A two state model for noise-induced resonance in bistable systems with delay*

MARKUS FISCHER AND PETER IMKELLER
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
Germany

3rd January 2005

Abstract

The subject of the present paper is a simplified model for a symmetric bistable system with memory or delay, the reference model, which in the presence of noise exhibits a phenomenon similar to what is known as stochastic resonance. The reference model is given by a one dimensional parametrized stochastic differential equation with point delay, basic properties whereof we check.

With a view to capturing the effective dynamics and, in particular, the resonance-like behavior of the reference model we construct a simplified or reduced model, the two state model, first in discrete time, then in the limit of discrete time tending to continuous time. The main advantage of the reduced model is that it enables us to explicitly calculate the distribution of residence times which in turn can be used to characterize the phenomenon of noise-induced resonance.

Drawing on what has been proposed in the physics literature, we outline a heuristic method for establishing the link between the two state model and the reference model. The resonance characteristics developed for the reduced model can thus be applied to the original model.

2000 AMS subject classifications: primary 34K50, 60H10; secondary 60G17, 34K11, 34K13, 34K18, 60G10.

Key words and phrases: stochastic differential equation; delay differential equation; stochastic resonance; effective dynamics; Markov chain; stationary process; stochastic synchronization.

1 Introduction

Stochastic resonance in a narrower sense is the random amplification of a weak periodic signal induced by the presence of noise of low intensity such that the signal amplification is maximal at a certain optimal non-zero level of noise. In addition to weak additive noise and a weak periodic input signal there is a third ingredient in systems where stochastic resonance can occur, namely a threshold or a barrier that induces several (in our case two) macroscopic states in the output signal.

Consider a basic, yet fundamental example. Let V be a symmetric one dimensional double well potential. A common choice for V is the standard quartic potential, see Fig. 1 a). The barrier mentioned above is in this case the potential barrier of V separating the two local minima. Assume that the periodic input signal is sinusoidal and the noise white. The output of such a system is described by the stochastic differential equation (SDE)

$$(1)\qquad dX(t) \;=\; -\Big(V'\big(X(t)\big) + a\cdot\sin\big(\tfrac{2\pi}{T}t\big)\Big)dt \;+\; \sigma\cdot dW(t), \qquad t\geq 0,$$

where W is a standard one dimensional Wiener process, $\sigma \geq 0$ a noise parameter, V' the first order derivative of the double well potential V, $a \geq 0$ the amplitude and $T > 0$ the period of the input signal.

*This work was partially supported by the DFG research center Matheon (FZT 86) in Berlin.

As an alternative to the system view, equation (1) can be understood as describing the overdamped motion of a small particle in the potential landscape V in the presence of noise and under the influence of an exterior periodic force. It was originally proposed by Benzi et al. (1981, 1982) and Nicolis (1982) as an energy balance model designed to explain the succession of ice and warm ages in paleoclimatic records as a phenomenon of quasi periodicity in the average global temperature on Earth.

If $a = 0$, i. e. in the absence of a periodic signal, equation (1) reduces to an autonomous SDE which has two metastable states corresponding to the two local minima of V. With $\sigma > 0$ sufficiently small, the diffusion will spend most of its time near the positions of these minima. In the presence of weak noise, there are two distinct time scales, a short one corresponding to the quadratic variation of the Wiener process, and a long one proportional to the average time it takes the diffusion to travel from one of the metastable states to the other.

The fact that the time scale induced by the noise process is small in comparison with the mean residence time as σ tends to zero should allow us to disregard small intrawell fluctuations when we are interested in the interwell transition behavior.

Suppose $a > 0$ small enough so that there are no interwell transitions in case $\sigma = 0$, i. e. in the deterministic case. The input signal then slightly and periodically tilts the double well potential V. We now have two different mean residence times, namely the average time the particle stays in the shallow well and the average time of residence in the deep well. Of course, both time scales also depend on the noise intensity.

Notice that deep and shallow well change roles every half period $\frac{T}{2}$. Given a sufficiently long period T, the noise intensity can now be tuned in such a way as to render the occurence of transitions from the shallow to the deep well probable within one half period, while this time span is too short for the occurrence of transitions in the opposite direction. At a certain noise level the output signal will exhibit quasi periodic transition behavior, thereby inducing an amplification of the input signal.

For a more comprehensive description of stochastic resonance, examples, variants and applications thereof see Gammaitoni et al. (1998) or Anishchenko et al. (1999). What models that exhibit stochastic resonance have in common is the quasi periodicity of the output at a certain non-zero noise level. More generally, stochastic resonance is an instance of noise-induced order.

In view of the fact that the system given by equation (1) can work as a random amplifier it seems natural to take the frequency spectrum of the output signal as basis for a measure of resonance. The most common measure of this kind is the spectral power amplification (SPA) coefficient. Another measure of resonance based on the frequency spectrum is the signal to noise ratio (SNR). For a detailed analysis see Pavlyukevich (2002).

In general, when measuring stochastic resonance, it is assumed that the solution is in a "stationary regime". Since equation (1) is time dependent for $a > 0$ we cannot expect $(X(t))_{t\geq 0}$ to be a stationary process. Transforming the non-autonomous SDE (1) into an autonomous SDE with state space $\mathbb{R} \times S^1$, one can recover the time homogeneous Markov property and a unique invariant probability measure exists, cf. Imkeller and Pavlyukevich (2002). In section 3 we will make use of the same idea of appropriately enlarging the state space in order to regain a time homogeneous Markov model.

A different starting point for a measure of resonance – the one that will be adopted here – is the distribution of intrawell residence times. Observe that the roles of the two potential wells are interchangeable.

A third class of measures of resonance is provided by methods of quantifying (un)certainty, in particular by the entropy of a distribution. This agrees well with the view of stochastic resonance as an instance of noise-induced order.

The fact that with $\sigma > 0$ and $a > 0$ small a typical solution to (1) spends most of its time near the positions of the two minima of the double well potential V suggests to identify the two potential wells with their respective minima. The state space $\mathbb{R}$ of the non-autonomous SDE thus gets reduced to two states, say -1 and 1, corresponding to the left and the right well, respectively.

According to an idea of McNamara and Wiesenfeld (1989) the effective dynamics of equation (1) can be captured in the two state model by constructing a $\{-1, 1\}$-valued time inhomogeneous Markov chain with certain (time dependent) transition rates. These rates are determined as the rates of escape from the potential well of the tilted double well potential which corresponds to the reduced state in question. An approximation of the rate of escape from a parabolic potential well is given in the limit of small noise by the Kramers formula, cf. section 5.

In the physics literature, a standard ansatz for calculating the two state process given time dependent transition rates is to solve an associated differential equation for the probabilities of occupying state ± 1 at time t, a so-called master equation (cf. Gammaitoni et al., 1998).

An advantage of the reduced model is its simplicity. It should be especially useful in systems with more than two meta-stable states. Although it is intuitively plausible to apply a two state filter, there is possibly a problem with the measure of resonance, for it might happen that with the same notion of tuning stochastic resonance would be detected in the two state model, while no optimal noise level, i. e. no point of stochastic resonance, exists in the continuous case. This is, indeed, a problem for the SPA coefficient and related measures, see Pavlyukevich (2002). The reason is that in passing to the reduced model small intrawell fluctuations are "filtered out", while they decisively contribute to the SPA coefficient in the original model.

Measures of resonance based on the distribution of intrawell residence times, however, do not have this limitation, that is they are robust under model reduction as Herrmann et al. (2003) show.

In equation (1) replace the term that represents the periodic input signal with a term that corresponds to a force dependent on the state of the solution path a fixed amount of time into the past, that is replace the periodic signal with a point delay. This yields what will be our reference model, see equation (2).

The idea to study such equations with regard to noise-induced resonance seems to originate with Ohira and Sato (1999). Their analysis, though, is of limited use, because they make too strong assumptions on independence between the components of the reduced model which they consider in discrete time only.

A better analysis of the reduced model for an important special choice of equation (2) can be found in Tsimring and Pikovsky (2001). The same model is the object of recent studies by Masoller (2003), Houlihan et al. (2004) and Curtin et al. (2004), and it will be our standard example, too.

While the measure of resonance applied by Tsimring and Pikovsky (2001) is essentially the first peak in the frequency spectrum, in the other articles focus is laid on the residence time distribution in the reduced model, which is compared with numerical simulations of the original dynamics. Under certain simplifying assumptions, approximative analytical results are obtained via a master equation approach, where the master equation is a DDE instead of an ODE.

In section 5 we follow Tsimring and Pikovsky (2001) in establishing the link between the reduced and the reference model. Results by Masoller (2003) show that the density of the residence time distribution has a characteristic jump. She proposes to take the height of this jump as a measure of resonance, and we will follow her proposal, supplementing it by an alternative.

Our approach is different, though, in that we do not use any kind of master equation. Instead, we construct a reduced model with enlarged state space, which has the Markov property and which allows us to explicitly calculate the stationary distributions as well as the residence time distributions.

2 The reference model

Consider the one dimensional motion of a small particle in the presence of large friction and additive white noise subject to the influence of two additional forces: one dependent on the current position of the particle and corresponding to a symmetric double well potential V, the other dependent on the position of the particle a certain amount of time r in the past and corresponding to a symmetric single well potential U, where the position of the extremum of U coincides with the position of the saddle point of V.

Without loss of generality we may assume that the saddle point of the potential V is at the origin and the extrema are located at $(-1,-L)$ and $(1,-L)$ respectively, where $L>0$ is the height of the potential barrier. A standard choice for V is the quartic potential $x \mapsto L(x^4-2x^2)$.

Instead of U we will consider $\beta \cdot U$, where β is a scalar, that serves to "adjust" explicitly the strength of the delay force. An admissible function for U is the parabola $x \mapsto \frac{1}{2}x^2$. In fact, with this choice of U and taking as potential V the quartic potential with $L=\frac{1}{4}$ we find ourselves in the setting that was studied by Tsimring and Pikovsky (2001).[1]

The dynamics that govern the motion of a Brownian particle as described above can be expressed by the following stochastic delay differential equation describing our reference model

$$dX(t) \;=\; -\big(V'\big(X(t)\big)+\beta\cdot U'\big(X(t-r)\big)\big)\,dt \;+\; \sigma\cdot dW(t), \qquad t\geq 0, \tag{2}$$

[1] Our notation is slightly different from that of equation (1) in Tsimring and Pikovsky (2001: p. 1). In particular, their parameter ϵ, indicating the "strength of the feedback", corresponds to $-\beta$, here.

Figure 1: Graphs on the interval $[-2, 2]$ of a) quartic double well potential V, b) quadratic delay potential $U : x \mapsto \frac{1}{2}x^2$, c) absolute value delay potential $U : x \mapsto |x|, x \in \mathbb{R} \setminus (-\delta, \delta)$, smoothly continued on $(-\delta, \delta)$.

where $W(.)$ is a standard one dimensional Wiener process, $r > 0$ the time delay, V', U' are the first derivatives of V and U, respectively, $\beta \in \mathbb{R}$ is a parameter regulating the intensity of the delay force and $\sigma \geq 0$ a noise parameter. In the special case $\sigma = 0$ equation (2) becomes a DDE, while in case $\beta = 0$ we have an ordinary SDE.

The above description of the two potentials is compatible with the following conditions on V and U:

(3a)	$V, U \in \mathbf{C}^2(\mathbb{R})$,	
(3b)	$V(x) = V(-x)$,	$U(x) = U(-x)$ for all $x \in \mathbb{R}$,
(3c)	$V'(x) = 0$ iff $x \in \{-1, 0, 1\}$,	$U'(x) = 0$ iff $x = 0$,
(3d)	$V''(-1) = V''(1) > 0$,	
(3e)	$\sup\{V'(x) \mid x \in (-\infty, -1) \cup (0, 1)\} \leq 0$,	$\sup\{U'(x) \mid x \in (-\infty, 0)\} \leq 0$.

A rather technical restriction on the geometry of V and U is the following: We assume that a constant R_{pot} greater than the positive root of V exists such that V and U are linear on $\mathbb{R} \setminus (-R_{pot}, R_{pot})$. In view of the symmetry of V and U it is sufficient to require that

(4) there exists $R_{pot} > 1$ such that for all $x \in [R_{pot}, \infty)$:
$$V(x) = V'(R_{pot}) \cdot (x - R_{pot}) + V(R_{pot}) \quad \text{and} \quad U(x) = U'(R_{pot}) \cdot (x - R_{pot}) + U(R_{pot}).$$

This condition is needed in order to ensure the existence of an invariant probability measure for X according to Scheutzow Scheutzow (1983, 1984) which is stated in this setting. We conjecture that this existence persists provided that the potential gradients of V and U have at least linear growth at infinity and the growth of $-V'$ is dominated by the growth of $\beta U'$, i. e. $\limsup_{R\to\infty} -\frac{V'(R)}{\beta U'(R)} < 1$ (see Proposition 1 below). Henceforth, whenever the reference model is concerned, we will suppose that conditions (3) and (4) are satisfied.

By applying results from the literature we check some important properties of equation (2). See Mohammed (1984, 1996) with regard to existence and uniqueness of solutions and the Markov property of the segment process $X_t = [s \mapsto X(t - s), s \in [-r, 0]], t \geq 0$, Scheutzow (1983, 1984), where conditions can be found that guarantee the existence of a stationary distribution, and Larssen (2002) for a version of the Yamada-Watanabe theorem on weak uniqueness.

Proposition 1. *Suppose that V, U satisfy conditions* (3) *and* (4)*. Let $\sigma \geq 0$, $\beta \in \mathbb{R}$ be given. Then the following holds for equation* (2)*:*

1. *Strong and weak solutions exist for every probability measure on $\mathcal{B}(\mathbf{C}([-r, 0]))$ as initial distribution, and the solutions are unique in the respective sense.*

2. *The segment solution processes enjoy the strong Markov property.*

3. *If $\sigma > 0$ and $\beta > -\frac{V'(R_{pot})}{U'(R_{pot})}$, then a unique invariant probability measure π exists for the segment process, which converges in total variation to π for every initial distribution.*

Let us have a look at basic parameter settings for equation (2). The simplest and least interesting choice of parameters is $\sigma = 0$ and $\beta = 0$, i. e. no noise and no delay. In this case, (2) reduces to a one dimensional ordinary differential equation with two stable solutions, namely -1 and 1, and an instable trivial solution.

The dynamics of the general deterministic delay equation, i. e. $\sigma = 0$, $\beta \neq 0$, is not obvious for all combinations $\beta \in \mathbb{R}$, $r > 0$. In Redmonda et al. (2002) stabilization of the trivial solution and the corresponding bifurcation points are studied. The parameter region such that the zero solution is stable is contained in $\beta \geq 1$, $r \in [0, 1]$.[2] This is not the parameter region we are interested in, here. Recall from section 1 that stochastic resonance is a phenomenon concerned with an increase of order in the presence of weak non-zero noise. For large $|\beta|$ the delay force would be predominant. Similarly, with r small the noise would not have enough time to influence the dynamics.

Indeed, we must be careful in our choice of β lest we end up with a randomly perturbed deterministic oscillator. Solutions to equation (2) exhibit periodic behaviour even for $\beta > 0$ comparatively small.

If $\beta = 0$ and $\sigma > 0$, then our SDDE (2) reduces to an ordinary SDE. Of interest is again the case of small noise. A Brownian particle moving along a solution trajectory spends most of its time fluctuating near the position of the minimum of one or the other potential well, while interwell transitions only occasionally occur.

Now, let $\sigma > 0$ and $|\beta|$ be small enough so that the corresponding deterministic system does not exhibit oscillations. Let us suppose first that β is positive. Then the effect of the delay force should be that of favouring interwell transitions whenever the Brownian particle is currently in the same potential well it was in r units of time in the past, while transitions should become less likely whenever the particle is currently in the well opposite to the one it was in before. Notice that the influence of the delay force alone is insufficient to trigger interwell transitions. In fact, with $\sigma > 0$ not too big, transitions are rare and a typical solution trajectory will still be found near the position of one or the other minimum of V with high probability.

Consider what happens if the noise intensity increases. Of course, interwell transitions become more frequent, while at the same time the intrawell fluctuations increase in strength. But there is an additional effect: As we let the noise grow stronger interwell transitions occur at time intervals of approximately the same length, namely at intervals between r and $2r$, with high probability. The solution trajectories exhibit quasi-periodic switching behaviour at a non-zero noise level. This is what we may call an instance of stochastic resonance.

Further increasing the noise intensity leads to ever growing intrawell fluctuations which eventually destroy the quasi-periodicity of the interwell transitions. When the noise is too strong, the potential barrier of V has no substantial impact anymore and random fluctuations easily crossing the barrier are predominant.

Suppose β is negative. The effect of the delay force, now, is that of pushing the Brownian particle out of the potential well it is currently in whenever the particle's current position is on the side of the potential barrier opposite to the one remembered in the past. Sojourns of duration longer than r, on the other hand, become prolonged due to the influence of the delay which in this case renders transitions less likely.

In order to obtain some kind of regular transition behaviour a higher noise level as compared to the case of positive β is necessary. Of course, one could change time scales by increasing the delay time r, thereby allowing for lower noise intensities. In section 5 we will state more precisely what regular transition behaviour means in case $\beta < 0$, yet we will not subsume it under the heading of stochastic resonance.

3 The two state model in discrete time

Applying the ideas sketched in section 1, we develop a reduced model with the aim of capturing the effective dynamics of the reference model. To simplify things further we start with discrete time. As the segment process associated with the unique solution to (2), the reference model equation, enjoys the strong Markov property, it is reasonable to approximate the transition behaviour of that solution by a sequence of Markov chains. One unit of time in the discrete case corresponds to r/M time units in the original model, where the delay interval $[-r, 0]$ is divided into $M \in \mathbb{N}$ equally spaced subintervals.

[2]Our parameter β corresponds to $-\alpha$ in equation (1.3) of Redmonda et al. (2002).

After defining the approximating Markov chains we obtain an explicit formula for their stationary distributions which will be useful in calculating, for each $M \in \mathbb{N}$, the residence time distribution in the stationary regime and deriving its density function in the limit of discrete time tending to continuous time. Finally, based on the residence time distributions, we introduce two simple measures of resonance.

The results on Markov chains we need are elementary and can be found, for example, in Brémaud (1999), which will be our standard reference.

3.1 A sequence of Markov chains and stationary distributions

Let $M \in \mathbb{N}$ be the discretization degree, that is the number of subintervals of $[-r, 0]$. The current state of the process we have to construct can attain only two values, say -1 and 1, corresponding to the positions of the two minima of the double-well potential V. Now, there are $M+1$ lattice points in $[-r, 0]$ that delimit the M equally spaced subintervals, giving rise to 2^{M+1} possible states in the enlarged state space.

Let $\mathrm{S}_M := \{-1, 1\}^{M+1}$ denote the state space of the Markov chain with time unit r/M. Elements of S_M will be written as $(M+1)$-tuples having $\{-1, 1\}$-valued entries indexed (from left to right) from $-M$ to 0. This choice of the index range serves as a mnemonic device to recall how we have discretized the delay interval $[-r, 0]$. Thus, $l \in \{-M, \ldots, 0\}$ corresponds to the point $l \cdot r/M$ in continuous time.

To embed the discrete into the time continuous model, let α, γ be positive real numbers. If $X(.)$ is the unique solution to (2) in the case of "interesting" noise parameter σ and delay parameter β, one may think of α as the escape rate of $X(.)$ from one of the two potential wells under the condition $X(t) \approx X(t-r)$ and of γ as the escape rate of $X(.)$ under the condition $X(t) \approx -X(t-r)$. All of the parameters of the reference model, including the delay length and the geometry of the potentials U and V, will enter the discrete model through the transition rates α and γ, cf. section 5.

In the discrete model of degree M, instead of two different transition rates we have two different transition probabilities α_M and γ_M with $\alpha_M = R_{sc}(\alpha, M)$, $\gamma_M = R_{sc}(\gamma, M)$, where R_{sc} is an appropriate scaling function. In analogy to the time discretization of a Markov process we set

$$R_{sc} : \{\alpha, \gamma\} \times \mathbb{N} \ni (\eta, N) \mapsto \tfrac{\eta}{\alpha+\gamma} \cdot (1 - e^{-\frac{\alpha+\gamma}{N}}) \in (0, 1). \tag{5}$$

Let $Z = (Z^{(-M)}, \ldots, Z^{(0)})$, $\tilde{Z} = (\tilde{Z}^{(-M)}, \ldots, \tilde{Z}^{(0)})$ be elements of S_M. A transition from Z to $\tilde{Z}$ shall have positive probability only if the following *shift condition* holds:

$$\forall\, l \in \{-M, \ldots, -1\} : \quad \tilde{Z}^{(l)} = Z^{(l+1)}. \tag{6}$$

Example. Take the element $(-1, 1, -1) \in \mathrm{S}_2$. According to the shift condition, starting from $(-1, \mathbf{1}, \mathbf{-1})$ there are at most two transitions with positive probability, namely to the elements $(\mathbf{1}, \mathbf{-1}, 1)$ and $(\mathbf{1}, \mathbf{-1}, -1)$. $\Diamond$

If (6) holds for Z and $\tilde{Z}$ then there are two cases to distinguish which correspond to the conditions $X(t) \approx X(t-r)$ and $X(t) \approx -X(t-r)$, respectively. Denote by $p^M_{Z\tilde{Z}}$ the probability to get from state Z to state $\tilde{Z}$. Under condition (6) we must have

$$\begin{aligned}
Z^{(0)} = Z^{(-M)} \quad \text{then} \quad p^M_{Z\tilde{Z}} &= \begin{cases} \alpha_M & \text{if} \quad \tilde{Z}^{(0)} \neq Z^{(0)}, \\ 1 - \alpha_M & \text{if} \quad \tilde{Z}^{(0)} = Z^{(0)}, \end{cases} \\
Z^{(0)} \neq Z^{(-M)} \quad \text{then} \quad p^M_{Z\tilde{Z}} &= \begin{cases} \gamma_M & \text{if} \quad \tilde{Z}^{(0)} \neq Z^{(0)}, \\ 1 - \gamma_M & \text{if} \quad \tilde{Z}^{(0)} = Z^{(0)}. \end{cases}
\end{aligned} \tag{7}$$

The fact that – because of (5) – we always have $\alpha_M, \gamma_M \in (0, 1)$, implies

$$p^M_{Z\tilde{Z}} \neq 0 \quad \text{iff} \quad \text{shift condition (6) is satisfied.} \tag{8}$$

Now define $\mathbf{P}_M := (p^M_{Z\tilde{Z}})_{Z,\tilde{Z} \in \mathrm{S}_M}$. Clearly, $\mathbf{P}_M$ is a $2^{M+1} \times 2^{M+1}$ transition matrix. For every $M \in \mathbb{N}$ we may choose an S_M-valued process $(X^M_n)_{n \in \mathbb{N}_0}$ on some measurable space $(\Omega_M, \mathcal{F}_M)$ and probability measures P^M_Z, $Z \in \mathrm{S}_M$, on $\mathcal{F}_M$ such that under P^M_Z the discrete process X^M is a homogeneous Markov chain with transition matrix $\mathbf{P}_M$ and initial condition $P^M_Z(X^M_0 = Z) = 1$.

If ν is a probability measure on the power set $\wp(\mathrm{S}_M)$ then, as usual, P_ν^M will denote the probability measure on $\mathcal{F}_M$ such that X^M is a Markov chain with transition matrix $\mathbf{P}_M$ and initial distribution ν with respect to P_ν^M. When there is no ambiguity about the probability measure P^M, we will write P_ν instead of P_ν^M.

From relation (8), characterizing the non-zero entries of $\mathbf{P}_M$, it follows that $\mathbf{P}_M$ and the associated Markov chains are irreducible. They are also aperiodic, because the time of residence in state $(-1,\dots,-1)$, for example, has positive probability for any finite number of steps. Since the state space S_M is finite, irreducibility implies positive recurrence, and these two properties together are equivalent to the existence of a uniquely determined stationary distribution on the state space, cf. Brémaud (1999: pp. 104-105). Therefore, for every $M \in \mathbb{N}$, we have a uniquely determined probability measure π_M on $\wp(\mathrm{S}_M)$ such that [3]

$$\pi_M(\tilde{Z}) \quad = \quad \sum_{Z\in\mathrm{S}_M} \pi_M(Z)\, p^M_{Z\tilde{Z}} \qquad \text{for all } \tilde{Z} \in \mathrm{S}_M. \tag{9}$$

There is a simple characterization of the stationary distribution π_M in terms of the number of "jumps" of the elements of S_M.[4] Let $Z = (Z^{(-M)},\dots,Z^{(0)})$ be an element of S_M, and define the number of jumps of Z as

$$\mathcal{J}(Z) := \#\left\{ j \in \{-M+1,\dots,0\} \mid Z^{(j)} \neq Z^{(j-1)} \right\}.$$

The global balance equations (9) then lead to

Proposition 2 (Number of jumps formula). *Let $M \in \mathbb{N}$. Set $\tilde{\alpha}_M := \frac{\alpha_M}{(1-\gamma_M)}$, $\tilde{\gamma}_M := \frac{\gamma_M}{(1-\alpha_M)}$, $\tilde{\eta}_M := \tilde{\alpha}_M \cdot \tilde{\gamma}_M$. Then for all $Z \in \mathrm{S}_M$ the following formula holds*

$$\pi_M(Z) \;=\; \frac{1}{c_M}\, \tilde{\alpha}_M^{\lfloor \frac{\mathcal{J}(Z)+1}{2} \rfloor}\, \tilde{\gamma}_M^{\lfloor \frac{\mathcal{J}(Z)}{2} \rfloor} \;=\; \frac{1}{c_M}\, \tilde{\alpha}_M^{\mathcal{J}(Z) \bmod 2}\, \tilde{\eta}_M^{\lfloor \frac{\mathcal{J}(Z)}{2} \rfloor}, \tag{10}$$

where $c_M := 2 \cdot \sum_{j=0}^{M} \binom{M}{j}\, \tilde{\alpha}_M^{j \bmod 2}\, \tilde{\eta}_M^{\lfloor \frac{j}{2} \rfloor}$.

Let c_M be the normalizing constant from proposition 2. By splitting up the sum in the binomial formula we see that

$$c_M \quad = \quad \Big(1 + \sqrt{\tfrac{\tilde{\alpha}_M}{\tilde{\gamma}_M}}\Big)\big(1 + \sqrt{\tilde{\eta}_M}\big)^M + \Big(1 - \sqrt{\tfrac{\tilde{\alpha}_M}{\tilde{\gamma}_M}}\Big)\big(1 - \sqrt{\tilde{\eta}_M}\big)^M. \tag{11}$$

3.2 Residence time distributions

Let Y^M be the $\{-1,1\}$-valued sequence of current states of X^M, that is [5]

$$Y_n^M := \begin{cases} (X_n^M)^{(0)} & \text{if} \quad n \in \mathbb{N}, \\ (X_0^M)^{(n)} & \text{if} \quad n \in \{-M,\dots,0\}. \end{cases}$$

Denote by $L_M(k)$ the probability to remain exactly k units of time in the same state conditional on the occurrence of a jump, that is

$$L_M(k) \;=\; P_{\pi_M}(Y_n^M = 1,\dots,Y_{n+k-1}^M = 1, Y_{n+k}^M = -1 \mid Y_{n-1}^M = -1, Y_n^M = 1), \quad k \in \mathbb{N}, \tag{12}$$

where $n \in \mathbb{N}$ is arbitrary. The above conditional probability is well defined, because

$$P_{\pi_M}\big(Y_{-1}^M = -1, Y_0^M = 1\big) \;=\; \pi_M\big(\{(*,\dots,*,-1,1)\}\big) \;>\; 0.$$

Here, $\{(*,\dots,*,-1,1)\}$ denotes the set $\{Z \in \mathrm{S}_M \mid Z^{(-1)} = -1, Z^{(0)} = 1\}$. By symmetry the roles of -1 and 1 in (12) are interchangeable. Under P_{π_M} not only X^M is a stationary process, but – as a coordinate

[3] For the probability of a singleton $\{Z\}$ under a discrete measure ν we just write $\nu(Z)$.

[4] At the moment, "number of changes of sign" would be a label more precise for $\mathcal{J}(Z)$, but cf. section 4.

[5] Recall the tuple notation for elements of S_M.

projection – Y^M is stationary, too, although it does not, in general, enjoy the Markov property. We note that $L_M(k)$, $k \in \mathbb{N}$, gives the residence time distribution of the sequence of current states of X^M in the stationary regime.

Observe that $L_M(.)$ has a "geometric tail". To make this statement precise set

$$K_M := P_{\pi_M}\big(Y_0^M = -1, Y_1^M = 1, \ldots, Y_M^M = 1 \mid Y_0^M = -1, Y_1^M = 1\big). \tag{13}$$

In view of the "extended Markov property" of Y^M, that is the Markov property of the segment chain X^M, we have

$$L_M(k) \quad = \quad (1-\gamma_M)\cdot K_M \cdot \alpha_M \cdot (1-\alpha_M)^{k-M-1}, \qquad k \geq M+1, \tag{14}$$

where $(1-\gamma_M)\cdot K_M$ is the probability mass of the geometric tail. Stationarity of P_{π_M} implies

$$K_M \quad = \quad \frac{\pi_M\big((-1,1,\ldots,1)\big)}{\pi_M\big(\{(*,\ldots,*,-1,1)\}\big)}.$$

From proposition 2 we see that

$$\pi_M\big((-1,1,\ldots,1)\big) \;=\; \frac{\tilde{\alpha}_M}{c_M},$$

and arranging the elements of $\{(*,\ldots,*,-1,1)\}$ according to their number of jumps we obtain

$$\pi_M\big(\{(*,\ldots,*,-1,1)\}\big) \quad = \quad \frac{\tilde{\alpha}_M}{2c_M}\cdot\Big(\Big(1+\sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1+\sqrt{\tilde{\eta}_M}\big)^{M-1} + \Big(1-\sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1-\sqrt{\tilde{\eta}_M}\big)^{M-1}\Big).$$

We therefore have

$$K_M \quad = \quad \frac{2}{\Big(1+\sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1+\sqrt{\tilde{\eta}_M}\big)^{M-1} + \Big(1-\sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1-\sqrt{\tilde{\eta}_M}\big)^{M-1}}. \tag{15}$$

In a similar fashion we can calculate $L_M(k)$ for $k \in \{1,\ldots,M\}$. We obtain

$$L_M(M) \quad = \quad \frac{P_{\pi_M}\big(Y_0^M=-1, Y_1^M=1,\ldots,Y_M^M=1, Y_{M+1}^M=-1\big)}{\pi_M\big(\{(*,\ldots,*,-1,1)\}\big)} \quad = \quad \gamma_M\cdot K_M, \tag{16a}$$

and for $k \in \{1,\ldots,M-1\}$

$$\begin{aligned} L_M(k) \quad = \quad \frac{\sqrt{\tilde{\gamma}_M}}{2}\cdot K_M\cdot\Big(&\sqrt{\tilde{\gamma}_M}\,\big((1+\sqrt{\tilde{\eta}_M})^{M-1-k} + (1-\sqrt{\tilde{\eta}_M})^{M-1-k}\big) \\ &+ \sqrt{\tilde{\alpha}_M}\,\big((1+\sqrt{\tilde{\eta}_M})^{M-1-k} - (1-\sqrt{\tilde{\eta}_M})^{M-1-k}\big)\Big). \end{aligned} \tag{16b}$$

More interesting than the residence time distribution in the case of discrete time is to know the behaviour of this distribution in the limit of discretization degree M tending to infinity.

Recall the definition of scaling function R_{sc} according to equation (5) for some numbers $\alpha, \gamma > 0$. If $\alpha_M = R_{sc}(\alpha, M)$ and $\gamma_M = R_{sc}(\gamma, M)$ for all $M \in \mathbb{N}$, then – with the usual notation $\mathcal{O}(.)$ for the order of convergence – we have

$$\alpha_M = \frac{\alpha}{M} + \mathcal{O}(\frac{1}{M^2}), \qquad\qquad \gamma_M = \frac{\gamma}{M} + \mathcal{O}(\frac{1}{M^2}). \tag{17}$$

Indeed, if condition (17) holds between the transition probabilities α_M, γ_M, $M \in \mathbb{N}$, and some positive transition rates α, γ, then we can calculate the normalizing constant c_M, the "tail constant" K_M and the density function of the residence time distribution in the limit $M \to \infty$.

Proposition 3. *Let $\alpha_M, \gamma_M \in (0,1)$, $M \in \mathbb{N}$. Suppose that the sequences $(\alpha_M)_{M\in\mathbb{N}}$, $(\gamma_M)_{M\in\mathbb{N}}$ satisfy relation* (17) *for some positive real numbers α, γ. Then c_M and K_M converge to c_∞ and K_∞, respectively, as $M \to \infty$, where*

$$c_\infty := \lim_{M\to\infty} c_M \;=\; \Big(1+\sqrt{\tfrac{\alpha}{\gamma}}\Big)e^{\sqrt{\alpha\gamma}} + \Big(1-\sqrt{\tfrac{\alpha}{\gamma}}\Big)e^{-\sqrt{\alpha\gamma}} \;=\; 2\cdot\sum_{k=0}^{\infty}\frac{1}{k!}\,\alpha^{k \bmod 2}\,(\alpha\gamma)^{\lfloor \frac{k}{2}\rfloor}, \tag{18}$$

$$K_\infty := \lim_{M\to\infty} K_M \;=\; \frac{2}{(1+\sqrt{\frac{\gamma}{\alpha}})e^{\sqrt{\alpha\gamma}} + (1-\sqrt{\frac{\gamma}{\alpha}})e^{-\sqrt{\alpha\gamma}}} \;=\; \frac{\sqrt{\alpha}}{\sqrt{\alpha}\cosh(\sqrt{\alpha\gamma})+\sqrt{\gamma}\sinh(\sqrt{\alpha\gamma})}. \tag{19}$$

Define a function $f_L : (0,\infty) \mapsto \mathbb{R}$, $\quad q \mapsto f_L(q) := \lim_{M\to\infty} M\cdot L_M\big(\lfloor qM \rfloor\big)$. Then

$$f_L(q) = \begin{cases} \sqrt{\gamma}\cdot K_\infty \cdot \Big(\sqrt{\gamma}\,\cosh\big(\sqrt{\alpha\gamma}(1-q)\big) + \sqrt{\alpha}\,\sinh\big(\sqrt{\alpha\gamma}(1-q)\big)\Big) & \textit{if} \quad q \in (0,1], \\ K_\infty \cdot \alpha \cdot \exp\big(-\alpha(q-1)\big) & \textit{if} \quad q > 1. \end{cases} \tag{20}$$

Observe that f_L as defined in proposition 3 is indeed the density of a probability measure on $(0,\infty)$. In case $\alpha = \gamma$ this probability measure is just an exponential distribution with parameter α $(=\gamma)$. If $\alpha \neq \gamma$ then f_L has a discontinuity at position 1, where the height of the jump is

$$f_L(1+) - f_L(1-) \quad = \quad K_\infty \cdot (\alpha - \gamma). \tag{21}$$

Clearly, the restrictions of f_L to $(0,1]$ and $(1,\infty)$, respectively, are still strictly decreasing functions, and $f_L(q)$, $q \in (1,\infty)$, is again the density of an exponential distribution, this time with parameter α $(\neq \gamma)$ and total probability mass K_∞. The function $f_L(q)$, $q \in (0,1)$, is the density of a mixture of two "hyperbolic" distributions with the geometric mean $\sqrt{\alpha\gamma}$ of α and γ as parameter and total probability mass $1 - K_\infty$. The ratio between the hyperbolic cosine and the hyperbolic sine density is $\sqrt{\gamma}$ to $\sqrt{\alpha}$.

Recall how at the beginning of this section we interpreted the discretization degree M as the number of subintervals of $[-r, 0]$, where $r > 0$ is the length of the delay that appears in equation (2). Let us assume that the numbers α, γ are functions of the parameters of our reference model, in particular of the noise parameter σ and the length of the delay r. Then we should interpret the density f_L as being defined on normalized time, that is one unit of time corresponds to r units of time in the reference model. The density of the residence time distribution for the two state model in continuous time should therefore read

$$\tilde{f}_L(t) := \tfrac{1}{r} f_L\big(\tfrac{t}{r}\big), \qquad t \in (0,\infty). \tag{22}$$

Before we may call $\tilde{f}_L$ the density of a residence time distribution, we have to justify the passage to the limit $M \to \infty$ at the level of distributions of the Markov chains X^M, which underlie the definition of L_M. We return to this issue in section 4.

3.3 Two measures of resonance

Drawing on the residence time distribution of the Markov chain X^M we introduce simple characteristics that provide us with a notion of quality of tuning for the reduced model in discrete time.

We consider X^M and the resonance characteristics to be defined in the stationary regime only, because by doing so we can guarantee that an eventual resonance behaviour of the trajectories of X^M is independent of transitory behaviour. We know that $\mathbf{P}_M$ is a positive recurrent, irreducible and aperiodic transition matrix and, therefore, the distribution of X_n^M converges to π_M in total variation as $n \to \infty$ for every initial distribution of X_0 (Brémaud, 1999: p. 130). In section 2 we saw an analogous result for the segment process of a solution to equation (2).

Assume that the transition probabilities α_M, γ_M are related to some transition rates α, γ by means of a smooth scaling function like (5), for example, such that condition (17) is satisfied. Under this assumption we let the discretization degree M tend to infinity. Assume further that α, γ are functions of the parameters of the reference model, in particular, that $\alpha = \alpha(\sigma)$, $\gamma = \gamma(\sigma)$ are $\mathbf{C}^2$-functions of the noise parameter $\sigma \in (0, \infty)$. The resonance characteristics can then be understood as functions of σ.

Recall that the residence time distribution L_M has a geometric tail in the sense that $L_M(k)$, $k \geq M+1$, renormalized by the factor $(1-\gamma_M) \cdot K_M$ is equivalent to a geometric distribution on $\mathbb{N} \setminus \{1, \dots, M\}$ with K_M as defined by (13). The distribution which L_M induces on $\{1, \dots, M\}$ is given – up to a renormalizing factor – by equations (16b) and (16a). A natural characteristic seems to be the jump in the density of the residence time distribution f_L obtained above. In discrete time, i. e. with discretization degree $M \in \mathbb{N}$, we set

$$\upsilon_M := M \cdot \big(L_M(M+1) - L_M(M) \big). \tag{23}$$

Because of (14), (16a) and (21) we have

$$\upsilon_M \; = \; M \cdot K_M \cdot \big((1-\gamma_M) \cdot \alpha_M - \gamma_M \big), \qquad \upsilon_\infty := \lim_{M \to \infty} \upsilon_M \; = \; K_\infty \cdot (\alpha - \gamma). \tag{24}$$

To consider the height of the discontinuity of f_L as a measure of resonance has already been proposed by Masoller (2003). Following her proposal we define what stochastic resonance means according to the jump characteristic.

Definition 1. Let $M \in \mathbb{N} \cup \{\infty\}$, and suppose that the following conditions hold:

(i) υ_M as a function of the noise parameter σ is twice continuously differentiable,

(ii) $\lim_{\sigma \downarrow 0} \upsilon_M(\sigma) = 0$,

(iii) υ'_M has a smallest root $\sigma_{opt} \in (0, \infty)$.

If υ_M has a global maximum at σ_{opt}, then let us say that the Markov chain X^M or, in case $M = \infty$, the reduced model defined by the family $(X^N)_{N \in \mathbb{N}}$ exhibits *stochastic resonance* and call σ_{opt} the *resonance point*. If υ_M has a global minimum at σ_{opt}, then let us say that the Markov chain X^M (or, in case $M = \infty$, the reduced model) exhibits *pseudo-resonance* and call σ_{opt} the *pseudo-resonance point*.

Alternatively, we may take the probability of transitions in a certain time window as characteristic of the resonance effect. For $M \in \mathbb{N}$ and $q \in (0, 1]$ define

$$\hat{\kappa}_M := \sum_{k=1}^{M} L_M(k), \qquad \kappa_M^{(q)} := \sum_{k=M+1}^{\lfloor (q+1)M \rfloor} L_M(k). \tag{25}$$

By summation over k we see from (14) that

$$\hat{\kappa}_M \; = \; 1 - (1-\gamma_M) \cdot K_M, \qquad \kappa_M^{(q)} \; = \; (1-\gamma_M) \cdot K_M \cdot \big(1 - (1-\alpha_M)^{\lfloor qM \rfloor} \big), \tag{26}$$

and letting M tend to infinity we get

$$\hat{\kappa}_\infty := \lim_{M \to \infty} \hat{\kappa}_M \; = \; 1 - K_\infty, \qquad \kappa_\infty^{(q)} := \lim_{M \to \infty} \kappa_M^{(q)} \; = \; K_\infty \cdot (1 - e^{-q \cdot \alpha}). \tag{27}$$

Recall that M steps in time of the chain X^M or the $\{-1, 1\}$-valued process Y^M correspond to an amount of time r in the reference model. Thus, $\hat{\kappa}_M$ corresponds to the probability of remaining at most time r in one and the same state, while $\kappa_M^{(q)}$ approximates the probability of state transitions occuring in a time window corresponding to $(r, (q+1)r]$ of length $q \cdot r$ given a transition at time zero.

In (25) we could have allowed for a "window width" $q > 1$. The interesting case, however, is a small time window, because then $\kappa_M^{(q)}$ measures the probability of transitions within the second delay interval. For $q = 1$ the two components of our resonance measure correspond to time windows of equal length, that is $\hat{\kappa}_M$ gives the probability of transitions within the first delay interval, while $\kappa_M^{(1)}$ is the probability of hopping events occurring in the second delay interval. Since L_M is geometrically distributed on

$\mathbb{N} \setminus \{1, \ldots, M\}$, $\kappa_M^{(1)}$ majorizes the transition probability for all time windows of the same length starting after the end of the first delay interval. Let us write κ_M for $\kappa_M^{(1)}$.

The idea of the following definition is to maximize quasi-periodicity by finding a noise level such that sojourns in the same state become neither too long nor too short. Here, short sojourns are those that last less than the length of one delay interval, long sojourns those that last longer than the length of two delay intervals. Observe that if the current state of X^M remains the same for more than M steps in discrete time, then the influence of the delay will be constant until a transition occurs.

Definition 2. Let $M \in \mathbb{N} \cup \{\infty\}$, and suppose that the following conditions hold:

(i) κ_M as a function of the noise parameter σ is twice continuously differentiable with values in the unit interval,

(ii) $\lim_{\sigma \downarrow 0} (\hat{\kappa}_M + \kappa_M)(\sigma) = 0$,

(iii) κ_M has a unique global maximum at $\sigma_{opt} \in (0, \infty)$.

If $\kappa_M(\sigma_{opt}) > \hat{\kappa}_M$, then let us say that the Markov chain X^M or, in case $M = \infty$, the reduced model defined by the family $(X^N)_{N \in \mathbb{N}}$ exhibits *stochastic resonance* of strength $\kappa_M(\sigma_{opt})$, and call σ_{opt} the *resonance point*, else let us speak of *pseudo-resonance* and call σ_{opt} the *pseudo-resonance point.*

In the above definition we might have taken a shorter time window than the second delay interval. A natural choice would have been the probability of transitions occuring in a time window corresponding to $(r, (1+q)r]$ normalized by the window width. In the limit $M \to \infty$ we obtain

$$\lim_{q \downarrow 0} \frac{1}{q} \cdot \kappa_\infty^{(q)} \quad = \quad K_\infty \cdot \alpha \quad = \quad f_L(1+). \tag{28}$$

Here, f_L is the density of the residence time distribution from proposition 3 and $f_L(1+)$ is the right-hand limit appearing in equation (21), which gives the height of the discontinuity of f_L.

Of course, definition 2 could be modified in other ways, most importantly by allowing the time window that corresponds to κ_M to float. This would be necessary for a distributed delay. Suppose that in the reference model instead of the point delay we had a delay supported on $[-r, -\delta]$ for some $\delta > 0$. Then a reasonable starting point for a measure of resonance could be a time window of length r with its left boundary floating from δ to r. Notice that a distributed delay (in the reference or in the reduced model) can be chosen in such a way as to render continuous the density f_L.

4 The two state model in continuous time

Our aim in this section is to justify the passage from time discretization degree M to the limit $M \to \infty$. Recall that we proved convergence for the residence times as $M \to \infty$ to reasonable limiting quantities. So the section will discuss limiting processes for which these are the residence times in equilibrium. For $M \in \mathbb{N}$ the Markov chain X^M takes its values in the finite space S_M with cardinality 2^{M+1}. The first thing to be done, therefore, is to choose a common state space for the Markov chains. This will be $D_0 := D_{\{-1,1\}}([-r, 0])$, the space of all $\{-1, 1\}$-valued cadlag functions, i. e. right-continuous functions with left limits, on the interval $[-r, 0]$, endowed with the Skorokhod topology.[6]

Recall how in section 3.1 we partitioned the delay interval $[-r, 0]$. Time step $n \in \{-M, -M+1, \ldots\}$ with respect to the chain X^M was said to correspond to point $n \cdot \frac{r}{M}$ in continuous time. Keeping in mind this correspondence we embed the spaces S_M, $M \in \mathbb{N}$, into D_0, which allows us to look upon the stationary distributions π_M as being probability measures on $\mathcal{B}(D_0)$ and to view the random sequences X^M as being D_0-valued Markov chains.

Now, because of shift condition (6) one may regard X^M as being a process with trajectories in $D_\infty := D_{\{-1,1\}}([-r, \infty))$, the space of all $\{-1, 1\}$-valued cadlag functions on the infinite interval $[-r, \infty)$,

[6] Cf. Billingsley (1999: Ch. 2).

endowed with the Skorokhod topology.[7] If the discretization of time is taken into account, then the chain X^M induces a probability measure on $\mathcal{B}(D_\infty)$ for every initial distribution over $\mathrm{S}_M \subset D_0$.

First, one establishes weak convergence of the stationary distributions or, equivalently, convergence of the π_M with respect to the Prohorov metric induced by the Skorokhod topology d_S° on D_0, then weak convergence of the laws on $\mathcal{B}(D_\infty)$.

The embedding of S_M, the state space of the Markov chain X^M, into D_0 in a sense reverses what one does when approximating solutions to stochastic delay differential equations by Markov chains in discrete time.[8] Approximation results of this kind can be found in Scheutzow (1983, 1984). The method is quite powerful as Lorenz (2003) shows, where weak convergence of the approximating processes to solutions of multi-dimensional autonomous SDDEs is related to a martingale problem that can be associated with the coefficients of the target equation.

Of course, D_0 is a toy space compared to $\mathbf{C}([-r,0],\mathbb{R}^d)$. Notice, however, that linear interpolation as in the case of $\mathbf{C}([-r,0],\mathbb{R}^d)$ is excluded, because the only continuous functions in D_0 are the two constant functions -1 and 1.

Let $M \in \mathbb{N}$, $Z \in \mathrm{S}_M$, and associate with $Z = (Z^{(-M)},\ldots,Z^{(0)})$ a function $f_Z : [-r,0] \to \{-1,1\}$ defined by

$$f_Z(t) := Z^{(0)} \cdot \mathbf{1}_{\{0\}}(t) + \sum_{i=-M}^{-1} Z^{(i)} \cdot \mathbf{1}_{[i\frac{r}{M},(i+1)\frac{r}{M})}(t), \quad t \in [-r,0].$$

Clearly, $f_Z \in D_0$. Hence, $\tilde{\imath}_M : Z \mapsto f_Z$ defines a natural injection $\mathrm{S}_M \hookrightarrow D_0$, which induces the following embedding of probability measures on $\wp(\mathrm{S}_M)$ into the set of probability measures on $\mathcal{B}(D_0)$.

$$\mathcal{M}_+^1(\mathrm{S}_M) \ni \mu \mapsto \tilde{\mu} := \sum_{Z \in \mathrm{S}_M} \mu(Z) \cdot \delta_{f_Z} \in \mathcal{M}_+^1(D_0),$$

where δ_f is the Dirac or point measure concentrated on $f \in D_0$.

Denote by $\tilde{\pi}_M$ the probability measure on $\mathcal{B}(D_0)$ associated with the stationary distribution π_M for the chain X^M, and write $\tilde{X}^M$ for the corresponding D_0-valued Markov chain. Since all we have done so far is a reinterpretation of the state space the results obtained in section 3 regarding X^M are also valid for $\tilde{X}^M$.

Let $M \in \mathbb{N}$. Let ν be a distribution on $\wp(\mathrm{S}_M)$ and denote by P_ν^M the probability measure on $\mathcal{F}_M$ such that X^M is a Markov chain with transition matrix $\mathbf{P}_M$ and initial distribution $X_0^M \overset{d}{\sim} \nu$. For a "point distribution" on $Z \in \mathrm{S}_M$ write P_Z^M.

For $f \in D_0$ let $Z(f)$ be the element of S_M such that $Z^{(i)} = f(\frac{r}{M} \cdot i)$ for all $i \in \{-M,\ldots,0\}$. Let $(Y_n^M)_{n\in\{-M,-M+1,\ldots\}}$ be the sequence of current states of X^M as defined at the beginning of section 3.2. Write

$$\tilde{Y}^M(t) := Y^M_{\lfloor \frac{t}{r} M \rfloor}, \quad t \geq -r.$$

For $A \in \mathcal{B}(D_\infty)$ set

$$\tilde{P}_f^M(A) := P_{Z(f)}^M\big(\tilde{Y}^M \in A\big), \qquad \tilde{P}_M(A) := P_{\pi_M}^M\big(\tilde{Y}^M \in A\big),$$

thereby defining probability measures on $\mathcal{B}(D_\infty)$. Note that $\tilde{P}_f^M$, $\tilde{P}_M$ are well defined and correspond to the distribution of X^M with $X_0^M \overset{d}{\sim} Z(f)$ and $X_0^M \overset{d}{\sim} \pi_M$, respectively.

The proof of convergence follows the usual strategy for this kind of problem. First, we check that the closure of $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$ is compact in $\mathcal{M}_+^1(D_0)$ with respect to the Prohorov topology. Now, (D_0, d_S°) is also complete. According to the Prohorov compactness criterion it is therefore sufficient to show that the set $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$ is tight.

For the second step, choose any limit point $\tilde{\pi} \in \mathcal{M}_+^1(D_0)$ of $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$, which exists according to the first step. It remains to show that $\tilde{\pi}$ is the unique limit point of $(\tilde{\pi}_M)$.

[7] Cf. Billingsley (1999: § 16) or Ethier and Kurtz (1986: Ch. 3 § 5).

[8] Under suitable conditions the approximating time series converge in distribution to the (weakly unique) solution of the SDDE.

Proposition 4. *Suppose the sequences of transition probabilities* $(\alpha_M)_{M\in\mathbb{N}}$, $(\gamma_M)_{M\in\mathbb{N}}$ *satisfy relation* (17) *for some transition rates* $\alpha, \gamma > 0$. *Then there is a probability measure* $\tilde{\pi}$ *on* $\mathcal{B}(D_0)$ *such that* $\tilde{\pi}_M$ *converges weakly to* $\tilde{\pi}$ *as* M *tends to infinity.*

Weak convergence of the sequence $(\tilde{P}_M)$ in $\mathcal{M}^1_+(D_\infty)$ can be demonstrated applying results from semimartingale theory as developed in Jacod and Shiryaev (1987).

Proposition 5. *Suppose scaling relation* (17) *holds. Let* $\tilde{P}_M$, $M \in \mathbb{N}$, *be defined as above, and let* $\tilde{\pi}$ *be the weak limit of* $(\pi_M)_{M\in\mathbb{N}}$ *according to proposition 4. Then there is a probability measure* $\tilde{P} \in \mathcal{M}^1_+(D_\infty)$ *such that* $\tilde{P}_M \xrightarrow{w} \tilde{P}$ *and the restriction of* $\tilde{P}$ *to* $\mathcal{B}(D_0)$ *is given by* $\tilde{\pi}$.

Let $M \in \mathbb{N}$ and $L_M(k)$, $k \in \mathbb{N}$, be the residence time distribution from section 3.2. One first checks that

$$L_M(k) \quad = \quad \tilde{P}_M\big(\xi = \tfrac{r}{M}\cdot k\big), \quad k \in \mathbb{N}, \tag{29}$$

where ξ is the interjump time. Suppose that scaling relation (17) is satisfied. Let $\tilde{P} \in \mathcal{M}^1_+(D_\infty)$ be the weak limit of $(\tilde{P}_M)$ according to proposition 5, and let $\tilde{f}_L$ be as defined by (22). Then it can be shown that $\tilde{f}_L$ is the density of the residence time distribution associated with $\tilde{P}$.

5 Connection between the reduced and the reference model

The aim of this section is to provide a heuristic way of establishing the missing link between our original model, which is given by equation (2), and the reduced model developed in section 3. The situation here is quite similar to the one that was studied by Tsimring and Pikovsky (2001), and we will closely follow their approach in deriving a relation between the transition rates α, γ and the parameters of the original model.

The main ingredient in finding such a relation is the so-called Kramers rate, which gives an asymptotic approximation of the time a Brownian particle needs in order to escape from a parabolic potential well in the presence of white noise only as the noise intensity tends to zero. By means of the Kramers rate we calculate escape rates from potentials that should mirror the "effective dynamics" of solutions to equation (2). The resonance characteristics defined in subsection 3.3 can then be written down explicitly as functions of the noise parameter σ, which allows us to numerically calculate the resonance point and to compare the optimal noise intensity according to the two state model with the behaviour of the original model.

Let $\mathcal{U}$ be a smooth double well potential with the positions of the two local minima at x_{left} and x_{right}, respectively, $x_{left} < x_{right}$, the position of the saddle point at $x_{max} \in (x_{left}, x_{right})$ and such that $\mathcal{U}(x) \to \infty$ as $|x| \to \infty$. An example for $\mathcal{U}$ is the double well potential V from sections 1 and 2. Consider the SDE

$$dX(t) \;=\; -\mathcal{U}'\big(X(t)\big)dt \;+\; \sigma\cdot dW(t), \qquad t \geq 0, \tag{30}$$

where $W(.)$ is a standard one dimensional Wiener process with respect to a probability measure P and $\sigma > 0$ is a noise parameter. Denote by $X^{x,\sigma}$ a solution of equation (30) starting in $X^{x,\sigma}(0) = x$, $x \in \mathbb{R}$. With $y \in \mathbb{R}$ let $\tau_y(X^{x,\sigma})$ be the first time $X^{x,\sigma}$ reaches y, that is we set

$$\tau_y(X^{x,\sigma}) := \inf\{t \geq 0 \mid X^{x,\sigma} = y\}.$$

Since we are interested in the transition behaviour of the diffusion, we need estimates for the distribution of $\tau_y(X^{x,\sigma})$ when x and y belong to different potential wells.

In the limit of small noise the Freidlin-Wentzell theory of large deviations (Freidlin and Wentzell, 1998) allows to determine the exponential order of $\tau_y(X^{x,\sigma})$ by means of the so-called quasipotential $Q(x, y)$ associated with the double well potential $\mathcal{U}$. One may think of $Q(x, y)$ as measuring the work a Brownian particle has to do in order to get from position x to position y. The following transition law holds.

Theorem 1 (Freidlin-Wentzell). *Let Q be the quasipotential associated with $\mathcal{U}$, let $x \in (-\infty, x_{max})$, $y \in (x_{max}, x_{right}]$. Set $q_l := Q(x_{left}, x_{max})$. Then*

$$\lim_{\sigma \downarrow 0} \sigma^2 \cdot \ln\left(\mathrm{E}_P\left(\tau_y(X^{x,\sigma})\right)\right) \quad = \quad q_l, \tag{31a}$$

$$\lim_{\sigma \downarrow 0} P\left(\exp\left(\frac{q_l - \delta}{\sigma^2}\right) < \tau_y(X^{x,\sigma}) < \exp\left(\frac{q_l + \delta}{\sigma^2}\right)\right) \quad = \quad 1 \qquad \textit{for all} \quad \delta > 0. \tag{31b}$$

Moreover, $Q(x_{left}, x_{max}) = 2\big(\mathcal{U}(x_{max}) - \mathcal{U}(x_{left})\big)$. If $x \in (x_{max}, \infty)$, $y \in [x_{left}, x_{max})$ then q_l has to be replaced with $q_r := Q(x_{right}, x_{max})$.

We notice that in travelling from position x in the left potential well to $y \in (x_{max}, x_{right}]$, a position in the downhill part of the right well, the transition time in the limit of small noise is determined exclusively by the way up from position x_{left} of the left minimum to position x_{max} of the potential barrier.

A typical path of $X^{x,\sigma}$, if $\sigma > 0$ is small, will spend most of its time near the positions of the two minima of the double well potential. Typically, the diffusion will reach the minimum of the potential well where it started, before it can cross the potential barrier at x_{max} and enter the opposite well.

Theorem 1 implies the existence of different time scales for equation (30). On the one hand, there is the time scale induced by the Wiener process, where one unit of time can be chosen as $\frac{1}{\sigma^2}$, that is the time it takes the quadratic variation process associated with $\sigma W(.)$ to reach 1. On the other hand, there is the mean escape time given by (31a), which is proportional to $\exp\left(\frac{2L}{\sigma^2}\right)$, where $L > 0$ is the height of the potential barrier. Clearly, with $\sigma > 0$ small, the time scale induced by the white noise is negligible in comparison with the escape time scale.

Moreover, if $\mathcal{U}(x_{left}) \neq \mathcal{U}(x_{right})$, then there are two different heights L_l and L_r for the potential barrier depending on where the diffusion starts. Suppose, for example, that $L_l < L_r$. According to (31b), waiting a time of order $\exp\left(\frac{2L_l + \delta}{\sigma^2}\right)$ with $0 < \delta < 2(L_r - L_l)$ one would witness transitions from the left well to the right well, but no transition in the opposite direction. If the waiting time was of an exponential order less than $\exp\left(\frac{2L_l}{\sigma^2}\right)$, there would be no interwell transitions at all, where "no transitions" means that the probability of a transition occurring tends to zero as $\sigma \to 0$. Thus, by slightly and periodically tilting a symmetric double well potential quasi-periodic transitions can be enforced provided the tilting period is of the right exponential order. This is the mechanism underlying stochastic resonance.

Now, let us suppose that $\tau_y(X^{x,\sigma})$, where $x < x_{max}$ and $y \in (x_{max}, x_{right}]$, is exponentially distributed with rate $\mathrm{r_K} > 0$ such that

$$\mathrm{r_K} \sim \exp\left(-\frac{2\big(\mathcal{U}(x_{max}) - \mathcal{U}(x_{left})\big)}{\sigma^2}\right). \tag{32}$$

Equations (31a) and (31b) of theorem 1 would be fulfilled. In the physics literature it is generally assumed that $\tau_y(X^{x,\sigma})$ obeys an exponential distribution with rate $\mathrm{r_K}$ provided $\sigma > 0$ is sufficiently small. This is known as *Kramers' law*, and $\mathrm{r_K}$ is accordingly called the *Kramers rate* of the respective potential well. It is, moreover, assumed that the proportionality factor missing in (32) can be specified as a function of the curvature of $\mathcal{U}$ at the positions of the minimum and the potential barrier, respectively. The Kramers rate thus reads

$$\mathrm{r_K} = \mathrm{r_K}(\sigma, \mathcal{U}) \quad = \quad \frac{\sqrt{|\,\mathcal{U}''(x_{left})\mathcal{U}''(x_{max})\,|}}{2\pi} \exp\left(-\frac{2|\,\mathcal{U}(x_{left}) - \mathcal{U}(x_{max})\,|}{\sigma^2}\right). \tag{33}$$

Observe that both the assumption of exponentially distributed interwell transition times and formula (33) for the Kramers rate are empirical approximations, where the noise parameter σ is supposed to be sufficiently small.

Well known results for one-dimensional diffusions, extended to the multi-dimensional framework in recent papers by Bovier et al. (2002a,b), show that in the limit of small noise the distribution of the interwell transition time indeed approaches an exponential distribution with a noise-dependent rate that asymptotically satisfies relation (32). The order of the approximation error can also be quantified. For our purposes, however, Kramers' law and the Kramers rate as given by equation (33) will be good enough.

In subsection 3.1 we introduced the transition rates α, γ as being switching rates in the two state model conditional on whether or not the current state agrees with the last remembered state. The idea, now, is to find two "effective" potentials $\mathcal{U}_\alpha$, $\mathcal{U}_\gamma$ such that α is proportional to the Kramers rate describing the escape time distribution from potential $\mathcal{U}_\alpha$, while γ is proportional to the Kramers rate for potential $\mathcal{U}_\gamma$, where the Kramers rate is given by formula (33). More precisely, we must have

$$\alpha = \alpha(\sigma) \;=\; r \cdot \mathrm{r_K}(\sigma, \mathcal{U}_\alpha), \qquad \gamma = \gamma(\sigma) \;=\; r \cdot \mathrm{r_K}(\sigma, \mathcal{U}_\gamma). \tag{34}$$

Note that the inclusion of the delay time r as a proportionality factor is necessary, because in the construction of our two state model we took one unit of time as equivalent to the length of the interval $[-r, 0]$.

There is an important point to be made here. In the discussion of section 3 we assumed that $X(t) \approx 1$ or $X(t) \approx -1$. The error of this approximation is of first order in β, and its contribution to the delay force is proportional to $\beta^2\, U''(1) + \mathcal{O}(\beta^3)$, i. e. of the second order in β. As long as we content ourselves with an approximation of first order in β, two states corresponding to the positions of the minima around -1 and 1 should be enough in order to model the effective dynamics of the reference equation. If we wanted to capture the influence of second order terms in the delay force, we would have to build up a model of four states corresponding to the positions $\pm\underline{x}_\alpha$, $\pm\underline{x}_\gamma$ of the minima of the distorted potential V.[9]

The problem disappears, of course, if U' is constant except on a small symmetric interval $(-\delta, \delta)$ around the origin (see Fig. 1 c), for in this case the delay force would not depend on the particular value of $X(t-r)$ provided $|X(t-r)| \geq \delta$.

Let $L := V(0) - V(1)$ be the height of the potential barrier of V. Set $\rho := |V''(0)V''(1)|$, $\eta := \frac{V'''(1)\,U'(1)}{(V''(1))^2}$, $\tilde{\eta} := \frac{U'(1)}{L}$. Neglecting terms of order higher than one, from (33) we obtain

$$\alpha = \alpha(\sigma) \quad\approx\quad r \cdot \frac{\sqrt{\rho\,(1-\eta\,\beta)}}{2\pi}\, \exp\left(-\frac{2L\,(1-\tilde{\eta}\,\beta)}{\sigma^2}\right), \tag{35a}$$

$$\gamma = \gamma(\sigma) \quad\approx\quad r \cdot \frac{\sqrt{\rho\,(1+\eta\,\beta)}}{2\pi}\, \exp\left(-\frac{2L\,(1+\tilde{\eta}\,\beta)}{\sigma^2}\right). \tag{35b}$$

Recall that the Kramers rate is exact only in the small noise limit. Thus, for the formulae (35a) to become the actual rates of escape it is necessary that σ tends to zero. If the rates α, γ as functions of r and σ are to converge to some finite non-zero values, we must have $\sigma \to 0$ and $r \to \infty$ such that $\frac{1}{\sigma^2}$ and $\ln(r)$ are of the same order. There remain errors due to the first order approximations of V, V'' and U, which make sense only if V, U are sufficiently regular and the delay parameter β is of small absolute value.

In subsection 3.3 we defined two measures of resonance, namely the jump height v_M of the residence time distribution density and the probabilities $\hat{\kappa}_M$, κ_M of transitions within the first and second delay interval, respectively.[10] Recall that $M \in \mathbb{N} \cup \{\infty\}$ is the degree of discretization, where $M = \infty$ denotes the limit $M \to \infty$. We restrict attention to the case $M = \infty$, that is to the two state model in continuous time.

Suppose the transition rates α, γ are functions of the reference model parameters as given by (35a) read as equalities. In particular, α, γ are functions of the delay length r and the noise parameter σ. Let us further suppose that the delay parameter β is of small absolute value, $r > 0$ is big enough so that the critical parameter region for σ lies within the scope of formula (35a), and that the remaining parameters are sufficiently nice.

As a consequence of the exponential form the Kramers rate possesses, we notice that

$$\sqrt{\alpha\gamma} \quad=\quad r \cdot \frac{\sqrt{\rho}}{2\pi} \cdot \sqrt[4]{1-\eta^2\beta^2}\, \exp\left(-\tfrac{2L}{\sigma^2}\right) \quad\approx\quad r \cdot \frac{\sqrt{\rho}}{2\pi}\, \exp\left(-\tfrac{2L}{\sigma^2}\right).$$

[9] Cf. also the numerical results in Curtin et al. (2004).

[10] The jump height measure corresponds to a measure of resonance proposed by Masoller (2003).

In first order of β, the geometric mean $\sqrt{\alpha\gamma}$ of α, γ coincides with the transition rate arising in case $\beta = 0$, that is when there is no delay. Compare this with proposition 3, which states that the residence time density $\tilde{f}_L$ is distributed on the first delay interval according to a mixed hyperbolic sine - cosine distribution with parameter $\sqrt{\alpha\gamma}$.

The conditions of definitions 1 and 2 are satisfied. If $\beta > 0$, then the reduced model exhibits stochastic resonance according to both definitions. According to the jump height measure there is no effect in case $\beta = 0$ and pseudo-resonance in case $\beta < 0$, while the time window measure does not distinguish between $\beta = 0$ and $\beta < 0$, classifying both cases as pseudo-resonance.

Let us specify the potentials V and U according to the model studied by Tsimring and Pikovsky (2001), that is V is the standard quartic potential and U a parabola, see Fig. 1. For the constants appearing in formula (35a) we have

$$L = \tfrac{1}{4}, \qquad \rho = 2, \qquad \eta = \frac{3}{2}, \qquad \tilde{\eta} = 4.$$

With $r = 500$, $\beta = 0.1$, for example, we obtain the resonance point $\sigma_\upsilon \approx 0.32$ according to the jump height measure, while the time window measure yields $\sigma_\kappa \approx 0.29$ with probability $\kappa_\infty(\sigma_\kappa) \approx 0.88$ for transitions occurring in the second delay interval.

Assume β is negative. Again, both measures yield an optimal noise level. With $\beta = -0.1$ we have $\sigma_\upsilon \approx 0.30$ as the noise level that maximizes the jump height in $\tilde{f}_L$. According to the time window measure optimal noise level is at $\sigma_\kappa \approx 0.34$, but $\kappa_\infty(\sigma_\kappa) \approx 0.02$, that is sojourns of duration between r and $2r$ are rare.

Figure 2: Graphs on $[0, 2]$ of the density f_L of the residence time distribution in normalized time. Parameters of the original model: $r = 500$, a) $\sigma = 0.30$, $\beta = 0.1$, b) $\sigma = 0.30$, $\beta = -0.1$, c) $\sigma = 0.35$, $\beta = -0.1$.

There seems to be a discrepancy, now, between the predicted optimal noise level and the level of "most regular" transition behaviour which one would expect from numerical simulation. This is true especially with regard to the jump height measure, the pseudo-resonance point σ_υ being too low.

The problem is that the expected residence time at the level of optimal noise in case $\beta < 0$ is long compared with r. In spite of the fact that long residence times are rare, there is a high probability of finding a solution path remaining in one and the same state for the length of many delay intervals or of witnessing a quasi-periodic transition behaviour break down.

For example, let $\sigma = 0.30$, $\beta = 0.1$. The expected residence time is then about $1.16r$, while with $\sigma = 0.30$ and $\beta = -0.1$ the expected residence time is around $4.62r$. Moreover, with β negative the exponential part of the residence time distribution has a "heavy tail" in the sense that long sojourns receive a relatively high probability, cf. Fig. 2.

These properties of the residence time distribution support the distinction made in definitions 1 and 2 between stochastic resonance and pseudo-resonance.

6 Conclusions and open questions

The main advantage of the two state model which has been our concern for most of this work is that it provides a tool for the analysis of the phenomenon of noise-induced resonance in systems with delay.

The reference model introduced in section 2 is a more elaborate system exhibiting stochastic resonance. Basic features of this model are the extended Markov property and the existence of an invariant probability measure. Both properties carry over to the two state model.

By first studying the two state model in discrete time we obtained an explicit characterization of its stationary distribution. It was thus possible to calculate the residence time distribution which in turn served as starting point for the definition of two simple measures of resonance. The characterization of the stationary distributions in discrete time together with the passage to the time limit also allows to calculate measures of resonance different from those considered here, for example the entropy of a distribution.

In section 5 a heuristic link between the reference and the two state model was outlined. The two state model seems to reliably mirror those aspects of the reference model that are responsible for the phenomenon of stochastic resonance. Observe that we did not show whether the dynamics of the original model in the limit of small noise is reducible to the two state model nor whether the resonance measures considered here are indeed robust under model reduction.

There are different ways in which to proceed. The reference model could be modified, for example, by substituting a distributed delay for the point delay. Clearly, the white noise could be replaced with noise of different type, and higher dimensional equations may be considered.

Lastly, the passage to continuous time as addressed in section 4 should be a special case of more general convergence results for continuous time Markov chains with delay.

References

V. Anishchenko, A. Neiman, F. Moss, and L. Schimansky-Geier. Stochastic resonance: noise-enhanced order. *Physics – Uspekhi*, 42(1):7–36, 1999.

R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. The mechanism of stochastic resonance. *Journal of Physics A*, 14:453–457, 1981.

R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. Stochastic resonance in climatic change. *Tellus*, 34: 10–16, 1982.

P. Billingsley. *Convergence of probability measures.* John Wiley & Sons, New York, 2nd edition, 1999.

A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes I. sharp asymptotics for capacities and exit times. Technical Report 767, Weierstrass-Institut für angewandte Analysis und Stochastik (WIAS), 2002a.

A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes II. precise asymptotics for small eigenvalues. Technical Report 768, Weierstrass-Institut für angewandte Analysis und Stochastik (WIAS), 2002b.

P. Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues.* Springer-Verlag, New York, 1999.

D. Curtin, P. Hegarty, S. D. Goulding, J. Houlihan, T. Busch, C. Masoller, and G. Huyet. Distribution of residence times in bistable noisy systems with time-delayed feedback. February 2004.

S. Ethier and T. Kurtz. *Markov Processes. Characterization and Convergence.* John Wiley & Sons, New York, 1986.

M. Freidlin and A. Wentzell. *Random perturbations of dynamical systems*, volume 260 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, New York, 1998.

L. Gammaitoni, P. Hänggi, P. Jung, and F. Marchesoni. Stochastic resonance. *Reviews of Modern Physics*, 70(1):223–287, 1998.

S. Herrmann, P. Imkeller, and I. Pavlyukevich. Stochastic resonance: non-robust and robust tuning notions. February 2003.

J. Houlihan, D. Goulding, T. Busch, C. Masoller, and G. Huyet. Experimental investigation of a bistable system in the presence of noise and delay. *Physical Review Letters*, 92(5):050601, 2004.

P. Imkeller and I. Pavlyukevich. Model reduction and stochastic resonance. *Stochastics and Dynamics*, 2(4):463–506, 2002.

J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin - Heidelberg - New York, 1987.

B. Larssen. Dynamic programming in stochastic control of systems with delay. *Stochastics and Stochastics Reports*, 74(3-4):651–673, 2002.

R. Lorenz. Weak approximation of stochastic delay differential equations with time series. 2003.

C. Masoller. Distribution of residence times of time-delayed bistable systems driven by noise. *Physical Review Letters*, 90(2):020601, 2003.

B. McNamara and K. Wiesenfeld. Theory of stochastic resonance. *Physical Review A*, 39(9):4854–4869, 1989.

S.-E. A. Mohammed. *Stochastic functional differential equations*. Pitman Publishing, London, 1984.

S.-E. A. Mohammed. Stochastic differential systems with memory: Theory, examples and applications. In L. Decreusefond, J. Gjerde, B. Øksendal, and A. Üstünel, editors, *Stochastic Analysis and Related Topics VI. Papers from the Sixth Oslo-Silivri Workshop on Stochastic Analysis held in Geilo*, Boston, 1996. Birkhäuser.

C. Nicolis. Stochastic aspects of climatic transitions – responses to periodic forcing. *Tellus*, 34:1–9, 1982.

T. Ohira and Y. Sato. Resonance with noise and delay. *Physical Review Letters*, 82:2811–2815, 1999.

I. E. Pavlyukevich. *Stochastic Resonance*. PhD thesis, Humboldt-Universität zu Berlin, 2002.

B. Redmonda, V. LeBlanc, and A. Longtin. Bifurcation analysis of a class of first-order nonlinear delay-differential equations with reflectional symmetry. *Physica D*, 166:131–146, 2002.

M. Scheutzow. *Qualitatives Verhalten der Lösungen von eindimensionalen nichtlinearen stochastischen Differentialgleichungen mit Gedächtnis*. PhD thesis, Fachbereich Mathematik der Universität Kaiserslautern, 1983.

M. Scheutzow. Qualitative behaviour of stochastic delay equations with a bounded memory. *Stochastics*, 12(1):41–80, 1984.

L. Tsimring and A. Pikovsky. Noise-induced dynamics in bistable systems with delay. *Physical Review Letters*, 87(25):250602, 2001.