New York: Wiley; 1986.
- Jacobsen M: *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Boston: Birkhäuser; 2006.
- Kotelenez P: **Law of large numbers and central limit theorem for linear chemical reactions with diffusion.** *Ann Probab* 1986, **14**(1):173-193.
- Kotelenez P: **Linear parabolic differential equations as limits of space-time jump Markov processes.** *J Math Anal Appl* 1986, **116**(1):42-76.
- Kurtz TG: **Solutions of ordinary differential equations as limits of pure jump Markov processes.** *J Appl Probab* 1970, **7**:49-58.
- Kurtz TG: **Limit theorems for a sequence of jump Markov processes approximating ordinary differential equations.** *J Appl Probab* 1971, **8**:344-356.
- Mélykúti B, Burrage K, Zygalkakis KC: **Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical Langevin equation.** *J Chem Phys* 2010, **132**:164109.
- Ohira T, Cowan JD: **Stochastic neurodynamics and the system size expansion.** In *Mathematics of Neural Networks: Models, Algorithms, and Applications*. Edited by Ellacott S, Mason JC, Anderson IC. Norwell: Kluwer Academic; 1997:290-294.
- Payne LE, Weinberger HF: **An optimal Poincaré inequality for convex domains.** *Arch Ration Mech Anal* 1960, **5**:286-292.
- Pothast R, Beim Graben P: **Existence and properties of solutions for neural field equations.** *Math Methods Appl Sci* 2010, **33**(8):935-949.
- Prévôt C, Röckner M: *A Concise Course on Stochastic Partial Differential Equations*. Berlin: Springer; 2007.
- Riedler MG: **Spatio-temporal stochastic hybrid models of excitable biological membranes.** *PhD thesis*. Heriot-Watt University; 2011.
- Riedler MG, Thieullen M, Wainrib G: **Limit theorems for piecewise deterministic processes in infinite dimensions with applications to diffusion approximation of spatially extended models of neuronal membranes.** *Electron J Probab* 2012, **17**:55.
- Sharpe M: *General Theory of Markov Processes*. San Diego: Academic Press; 1988.
- Topsoe F: **Preservation of weak convergence under mappings.** *Ann Math Stat* 1967, **38**(6):1661-1665.

30. Touboul J: **Propagation of chaos in neural fields**. arXiv:1108.2414v3; 2011.
31. Touboul J, Ermentrout B: **Finite-size and correlation-induced effects in mean-field dynamics**. *J Comput Neurosci* 2011, **31**(3):453-484.
32. Touboul J, Hermann G, Faugeras O: **Noise-induced behaviors in neural mean field dynamics**. *SIAM J Appl Dyn Syst* 2012, **11**(1):49-81.
33. van Kampen NG: *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland; 1981.
34. Veltz R, Faugeras O: **Local/global analysis of the stationary solutions of some neural field equations**. *SIAM J Appl Dyn Syst* 2010, **9**:954-998.
35. Wloka J: **Vektorwertige Sobolev-Slobodeckijische Distributionen**. *Math Z* 1967, **98**:303-318.
36. Yin GG, Zhang Q: *Continuous-Time Markov Chains and Applications*. New York: Springer; 1998.