

## Part A

# Phase-space distribution methods for many-mode quantum mechanics

## Chapter 3

# Generalized quantum phase-space representations

### 3.1 Introduction

In this chapter a formalism encompassing very general phase-space distributions describing quantum states and their evolution is presented. The correspondence made is between a quantum density operator and a distribution of operators, or “kernels”, as in (1.4). The motivation for this is to investigate what distribution properties are essential for tractable mesoscopic first-principles simulations, and to provide a systematic framework in which comparison between methods is simplified. To this end, the following two conditions on the distribution will be kept in mind throughout:

- **Exact correspondence:** A statistical sample of the kernel operators chosen according to the phase-space distribution must approach the exact quantum density matrix in the limit of infinite samples. This property must be maintained during evolution via stochastic equations.
- **Variable number linear in system size:** The number of variables needed to specify a kernel (and hence the number of stochastic equations needed to model quantum evolution) must scale linearly with  $N$ , the number of subsystems.

The formalism is introduced in Section 3.2, and correspondence between density matrix and distribution considered. Subsequently, it is shown how the density matrix corresponds to a statistical sample of kernel operators in Section 3.3, while in 3.4 it is shown how the quantum evolution corresponds to a set of stochastic equations for these. Along the way, examples are given using the positive P representation, commonly used in quantum optics. The resulting requirements on the operator kernels for such a scheme are summarized in Section 3.6.

It is also shown that the phase-space distribution methods have two generic properties that are convenient for practical simulations:

- **Parallel sample evolution:** The individual stochastic realizations of the kernel operator (which are later to be averaged to obtain observables) evolve independently of each other. This allows straightforward and efficient parallel computation (more detail in Section 3.5).
- **All observables in one simulation:** One algorithm is capable of giving estimates for any/all observable averages (see Section 3.3.1).

Some previous generalizations of phase-space representations such as generalized P representations[11], and the (already quite general) discussion of phase-space representations in the article by Drummond and Deuar[56], are contained as special cases of the discussion in this chapter.

## 3.2 Representation of a density matrix

We expand the density matrix

$$\hat{\rho} = \int P(C)\hat{\Lambda}(C)dC, \quad (3.1)$$

as in (1.4), where  $C$  is a set of configuration variables specifying the kernel operator  $\hat{\Lambda}(C)$ . The idea is that if  $P$  is real and positive, it can be considered a probability distribution, and we then can approximate the density matrix using  $\mathcal{S}$  samples  $\hat{\Lambda}(C)$ . Each such sample is fully defined by its set of configuration parameters  $C$ .

### 3.2.1 Properties of the distribution

To be able to interpret the function  $P(C)$  as a probability distribution, we require  $P$  to be

- REAL.
- NON-NEGATIVE:  $P \geq 0$ .
- NORMALIZABLE: i.e.  $\int P(C)dC$  converges to a finite value.
- NON-SINGULAR. The primary reason why a singular distribution is a problem is that it may be incapable of being sampled in an unbiased way by a finite number of samples. For example, if one initially has a non-singular distribution  $P(t)$  but singularities arise through some dynamical process after a time  $t_{\text{sing}}$ , then a set of samples that were unbiased estimators of  $P(t)$  initially ( $t \ll t_{\text{sing}}$ ), will generally be incapable of sampling the singularities at  $t \geq t_{\text{sing}}$  correctly.

Some singular behavior can, however, be tolerated in initial distributions (e.g.  $P(v) = \sum_{j \ll \mathcal{S}} \delta(v - v_0^{(j)})$  for some real variable  $v$ ) provided the number of singularities is finite and much less than the number of samples  $\mathcal{S}$ . This can even be desirable to achieve a starting set of samples that is compact.

If these conditions are satisfied, then when each sample (labeled by  $j$ ) is defined by its own configuration parameters  $C^{(j)}$ ,

$$\lim_{\mathcal{S} \rightarrow \infty} \sum_{j=1}^{\mathcal{S}} \hat{\Lambda}(C^{(j)}) = \hat{\rho}. \quad (3.2)$$

Many well-known distributions of the general form (3.1) do not satisfy these conditions for general quantum states. For example the Wigner distribution[57] is commonly negative in some regions of phase space when the system exhibits nonclassical statistics, the complex P distribution[11] is not real, and the Glauber-Sudarshan P distribution[58, 59] can be singular[60] (also when nonclassical statistics are present). These distributions are often very useful, but not for many-body stochastic simulations.

### 3.2.2 Splitting into subsystems

A mesoscopic system will consist of some number  $N$  of subsystems. In this thesis the subsystems are usually spatial or momentum modes, although other common approaches are to label as subsystems individual particles (if their number is conserved), or orbitals.

A crucial aim, as pointed out at the beginning of this chapter and in Section 1.3, is to have system configurations  $C$  specifying each sample contain a number of parameters (variables) that is only linear in  $N$ . Otherwise any calculations with macro- or mesoscopic  $N$  will become intractable due to the sheer number of variables. Hence, the kernel  $\widehat{\Lambda}$  should be a separable tensor product of subsystem operators<sup>1</sup>:

$$\widehat{\Lambda} = f_{\text{glob}}(C_{\text{glob}}) \otimes_{k=1}^N \widehat{\Lambda}_k(C_k, C_{\text{glob}}). \quad (3.3)$$

Here the *local kernel* for each  $k$ th subsystem is described by its own set of local configuration variables  $C_k$ . There may also be some additional “global” variables (of number  $\ll \mathcal{O}(N)$ ) in the set  $C_{\text{glob}}$  affecting global factors  $f_{\text{glob}}(C_{\text{glob}})$  or several subsystems. The full configuration variable set for the kernel is  $C = \{C_{\text{glob}}, C_1, \dots, C_M\}$ .

A separate issue altogether is the choice of basis (i.e. the choice of  $\widehat{\Lambda}_k$ ) for each subsystem. This chapter aims to stay general, and choosing  $\widehat{\Lambda}_k$  is deferred to Chapter 5 and later, apart from discussing general features of the  $\widehat{\Lambda}$ s necessary for a successful simulation.

### 3.2.3 Dual configuration space of off-diagonal kernels

Here, arguments will be put forward that for non-trivial simulations the local kernels  $\widehat{\Lambda}_k$  are best chosen to include off-diagonal operators.

A density matrix can always be diagonalized in some orthogonal basis, however there are some basic reasons why such a basis is not usually suitable for many-mode simulations of the type considered here.

1. Firstly, a general quantum state will contain entanglement between subsystems, which precludes writing it as a distribution in separable form (3.3) with

---

<sup>1</sup>Strictly speaking,  $\widehat{\Lambda}$  could also be a finite sum of ( $\ll N$ ) tensor products of subsystem operators. This may be useful in some situations.

only diagonal kernels.

2. Secondly, while for some situations the state will, after all, be separable, and could be written using only locally diagonal kernel operators, generally the basis that diagonalizes the density matrix will change with time in a nontrivial way. The kernels, on the other hand, do not change. This means that while the initial separable state could, in this case, be sampled with diagonal local kernels, the subsequent exact quantum evolution could not be simulated using those kernels.
3. Thirdly, exceptions to the above arguments occur if one has a system that remains separable while it evolves. Alternatively, if its inseparability has simple time-evolution that could be found exactly by other means, then this time-evolution could be hardwired into a time-dependent kernel  $\widehat{\Lambda}(t)$  in such a way that (3.3) continues to hold with diagonal (now-time-dependent) local kernels. In such a case, however, there is no point in carrying out time-consuming stochastic simulations of the whole system, when one could just investigate each subsystem separately.

Still, the above arguments are not a rigorous proof, primarily because non-orthogonal basis sets have not been considered. However, no diagonal distribution that can be used to represent completely general quantum states is presently known. For example the Glauber-Sudarshan P distribution[58, 59], which has  $\widehat{\Lambda}_k$  as projectors onto local coherent states (which are non-orthogonal), is defined for all quantum states, but it has been shown that when some nonclassical states are present, the distribution  $P$  becomes singular[60], and not amenable to unbiased stochastic simulations.

In summary, it appears that to allow for off-diagonal entangling coherences between subsystems, the local kernels should be of an off-diagonal form

$$\widehat{\Lambda}_k(C_k, C_{\text{glob}}) = |C'_k, C_{\text{glob}}\rangle \langle \widetilde{C}'_k, C_{\text{glob}}|, \quad (3.4)$$

where the  $C'_k$  and  $\widetilde{C}'_k$  are subsets of *different* independent parameters, and the local parameter set for the kernel is  $C_k = \{C'_k, \widetilde{C}'_k\}$ .

Finally, an exception to this off-diagonal conjecture might be the dynamics of closed systems — which remain as pure states for the whole duration of a simulation. In such a case, one might try to expand the state vector (rather than the density matrix)  $|\psi\rangle$  directly, along the lines of

$$|\psi\rangle = \int P_\psi(C_\psi, \theta) e^{i\theta} |C_\psi\rangle dC_\psi d\theta. \quad (3.5)$$

The global phase factor  $e^{i\theta}$  is required to allow for superpositions with a real positive  $P_\psi$ .

### 3.2.4 An example: the positive P distribution

So that the discussion does not become too opaque, let us make a connection to how this looks in a concrete example, the positive P distribution[10, 11]. This distribution has widely used with success in quantum optics, and with Bose atoms as well[15, 61] in mode-based calculations. It also forms the basis of the gauge P distribution, which will be explained and investigated in Chapter 5 and used throughout this thesis. The positive P local kernel at a lattice point (spatial mode — of which there are  $N$ ) is

$$\widehat{\Lambda}_k = |\alpha_k\rangle_k \langle\beta_k^*|_k \exp\left(-\alpha_k\beta_k + \frac{1}{2}|\alpha_k|^2 + \frac{1}{2}|\beta_k|^2\right), \quad (3.6)$$

where the states  $|\alpha_k\rangle_k$  are normalized coherent states at the  $k$ th lattice point with amplitude  $\alpha_k$ , their form given by

$$|\alpha_k\rangle_k = \exp\left[\alpha_k \widehat{a}_k - \frac{1}{2}|\alpha_k|^2\right] |0\rangle_k. \quad (3.7)$$

Note: the subscript  $k$  on state vectors indicates that they are local to the  $k$ th subsystem, and  $|0\rangle_k$  is the vacuum. The global function is just  $f_{\text{glob}}(C_{\text{glob}}) = 1$ , and the entire kernel can be written in terms of vectors of parameters  $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_N\}$  using

$$|\boldsymbol{\alpha}\rangle = \otimes_{k=1}^N |\alpha_k\rangle_k \quad (3.8)$$

as

$$\widehat{\Lambda} = |\boldsymbol{\alpha}\rangle \langle\boldsymbol{\beta}^*| \exp\left(-\boldsymbol{\alpha} \cdot \boldsymbol{\beta} + \frac{1}{2}|\boldsymbol{\alpha}|^2 + \frac{1}{2}|\boldsymbol{\beta}|^2\right). \quad (3.9)$$

The sets of parameters are  $C_{\text{glob}} = \{\}$ ,  $C_k = \{\alpha_k, \beta_k\}$ . The kernel is in “dual” off-diagonal operator space, and it has been shown[11] that any density matrix can be written with real, positive  $P(C)$  using this kernel. It has also been shown constructively[11] that a non-singular positive P distribution exists, and is given by

$$P_+(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{(4\pi^2)^N} \exp\left(-\frac{1}{4}|\boldsymbol{\alpha} - \boldsymbol{\beta}^*|^2\right) \left\langle \frac{\boldsymbol{\alpha} + \boldsymbol{\beta}^*}{2} \middle| \hat{\rho} \middle| \frac{\boldsymbol{\alpha} + \boldsymbol{\beta}^*}{2} \right\rangle. \quad (3.10)$$

There are four real variables per lattice point.

### 3.3 Stochastic interpretation of the distribution

To efficiently sample the quantum state we make two correspondences. Firstly, as discussed in Section 3.2, the distribution  $P(C)$  corresponds to the density matrix  $\hat{\rho}$ . The second correspondence, as per (3.2), is between the distribution  $P(C)$  that exists in some high-dimensional space, and  $\mathcal{S}$  samples  $C^{(j)}$  distributed according to  $P(C)$ .

#### 3.3.1 Calculating observables

Quantum mechanics concerns itself with calculating expectation values of observables, so the equivalence between it and the stochastic equations for the variables in  $C$  rests solely on obtaining the same evolution of observable averages.

Suppose one wishes to calculate the expectation value of observable  $\hat{O}$  given a set of  $\mathcal{S}$  operator samples  $\{\hat{\Lambda}(C^{(j)})\}$ , with  $j = 1, \dots, \mathcal{S}$ . (Operationally, one actually has the set of  $\mathcal{S}$  operator parameter sets  $\{C^{(j)}\}$ .) For a normalized density matrix, the expectation value is  $\langle \hat{O} \rangle = \text{Tr} [\hat{O}\hat{\rho}]$ , however e.g. in the thermodynamic evolution of Section 2.6 one has un-normalized density matrices, in which case

$$\langle \hat{O} \rangle = \frac{\text{Tr} [\hat{O}\hat{\rho}_u]}{\text{Tr} [\hat{\rho}_u]}. \quad (3.11)$$

Using the representation (3.1) this leads to

$$\langle \hat{O} \rangle = \frac{\int P(C) \text{Tr} [\hat{O}\hat{\Lambda}(C)] dC}{\int P(C) \text{Tr} [\hat{\Lambda}(C)] dC}. \quad (3.12)$$



We can make use of the Hermitian nature of both density matrices and observables to get extra use out of non-Hermitian (off-diagonal) kernels, giving

$$\langle \widehat{O} \rangle = \frac{\int P(C) \left( \text{Tr} \left[ \widehat{O} \widehat{\Lambda}(C) \right] + \text{Tr} \left[ \widehat{O} \widehat{\Lambda}^\dagger(C^*) \right] \right) dC}{\int P(C) \text{Tr} \left[ \widehat{\Lambda}(C) + \widehat{\Lambda}^\dagger(C^*) \right] dC}. \quad (3.13)$$

Actually, one could impose hermiticity on the kernel at the representation level via  $\widehat{\Lambda} \rightarrow \frac{1}{2}(\widehat{\Lambda} + \widehat{\Lambda}^\dagger)$ , but this can complicate the resulting stochastic equations for  $C$  — so, let us impose this only at the level of the observable moments.

When samples are taken  $P(C)$  is interpreted as a probability distribution, and, lastly, noting that the expectation values of observables are real (since  $\widehat{O}$  is Hermitian), one obtains that the expectation value of an arbitrary observable  $\widehat{O}$  can be estimated from the samples by the quantity

$$\bar{O} = \frac{\left\langle \text{Re} \left\{ \text{Tr} \left[ \widehat{O} \widehat{\Lambda} \right] \right\} \right\rangle_{\text{stoch}} + \left\langle \text{Re} \left\{ \text{Tr} \left[ \widehat{O} \widehat{\Lambda}^\dagger \right] \right\} \right\rangle_{\text{stoch}}}{2 \left\langle \text{Re} \left\{ \text{Tr} \left[ \widehat{\Lambda} \right] \right\} \right\rangle_{\text{stoch}}}, \quad (3.14)$$

where stochastic averages over samples are indicated by  $\langle \cdot \rangle_{\text{stoch}}$ . The correspondence<sup>2</sup> is

$$\lim_{\mathcal{S} \rightarrow \infty} \bar{O} = \langle \widehat{O} \rangle. \quad (3.15)$$

### 3.3.2 Assessing estimate accuracy

One expects that the accuracy of the averages in the (3.14) expression using  $\mathcal{S}$  samples will improve as  $\sqrt{\mathcal{S}}$  via the Central Limit Theorem (CLT), since they are just normalized sums over  $\mathcal{S}$  terms. The uncertainty in a mean value  $\bar{v} = \langle v \rangle_{\text{stoch}}$  calculated with  $\mathcal{S}$  samples can then be estimated by

$$\Delta \bar{v} = \sqrt{\frac{\langle v^2 \rangle_{\text{stoch}} - \langle v \rangle_{\text{stoch}}^2}{\mathcal{S}}} \quad (3.16)$$

at the one  $\sigma$  confidence level.

There is, however, a subtlety when several averages over the same variables  $C$  are combined as in (3.14), because the quantities averaged may be correlated.

---

<sup>2</sup>Provided there are no boundary term errors (see Chapter 6).

Then the accuracy estimate (3.16) may be either too large or too small. A way to overcome this is **subensemble averaging**. While the best estimate  $\bar{O}$  is still calculated from the full ensemble  $\mathcal{S}$  as in (3.14), the  $\mathcal{S}$  samples are also binned into  $\mathcal{S}_E$  subensembles with  $s$  samples in each ( $\mathcal{S} = s\mathcal{S}_E$ ). The  $s$  elements of each subensemble are used as in (3.14) to obtain independent estimates of the observable average  $\{\bar{O}^{(1)}, \bar{O}^{(2)}, \dots, \bar{O}^{(\mathcal{S}_E)}\}$ , distributed around  $\bar{O}$ . One has (approximately)

$$\bar{O} \approx \frac{1}{\mathcal{S}_E} \sum_j \bar{O}^{(j)}. \quad (3.17)$$

Given this, the CLT can be applied to the  $\bar{O}^{(j)}$  to estimate the uncertainty in the observable estimate at the one  $\sigma$  level:

$$\Delta\bar{O} = \Delta\langle\hat{O}\rangle = \sqrt{\frac{\frac{1}{\mathcal{S}_E} \sum_j (\bar{O}^{(j)})^2 - \frac{1}{\mathcal{S}_E^2} (\sum_j \bar{O}^{(j)})^2}{\mathcal{S}_E}}. \quad (3.18)$$

A practical issue to keep in mind is that both the number of samples in a subensemble  $s$  and the number of subensembles themselves should be large enough so that both: 1) The subensemble estimates  $\bar{O}^{(j)}$  are reasonably close to the accurate value  $\bar{O} \approx \langle\hat{O}\rangle$  so that (3.17) is true, and also 2) that there is enough of them that the right hand side of (3.17) has an approximately Gaussian distribution, and the CLT can be applied to obtain (3.18).

An issue that arises when  $\bar{O}$  involves a quotient of random variable averages as in (3.14), is that the subensemble size  $s$  should be large enough that the denominator is far from zero for all subensembles. Otherwise subensembles for which this denominator is close to zero have an inordinate importance in the final estimate of the mean. More details in Appendix C.

### 3.3.3 Calculating non-static observables

The observable estimate (3.14) of Section 3.3.1 implicitly assumes that the explicit form of  $\hat{O}$  is known *a priori*. When comparing with experiment, these are usually all the observable averages one needs.

Nevertheless, in many theoretical works some other observables that do not fit this mould are considered. Perhaps the most common example of these is the mutual

fidelity between two states  $\hat{\rho}_1$  and  $\hat{\rho}_2$

$$F(\hat{\rho}_1, \hat{\rho}_2) = \text{Tr} [\hat{\rho}_1 \hat{\rho}_2] = \frac{\text{Tr} [\hat{\rho}_{1u} \hat{\rho}_{2u}]}{\text{Tr} [\hat{\rho}_{1u}] \text{Tr} [\hat{\rho}_{2u}]}, \quad (3.19)$$

which lies between zero (for two orthogonal pure states) and unity for pure  $\hat{\rho}_1 = \hat{\rho}_2$ . A commonly considered special case is the purity  $F_0(\hat{\rho}) = F(\hat{\rho}, \hat{\rho})$ .

A difficulty with evaluating such quantities as  $F_0(\hat{\rho})$  or  $F(\hat{\rho}_1, \hat{\rho}_2)$  on states calculated with stochastic methods is that the density matrices are not known exactly, only their estimates in the form of  $\mathcal{S}$  kernel samples. However, one can still proceed by expanding using (3.1) as  $\hat{\rho}_j = \int P_j(C_j) \hat{\Lambda}_j(C_j) dC_j$  ( $j = 1, 2$ ) to give

$$F(\hat{\rho}_1, \hat{\rho}_2) = \frac{\int P_1(C_1) P_2(C_2) \text{Tr} [\hat{\Lambda}_1(C_1) \hat{\Lambda}_2(C_2)] dC_1 dC_2}{\int P_1(C_1) \text{Tr} [\hat{\Lambda}_1(C_1)] dC_1 \int P_2(C_2) \text{Tr} [\hat{\Lambda}_2(C_2)] dC_2}. \quad (3.20)$$

(Note that the kernels ( $\hat{\Lambda}_1$  and  $\hat{\Lambda}_2$ , respectively) used for representing the two density matrices do not have to be of the same form. The utility of doing so, however, depends entirely on whether the trace of the product of the two kinds of kernels can be evaluated in closed form for arbitrary sets of parameters  $C_1$  and  $C_2$ .)

As before, off-diagonal kernels can be used twice due to the hermiticity of density matrices, giving

$$F(\hat{\rho}_1, \hat{\rho}_2) = \frac{\int P_1(C_1) P_2(C_2) \text{Tr} [(\hat{\Lambda}_1(C_1) + \hat{\Lambda}_1^\dagger(C_1^*)) (\hat{\Lambda}_2(C_2) + \hat{\Lambda}_2^\dagger(C_2^*))] dC_1 dC_2}{\int P_1(C_1) \text{Tr} [\hat{\Lambda}_1(C_1) + \hat{\Lambda}_1^\dagger(C_1^*)] dC_1 \int P_2(C_2) \text{Tr} [\hat{\Lambda}_2(C_2) + \hat{\Lambda}_2^\dagger(C_2^*)] dC_2}. \quad (3.21)$$

The  $P_1(C_1)$  and  $P_2(C_2)$  can be interpreted as probability distributions of *independently realized* configuration samples. Denoting the first set of  $\mathcal{S}_1$  samples as  $\{C_1^{(j_1)}\}$ , and the second set of  $\mathcal{S}_2$  as  $\{C_2^{(j_2)}\}$ , one obtains an estimate of the mutual fidelity between the two states represented by those two sets of samples as

$$\bar{F} = \frac{\sum_{j_1, j_2} \text{Re} \left\{ \text{Tr} \left[ \hat{\Lambda}_1(C_1^{(j_1)}) \left\{ \hat{\Lambda}_2(C_2^{(j_2)}) + \hat{\Lambda}_2^\dagger([C_2^{(j_2)}]^*) \right\} \right] \right\}}{2 \sum_{j_1} \text{Re} \left\{ \text{Tr} \left[ \hat{\Lambda}_1(C_1^{(j_1)}) \right] \right\} \sum_{j_2} \text{Re} \left\{ \text{Tr} \left[ \hat{\Lambda}_2(C_2^{(j_2)}) \right] \right\}} \quad (3.22)$$

(remembering that observable averages are real), with the large sample limit

$$\lim_{\mathcal{S}_1 \rightarrow \infty, \mathcal{S}_2 \rightarrow \infty} \bar{F} = F(\hat{\rho}_1, \hat{\rho}_2). \quad (3.23)$$

Note that the sums in (3.22) are carried out over *all pairs* of samples  $C_1^{(j_1)}$  and  $C_2^{(j_2)}$ .

If the two states whose mutual fidelity is to be calculated arise in the same simulation, then one can ensure independence of the two sets of samples labeled by  $j_1$  and  $j_2$  by setting aside half of all simulated samples for the set  $\{C_1^{(j_1)}\}$ , the other for  $\{C_2^{(j_2)}\}$ . This separation of samples is always necessary if one wants to estimate purity  $F_0$ .

Uncertainty estimates can be obtained in a similar subensemble averaging manner as outlined in Section 3.3.2, by working out (3.22) for each subensemble, then using the CLT to estimate uncertainty in the mean of the subensemble estimates, by the same procedure as in (3.18).

Finally, a major disadvantage of trying to work out fidelity estimates in the above manner from such stochastic simulations is that one must keep all the trajectories in computer memory, so that after all samples have been produced they can be combined in all possible pairs to evaluate the quantities  $\text{Tr} \left[ \widehat{\Lambda}_1(C_1^{(j_1)}) \widehat{\Lambda}_2(C_2^{(j_2)}) \right]$ . Compared to a simulation that only considers static observables as per (3.14), this increases the space required by a factor of  $\mathcal{S}$ .

### 3.3.4 Overcomplete vs. orthogonal bases

Consider observable calculations when using local orthogonal bases for the local kernel operators  $\widehat{\Lambda}_k$ . In a mode formulation, this would imply that the local parameters  $C_k$  consist of discrete quantum numbers for the  $k$ th mode (e.g. occupation), while for a particle formulation  $C_k$  could, for example, consist of (continuous) positions of the  $k$ th particle, since all position eigenstates are orthogonal. Denoting this basis as  $|C'_k\rangle_k$ , and remembering that off-diagonal kernels should be used (see Section 3.2.3), a typical local kernel (omitting global parameters  $C_{\text{glob}}$ ) will have the form  $\widehat{\Lambda}_k = |C'_k\rangle_k \langle \widetilde{C}'_k|_k$  as in (3.4), and the local parameter set is  $C_k = \{C'_k, \widetilde{C}'_k\}$ .

Suppose one wants to calculate the expectation value of a local observable  $\widehat{O} = \widehat{O}_k \otimes_{k' \neq k} \widehat{I}_{k'}$  that is diagonal<sup>3</sup> in this  $|C'_k\rangle_k$  basis. The observable estimate expression

---

<sup>3</sup>Or that contains some components diagonal in the basis  $|C'_k\rangle_k$ .

(3.14) is proportional to averages of quantities like

$$\mathrm{Tr} \left[ \widehat{O} \widehat{\Lambda} \right] = \mathrm{Tr} \left[ \widehat{O}_k \widehat{\Lambda}_k \right] f_{\mathrm{glob}}(C_{\mathrm{glob}}) \prod_{k' \neq k} \mathrm{Tr} \left[ \widehat{\Lambda}_{k'} \right]. \quad (3.24)$$

If *any one* of the local kernel samples  $\widehat{\Lambda}_{k'}$  or  $\widehat{\Lambda}_k$  are non-diagonal, then the whole  $N$ -subsystem sample contributes nothing to the observable estimate, even if the off-diagonal kernel samples are for a different subsystem than the one in which the local observable is being considered ( $k$ ). In a large system, *practically all* samples are likely to have at least one subsystem in which the local kernel is non-diagonal<sup>4</sup>. For example, if a proportion  $p_{\mathrm{od}} \leq 1$  of samples of each subsystem are off-diagonal then only a proportion  $(1 - p_{\mathrm{od}})^N$  will contribute to observable estimates. The result — no reasonable estimate of observables despite lots of calculation. Similar effects occur for off-diagonal local observables (in this case for a sample to contribute all  $k' \neq k$  subsystems must have diagonal kernel samples, while the  $k$ th subsystem sample must have diagonal  $\widehat{O}_k \widehat{\Lambda}_k$ ), and for observables  $\widehat{O} = \otimes_j \widehat{O}_j$  spanning several subsystems (in this case  $\widehat{O}_j \widehat{\Lambda}_j$  must be diagonal for all  $j$ ).

A related issue arises in path integral methods (based on particle positions, for example), and it is the reason that the  $\mathcal{M}$  configurations in one “sample” must form a ring polymer structure ( $C_{(0)} = C_{(\mathcal{M})}$ ) if one wants to calculate diagonal observables. Off-diagonal observables (e.g. momentum distributions in position eigenstate bases) require a separate simulation, which contains some open polymer structures.

This inefficient sampling is the main reason why simulations based on particle positions, or occupation numbers at lattice sites, are not usually successful in phase-space distribution based methods. In both cases the position eigenstates or Fock occupation number states used are orthogonal.

To deal with this sampling problem, one can use overcomplete basis sets (in which the basis vectors are not orthogonal) for the description of the subsystems. Then the local off-diagonal kernels  $\widehat{\Lambda}_k$  can usually be explicitly normalized so that  $\mathrm{Tr} \left[ \widehat{\Lambda}_k \right] = 1$ . The traces averaged in (3.14) become simply

$$\mathrm{Tr} \left[ \widehat{O} \widehat{\Lambda} \right] = \mathrm{Tr} \left[ \widehat{O}_k \widehat{\Lambda}_k \right] f_{\mathrm{glob}}(C_{\mathrm{glob}}), \quad (3.25)$$

---

<sup>4</sup>Apart from some special cases, of course.

and only local kernel parameters  $C_k$  are relevant for the calculation of local observables. (Possibly apart from some global factors, which do not lead to sample contributions decaying exponentially with  $N$ ).

The majority of this thesis considers methods based on coherent state expansions, which are of this kind. A non-orthogonal approach for particle-conserving systems that does use a position wavefunction is to write the density matrix as a distribution over off-diagonal projectors onto orbitals (with arbitrary wavefunctions, not necessarily orthogonal) occupied by all  $N$  particles, as recently developed by Carusotto *et al*[1].

Overcomplete expansions go a long way towards working around the problem, but not all non-orthogonal kernels that one might want to use can be normalized. A case in point are kernels that have zero trace for a set of parameters of measure zero. In such a case one should definitely *not* normalize (at least not fully) as boundary term errors tend to result (see Chapter 6). A partial normalization that avoids any singularities in the resulting kernel (potentially caused by normalizing a kernel whose trace tends to zero) may be successful in such cases. However, the inefficient sampling can then recur. The reason is that there is then an effective “overlap” range in configuration space, such that if the “ket” and “bra” parameters  $C'_k$  and  $\tilde{C}'_k$  differ by more than this range, the resulting kernel will have minimal weight in observable averages in comparison with more diagonal kernels. If the phase-space distributions of  $C'_k$  and  $\tilde{C}'_k$  have spreads significantly wider than the overlap range, a proportion  $p_{\text{od}} \approx \mathcal{O}(1)$  of samples again have largely zero weight in moment calculations, leading again to only  $\approx (1 - p_{\text{od}})^N$  significant samples overall. This situation can be additionally insidious because the weight of the samples is never exactly zero. This leads superficially to finite averages in the observable expression (3.14), however if  $\mathcal{S}(1 - p_{\text{od}})^N \lesssim 1$  there will be rare “very diagonal” configurations that are not sampled at all, but contribute a great deal to the averages in the limit  $\mathcal{S} \rightarrow \infty$ . Conclusion: Care must be taken with un-normalized kernels to make sure no sampling bias is introduced.

Summarizing, barring possible developments that may be able to work around the sampling problem discussed in this section, it appears to be necessary to use

overcomplete basis sets for the local kernels  $\widehat{\Lambda}_k$  to avoid wasting a lot of computing resources on samples that do not end up contributing to observable estimates. It also appears desirable to make the trace of the local kernels normalized to unity if possible.

### 3.3.5 Positive P distribution example continued

Returning to the positive P distribution example of Section 3.2.4, all operators of supported states can be written as linear combinations of the moments of the local annihilation and creation operators  $\widehat{a}_k, \widehat{a}_k^\dagger$ . Thus, to evaluate any observable it suffices to know how to evaluate an expectation value of a general product of the form  $\otimes_j \widehat{a}_{L_j}^\dagger \otimes_k \widehat{a}_{L'_k}$  (The  $L_j$  and  $L'_k$  are subsystem labels (not necessarily unique), while the  $j$  and  $k$  are “subsystem label counters”).

The coherent states forming the overcomplete kernels are eigenstates of the annihilation operators  $\widehat{a}_k$ :

$$\widehat{a}_k |\alpha_k\rangle_k = \alpha_k |\alpha_k\rangle_k, \quad (3.26)$$

and kernels are normalized:

$$\text{Tr} [\widehat{\Lambda}_k] = 1. \quad (3.27)$$

Now the same procedure as in Section 3.3.1 for (3.14) if followed, omitting only taking the real part of the numerator and denominator because  $\otimes \widehat{a}^\dagger \otimes \widehat{a}$  is (for convenience) not necessarily a strictly Hermitian observable. This leads to the expectation value estimate

$$\begin{aligned} \langle \otimes_j \widehat{a}_{L_j}^\dagger \otimes_k \widehat{a}_{L'_k} \rangle &= \lim_{S \rightarrow \infty} \overline{\{L_j, L'_k\}} \\ \overline{\{L_j, L'_k\}} &= \frac{1}{2} \left\langle \prod_j \beta_{L_j} \prod_k \alpha_{L'_k} + \prod_j \alpha_{L_j}^* \prod_k \beta_{L'_k}^* \right\rangle_{\text{stoch}}. \end{aligned} \quad (3.28)$$

Hermitian observables are constructed by combining the operator products and stochastic averages of (3.28) with their adjoints and complex conjugates, respectively.

For the fidelity calculations of Section 3.3.3, one finds

$$\mathrm{Tr} \left[ \widehat{\Lambda}(\boldsymbol{\alpha}_1, \boldsymbol{\beta}_1) \widehat{\Lambda}(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_2) \right] = \prod_k \mathrm{Tr} \left[ \widehat{\Lambda}_k(\alpha_{1k}, \beta_{1k}) \widehat{\Lambda}_k(\alpha_{2k}, \beta_{2k}) \right] \quad (3.29a)$$

$$\mathrm{Tr} \left[ \widehat{\Lambda}_k(\alpha_{1k}, \beta_{1k}) \widehat{\Lambda}_k(\alpha_{2k}, \beta_{2k}) \right] = \exp [ -(\alpha_{1k} - \alpha_{2k})(\beta_{1k} - \beta_{2k}) ] \quad (3.29b)$$

$$\mathrm{Tr} \left[ \widehat{\Lambda}_k(\alpha_{1k}, \beta_{1k}) \widehat{\Lambda}_k^\dagger(\alpha_{2k}^*, \beta_{2k}^*) \right] = \exp [ -(\alpha_{1k} - \beta_{2k}^*)(\beta_{1k} - \alpha_{2k}^*) ] \quad (3.29c)$$

for two kernel samples described by variables  $\alpha_{1k}$ ,  $\beta_{1k}$  and  $\alpha_{2k}$ ,  $\beta_{2k}$  respectively.

The coherent state basis is non-orthogonal:

$$\langle \beta^* | \alpha \rangle = e^{\alpha\beta - \frac{1}{2}|\alpha|^2 - \frac{1}{2}|\beta|^2}, \quad (3.30)$$

and overcomplete:

$$\int |\alpha\rangle\langle\alpha| d^2\alpha = \widehat{I}\pi. \quad (3.31)$$

(Note: the notation  $d^2v = d\mathrm{Re}\{v\} d\mathrm{Im}\{v\}$  is used.)

### 3.3.6 Non-uniqueness of distributions

When kernels with distinct parameters are non-orthogonal, as is the case when using local overcomplete basis sets, a given quantum state  $\widehat{\rho}$  can be described by a whole family of different distributions  $P(C)$ .

A simple example is the vacuum state with the aforementioned positive P distribution. Any density matrix that satisfies

$$\langle 0 | \widehat{\rho} | 0 \rangle = \mathrm{Tr} [\widehat{\rho}] \quad (3.32)$$

must be the vacuum state  $|0\rangle\langle 0|$ , since (3.32) specifies that the population of any non-vacuum states must be zero. Expanding as in (3.1) gives

$$\int P(C) \langle 0 | \widehat{\Lambda}(C) | 0 \rangle dC = \int P(C) \mathrm{Tr} [\widehat{\Lambda}(C)] dC, \quad (3.33)$$

which for the positive P distribution is (using  $\int P(C) dC = 1$ )

$$\int P_+(\boldsymbol{\alpha}, \boldsymbol{\beta}) e^{-\boldsymbol{\alpha}\cdot\boldsymbol{\beta}} d^{2N}\boldsymbol{\alpha} d^{2N}\boldsymbol{\beta} = 1, \quad (3.34)$$



A class of P distributions that satisfy this are Gaussian ensembles of  $\alpha_k = \beta_k$  for each subsystem  $k$ :

$$\begin{aligned} P_+(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \prod_k P_k(\alpha_k, \beta_k), \\ P_{+k}(\alpha_k, \beta_k) &= \delta^2(\alpha_k - \beta_k) \left( \frac{1+2\sigma_k^2}{2\pi\sigma_k^2} \right) e^{-|\alpha_k|^2/2\sigma_k^2} \end{aligned}$$

for arbitrary positive real standard deviations  $\boldsymbol{\sigma} = \{\sigma_1, \dots, \sigma_N\}$ . This can be checked by substitution into (3.34). The most compact such distribution is of course in the limit of  $\boldsymbol{\sigma} \rightarrow 0$ , when

$$P_+(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \delta^{2N}(\boldsymbol{\alpha})\delta^{2N}(\boldsymbol{\beta}). \quad (3.35)$$

One consequence of this is that the variables  $\alpha_k$  and  $\beta_k$  do not necessarily correspond to any physical observable. For example the ‘‘coherent state amplitudes’’  $\alpha_k$  may be highly nonzero, despite representing the vacuum. Only when they appear in the combinations allowed by (3.14) is there a physical meaning. As an example, the estimator of occupation number  $\langle \hat{n}_k \rangle$  is given, from (3.14), by

$$\begin{aligned} \bar{n}_k &= \langle \text{Re} \{ \alpha_k \beta_k \} \rangle_{\text{stoch}} \\ &= \int P_+(\boldsymbol{\alpha}, \boldsymbol{\beta}) \text{Re} \{ \alpha_k \beta_k \} d^{2N} \boldsymbol{\alpha} d^{2N} \boldsymbol{\beta} \\ &= \int P_{+k}(\alpha_k, \beta_k) \text{Re} \{ \alpha_k \beta_k \} d^2 \alpha_k d^2 \beta_k \\ &\propto \int e^{-|\alpha_k|^2/2\sigma_k^2} (\text{Re} \{ \alpha_k \}^2 - \text{Im} \{ \alpha_k \}^2) d^2 \alpha_k \\ &= \int e^{-\text{Re} \{ \alpha_k \}^2/2\sigma_k^2} e^{-\text{Im} \{ \alpha_k \}^2/2\sigma_k^2} (\text{Re} \{ \alpha_k \}^2 - \text{Im} \{ \alpha_k \}^2) d\text{Re} \{ \alpha_k \} d\text{Im} \{ \alpha_k \} \\ &= 0, \end{aligned} \quad (3.36)$$

the last line following from  $\text{Re} \{ \alpha_k \} \leftrightarrow \text{Im} \{ \alpha_k \}$  symmetry. In practice, however, it is not irrelevant which distribution (3.35) is used, because the low- $\boldsymbol{\sigma}$  distributions are much more compact and thus give much better accuracy when calculating observable estimates from a finite sample.

Generalizing, some consequences of the non-uniqueness of distributions based on overcomplete basis sets are expected to be:

- The variables in the sampled set  $C$  will not necessarily correspond to any physical observables. Only when they appear in the observable combinations

allowed by (3.14) is there a physical meaning. Because of this, such distributions are then often said to be in a “non-classical” phase space. (Also because of the number of variables (at least 4) specifying a local off-diagonal kernel (classically only two - e.g. position and momentum)).

- The same state may be represented by many different distributions of the variables. However, some of these will give more precise estimates of observables than others<sup>5</sup>.
- Some distributions representing the same state may be narrower than others in certain directions of phase space, while broader in other directions. This means that a distribution may be better than another for estimating one observable, but worse for a different one.
- Such differences in practical efficiency open the way for various optimizations, which here will be called stochastic gauges. See Chapters 4 and 7.
- It is generally more difficult to obtain an explicit expression for  $P(C)$  given an initial state. If the kernel is orthogonal in a basis  $|C'\rangle$ , thus  $\hat{\Lambda}(C = \{C', \tilde{C}'\}) = |C'\rangle \langle \tilde{C}'|$ , obtaining the unique  $P(\hat{\rho})$  is easy:  $P(C) = \langle C' | \hat{\rho} | \tilde{C}' \rangle$ . In an over-complete basis, one can obtain expressions such as (3.10), but these are usually only one of many possible.

### 3.4 Equations of motion

To simulate the evolution of the state one must make correspondences to the quantum master equation<sup>6</sup>, analogous to those made for the density operator in the previous sections. The master equation results in a differential equation<sup>7</sup> for  $P(C)$ . This can then often be made to correspond to stochastic equations for the configuration variables  $C$ , which specify the evolution of the  $\mathcal{S}$  initial samples.

<sup>5</sup>Given the same number of samples.

<sup>6</sup>Or, possibly different non-Markovian kinds of quantum evolution equations. In this thesis, however, only master equations are considered.

<sup>7</sup>A partial differential equation if  $C$  is continuous, or a set of coupled (possibly partial) differential equations for the various probabilities  $P(C)$  if  $C$  is discrete.

Now, while some equation for  $P(C)$  can always be obtained, in its raw form it is not very useful, because solving it directly for  $P(C, t)$  may be as (or even more) computationally intensive as solving the master equation directly. For example for continuous variables, one would try to evaluate the evolution of  $P(C)$  on a lattice of (say)  $M_{\text{latt}}$  points per variable. Since the number of variables in the set  $C$  is linear in  $N$ , the number of lattice points on which to evaluate the evolution of  $P(C)$  would scale as  $M_{\text{latt}}^N$  — exponential in  $N$ , just like a brute force density matrix element approach.

One needs to make a correspondence between the evolution of  $P(C)$  and stochastic evolution of  $C$ , its samples. This is not in general a trivial (or perhaps even a feasible) exercise, but when  $P(C)$  obeys one of the broad class of Fokker-Planck second order partial differential equations (FPE), there is a well-known “textbook” method[53, 62] to obtain stochastic equations for  $C$ . This is the approach that will be henceforth considered here.

It is worth noting that there may also be ways of obtaining stochastic equations for the configuration variables  $C$  for different kinds of equations. One direction that has been recently investigated by Olsen *et al*[63] are differential equations for  $P(C)$  involving third order partial derivative terms. The resulting equations have stochastic terms with different statistics than the Wiener increments that will be used here. The results are promising but there appear to be serious numerical stability problems.

### 3.4.1 Master equation to Fokker-Planck equation

Let us investigate in more detail what exactly is required to obtain an FPE for  $P$ . These are of the general form

$$\frac{\partial}{\partial t}P(C) = - \sum_j \frac{\partial}{\partial C_j} [A_j(C)P(C)] + \frac{1}{2} \sum_{jk} \frac{\partial^2}{\partial C_j \partial C_k} [D_{jk}(C)P(C)]. \quad (3.37)$$

A point on notation: here, the indices  $j, k$  will label all system configuration variables  $C_j$  consecutively, *not* implying any relationship between  $C_j$  and a  $j$ th subsystem.

The first (rather trivial, but significant) comment is that since an FPE has terms

of the differential form  $\partial P(C)/\partial C_j$ , the system variables  $C_j$  should all be continuous so that partial derivatives can be defined.

Additionally, not all combinations of master equations and operator kernels can lead to Fokker-Planck equations. Let us investigate what the exact requirements are. All master equations consist of terms  $\mathcal{T}_l$  linear in the density operator, and can be written as

$$\frac{\partial \hat{\rho}}{\partial t} = \sum_l \mathcal{T}_l[\hat{\rho}], \quad (3.38)$$

with the  $\mathcal{T}_l$  composed of combinations of system operators  $\hat{A}_j$  and linear in<sup>8</sup>  $\hat{\rho}$ . When the kernel depends on continuous parameters, such system operators and the kernel can often be found to satisfy mutual second-order differential identities of the forms

$$\hat{A}_j \hat{\Lambda}(C) = \left[ A_j^{(0)}(C) + \sum_k A_{jk}^{(1)}(C) \frac{\partial}{\partial C_k} + \sum_{kl} A_{jkl}^{(2)}(C) \frac{\partial}{\partial C_k} \frac{\partial}{\partial C_l} \right] \hat{\Lambda}(C) \quad (3.39a)$$

$$\hat{\Lambda}(C) \hat{A}_j = \left[ \tilde{A}_j^{(0)}(C) + \sum_k \tilde{A}_{jk}^{(1)}(C) \frac{\partial}{\partial C_k} + \sum_{kl} \tilde{A}_{jkl}^{(2)}(C) \frac{\partial}{\partial C_k} \frac{\partial}{\partial C_l} \right] \hat{\Lambda}(C) \quad (3.39b)$$

where all differential operators are henceforth defined to act on the right. (a well-known example is in (3.57) for the positive P representation.) If, however, higher than second order differential terms must also be included, no FPE can be obtained.

Assuming (3.39) hold, by substitution of them into the forms of the  $\mathcal{T}_l(\{\hat{A}_j\}, \hat{\rho})$ , differential expressions for  $\mathcal{T}_l[\hat{\Lambda}]$  can be found (note that now  $\mathcal{T}_l$  acts on the kernel rather than the density matrix itself). Provided that the master equation does not contain terms of too high order in operators for which the differential terms are nonzero, then the operation of each term on the kernel will be able to (after some algebra) be written in a form:

$$\mathcal{T}_l[\hat{\Lambda}] = \left[ T_l^{(0)}(C) + \sum_j T_{lj}^{(1)}(C) \frac{\partial}{\partial C_j} + \frac{1}{2} \sum_{jk} T_{ljk}^{(2)}(C) \frac{\partial}{\partial C_j} \frac{\partial}{\partial C_k} \right] \hat{\Lambda}. \quad (3.40)$$

The factor  $\frac{1}{2}$  in second order terms is introduced for later convenience.

It is crucial that this expansion in partial derivatives terminates at second order for all terms  $\mathcal{T}_l$ , otherwise the final partial differential equation for  $P$  will not be of the

---

<sup>8</sup>Or, equally well, linear in an un-normalized  $\hat{\rho}_u$ .

FPE form (3.37). This usually creates a restriction on the complexity of processes in the model that can be considered. For example in the gauge P representation used in later parts of this thesis, one-particle and two-particle interactions are fine, but three-particle collisions cannot be treated. This restriction can often be surmounted if one increases the complexity of the kernel, at the cost of additional complexity in the equations.

Having confirmed that the identities (3.40) hold with no 3rd or higher derivative terms, the FPE for  $P$  can be derived as follows. Firstly, note that only kernels that are not explicitly time-dependent are being considered here, although time-dependent kernels can also be treatable by a similar, but more voluminous approach. Having said this, the master equation expanded as (3.38) can be written using (3.40) as

$$\int \hat{\Lambda}(C) \frac{\partial P(C)}{\partial t} dC = \sum_l \int P(C) \times \left[ T_l^{(0)}(C) + \sum_j T_{lj}^{(1)}(C) \frac{\partial}{\partial C_j} + \frac{1}{2} \sum_{jk} T_{ljk}^{(2)}(C) \frac{\partial}{\partial C_j} \frac{\partial}{\partial C_k} \right] \hat{\Lambda}(C) dC. \quad (3.41)$$

We now integrate the right hand side by parts, which gives

$$\int \hat{\Lambda}(C) \frac{\partial P(C)}{\partial t} dC = \sum_l \int \hat{\Lambda}(C) \times \left[ T_l^{(0)}(C) - \sum_j \frac{\partial}{\partial C_j} T_{lj}^{(1)}(C) + \frac{1}{2} \sum_{jk} \frac{\partial}{\partial C_k} \frac{\partial}{\partial C_j} T_{ljk}^{(2)}(C) \right] P(C) dC \quad (3.42)$$

*provided that boundary terms can be discarded.* This last is unfortunately not always true, most typically when the distribution  $P(C)$  has power-law tails as the  $|C_j|$  head to infinity. Chapter 6 is devoted to this issue and ways to ensure that these boundary terms are forced to zero using the stochastic gauge technique. Section 6.1.1 derives the exact form of the boundary terms (6.1), which must be zero for the derivation to succeed. Also, there have been developed numerical tests that allow one to check if nonzero boundary terms may be a problem[64].

Now, there may be many  $P(C)$  that satisfy (3.42), but *one* certainly is

$$\frac{\partial P(C)}{\partial t} = \sum_l \left[ T_l^{(0)}(C) - \sum_j \frac{\partial}{\partial C_j} T_{lj}^{(1)}(C) + \frac{1}{2} \sum_{jk} \frac{\partial}{\partial C_k} \frac{\partial}{\partial C_j} T_{ljk}^{(2)}(C) \right] P(C). \quad (3.43)$$

This is almost an FPE, except for the possible terms constant in  $P$ . These typically appear when the normalization of  $\hat{\rho}$  is not conserved. It turns out that in many cases the constant terms are zero, and one can directly obtain a Fokker-Planck equation of the form (3.37), noting the correspondences for the drift vector and diffusion matrices

$$A_j(C) = \sum_l T_{lj}^{(1)}(C) \quad (3.44a)$$

$$D_{jk}(C) = \sum_l T_{ljk}^{(2)}(C). \quad (3.44b)$$

Nonzero constant terms can be easily treated using a gauge method described in Section 4.2.

### 3.4.2 Fokker-Planck equation to stochastic equations

An appropriate set of diffusive stochastic Langevin equations for the variables  $C_j$  is known to be equivalent to the FPE for the distribution  $P(\{C_j\})$  in the limit of infinitely many samples of the variables  $C_j$ . This textbook topic is described in detail e.g. in Gardiner[53, 62]. An important qualification is that this correspondence only holds if the diffusion matrix  $D_{jk}(C)$  when written for real variables  $C_j$  is *positive semidefinite*<sup>9</sup>.

If this is the case, then Langevin equations for the (real)  $C_j$ , equivalent to the FPE (3.37), are

$$dC_j(t) = A_j(C, t) dt + \sum_k B_{jk}(C, t) dW_k(t) \quad (3.45)$$

in the Ito calculus. Here, the *noise matrices*  $B_{jk}$  are any matrices that satisfy the diffusion matrix decomposition

$$D = BB^T, \quad (3.46)$$

i.e. in terms of matrix elements

$$D_{jk} = \sum_l B_{jl} B_{kl}. \quad (3.47)$$

---

<sup>9</sup>i.e. has no eigenvalues with negative real part.

The  $dW_k(t)$  are real stochastic Wiener increments (i.e quantities satisfying the large sample ( $\mathcal{S} \rightarrow \infty$ ) averages)

$$\langle dW_j(t) \rangle_{\text{stoch}} = 0 \quad (3.48a)$$

$$\langle dW_j(t) dW_k(t') \rangle_{\text{stoch}} = \delta_{jk} \delta(t - t') dt^2. \quad (3.48b)$$

For the small but finite time steps  $\Delta t \rightarrow dt$  in a calculation,  $dW_k(t)$  are usually implemented by real Gaussian noises<sup>10</sup> independent for each  $k$  and each time step, having mean zero, and variance  $\Delta t$ . Some more detail of the numerical implementation of stochastic equations with Wiener increments can be found in Appendix B. Lastly, it was chosen to write the equations immediately in difference form, as this is the form in which actual computer calculations are made.

Now, clearly all the coefficients  $A_j$  and  $B_{jk}$  must be real, since the equations are for real variables  $C_j$ . However, the decomposition (3.46) in terms of real  $B$  is only possible if the symmetric diffusion matrix  $D(C)$  is positive semidefinite. This explains the reason for the positive semidefinite condition on  $D$ .

Finally, strictly speaking, a more general set of stochastic equations that remains equivalent to the FPE (3.37) is possible, and will be of use in later parts of this thesis. Consider that only the means and variances (3.48) of the Wiener increments are specified. If one writes (3.45) as

$$dC_j(t) = A_j(C, t) dt + dX_j(C, t) \quad (3.49)$$

then by the properties of Ito stochastic calculus, the only binding relationships for the stochastic terms  $dX_j$  are that

$$\langle dX_j(t) \rangle_{\text{stoch}} = 0, \quad (3.50a)$$

$$\langle dX_j(t) dX_k(t') \rangle_{\text{stoch}} = \langle D_{jk} \rangle_{\text{stoch}} \delta(t - t') dt^2. \quad (3.50b)$$

Indeed,  $dX_j(t)$  need not even be Markovian (i.e. independent of quantities at times  $t' \neq t$ ), provided the conditions (3.50) are satisfied. The relationship between the

---

<sup>10</sup>Other noise distributions are also possible, simply provided the conditions (3.48) are satisfied. Since by the CLT the sum of several infinitesimal noises will always approach an infinitesimal Gaussian-distributed noise anyway, it is usually more efficient to start with a Gaussian noise right away, rather than make one the hard way by summing several non-Gaussian noises.

Fokker-Planck equation (3.37) and stochastic equations (3.49) is considered in a more rigorous fashion and covered exhaustively in the texts by Gardiner[53, 62], but equations (3.45) through to (3.50) will be sufficient for the considerations of this thesis.

### 3.4.3 Ensuring positivity of diffusion for analytic kernels

In the important case of a kernel that can be written as analytically dependent on complex parameters  $z_j$  (rather than on real  $C_j$ )<sup>11</sup>, then we can use the freedom of interpretation of derivatives with respect to complex arguments to choose such a Fokker-Planck equation that the diffusion matrix  $D$  is always positive semidefinite, and Langevin stochastic equations (3.45) can always be derived. The procedure described below is simply an application of the same method used for the special case of a positive P distribution by Drummond and Gardiner[11].

This freedom of derivative choice arises from the property of all analytic functions  $f(z_j)$  that

$$\frac{\partial f(z_j)}{\partial z_j} = \frac{\partial f(z_j)}{\partial z'_j} = -i \frac{\partial f(z_j)}{\partial z''_j}. \quad (3.51)$$

(Where real and imaginary parts of the variable have been denoted for brevity by the shorthand  $z_j = z'_j + iz''_j$ .)

The derivation in Section 3.4.1 follows through formally without change if the real variables  $C_j$  are replaced with the complex  $z_j$ , the set of parameters now being  $C = \{z_j\}$ . Using (3.51) the terms in first order derivatives of the kernel in (3.40) can now be equated to

$$T_{lj}^{(1)}(C) \frac{\partial}{\partial z_j} \hat{\Lambda}(C) = \left[ \text{Re} \left\{ T_{lj}^{(1)}(C) \right\} \frac{\partial}{\partial z'_j} + \text{Im} \left\{ T_{lj}^{(1)}(C) \right\} \frac{\partial}{\partial z''_j} \right] \hat{\Lambda}(C). \quad (3.52)$$

To deal with the second order terms, one can decompose  $\sum_l T_{lj}^{(2)} = D_{jk} = \sum_p B_{jp} B_{kp}$  into the noise matrix forms (which will in general be complex), use the shorthand

---

<sup>11</sup>i.e.  $\hat{\Lambda}$  depends on  $z_j$  but not on any  $z_j^*$ .



$B = B' + iB''$ , and express them (all together) using (3.51) as

$$\begin{aligned} \sum_l T_{ljk}^{(2)}(C) \frac{\partial}{\partial z_j} \frac{\partial}{\partial z_k} \hat{\Lambda}(C) &= \sum_p B_{jp} B_{kp} \frac{\partial}{\partial z_j} \frac{\partial}{\partial z_k} \hat{\Lambda}(C) \\ &= \sum_p \left[ B'_{jp} B'_{kp} \frac{\partial}{\partial z'_j} \frac{\partial}{\partial z'_k} + B'_{jp} B''_{kp} \frac{\partial}{\partial z'_j} \frac{\partial}{\partial z''_k} + B''_{jp} B'_{kp} \frac{\partial}{\partial z''_j} \frac{\partial}{\partial z'_k} + B'_{jp} B''_{kp} \frac{\partial}{\partial z''_j} \frac{\partial}{\partial z''_k} \right] \hat{\Lambda}(C). \end{aligned} \quad (3.53)$$

Let us see what has happened to the diffusion matrix for real variables, as it is *this* that must be positive semidefinite for Langevin equations (3.45) to be obtained. While the indices  $j, k, p = 1, \dots, N_z$  have referred to complex variables (of number  $N_z$ ), let us also define real variables  $v_{j'}$  (with primed indices  $j', k', p' = 1, \dots, 2N_z$ ) by  $z'_j = v_j$  and  $z''_j = v_{(j+N_z)}$ . Denoting the diffusion and noise matrices of the real variables  $v_{j'}$  as  $D^{(v)}$  and  $B^{(v)}$ , respectively, we have  $D_{j',k'}^{(v)} = \sum_{p'=1}^{2N_z} B_{j',p'}^{(v)} B_{k',p'}^{(v)}$ , and from (3.53):

$$B_{j,p}^{(v)} = \operatorname{Re} \{ B_{jp} \}, \quad (3.54a)$$

$$B_{j+N_z,p}^{(v)} = \operatorname{Im} \{ B_{jp} \}, \quad (3.54b)$$

$$B_{j,p+N_z}^{(v)} = B_{j+N_z,p+N_z}^{(v)} = 0, \quad (3.54c)$$

which are all explicitly real. Thus, the diffusion matrix is explicitly positive semidefinite, and Langevin equations for all the  $v_{j'}$  can be obtained.

From (3.45), (3.52), (3.44a), and (3.54) the resulting stochastic equations for the real variables then are

$$dv_j(t) = \operatorname{Re} \{ A_j(C, t) \} dt + \sum_k B_{j,k}^{(v)}(C, t) dW_k(t) \quad (3.55a)$$

$$dv_{j+N_z}(t) = \operatorname{Im} \{ A_j(C, t) \} dt + \sum_k B_{j+N_z,k}^{(v)}(C, t) dW_k(t), \quad (3.55b)$$

which can be written in shorthand form for the complex variables  $z_j$  as

$$dz_j(t) = A_j(C, t) dt + \sum_k B_{jk}(C, t) dW_k(t). \quad (3.56)$$

Note that the Wiener increments  $dW_k$  remain real.

The only requirement (apart from obtaining an FPE) to carry out this procedure, which ensures a stochastic equation interpretation, is that the kernel be written as an analytic function of complex variables.

Finally, the procedure described in this section can be carried out on only a part of the total set of variables  $C$ , if the kernel can be made analytic in only this part. Provided the resulting combined noise matrix  $B^{(v)}$  for all the real variables of the system (real and imaginary parts of the complex variables, and the remaining real variables) is real, the diffusion matrix is positive semidefinite, and Langevin equations for all the variables can still be obtained in the same way as here.

### 3.4.4 Positive P example: dilute lattice Bose gas equations

Continuing with the positive P representation example, let us obtain the stochastic equations corresponding to the quantum dynamics of the lattice interacting Bose gas Hamiltonian (2.17) with single-particle losses to a zero temperature heat bath. The master equation is (2.20) with Linblad operators (2.21), and can be written in terms of  $\hat{\rho}$  and the creation and destruction operators  $\hat{a}_{\mathbf{n}}, \hat{a}_{\mathbf{n}}^\dagger$ , which can be identified with the  $\hat{A}_j$  of Section 3.4.1.

These can be easily verified (via (3.9) and (3.26), and the mutual commutation of local kernels  $\hat{\Lambda}_{\mathbf{n}}$ ) to obey the identities

$$\hat{a}_{\mathbf{n}}\hat{\Lambda} = \alpha_{\mathbf{n}}\hat{\Lambda}, \quad (3.57a)$$

$$\hat{a}_{\mathbf{n}}^\dagger\hat{\Lambda} = \left(\beta_{\mathbf{n}} + \frac{\partial}{\partial\alpha_{\mathbf{n}}}\right)\hat{\Lambda}, \quad (3.57b)$$

$$\hat{\Lambda}\hat{a}_{\mathbf{n}}^\dagger = \beta_{\mathbf{n}}\hat{\Lambda}, \quad (3.57c)$$

$$\hat{\Lambda}\hat{a}_{\mathbf{n}} = \left(\alpha_{\mathbf{n}} + \frac{\partial}{\partial\beta_{\mathbf{n}}}\right)\hat{\Lambda} \quad (3.57d)$$

for all  $N$  lattice points  $\mathbf{n}$ . It can be seen that to obtain an FPE there can be up to two operations of the kind  $\hat{a}^\dagger\hat{\rho}$  or  $\hat{\rho}\hat{a}$  per term in the master equation, and any number of  $\hat{a}\hat{\rho}$  or  $\hat{\rho}\hat{a}^\dagger$ . This allows any one- or two-particle processes but rules out three-particle processes, which contain terms of third order in  $\hat{a}$  or  $\hat{a}^\dagger$  in the Hamiltonian.

The terms in the master equation can be sorted (according to processes) as

$$\mathcal{T}_1[\hat{\rho}] = -i \sum_{\mathbf{nm}} \omega_{\mathbf{nm}} [\hat{a}_{\mathbf{n}}^\dagger \hat{a}_{\mathbf{m}} \hat{\rho} - \hat{\rho} \hat{a}_{\mathbf{n}}^\dagger \hat{a}_{\mathbf{m}}], \quad (3.58a)$$

$$\mathcal{T}_2[\hat{\rho}] = -i\chi \sum_{\mathbf{n}} [\hat{a}_{\mathbf{n}}^\dagger{}^2 \hat{a}_{\mathbf{n}}^2 \hat{\rho} - \hat{\rho} \hat{a}_{\mathbf{n}}^\dagger{}^2 \hat{a}_{\mathbf{n}}^2], \quad (3.58b)$$

$$\mathcal{T}_3[\hat{\rho}] = \sum_{\mathbf{n}} \gamma_{\mathbf{n}} \left[ \hat{a}_{\mathbf{n}} \hat{\rho} \hat{a}_{\mathbf{n}}^\dagger - \frac{1}{2} \hat{a}_{\mathbf{n}}^\dagger \hat{a}_{\mathbf{n}} \hat{\rho} - \frac{1}{2} \hat{\rho} \hat{a}_{\mathbf{n}}^\dagger \hat{a}_{\mathbf{n}} \right]. \quad (3.58c)$$

Which, using (3.57), gives

$$\mathcal{T}_1[\hat{\Lambda}] = -i \sum_{\mathbf{nm}} \omega_{\mathbf{nm}} \left( \alpha_{\mathbf{m}} \frac{\partial}{\partial \alpha_{\mathbf{n}}} - \beta_{\mathbf{n}} \frac{\partial}{\partial \beta_{\mathbf{m}}} \right) \hat{\Lambda}, \quad (3.59a)$$

$$\mathcal{T}_2[\hat{\Lambda}] = -i\chi \sum_{\mathbf{n}} \left( 2\alpha_{\mathbf{n}}^2 \beta_{\mathbf{n}} \frac{\partial}{\partial \alpha_{\mathbf{n}}} - 2\alpha_{\mathbf{n}} \beta_{\mathbf{n}}^2 \frac{\partial}{\partial \beta_{\mathbf{n}}} + \alpha_{\mathbf{n}}^2 \frac{\partial^2}{\partial \alpha_{\mathbf{n}}^2} - \beta_{\mathbf{n}}^2 \frac{\partial^2}{\partial \beta_{\mathbf{n}}^2} \right) \hat{\Lambda}, \quad (3.59b)$$

$$\mathcal{T}_3[\hat{\Lambda}] = -\frac{1}{2} \sum_{\mathbf{n}} \gamma_{\mathbf{n}} \left( \alpha_{\mathbf{n}} \frac{\partial}{\partial \alpha_{\mathbf{n}}} + \beta_{\mathbf{n}} \frac{\partial}{\partial \beta_{\mathbf{n}}} \right) \hat{\Lambda}. \quad (3.59c)$$

Proceeding in the same manner as in Section 3.4.3 (remember that the  $\alpha_{\mathbf{n}}$  and  $\beta_{\mathbf{n}}$  variables are complex), we obtain the (Ito) stochastic equations

$$d\alpha_{\mathbf{n}} = -i \sum_{\mathbf{m}} \omega_{\mathbf{nm}} \alpha_{\mathbf{m}} dt - 2i\chi \alpha_{\mathbf{n}}^2 \beta_{\mathbf{n}} dt - \frac{\gamma_{\mathbf{n}}}{2} \alpha_{\mathbf{n}} dt + i\alpha_{\mathbf{n}} \sqrt{2i\chi} dW_{\mathbf{n}}(t), \quad (3.60a)$$

$$d\beta_{\mathbf{n}} = i \sum_{\mathbf{m}} \omega_{\mathbf{mn}} \beta_{\mathbf{m}} dt + 2i\chi \alpha_{\mathbf{n}} \beta_{\mathbf{n}}^2 dt - \frac{\gamma_{\mathbf{n}}}{2} \beta_{\mathbf{n}} dt + \beta_{\mathbf{n}} \sqrt{2i\chi} d\widetilde{W}_{\mathbf{n}}(t). \quad (3.60b)$$

Here the  $2N$  real Wiener increments  $dW_{\mathbf{n}}(t)$  and  $d\widetilde{W}_{\mathbf{n}}(t)$  are all independent at each time step. We have used the “simplest” diagonal noise matrix decomposition  $B_{\mathbf{nm}} = B_{\mathbf{nn}} \delta_{\mathbf{nm}}$  here. More on the possibilities with these decompositions in Chapter 4.

### 3.5 Convenience for parallel computation

There is no cross-talk between separate realizations of the stochastic process in the simulation scheme described in this chapter. This is a highly desirable property for numerical realization because only one trajectory at a time need be stored in memory. In fact, many independent trajectories could be run at the same time on different computers in a cluster, or even on entirely separate computers, and then collected together at the end. The same applies to calculating observable averages

(3.14), where all one needs is to keep track of the running sum of averaged quantities and their number<sup>12</sup>.

In combination with the linear scaling of variable number this makes for a potentially powerful and efficient combination for many-body simulations.

An exception to this convenient scaling with trajectory number are non-observable quantities such as fidelity, estimated by methods like in Section 3.3.3. In that kind of calculation we have to keep in memory the variables for all  $s$  samples in a subensemble to use all pairs of them in the calculation (3.22). Memory requirements increase by a factor of  $s$  to hold the entire subensemble. The actual calculation of estimates like (3.22) also takes a time  $\propto s^2$  as opposed to  $\propto s$  for static observable estimations (3.14). Since for reasonable subensemble averages one typically needs  $\mathcal{O}(100)$  or more samples to a subensemble, this requires an increase of two orders of magnitude in computing resources — highly nontrivial. Fortunately, at least the linear scaling  $\propto N$  of the number of variables remains untouched.

## 3.6 Summary of representation requirements

This chapter has shown or strongly indicated that for a representation to be useful for quantum simulations of many-body systems, it must satisfy the requirements listed below:

1. **Non-negative real distribution:** To interpret  $P(C)$  as a probability distribution over kernel operators  $\hat{\Lambda}(C)$ , it must be real, normalizable to unity, and non-negative.
2. **Non-singular distribution:** Needed for finite probabilities and un-biased dynamic sampling, although delta functions in the initial state can be tolerated.
3. **Complete:** The representation must be able to describe the initial states, and any states reachable by the subsequent evolution, while satisfying the

---

<sup>12</sup>And possibly the sum of the squares of the averaged properties for assessment of uncertainty in the estimate as per (3.18).

above conditions. In practice this usually implies that the kernel must form a complete or overcomplete basis, and that the kernel be off-diagonal. See Section 3.2.3.

4. **Linear scaling:** The number of system variables in  $C$  must scale linearly with the system size  $N$  (Typically number of modes, particles, orbitals, ...). In practice this requires the kernel to be a tensor product over local kernels  $\widehat{\Lambda}_k$  for each subsystem, or a linear combination of a small ( $\ll N$ ) number of such tensor product terms. See Section 3.2.2.
5. **Most samples significant:** The majority of samples (i.e.  $\mathcal{O}(\mathcal{S})$ ) should contribute significantly to all observable estimates of interest, particularly observables local to the subsystems. This appears to require the use of an overcomplete basis, and kernels normalized locally (i.e.  $\text{Tr} [\widehat{\Lambda}_k] = 1$ ), perhaps apart from some global multiplying functions. These issues are discussed in Section 3.3.4.
6. **Finite kernel trace:** All configurations  $C$  must lead to a finite kernel trace, otherwise observable averages will be undefined for  $\mathcal{S} \rightarrow \infty$  and biased for finite  $\mathcal{S}$  due to divergence of the denominator in (3.14).

If the stochastic interpretation of the quantum evolution is obtained via a Fokker-Planck equation as in Section 3.4 (presently the only approach that has led to nontrivial simulations), then the following requirements must also be satisfied:

7. **Differential equation:** The configuration parameters  $C$  should be continuous to allow a differential equation for  $P(C)$ .
8. **Fokker Planck equation:** The mapping  $\frac{\partial \widehat{\rho}}{\partial t} \rightarrow \frac{\partial P(C)}{\partial t}$  must generate only first and second order partial derivatives of  $P(C)$ . See Section 3.4.1.
9. **Positive propagator:** The short time diffusion in phase space must be non-negative (i.e. the diffusion matrix for real variables  $D^{(v)}$  has no eigenvalues with negative real part), to allow a stochastic interpretation. This can be

guaranteed if the kernel is an analytic function of complex variables (see Section 3.4.3), although in some situations non-analytic kernels may also lead to satisfactory diffusion.

Furthermore, there are several additional requirements that appear generically in simulations, and must be satisfied to avoid bias and/or unmanageable noise in the observable estimates.

10. **Stable trajectories:** If trajectories in phase space diverge rapidly, so does the width of the distribution  $P(C)$ , and all useful precision in the estimates of observables is lost.
11. **Unbiased:** Pathological cases are common (particularly in nonlinear underdamped systems), where the distribution  $P(C)$  is too broad to allow unbiased sampling of some or all observable moments — typically when  $P(C)$  develops power-law tails. These “boundary term” errors are considered in detail in Chapter 6, and usually go in tandem with unstable trajectories.
12. **UV convergent:** Numerical simulations almost always require one to approximate space by a finite lattice. This does not change the physical predictions of the model provided the lattice spacing is much finer than any other significant length scale in the model. This, in turn, requires that the effect of vacuum fluctuations on the equations of motion and observable estimates abates at large momentum (i.e. at small length scales). A mode-based lattice calculation will typically be non-UV-convergent if there are noise terms in the equations that do not disappear at zero mode occupation.