1 Preliminaries

1.1 The class of models

In this section we will define the class of models that will be the primary concern of this work. They take the general form of a phonon term, an electron term and an interaction term,

\[ H = H_{\text{el}} + H_{\text{ph}} + H_{\text{int}}, \]

(1.1)
each of which we will specify in some more detail. This Hamiltonian is a special case of a more general one, as mentioned in the introduction. The derivation (or justification) of these Hamiltonians and in particular the expression of a many-body Hamiltonian in terms of many single body ones can be found in any textbook [10, 11, 13, 30] on solid state physics and will not be repeated here. We will introduce the necessary notational conveniences as we go along. Some generalities will be noted for future reference.

1.1.1 Electrons

The electron term in (1.1) is given by

\[ H_{\text{el}} = \sum_{ij} T_{ij} c_i^\dagger c_j, \]

(1.2)
a quadratic form in the anticommuting fermion annihilation and creation operators which satisfy

\[ \{ c_i^\dagger, c_j \} = \delta_{ij}, \quad \{ c_i^\dagger, c_j^\dagger \} = \{ c_i, c_j \} = 0. \]

(1.3)
Lattice and any other (spin, atomic orbital or band) structure may be hidden in the indices. In this case we split such a multi index as

\[ i = (\tilde{i}, i_m, i_\sigma) \]

(1.4)
into a spatial part \( \tilde{i} \), an (optional) orbital labeling part \( i_m \), and a spin part \( i_\sigma \) (also optional). The total number of lattice sites will be denoted by \( L \). Usually the \( i_m \) run over the labels of the spherical harmonics, and the \( i_\sigma \in \uparrow, \downarrow \).

All physical parameters are hidden in the matrix \( T \). For a typical one-dimensional model with only nearest neighbour interactions

\[ T_{\tilde{i}m, \tilde{j}m} (\mu, t) = -\delta_{ij} \left[ t (\delta_{ij} + \delta_{i+1,j}) + \mu \delta_{ij} \right], \]

(1.5)
with \( \tilde{i} \) the unit vector on the lattice on which the \( i \) live. We leave the vector signs over the indices even in the one dimensional case to distinguish spatial and spin parts. A slightly more general form is obtained by specifying that the spatial dependence of \( T \) can be diagonalized by a Fourier transformation from the real to the reciprocal lattice,

\[ T_{\tilde{i}m, \tilde{j}m} = -\delta_{ij} \left[ \sum_k (\epsilon_k - \mu) \exp \left[ ik \cdot (\tilde{i} - \tilde{j}) \right] \right], \]

(1.6)
and then specifying the spectrum $\varepsilon^*_k$. For $\varepsilon^*_k = -2t \cos 2\pi k/L$ one recovers the tight-binding model (1.5). In the opposite extreme, for $\varepsilon^*_k = \hbar^2k^2/2m$, the free system projected onto a discrete lattice appears.

Equations (1.5) and (1.6) have obvious generalizations to longer range hopping matrix elements, higher dimensions and spin dependencies. The chemical potential $\mu$ always multiplies minus the unit matrix with respect to the indices $i, j$. $T^0$ will denote the matrix $T$ without the chemical potential term, $T = T^0 - \mu \mathbb{I}$. We impose Hermiticity of $H_{el}$, which implies that $T$ should be Hermitian. In the absence of a magnetic field, $T$ can be taken to be real and symmetric. The eigenvalues of $T$ will be denoted by $\varepsilon_k$.

The case of a quadratic form in fermion operators will arise so often in this work that it is useful to introduce a compact 'dot' notation for summation over the fermion–indices $i, j$…

$$A \cdot B \overset{\text{def}}{=} \sum_i A_i B_i.$$  \hfill (1.7)

Now we can write

$$H_{el} = c^\dagger T c.$$  \hfill (1.8)

### 1.1.2 Phonons

The phonon term in the models we consider reads

$$H_{ph} = \sum_i \frac{p_i^2}{2M_i} + \frac{1}{2} \sum_{ij} x_i K_{ij} x_j.$$  \hfill (1.9)

The different masses $M_i$ can always be rescaled, by the canonical transformation

$$
\begin{align*}
p_i &\to p'_i = \frac{p_i}{\sqrt{M_i/M_0}} \\
x_i &\to x'_i = \sqrt{M_i/M_0} x_i \\
K_{ij} &\to K'_{ij} = \frac{M_0 K_{ij}}{\sqrt{M_i M_j}}.
\end{align*}
$$  \hfill (1.10)

after which

$$H_{ph} = \sum_i \frac{p'^2_i}{2M_0} + \frac{1}{2} \sum_{ij} x'_i K'_{ij} x'_j.$$  \hfill (1.11)

We will retain $M_0$ as a control parameter, and drop the primes. Now diagonalizing

$$K_{ij} = \sum_k S^\dagger_{ik} K_k S_{kj}$$  \hfill (1.12)

we put

$$
\begin{align*}
p'_i &\to p''_i = \sum_i p'_i S^\dagger_{ik} \\
x'_i &\to x''_i = \sum_i x'_i S^\dagger_{ik}
\end{align*}
$$  \hfill (1.13)

and arrive at

$$H_{ph} = \frac{1}{2} \sum_k (M_0^{-1} p_k^2 + \kappa_k x''_k x''_k).$$  \hfill (1.14)
1.1 The class of models

Transforming to the creation and annihilation operators

\[ a_k = \frac{1}{\sqrt{2\hbar M_0 \omega_k}} (M_0 \omega_k x_k + ip_k) \]
\[ a_k^\dagger = \frac{1}{\sqrt{2\hbar M_0 \omega_k}} (M_0 \omega_k x_k - ip_k) \]  

(1.15)

where \( M_0 \omega_k^2 = \kappa_k \), we arrive at the familiar diagonal form

\[ H_{ph} = \frac{1}{2} \sum_k \hbar \omega_k (a_k^\dagger a_k + a_k a_k^\dagger) \]  

(1.16)

with the boson operators satisfying

\[ [a_k, a_{k'}^\dagger] = \delta_{kk'}, \quad [a_k^\dagger, a_{k'}^\dagger] = [a_k, a_{k'}] = 0. \]  

(1.17)

The same remarks apply to the indices of these operators as to those of the fermions, although it need not be that the same spatial structure is present in both sets. Additional spin indices are not present in the case of ordinary phonons. We introduce a similar notation, using \( \pi \) rather than \( \omega \) to distinguish between the two sets of indices. Let \( \Omega \) be the diagonal matrix of \( \omega_k \)'s,

\[ H_{ph} = \frac{1}{2} \hbar (a^\dagger \pi \Omega a + a \cdot \Omega \pi a^\dagger) = \hbar (a^\dagger \pi \Omega a + \frac{1}{2} \text{Sp} \Omega). \]  

(1.18)

The symbol \( \text{Sp} \Omega \) denotes the trace of the \( L \times L \) matrix \( \Omega \). We shall not hesitate to mix “dot” and index notations freely when clarity demands.

The limit where the mass \( M_0 \rightarrow \infty \), while keeping the spring constants \( K_{ij} \) constant is called the \textit{adiabatic} limit. Then the phonons become infinitely heavy and the momenta disappear from the Hamiltonian.

1.1.3 Electron–phonon interactions

The interactions between the electrons and the phonons are described by the term

\[ H_{\text{int}} = \sum_{ijk} a_k c_i^\dagger \Lambda_{ij}^{k} c_j + a_k^\dagger c_i^\dagger \Lambda_{ij}^{k} c_j = a \cdot c^\dagger \cdot \Lambda \cdot c + a^\dagger \cdot c^\dagger \cdot \Lambda^\dagger \cdot c. \]  

(1.19)

As the coupling matrix \( \Lambda \) has three indices, one associated with the bosons, and two associated with the fermions, notation does tend to get a little awkward here. We adopt the convention that operations like transposing and taking Hermitian conjugates will only be denoted by the usual symbols \( \Lambda^\dagger \) and \( \Lambda^\dagger \) if meant with respect to the fermion indices.

The physical meaning of the \( \Lambda^\dagger \) can be understood when we consider their Hermitian and anti-Hermitian parts separately,

\[ H_{\text{int}} = \sum_k (a_k^\dagger \Lambda_k + \Lambda_k^t) c + (a_k^\dagger \Lambda_k - \Lambda_k^t) c. \]  

(1.20)

So we see that the Hermitian part of the \( \Lambda^\dagger \) couples the displacement operator \( x_k \) to some density of fermions. Similarly, the anti-Hermitian part of the \( \Lambda^\dagger \) couples the momenta \( p_k \) to
such fermion densities. The latter we will term interactions of the Fröhlich type, the former ones will be called of the Holstein type. Indeed, the model defined by
\[
\lambda_{ij}^k = \frac{\lambda_0 \delta_{ik} \delta_{jk} \delta_{i_r j_r}}{\sqrt{2M_\omega / \hbar}}
\]  
and dispersionless frequencies
\[
\omega_k = \omega_0,
\]
gives the Holstein Model
\[
\mathcal{H}_{\text{Holstein}} = \hbar \omega_0 \sum_k \left( a_k^\dagger a_k + \frac{1}{2} \right) + c^\dagger \cdot \mathbf{T} \cdot c + \lambda_0 \sum_{i\sigma} \chi_i n_{i\sigma} \tag{1.23}
\]

with the local electron densities \( n_i = c_i^\dagger c_i \). As discussed in the introduction, this model was introduced by Holstein[20] (albeit in a different notation) in 1959 to study the 'small polaron' problem. Loosely speaking, a polaron is a lattice distortion tied to an electron. The word 'small' here refers to the wavelength of these lattice distortions (phonons). The Holstein interaction describes how the presence of an electron at a lattice site shifts the equilibrium position of the oscillator. We will use this specific case as our primary example throughout this work.

The model defined by
\[
\lambda_{im}^k = -i \sqrt{\frac{1}{2} M \hbar \omega_k \lambda(\mathbf{k}) e^{i \mathbf{k} \cdot \mathbf{r}_i} \delta_{im} \delta_{j_r m_r}}, \tag{1.24}
\]
with \( k \)-dependent couplings \( \lambda(\mathbf{k}) \sim 1/|\mathbf{k}| \) and dispersionless long wavelength optical phonons with frequency \( \omega_{\text{LO}} \) leads to
\[
\mathcal{H}_{\text{Fröhlich}} = \hbar \omega_{\text{LO}} \sum_k \left( a_k^\dagger a_k + \frac{1}{2} \right) + c^\dagger \cdot \mathbf{T} \cdot c + \sum_{i\sigma} \lambda(\mathbf{k}) \left( e^{i \mathbf{k} \cdot \mathbf{r}_i} - e^{-i \mathbf{k} \cdot \mathbf{r}_i} a_k^\dagger \right) n_{i\sigma}. \tag{1.25}
\]

As discussed in the introduction, (1.25) was introduced by Fröhlich[19] to study the 'large polaron'.

Another special case that should be mentioned is the Su–Schrieffer–Heeger (SSH) model for trans-polyacetylene, a one dimensional chain where the phonons couple to the hopping as described by
\[
\lambda_{ij}^k = \alpha(\delta_{ik} \delta_{kj} + \delta_{ik+1} \delta_{kj+1}). \tag{1.26}
\]
The phonons form a harmonic chain and the total Hamiltonian reads
\[
\mathcal{H}_{\text{SSH}} = -\sum_{i\sigma} \left[ t - \alpha(\chi_{i+1} - \chi_i) \right] \left( c_{i\sigma}^\dagger c_{i+1\sigma} + c_{i+1\sigma}^\dagger c_{i\sigma} \right) + \sum_{i} \left[ \frac{\hbar^2}{2M} + \frac{K}{2}(\chi_{i+1} - \chi_i)^2 \right]. \tag{1.27}
\]

Many of the numerical and some of the analytical methods elaborated on in this thesis carry over with minor modifications to the SSH model (E.g. [39, 40, 41, 43], and references therein).
1.1.4 Hermitian quadratic forms of fermions

We already noticed that hermiticity of \( \mathbf{c}^\dagger \cdot \mathbf{M} \cdot \mathbf{c} \) implies hermiticity of \( \mathbf{M} \), both concepts being defined with respect to multiplication and inproduct on their appropriate spaces. This remark deserves some elaboration. It should be clear that the properties of a quadratic form \( \mathbf{c}^\dagger \cdot \mathbf{M} \cdot \mathbf{c} \) are closely related to, but should be clearly distinguished from, the matrix \( \mathbf{M} \). The former is an operator in a high-dimensional fermionic Fock space[11], eg. for a system of \( \mathbf{L} \) sites with spins and no further atomic-orbital related structure, it has \( 2^{(2\mathbf{s}+1)\mathbf{L}} \times 2^{(2\mathbf{s}+1)\mathbf{L}} \) matrix elements. The matrix \( \mathbf{M} \) is a much simpler thing, operating on a \( (2\mathbf{s}+1)\mathbf{L} \) dimensional vector space over \( \mathbf{R} \) or \( \mathbf{C} \). As a simple example illustrating this difference consider the spectrum of such a matrix \( \mathbf{M} \) which is diagonalized by a transformation \( \mathbf{S} \) and has eigenvalues \( \epsilon_k \),

\[
\mathbf{M}_{ij} = \sum_k S^\dagger_{ik} \epsilon_k S_{kj}.
\]

If we introduce new fermion operators \( \mathbf{c}'_k = (\mathbf{S} \cdot \mathbf{c})_k \) and corresponding number operators \( \mathbf{n}'_k = \mathbf{c}'_k^\dagger \mathbf{c}'_k \), this allows us immediately to diagonalize the quadratic form

\[
\mathbf{c}^\dagger \cdot \mathbf{M} \cdot \mathbf{c} = \sum_k (\mathbf{c}^\dagger \cdot \mathbf{S}^\dagger)_k \epsilon_k (\mathbf{S} \cdot \mathbf{c})_k = \sum_k \epsilon_k \mathbf{n}'_k.
\]

As the \( \mathbf{n}'_k \) commute, we see that the spectrum of the quadratic form consists of all \( 2^{(2\mathbf{s}+1)\mathbf{L}} \) possible sums of the \( (2\mathbf{s}+1)\mathbf{L} \) eigenvalues \( \epsilon_k \) of \( \mathbf{M} \).

Two very useful identities should be mentioned here. Let the elements \( \mathbf{A}_{ij} \) and \( \mathbf{B}_{ij} \) of the matrices \( \mathbf{A} \) and \( \mathbf{B} \) be real or complex numbers, then it takes little effort to show that

\[
\begin{align*}
\left\{ (\mathbf{A} \cdot \mathbf{c})_k, (\mathbf{c}^\dagger \mathbf{B})_j \right\} &= (\mathbf{A} \cdot \mathbf{B})_{ij} \quad (1.28) \\
[\mathbf{c}^\dagger \cdot \mathbf{A} \cdot \mathbf{c}, \mathbf{c}^\dagger \cdot \mathbf{B} \cdot \mathbf{c}] &= \mathbf{c}^\dagger \cdot [\mathbf{A}, \mathbf{B}] \cdot \mathbf{c}. \quad (1.29)
\end{align*}
\]

Again, note that strictly speaking, in this last identity the commutators are defined on completely different spaces. In general, context and notation will leave no ambiguity as to which space is involved in any expression or statement, but a word of caution seems in place.

It is occasionally useful to consider the symmetry of a quadratic fermion form under particle-hole symmetry, by replacing creation operators with annihilation operators and vice versa. In particular

\[
\mathbf{c}^\dagger \cdot \mathbf{M} \cdot \mathbf{c} \pm \mathbf{c} \cdot \mathbf{M} \cdot \mathbf{c}^\dagger = \sum_{ij} \mathbf{M}_{ij} [\mathbf{c}^\dagger_i \mathbf{c}_j \pm \mathbf{c}_i \mathbf{c}^\dagger_j]
\]

\[
= \sum_{ij} \mathbf{M}_{ij} [\mathbf{c}^\dagger_i \mathbf{c}_j \pm \left\{ \mathbf{c}_i \cdot \mathbf{c}^\dagger_j \right\} \mp \mathbf{c}^\dagger_i \mathbf{c}_j]
\]

\[
= \mathbf{c}^\dagger \cdot (\mathbf{M} \mp \mathbf{M}^\dagger) \cdot \mathbf{c} \pm \text{Sp} \mathbf{M}. \quad (1.30)
\]

In the case of a real and symmetric matrix, one of these (+) is trivial, and the other one (−) occurs rather frequently in this work and will get a special notation:

\[
: \mathbf{c}^\dagger \cdot \mathbf{M} \cdot \mathbf{c} : \overset{\text{def}}{=} \frac{1}{2} \sum_{ij} \mathbf{M}_{ij} (\mathbf{c}^\dagger_i \mathbf{c}_j - \mathbf{c}_i \mathbf{c}^\dagger_j) = \mathbf{c}^\dagger \cdot \frac{1}{2} \left[ \mathbf{M} + \mathbf{M}^\dagger \right] \cdot \mathbf{c} - \frac{1}{2} \text{Sp} \mathbf{M}. \quad (1.31)
\]

Like in (1.18) the symbol \( \text{Sp} \mathbf{M} \) denotes the trace of the matrix \( \mathbf{M} \), and we will reserve the symbol \( \text{Tr} \) for traces over state-spaces.
Another important case of the relation between the matrix $A$ and the quadratic form $c^\dagger A c$, is the unitary transformation operator $\exp(\imath c^\dagger A c)$. Differentiating the similarity transform with respect to $\alpha$:

$$\frac{\partial}{\partial \alpha} e^{\imath \alpha c^\dagger A c} c^\dagger e^{-\imath \alpha c^\dagger A c} = \imath e^{\imath \alpha c^\dagger A c} \left[ c^\dagger A c, c^\dagger \right] e^{-\imath \alpha c^\dagger A c} = \imath \sum_k e^{\imath \alpha c^\dagger A c} c^\dagger_k e^{-\imath \alpha c^\dagger A c} A_{kk},$$

provided the elements of the matrix $A$ all commute with each other. That this is a non-trivial requirement will become clear in sections 4.2 and 4.3 where the elements of $A$ contain the phonon operators $a^\dagger$ and $a$. If it is satisfied, then integrating both sides one obtains

$$e^{\imath \alpha c^\dagger A c} c^\dagger e^{-\imath \alpha c^\dagger A c} = c^\dagger e^{\imath \alpha A}.$$  

By restricting the terms of the model (1.1), as we have, to quadratic forms of fermions we have sacrificed generality for tractability in a way that definitely calls for justification.

It is fair to say that the simplicity of these Hermitian quadratic forms in comparison to the class of general operators on a Fock space is what makes the analysis in the rest of this thesis possible. In that respect, the following chapters should justify this course of action. On the other hand, an extensively used Hamiltonian like the one of the Hubbard model, which includes a quartic electron-electron interaction term, is excluded from treatment this way, a fact that for some would invalidate all further effort. Nevertheless, an important relationship between the Hubbard model and (1.1) will be treated in chapter 4.

### 1.1.5 Special cases

Models (1.1) related by transformations of the form

\[
\begin{align*}
a & \rightarrow S \bullet a \\
c & \rightarrow U \cdot c
\end{align*}
\]

naturally group into equivalence classes whose properties are equally easily related. So far no special conditions have been imposed, except Hermiticity of $\mathcal{H}$, which implies in this case, Hermiticity of its three constituting terms $\mathcal{H}_{\text{ph}}$, $\mathcal{H}_{\text{el}}$ and $\mathcal{H}_{\text{int}}$. The latter is Hermitian by its form as two conjugated terms, the other two are only if $\Omega$ (and therefore $K$) and $T$ are, which will be assumed henceforth.

The Holstein, the Fröhlich and the SSH Models have already been mentioned as special cases of the model (1.1). They are special in the sense that the $\Lambda^k$ are all Hermitian in the fermion indices, or can be brought to Hermitian form by a transformation $S$ working only on the phonon operators $a^\dagger_i$. Models like these will be said to satisfy

\[
\text{property A: } \exists S \ \forall i \ (S \bullet \Lambda)^\dagger = (S \bullet \Lambda)^\dagger.
\]

Unless explicitly mentioned, we will restrict ourselves to models which satisfy this property, because either the extension to the more general case is straightforward or it is needlessly complicated.

The Holstein and Fröhlich models are special in another sense: the coupling matrices $\Lambda^k$ all commute for different $i$. It should be clear that no generality is lost if $\Omega$ is taken to be diagonal. However, it is in general impossible to diagonalize the $\Lambda^k$ in the fermion indices for all
**1.2 Thermodynamics**

$k$ simultaneously, let alone simultaneously with $T$. If the latter is possible, the model becomes entirely trivial, as it can be written as the sum of non-interacting "sites". The former will be referred to throughout as

\[ \text{property } B: \quad \forall i, j \quad [\Lambda^i, \Lambda^j] = [\Lambda^i, \Lambda^{ij}] = 0. \]  

(1.35)

This property, like the previous is unaffected by similarity transformations of the phonon part of the system.

### 1.2 Thermodynamics

**Expectation values**

For a system described by a Hamiltonian $\mathcal{H}$, one can write down a partition function $Z$ and a free energy $^1 F$ defined by\[2, 3]\]

\[ Z \overset{\text{def}}{=} \text{Tr} \ e^{-\beta \mathcal{H}}, \]

(1.36)

\[ F \overset{\text{def}}{=} -\beta^{-1} \log \text{Tr} \ e^{-\beta \mathcal{H}} \]

(1.37)

with the trace $\text{Tr}$ interpreted in the quantum–mechanical way as the sum over all diagonal elements, and $\beta$ the inverse temperature, converted to units of energy with the Boltzmann constant $k_B$.

\[ \beta = 1/k_B T. \]  

(1.38)

The operator $Z^{-1} e^{-\beta \mathcal{H}}$, called the thermodynamical density operator, plays the rôle of the quantum–mechanical equivalent of the Boltzmann weight, to define the average of an operator $\Lambda$ as

\[ \langle \Lambda \rangle \overset{\text{def}}{=} Z^{-1} \text{Tr} \ e^{-\beta \mathcal{H}} \Lambda. \]  

(1.39)

The free energy, when considered as a function of the appropriate variables, is a generating function for all possible sorts of correlation functions, which describe the behaviour of a many-body system $\mathcal{H}$ in thermal equilibrium. Eg.

\[ -\beta^{-1} \left. \frac{\partial}{\partial \lambda} \right|_{\lambda=0} \log \text{Tr} \ e^{-\beta \mathcal{H} + \lambda \Lambda} = \langle \Lambda \rangle. \]  

(1.40)

The foundations of these formulae, and the basis of Thermodynamics and Statistical Mechanics is a highly non–trivial topic on which I have nothing to say. I will consider the rules and the object of the game as fixed: find the free energy and its derivatives for a given system in order to find out its thermodynamical behaviour. Unfortunately, the three terms in (1.1) in general do not commute, which makes it non–trivial to diagonalize $\mathcal{H}$. Yet it is in general impossible to calculate the matrix elements of $e^{-\beta \mathcal{H}}$, let alone the sum over all diagonal matrix elements, if one doesn't know how to diagonalize $\mathcal{H}$. Three methods to tackle this are referred to in this work and will be briefly introduced here.

\[^1\text{We will use the same symbols for the free energy of the canonical ensemble and the grand potential of the grand canonical ensemble, introduced in chapter 2.}\]
Approximation methods

One option to circumvent the problem is perturbation theory: write the Hamiltonian as

$$\mathcal{H} = \mathcal{H}_0 + \lambda \Lambda,$$

(1.41)

where $\mathcal{H}_0$ can be diagonalized easily and the matrix elements of $\Lambda$ can be calculated in the eigenbasis of $\mathcal{H}_0$. A Taylor expansion of $F$ and its derivatives in the 'small' parameter $\lambda$ then gives rise to integrals for which nifty diagrammatic techniques have been developed and borrowed from field theory. Unfortunately, this expansion need not converge to reality. For example, if $\lambda$ drives a phase transition, the series shouldn't converge near the transition, as a phase transition is a discontinuity in $F$ or its derivatives.

Another approach is based on a variational principle. Let $\mathcal{H}_0$ be some trial Hamiltonian, depending upon some parameters $\alpha$ and denote the thermodynamic average with respect to this Hamiltonian by $\langle \cdot \rangle_0$ etc. Then Jensen's inequality[5] gives

$$F = -\beta^{-1} \log \text{Tr} \ e^{-\beta \mathcal{H}}$$

$$= -\beta^{-1} \log \frac{\text{Tr} \ e^{-\beta \mathcal{H}_0} e^{-\beta \Lambda}}{\text{Tr} \ e^{-\beta \mathcal{H}_0}}$$

$$\leq -\beta^{-1} \log \langle \ e^{\beta \mathcal{H}_0} e^{-\beta \mathcal{H}} \rangle_0 + F_0(\alpha)$$

$$= \langle \mathcal{H} - \mathcal{H}_0 \rangle_0 + F_0(\alpha).$$

(1.42)

At zero temperature $\langle \mathcal{H}_0 \rangle_0 = F_0$, $F = \langle \mathcal{H} \rangle$, and minimization of (1.42) with respect to the parameters $\alpha$ is equivalent to a standard variational principle, with an 'implicit' trial state: the ground state of the trial Hamiltonian.

The idea is now that the ground state of a $\mathcal{H}_0(\alpha)$ giving a good upper bound to the ground state energy of $\mathcal{H}$, is a good approximation to the ground state of $\mathcal{H}$. There is no mathematical necessity involved in this argument whatsoever. In particular, in the case of a highly degenerate groundstate this procedure may be rather misleading[42].

A third approach, is based on the Trotter–Suzuki Formula[1, 26]. Let $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$, then

$$e^{-\beta \mathcal{H}} = \lim_{m \to \infty} \left( e^{-\beta \mathcal{H}_1/m} e^{-\beta \mathcal{H}_2/m} \right)^m,$$

(1.43)

which may be generalized to decompositions into more than two terms. It has been shown[8] that for finite $m$ the following bound applies.

$$\| e^{-\beta \mathcal{H}_1 + \mathcal{H}_2} - \left( e^{-\beta \mathcal{H}_1/m} e^{-\beta \mathcal{H}_2/m} \right)^m \|_\infty$$

$$\leq \frac{\beta^2}{2m} \| \mathcal{H}_1, \mathcal{H}_2 \|_\infty \exp \left[ \beta (\| \mathcal{H}_1 \|_\infty + \| \mathcal{H}_2 \|_\infty) \right],$$

(1.44)

where the $\| A \|_\infty$ is the matrix infinity norm, equal to the largest absolute eigenvalue of $A$. In fact, one can show that even for zero temperature, i.e. $\beta \to \infty$, with $\beta/m = \tau = \text{cst.}$, the error in the approximant is of order $\tau^3$. This can be further improved by symmetrizing the product as

$$e^{-\beta \mathcal{H}_1/2m} e^{-\beta \mathcal{H}_2/m} e^{-\beta \mathcal{H}_1/2m},$$
which is correct to second order in $\tau = \beta/m$. The applicability of (1.43) depends on the possibility to choose the decomposition in such a way that the individual terms are easily diagonalized.

The above methods can be applied to calculate various static and dynamic correlation functions, that yield valuable information on the physical properties of the model. The static correlation functions describe the ordering present in a certain thermodynamic state. The dynamic correlation functions are related to transport properties. The specific form of these functions, why they are of interest, and their derivation belong to the domain of Kubo's Dynamic Response theory[12]. Further introduction of these functions will be postponed until their use in the subsequent chapters.

**Finite systems**

Many textbook results concern infinite systems with finite density of particles, the so-called thermodynamic limit. Obviously, for numerical simulations in real space, like those introduced in the next chapter, this is not feasible, but this need not hamper the applicability of the theory to be developed in the remainder of this work. First of all it turns out that in many cases (especially in low dimension) the behaviour of large systems is already reached for pretty small systems (but we will also discover a huge finite size effect in section 2.4.3). Secondly, many real systems are not infinite, not even approximately. Polymer chains and mesoscopic lumps may contain as few as some tens of atoms. The presence of finite size effects does not mean that these systems are not well described by statistical physics.

### 1.3 Monte Carlo methods

Expectation values in statistical physics, as well as in a host of other branches of science lead to integrals that more often than not are not solvable in closed form, but are only amenable to numerical methods. A very well known, very effective and very elegant method for a rather broad variety of problems of this sort is known as "Monte Carlo Integration".

The central idea behind a "Monte Carlo" method is the following: The expectation value of a function $f$ under a proper weight function $w \geq 0$,

$$\langle f \rangle_w \triangleq \frac{\int w(x)f(x)dx}{\int w(x)dx}$$

may be obtained by a summation of $f$ over a set of $N$ points $x_i \in MC$

$$\langle f \rangle_w \approx N^{-1} \sum_{x_i \in MC} f(x_i) \quad (1.46)$$

where the points are distributed as $w$, ie. the probability for a point from a neighbourhood $X$ to be in the set $MC$, must be equal to $\int_X w(x)dx/\int w(x)dx$. In fact one shows that the error in (1.46) vanishes as $1/\sqrt{N}$. This is nothing but the Central Limit theorem[15] applied to the case of statistical sampling: average a sample to sample an average.

If the set $MC$ is interpreted as the trajectory of a Markov process $P(x \rightarrow x')$, this process should have $w$ as it's equilibrium distribution for (1.46) to hold. The problem is now to construct a Markov process, preferably one that is simulated efficiently on a computer, that has the desired
equilibrium distribution. Of course,
\[ P(x \rightarrow x') = \frac{w(x')}{\int w(x'')dx''}, \]
(1.47)
independent of \( x \) is always a solution. As a simple example, consider \( w \) to be the uniform distribution on a \( d \)-dimensional hypercube \( \mathbb{D}, 1 \)^\( d \). An obvious solution is \( P(x \rightarrow x') = 1 \), i.e. for any point \( x \), go to another point \( x' \) by uniformly random picking it from the hypercube.

In general however, the denominator in (1.47) is unknown. The advantage of the generality for any point \( x \) and that it can be used without explicitly calculating \( \int w(x'')dx'' \).

A distribution \( w \) is the equilibrium distribution of a Markov process \( P \) if and only if the condition is satisfied that:
\[ \forall x \sum_{x'} w(x)P(x \rightarrow x') = \sum_{x'} w(x')P(x' \rightarrow x). \]
(1.48)
The interpretation is intuitive: \( w(x)P(x \rightarrow x') \) is the amount of probability flowing from \( x \) to \( x' \). If the influx of probability at all points equals the outflux, nothing changes. In practice the easiest way to fulfill this is by making sure the sufficient condition of detailed balance is satisfied, which is nothing but termwise equality of the sums in (1.48):
\[ \forall x, x' \quad w(x)P(x \rightarrow x') = w(x')P(x' \rightarrow x). \]
(1.49)
Again the interpretation is intuitive: if between any two points \( x \) and \( x' \) the flux of probability one way is equal to that the other way, then (1.48) is satisfied.

In the case that \( w(x) \) is a Boltzmann weight \( \exp [-\beta E(x)] \), equation 1.49 leads to
\[ \frac{P(x \rightarrow x')}{P(x' \rightarrow x)} = \frac{w(x')}{w(x)} = \exp \{-\beta [E(x') - E(x)]\}. \]
(1.50)
If we take
\[ P(x \rightarrow x') = \min[1, \exp [-\beta E(x') - E(x)]] \]
(1.51)
then (1.49) will be satisfied. We see that a step from \( x \) to \( x' \) that lowers the energy is always taken, whereas an increase in energy is only accepted with a probability equal to the relative Boltzmann weight. This method was invented by Metropolis et.al.[4], and is called Metropolis Monte Carlo (MMC). A sequence \( x_t \) that satisfies (1.51) is generated by repeated execution of the following pseudo code:
\[ \begin{align*}
    x_{\text{try}} &\leftarrow x_t + \text{RND}([-\delta, \delta]^d) \\
    \text{if} &\quad \exp \{-\beta [E(x_{\text{try}}) - E(x_t)]\} > \text{RND}([0, 1]) \\
    \text{then} &\quad x_{t+1} \leftarrow x_{\text{try}} \\
    \text{else} &\quad x_{t+1} \leftarrow x_t
\end{align*} \]
(1.52)
Here, \( \text{RND} \) is a routine to draw a pseudo-random number uniformly from the set in its argument, and \( d \) is the dimension of \( x \)'s domain.

If the inverse temperature \( \beta \) is increased to infinity, the trial state will only be accepted if it has a lower energy than the current state, and the entire algorithm becomes a trial-and-error walk towards the ground state. In terms of computational efficiency this may not be the
optimal method to find the ground state, in which case steepest descent (SDM) or conjugate
gradient (CGM) methods may prove useful. All three methods, MMC, SDM, and CGM are
not foolproof. In particular, they may end up in a local minimum. The only way to check this
is to check the robustness of results under different initial conditions.

As a final thought on the subject of this section, I would like to mention that many ar-
guments leading to the postulation of the notion of an ensemble in statistical physics, state
(or paraphrase) in some way that a many particle system in a heat bath is a complicated
process[2], which, if properly described (eg. by equations of motion) satisfies the Markov prop-
erty. Frequently such arguments proceed by stating that not the trajectories of the equations of
motion but the properties of the equilibrium distribution are the proper object of investigation.
When we look upon the matter from this angle, we may say that a Monte Carlo simulation of
a system that has some underlying equations of motion, is nothing but replacing one Markov
process by another one that is easier to simulate, but has the same equilibrium properties.