



Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: <http://www.tandfonline.com/loi/tmph20>

Maximum probability domains for Hubbard models

Guillaume Acke, Stijn De Baerdemacker, Pieter W. Claeys, Mario Van Raemdonck, Ward Poelmans, Dimitri Van Neck & Patrick Bultinck

To cite this article: Guillaume Acke, Stijn De Baerdemacker, Pieter W. Claeys, Mario Van Raemdonck, Ward Poelmans, Dimitri Van Neck & Patrick Bultinck (2016): Maximum probability domains for Hubbard models, *Molecular Physics*, DOI: [10.1080/00268976.2016.1153742](https://doi.org/10.1080/00268976.2016.1153742)

To link to this article: <http://dx.doi.org/10.1080/00268976.2016.1153742>



Published online: 29 Feb 2016.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Maximum probability domains for Hubbard models

Guillaume Acke^a, Stijn De Baerdemacker^{a,b}, Pieter W. Claeys^b, Mario Van Raemdonck^a, Ward Poelmans^b, Dimitri Van Neck^b and Patrick Bultinck^a

^aDepartment of Inorganic and Physical Chemistry, Ghent University, Ghent, Belgium; ^bCenter for Molecular Modeling, Ghent University, Zwijnaarde, Belgium

ABSTRACT

The theory of maximum probability domains (MPDs) is formulated for the Hubbard model in terms of projection operators and generating functions for both exact eigenstates as well as Slater determinants. A fast MPD analysis procedure is proposed, which is subsequently used to analyse numerical results for the Hubbard model. It is shown that the essential physics behind the considered Hubbard models can be exposed using MPDs. Furthermore, the MPDs appear to be in line with what is expected from Valence Bond (VB) Theory-based knowledge.

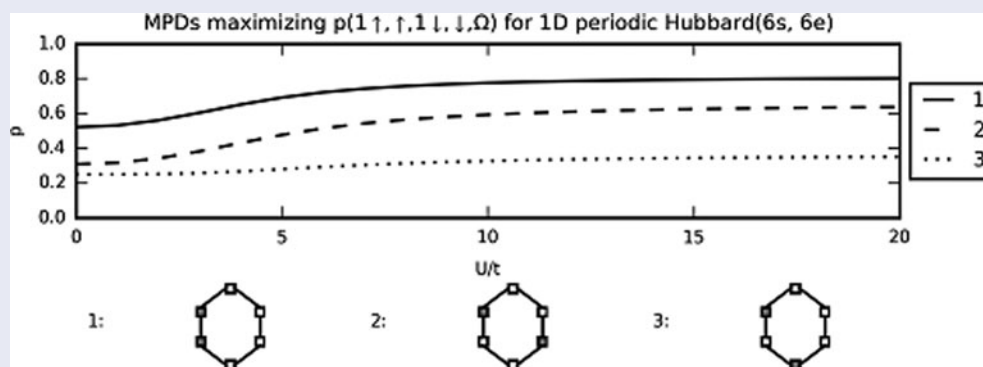
ARTICLE HISTORY

Received 9 November 2015

Accepted 3 February 2016

KEYWORDS

Maximum probability domains; Hubbard model; generating function; Hubbard benzene; Hubbard 1,3,5-hexatriene



1. Introduction

Chemists often rely on a set of concepts that are not directly derivable from quantum mechanical observables [1]. As such, these tools are not uniquely defined from first principles, and are highly prone to user bias. Nevertheless, despite the inherent arbitrariness behind them, these tools still constitute a major part of chemical theory as they have proven their value over and over again for decades.

One of these tools is the theory of Lewis structures, which was proposed by Lewis in 1916 [2], only 20 years after the discovery of the electron and 10 years before Schrödinger's seminal paper [3]. The visionary character of Lewis can hardly be overestimated as the far majority of electronic structures, especially in organic chemistry, are based on Lewis's rules. The combination of Lewis structures along with Valence Shell Electron Pair Repulsion (VSEPR) theory [4,5] also laid the foundations of structural chemistry.

In view of the marvel of Lewis's theory, major problems remain for modern-day theoretical chemists in explaining this success. Lewis structures rely on non-observable properties in a similar way as, e.g. the concepts of atoms-in-molecules (AIM) [6–9] or the chemical bond. In this context, 'non-observable' indicates that these concepts cannot be extracted uniquely from the wave function using a Hermitian operator based on the correspondence principle. This does not mean that these concepts are *per se* orthogonal to quantum mechanics but it does mean that there exist multiple ways to extract them from a wave function, in the best case using well argued but nonetheless biased approaches.

In the present work we discuss a method that does not introduce the concept of Lewis structures into quantum mechanics but rather lets them – when applicable – emanate from the structure of the wave function Ψ without reference to an AIM model. This central idea sets this method apart from methods based on, e.g. delocalisation

indices [10–12], also known as shared electron distribution indices (SEDIs) [13], where first an AIM method is chosen to identify two bonded atoms or atomic domains or more in the case of multicentre bonding [14,15].

According to the Copenhagen interpretation of quantum mechanics, a probability is obtained when the quantity $|\Psi|^2$ is multiplied by an infinitesimal volume element. For an N -electron wave function $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)$, where \mathbf{x} denotes a space–spin coordinate $\mathbf{x} = (\mathbf{r}, \sigma)$ with $\mathbf{r} \in \mathbb{R}^3$ and $\sigma \in \{\uparrow, \downarrow\}$, the probability $p(1, \Omega)$ of finding an electron in a volume $\Omega \subset \mathbb{R}^3$ is given by summing over all N spin coordinates and integrating one electron over the domain Ω while integrating the remaining $N - 1$ electrons over the entire space

$$p(1, \Omega) = N \sum_{\sigma_1, \dots, \sigma_N} \int_{\Omega} d\mathbf{r}_1 \underbrace{\int_{\mathbb{R}^3} d\mathbf{r}_2 \dots \int_{\mathbb{R}^3} d\mathbf{r}_N}_{N-1} |\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)|^2. \quad (1)$$

The prefactor N corrects for the permutational symmetry of the N electrons. A related, but different question is the probability $P(1, \Omega)$ of finding one (and only one) electron in a domain Ω with all other electrons outside this volume (or, equivalently, in the complementary domain $\bar{\Omega} = \mathbb{R}^3 \setminus \Omega$).

$$P(1, \Omega) = N \sum_{\sigma_1, \dots, \sigma_N} \int_{\Omega} d\mathbf{r}_1 \underbrace{\int_{\bar{\Omega}} d\mathbf{r}_2 \dots \int_{\bar{\Omega}} d\mathbf{r}_N}_{N-1} |\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)|^2. \quad (2)$$

This can be generalised to the probability $P(\nu, \Omega)$ of finding ν (and only ν) electrons in Ω , resulting in the probability

$$P(\nu, \Omega) = \binom{N}{\nu} \sum_{\sigma_1, \dots, \sigma_N} \underbrace{\int_{\Omega} d\mathbf{r}_1 \dots \int_{\Omega} d\mathbf{r}_\nu}_{\nu} \underbrace{\int_{\bar{\Omega}} d\mathbf{r}_{\nu+1} \dots \int_{\bar{\Omega}} d\mathbf{r}_N}_{N-\nu} |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)|^2. \quad (3)$$

When the domain Ω is related to some AIM, these formulae link directly to Domain Averaged Fermi Holes [16], SEDI [13] and multicentre indices [14,15,17–22]. Interestingly, Savin showed that chemical interpretations can also be associated with those domains whose shapes have been optimised so as to maximise these probabilities. The

resulting domains are called the ‘maximum probability domains’ (MPDs) [23]. An $\text{MPD}(\nu)$ is a domain Ω^* for which the probability of finding ν (and only ν) electrons is maximal

$$P(\nu, \Omega^*) = \max_{\Omega} P(\nu, \Omega), \quad (4)$$

where the maximality criterion is understood to include local maxima. This makes the domain the result of a well-defined criterion without inferring it from an AIM model. The underlying hope is that the resulting MPDs will capture those regions in space where the bonding pairs of the Lewis model would appear. The theory of MPDs therefore has a similar goal as the loge theory [24] and the Electron Localisation Function [25–27]. In 2004, Cancès *et al.* [28] devised an algorithm to perform this optimisation, spurring renewed interest in MPDs [29–37].

Despite many interesting results from probabilities computed from user-specified domains, only relatively little work has appeared where true MPDs are computed, i.e. where the domains are actually optimised under probability criteria. This is due to several bottlenecks. The optimisation procedure has currently only been derived for single determinant wave functions, which are known to fail for many interesting molecules where correlation is important. Furthermore, technical issues with the optimisation algorithms lead to numerical errors, which cloud the fundamental properties and understanding of MPDs. Such fundamental questions have only been addressed using analytical models [38,39].

In the present work, we aim to go beyond purely analytical models by performing full configuration interaction (FCI) calculations for a discrete Hubbard model system. In this model system, both the domain probabilities as well as the optimisation of the domains can be formulated much more succinctly because the bottlenecks associated with optimising probability domains in \mathbb{R}^3 evaporate by discretising the space on a lattice. In quantum chemical terms, Hubbard models can be described as networks of hydrogens where a minimal basis set is used and a zero differential overlap occurs between the basis functions, rendering them orthonormal. The diagonalisation of the model Hamiltonian may be performed for different values of U and t , known as the on-site and hopping terms in the Hubbard Hamiltonian [40] (see Section 2). The physics of the problem foretells what is to be expected for different limiting regimes and this model will thus serve well as a testing ground for MPD theory. For reasons explained below, MPDs are slightly generalised to a set-up where a separate domain for ν_{\uparrow} (and only ν_{\uparrow}) electrons with spin \uparrow is sought as well as a possibly different domain for ν_{\downarrow} (and only ν_{\downarrow}) electrons with spin \downarrow , for

which the probabilities are given by:

$$\begin{aligned}
& P(\nu_\uparrow, \uparrow, \Omega_\uparrow, \nu_\downarrow, \downarrow, \Omega_\downarrow) \\
&= \binom{N_\uparrow}{\nu_\uparrow} \binom{N_\downarrow}{\nu_\downarrow} \underbrace{\int_{\Omega_\uparrow} d\mathbf{r}_{1\uparrow} \dots \int_{\Omega_\uparrow} d\mathbf{r}_{\nu_\uparrow}}_{\nu_\uparrow} \underbrace{\int_{\bar{\Omega}_\uparrow} d\mathbf{r}_{\nu_\uparrow+1} \dots \int_{\bar{\Omega}_\uparrow} d\mathbf{r}_{N_\uparrow}}_{N_\uparrow-\nu_\uparrow} \\
&\quad \underbrace{\int_{\Omega_\downarrow} d\mathbf{r}_{N_\uparrow+1} \dots \int_{\Omega_\downarrow} d\mathbf{r}_{N_\uparrow+\nu_\downarrow}}_{\nu_\downarrow} \underbrace{\int_{\bar{\Omega}_\downarrow} d\mathbf{r}_{N_\uparrow+\nu_\downarrow+1} \dots \int_{\bar{\Omega}_\downarrow} d\mathbf{r}_{N_\uparrow+N_\downarrow}}_{N_\downarrow-\nu_\downarrow} \\
& |\Psi((\mathbf{r}_1, \uparrow), \dots, (\mathbf{r}_{N_\uparrow}, \uparrow), (\mathbf{r}_{N_\uparrow+1}, \downarrow), \dots, (\mathbf{r}_{N_\uparrow+N_\downarrow}, \downarrow))|^2,
\end{aligned} \tag{5}$$

with N_\uparrow and N_\downarrow the total number of spin- \uparrow and spin- \downarrow electrons, respectively.

The actual test case corresponds to a 6-site, 6-electron system with and without periodic boundary conditions, which chemically corresponds to a minimal Hubbard model of benzene and 1,3,5-hexatriene. Results of these calculations can be put in the context of the well-known work by Hückel, whose theory very closely resembles Hubbard models [41].

2. Methodology

2.1. Hubbard models

The most general Fermi-Hubbard model Hamiltonian allowing for intersite hopping between any L sites, together with a site-dependent on-site repulsion is given by

$$\hat{H} = - \sum_{i,j=1}^L \sum_{\sigma=\uparrow,\downarrow} t_{ij} \hat{a}_{i\sigma}^\dagger \hat{a}_{j\sigma} + \sum_{i=1}^L U_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \tag{6}$$

with t_{ij} an intersite-dependent, spin-independent hopping, U_i the site-dependent, on-site repulsion depending on the local charges, and $\hat{a}_{i\sigma}^\dagger$ and $\hat{a}_{i\sigma}$ the creation/annihilation operators of a particle with spin σ ($\sigma = \uparrow, \downarrow$) on site i . The number operator $\hat{n}_{i\sigma} = \hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma}$ counts the number of particles with spin σ on site i , with eigenvalues 0 and 1 because of the fermion anticommutation relations

$$\{\hat{a}_{i\sigma}, \hat{a}_{j\tau}^\dagger\} = \delta_{ij} \delta_{\sigma\tau}, \quad \{\hat{a}_{i\sigma}^\dagger, \hat{a}_{j\tau}^\dagger\} = 0, \quad \{\hat{a}_{i\sigma}, \hat{a}_{j\tau}\} = 0. \tag{7}$$

We will explicitly use hat notation for the number operators $\hat{n}_{i\sigma}$ to distinguish between the operators and their eigenvalues $n_{i\sigma}$.

Although the adjacency matrix t allows for hopping between any two sites in general, it usually follows the geometry of the lattice, which means that only hopping between neighbouring sites is considered. In the same vein, the on-site repulsion is usually chosen site-independent because of symmetry considerations. In the present paper (see Section 4), we will consider a Hubbard model on a one-dimensional (1D) chain with and without periodic boundary conditions, but it is clear that the following derivations are generally valid for Hubbard models.

The eigenstates of the Hamiltonian can be expanded in the orthonormal basis spanned by the eigenstates of the number operators $\hat{n}_{i\sigma}$,

$$|\Psi\rangle = \sum_{[n]} c_{[n]} |[n]\rangle, \tag{8}$$

with $c_{[n]}$ the coefficient belonging to a basis state labelled by the partitioning $[n] = [[n_\uparrow], [n_\downarrow]] = [[n_{1\uparrow}, n_{2\uparrow}, \dots, n_{L\uparrow}], [n_{1\downarrow}, n_{2\downarrow}, \dots, n_{L\downarrow}]]$ of n_\uparrow spin- \uparrow and n_\downarrow spin- \downarrow particles over L sites. Note that these basis states have in essence the same structure as the FCI basis vectors described in [42] and can be represented accordingly as bitstrings [43]. Both n_\uparrow and n_\downarrow are good quantum numbers for the eigenstates (8) due to the particle-number and spin-projection symmetry of the Hamiltonian (6). Explicitly, the basis states $|[n]\rangle$ can be expressed in Fock space

$$|[n]\rangle = \bigotimes_{i=1}^L |n_{i\uparrow}\rangle \bigotimes_{i=1}^L |n_{i\downarrow}\rangle = \prod_{i=1}^L (\hat{a}_{i\uparrow}^\dagger)^{n_{i\uparrow}} \prod_{i=1}^L (\hat{a}_{i\downarrow}^\dagger)^{n_{i\downarrow}} |\theta\rangle, \tag{9}$$

with $|\theta\rangle$ the particle vacuum. Using this representation, it is straightforward to verify that these states are eigenstates of the number operators $\hat{n}_{i\sigma}$

$$\hat{n}_{i\sigma} |[n]\rangle = n_{i\sigma} |[n]\rangle, \tag{10}$$

leading to the interpretation of the number operator $\hat{n}_{i\sigma}$ as a projection operator for fermionic systems. Indeed, if the site i is not occupied with a spin- σ particle ($n_{i\sigma} = 0$), then the state $|[n]\rangle$ is annihilated, whereas it is left unaltered otherwise ($n_{i\sigma} = 1$). More general, the number operator $\hat{n}_{i\sigma}$ projects a general state onto that part of the Hilbert space that has the i th site occupied with a spin- σ particle

$$\hat{n}_{i\sigma} \sum_{[n]} c_{[n]} |[n]\rangle = \sum_{[n]|n_{i\sigma}=1} c_{[n]} |[n]\rangle. \tag{11}$$

Consequently, the *probability* of finding a spin- σ electron at site i is also equivalent to the expectation value of the

corresponding number operator

$$\langle \Psi | \hat{n}_{i\sigma} | \Psi \rangle = \sum_{[n] | n_{i\sigma}=1} |c_{[n]}|^2. \quad (12)$$

Formally, the number operators $\hat{n}_{i\sigma}$ fulfil the requirements of projection operators thanks to the fermionic anticommutation relations

$$\hat{n}_{i\sigma}^2 = \hat{n}_{i\sigma}, \quad (13)$$

however they do not partition the Hilbert space into disjoint orthogonal subspaces due to the fact that, at the operator level

$$\hat{n}_{i\sigma} \hat{n}_{j\tau} \neq 0, \quad \forall i, j, \sigma, \tau. \quad (14)$$

Note that this does not imply that for any state $|\psi\rangle$,

$$\hat{n}_{i\sigma} \hat{n}_{j\tau} |\psi\rangle \neq 0, \quad \forall i, j, \sigma, \tau. \quad (15)$$

The orthogonal complement of each projection operator $\hat{n}_{i\sigma}$ has a physical interpretation

$$1 - \hat{n}_{i\sigma} = 1 - \hat{a}_{i\sigma}^\dagger \hat{a}_{i\sigma} = \hat{a}_{i\sigma} \hat{a}_{i\sigma}^\dagger, \quad (16)$$

as the number operator counting the number of spin- σ holes at site i . Like its complementary particle number operator, the hole number operator has the structure of a projection operator, so we will use the following definitions in the remainder of the paper for ease of notation

$$\hat{O}_{i\sigma} := \hat{n}_{i\sigma} \quad \hat{U}_{i\sigma} := 1 - \hat{n}_{i\sigma}, \quad (17)$$

projecting against the orthogonal subspaces that are occupied (O) and unoccupied (U) with a spin- σ , respectively. Note that $\hat{O}_{i\sigma} + \hat{U}_{i\sigma} = 1$ and $\hat{O}_{i\sigma} \hat{U}_{i\sigma} = 0$, $\forall i, \sigma$.

2.2. MPDs using projection operators

In contrast to the intricate definition for continuous domains (see Section 1), in discrete lattice models such as the Hubbard model, a *domain* Ω can be defined as a subset of the total set of sites. So, a domain Ω of size M within a Hubbard model of size L can also be represented as a partitioning

$$\Omega = [\omega] = [\omega_1, \omega_2, \dots, \omega_L], \quad (18)$$

with the restriction that $\omega_i = 0$ or 1 ($\forall i$), and $\sum_{i=1}^L \omega_i = M$. A domain Ω does not necessarily have to be connected, so there are $\binom{L}{M}$ different possibilities to pick a domain of size M , adding up to 2^L different domains of variable size in total. Often it is more convenient to

use a different notation for Ω , as the subset $\{i_1, i_2, \dots, i_M\}$ of the M sites constituting the domain, so for which $\omega_i = 1$. For instance, a domain Ω of size $M = 2$, consisting of the first and third site within a total of $L = 5$ sites can be represented in two different ways: $\Omega = [1, 0, 1, 0, 0] = \{1, 3\}$.

Although the present definition of MPDs (Equation (4)) is blind to the spin degree of freedom in the model, it is possible to resolve the spin dependency. Not to complicate matters, we first focus on same-spin particles. The probability $P(\nu_\sigma, \sigma, \Omega)$ of finding ν_σ (and only ν_σ) spin- σ particles within a domain Ω can be conveniently formulated using the projection operators (17). Before introducing the general formula, it is instructive to discuss some special cases. As pointed out in the previous section, the probability $P(1, \sigma, \{i\})$ of finding one (and only 1) spin- σ particle at a single site i (or domain $\Omega = \{i\}$ of size $M = 1$) is simply the expectation value of the number of spin- σ particles at that site, or equivalently, the expectation value of the projection operator

$$P(1, \sigma, \{i\}) = \langle \Psi | \hat{n}_{i\sigma} | \Psi \rangle = \langle \Psi | \hat{O}_{i\sigma} | \Psi \rangle. \quad (19)$$

Equivalent, the probability of finding $\nu_\sigma = 0$ particles within $\Omega = \{i\}$ is

$$P(0, \sigma, \{i\}) = \langle \Psi | 1 - \hat{n}_{i\sigma} | \Psi \rangle = \langle \Psi | \hat{U}_{i\sigma} | \Psi \rangle. \quad (20)$$

The situation is a little more involved for the $M = 2$ case. The probability of finding exactly $\nu_\sigma = 0, 1$ or 2 can be written, respectively, as

$$P(0, \sigma, \{i_1, i_2\}) = \langle \Psi | \hat{U}_{i_1\sigma} \hat{U}_{i_2\sigma} | \Psi \rangle, \quad (21)$$

$$P(1, \sigma, \{i_1, i_2\}) = \langle \Psi | (\hat{O}_{i_1\sigma} \hat{U}_{i_2\sigma} + \hat{U}_{i_1\sigma} \hat{O}_{i_2\sigma}) | \Psi \rangle, \quad (22)$$

$$P(2, \sigma, \{i_1, i_2\}) = \langle \Psi | \hat{O}_{i_1\sigma} \hat{O}_{i_2\sigma} | \Psi \rangle. \quad (23)$$

Indeed, the projection operators $\hat{O}_{i\sigma}$ and $\hat{U}_{i\sigma}$ make sure that only those components $[[n]]$ of the state $|\Psi\rangle$ survive that correspond to $\nu_\sigma = 1$ within the domain $\{i_1, i_2\}$. The combined use of \hat{O} and \hat{U} projection operators in $P(1, \{i_1, i_2\})$ is essential to ascertain that one (and only one) particle is contained within the domain. From these examples, the general formulation of the probabilities in terms of projection operators comes naturally. Explicitly, the probability to find ν_σ (and only ν_σ) particles within a domain $\Omega = \{i_1, \dots, i_M\}$ is given by

$$P(\nu_\sigma, \sigma, \Omega) = \langle \Psi | \hat{P}(\nu_\sigma, \sigma, \Omega) | \Psi \rangle, \quad (24)$$

with the probability domain operator defined as

$$\hat{P}(v_\sigma, \sigma, \Omega) = \frac{1}{v_\sigma!(M - v_\sigma)!} \sum_{\pi \in S_M} \prod_{k=1}^{v_\sigma} \hat{O}_{i_{\pi(k)}\sigma} \prod_{k=v_\sigma+1}^M \hat{U}_{i_{\pi(k)}\sigma}, \quad (25)$$

and π a permutation of the M sites within the domain Ω (π is an element of S_M , the permutation group over M elements). The prefactor corrects for double counting. Because the projection operators are consistent with the interpretation as probabilities, we will omit the expectation value notation, and continue with the projection operator formulation (25).

2.3. Generating function

2.3.1. General case

The generating function operator

$$\hat{G}(\sigma, \Omega, t) = \sum_{v_\sigma=0}^M \hat{P}(v_\sigma, \sigma, \Omega) t^{v_\sigma}, \quad (26)$$

condenses all possible probabilities for a given domain Ω into one polynomial function in t of degree M . It is straightforward to prove (see Section A.1 in the Appendix section) that the generating function operator for a domain $\Omega = \{i_1, \dots, i_M\}$ of size M can be factorised as

$$\hat{G}(\sigma, \Omega, t) = \prod_{k=1}^M (\hat{U}_{i_k\sigma} + t\hat{O}_{i_k\sigma}), \quad (27)$$

or, equivalently

$$\hat{G}(\sigma, \Omega, t) = t^{\hat{n}_\sigma(\Omega)}, \quad (28)$$

in which $\hat{n}_\sigma(\Omega)$ counts the number of sites in the domain Ω that are occupied by a spin- σ particle. When evaluating the generating function operator for a quantum state $|\Psi\rangle$, it is convenient to express $|\Psi\rangle$ in the site basis spanned by the basis states (9). We obtain

$$\begin{aligned} G(\sigma, \Omega, t) &= \langle \Psi | \hat{G}(\sigma, \Omega, t) | \Psi \rangle \\ &= \sum_{[n],[m]} c_{[n]}^* c_{[m]} \langle [n] | \hat{G}(\sigma, \Omega, t) | [m] \rangle. \end{aligned} \quad (29)$$

All projection operators (17) are diagonal in the site basis, so the generating function reduces to diagonal elements only,

$$G(\sigma, \Omega, t) = \sum_{[n]} |c_{[n]}|^2 \langle [n] | \hat{G}(\sigma, \Omega, t) | [n] \rangle. \quad (31)$$

The factorised expression for the generating function operator (27) now comes in handy, because one can evaluate each factor separately. Making use of the explicit notation for $|[n]\rangle$ in Equation (9), one obtains

$$\begin{aligned} \langle [n] | \hat{G}(\sigma, \Omega, t) | [n] \rangle &= \prod_{k=1}^M \langle n_{i_k\sigma} | \hat{U}_{i_k\sigma} + t\hat{O}_{i_k\sigma} | n_{i_k\sigma} \rangle \\ &= \prod_{k=1}^M (1 - n_{i_k\sigma} + t n_{i_k\sigma}), \end{aligned} \quad (32)$$

in which we have used the diagonal property of the projector operators in the site basis. Because $n_{i_k\sigma}$ is either 1 (the site i_k in Ω is occupied with a spin- σ particle in state $|[n]\rangle$) or 0 (the site i_k in Ω is *not* occupied with a spin- σ particle in state $|[n]\rangle$), the formula reduces to

$$\langle [n] | \hat{G}(\sigma, \Omega, t) | [n] \rangle = t^{n_\sigma(\Omega)}, \quad (33)$$

in which $n_\sigma(\Omega)$ is the number of sites in the domain Ω that are occupied by a spin- σ particle in the state $|[n]\rangle$, consistent with the exponential form of the generating function operator (28). For a general state, one gets the result

$$G(\sigma, \Omega, t) = \sum_{[n]} |c_{[n]}|^2 t^{n_\sigma(\Omega)}. \quad (34)$$

It is now possible to extract the probabilities $P(v_\sigma, \sigma, \Omega)$ by filtering those coefficients that have $v_\sigma = n_\sigma(\Omega)$ particles in the domain Ω

$$P(v_\sigma, \sigma, \Omega) = \sum_{[n]} |c_{[n]}|^2 \delta_{v_\sigma, n_\sigma(\Omega)}. \quad (35)$$

2.3.2. Slater determinants

The factorisation of the generating function (27) holds for arbitrary states and is not necessary limited to eigenstates of a Hamiltonian. However, as can be seen from (34), an exact evaluation of the generating function requires a summation over the whole Hilbert space, limiting the practical applicability to small systems. This problem was previously encountered for MPDs in continuous space. Nevertheless, it was shown how the generating function and resulting probabilities can be efficiently evaluated for a Slater determinant [28]. These results can be immediately generalised towards lattice models. Taking $|\Phi\rangle$ as a Slater determinant with $N = n_\uparrow + n_\downarrow$ orthonormal occupied states, the generating function can be shown to reduce to

$$\langle \Phi | \prod_{k=1}^M (\hat{U}_{i_k\sigma} + t\hat{O}_{i_k\sigma}) | \Phi \rangle = \det \{ [I_N - S_\sigma(\Omega)] + tS_\sigma(\Omega) \}, \quad (36)$$

with I_N the $N \times N$ identity matrix and $S_\sigma(\Omega)$ the $N \times N$ symmetric matrix defined as

$$S_\sigma(\Omega)_{\alpha\beta} = \langle \alpha | \sum_{k=1}^M \hat{n}_{k\sigma} | \beta \rangle = \langle \alpha | \hat{n}_\sigma(\Omega) | \beta \rangle, \quad (37)$$

where $\alpha, \beta = 1 \dots N$ label the occupied single-particle states in the Slater determinant. A dual expression can be found as

$$\langle \Phi | \prod_{k=1}^M (\hat{U}_{k\sigma} + t \hat{O}_{k\sigma}) | \Phi \rangle = \det \{ [I_M - T_\sigma(\Omega)] + t T_\sigma(\Omega) \}, \quad (38)$$

with I_M the $M \times M$ identity matrix and $T_\sigma(\Omega)$ the $M \times M$ symmetric matrix defined as

$$T_\sigma(\Omega)_{ij} = \langle i\sigma | \sum_{\alpha=1}^N \hat{n}_\alpha | j\sigma \rangle, \quad (39)$$

where $|i\sigma\rangle = \hat{a}_{i\sigma}^\dagger |\theta\rangle$, with i, j occupied sites in Ω , and \hat{n}_α is the occupation number of the orbital α . Starting from the factorised expression for the generating function, the proof for these identities is analogous to the one presented previously [28] and is given in Section A.2 in the Appendix section.

The size of these matrices scales with the number of occupied orbitals and the size of Ω , respectively, and both the matrix elements and the determinant can be calculated in a polynomial time. Furthermore, all probabilities $P(v_\sigma, \sigma, \Omega)$ can be expressed in the eigenvalues of these matrices [28]. However, it should be stressed that these results present a trade-off: starting from a Slater determinant allows for an efficient calculation of the MPDs of an approximate state, whereas starting from the exact solution allows for an involved calculation of the MPDs of the exact state. In the present paper, we will limit ourselves to calculations using the exact ground state.

2.4. Spin dependency of the MPD

The extra spin degree of freedom complicates matters because a single site i within a domain Ω can now be occupied with a spin- \uparrow particle, a spin- \downarrow particle, or both. Consequently, the site number operator $\hat{n}_i = \hat{n}_{i\uparrow} + \hat{n}_{i\downarrow}$ loses its interpretation as a projection operator

$$\hat{n}_i^2 = (\hat{n}_{i\uparrow} + \hat{n}_{i\downarrow})^2 = \hat{n}_{i\uparrow} + \hat{n}_{i\downarrow} + 2\hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \neq \hat{n}_i, \quad (40)$$

so the probabilities need to be broken down into its spin projections $\sigma = \uparrow, \downarrow$. In its most general form, the probability of finding v_\uparrow (and only v_\uparrow) and v_\downarrow (and only v_\downarrow) in respective domains Ω_\uparrow and Ω_\downarrow is given at the operator

level

$$\hat{P}(v_\uparrow, \Omega_\uparrow; v_\downarrow, \Omega_\downarrow) = \hat{P}(v_\uparrow, \uparrow, \Omega_\uparrow) \hat{P}(v_\downarrow, \downarrow, \Omega_\downarrow). \quad (41)$$

Note that in general the domains Ω_\uparrow and Ω_\downarrow need not be the same; however, it is often in the interest of interpretation to do so ($\Omega_\uparrow = \Omega_\downarrow = \Omega$). The differentiation between the two spin degrees of freedom necessitates the introduction of an additional degree of freedom in the generating function operator

$$\hat{G}(\Omega_\uparrow, t; \Omega_\downarrow, s) = \sum_{v_\uparrow=0}^{M_\uparrow} \sum_{v_\downarrow=0}^{M_\downarrow} \hat{P}(v_\uparrow, \Omega_\uparrow; v_\downarrow, \Omega_\downarrow) t^{v_\uparrow} s^{v_\downarrow}. \quad (42)$$

The probabilities are nicely factorised (41), which easily extends to the generating function

$$\hat{G}(\Omega_\uparrow, t; \Omega_\downarrow, s) = \sum_{v_\uparrow=0}^{M_\uparrow} \hat{P}(v_\uparrow, \uparrow, \Omega_\uparrow) t^{v_\uparrow} \sum_{v_\downarrow=0}^{M_\downarrow} \hat{P}(v_\downarrow, \downarrow, \Omega_\downarrow) s^{v_\downarrow} \quad (43)$$

$$= \hat{G}(\uparrow, \Omega_\uparrow, t) \hat{G}(\downarrow, \Omega_\downarrow, s), \quad (44)$$

implying that many properties of the generating function (such as the exponential form (28), or the expectation value for a single Slater determinant) can be simply transferred from the single spin- σ case.

Multiple ‘spin-averaged’ properties can now be extracted from the spin-resolved generating function (42). For ease of interpretation, we will choose $\Omega_\uparrow = \Omega_\downarrow = \Omega$, but everything can be formulated with distinct domains

- The probability of finding v particles (*independent* of the spin, so $v = v_\uparrow + v_\downarrow$) in a domain Ω can be found by putting $s = t$ in the generating function, and reorganising the summation

$$\hat{G}(\Omega, t; \Omega, t) = \sum_{v=0}^{2M} \left(\sum_{v_\uparrow=\max(v-M,0)}^{\min(M,v)} \hat{P}(v_\uparrow, \uparrow, \Omega) \hat{P}(v - v_\uparrow, \downarrow, \Omega) \right) t^v. \quad (45)$$

The probability of finding v (and only v) particles is obtained by taking the coefficient of t^v (assume v below or equal to half-filling of the domain $v \leq M$, the other case is symmetric)

$$\hat{P}_{\uparrow\downarrow}(v, \Omega) := \sum_{v_\uparrow=0}^v \hat{P}(v_\uparrow, \uparrow, \Omega) \hat{P}(v - v_\uparrow, \downarrow, \Omega). \quad (46)$$

Note that this definition of $\hat{P}_{\uparrow\downarrow}(v, \Omega)$ coincides with the original definition of MPDs, independent of the spin of the particles that are in the domain.

- An interesting variant of the previous definition is the probability of finding v_{\uparrow} spin- \uparrow particles in a domain Ω , irrespective of how many spin- \downarrow particles there are. This can be found by substituting $s = 1$ in the generating function (42)

$$\begin{aligned}\hat{G}(\Omega, t; \Omega, 1) &= \hat{G}(\uparrow, \Omega_{\uparrow}, t) \hat{G}(\downarrow, \Omega_{\downarrow}, 1) \\ &= \hat{G}(\uparrow, \Omega_{\uparrow}, t).\end{aligned}\quad (47)$$

The probability of finding v_{\uparrow} spin- \uparrow particles then coincides with the definition of the spin-dependent probabilities defined previously (25)

$$\hat{P}_{\uparrow}(v_{\uparrow}, \Omega) := \hat{P}(v_{\uparrow}, \uparrow, \Omega).\quad (48)$$

The same argument holds for the spin- \downarrow probabilities when taking $t = 1$.

3. Implementation

For the calculation of the MPDs, we assume that an eigenvector $|\Psi\rangle$ of the relevant Hubbard model Hamiltonian \hat{H} has been determined and can be expressed in the site basis $\{|[n]\rangle\}$ with coefficients $\{c_{[n]}\}$. If we constrain all spin-resolved domains $\Omega_{\uparrow}, \Omega_{\downarrow}$ to occupy the same sites, such that for all spin-unresolved domains Ω the following is valid: $\forall \Omega: \Omega_{\uparrow} = \Omega_{\downarrow} = \Omega$, then we can construct 2^L domains in total, with L the number of sites. Since we can represent such domains $\Omega \in \{\Omega\}$, $|\{\Omega\}| = 2^L$ in the same site basis $\{|[n]\rangle\}$, we can store them as bitstrings (Section 2.2). As such, we can determine the associated probabilities as follows:

```
Data:  $\{|[n]\rangle\}$  (explicit or via addressing scheme),  

 $\{c_{[n]}\}, \{\Omega\}$   

Result:  $P(v, \Omega) (\forall \Omega \in \{\Omega\})$   

for  $|[n]\rangle$  in  $\{|[n]\rangle\}$  do  

  for  $\Omega$  in  $\{\Omega\}$  do  

    if  $\text{POPCNT}(|[n]\rangle \wedge \Omega) == v$  then  

       $P(v, \Omega) += |c_{[n]}|^2$   

    end  

  end  

end
```

In this algorithm, POPCNT counts the number of ‘1’ bits in the passed bitstring (this functionality is available as an intrinsic procedure in most Fortran and C/C++ compilers).

Since the definition of MPDs (Equation (4)) allows for local maxima, we can define the set of MPDs for v electrons in the present context as follows:

$$\text{MPDs}(v) = \{\Omega^* | \forall \Omega \in \epsilon_{\Omega^*} : P(v, \Omega^*) \geq P(v, \Omega)\},\quad (49)$$

where ϵ_{Ω^*} defines the set of domains which are within a certain ‘distance’ ϵ from the domain Ω^*

$$\epsilon_{\Omega^*} = \{\Omega | \|\Omega^* - \Omega\| \leq \epsilon\}.\quad (50)$$

In this work, we use the Hamming distance as a distance measure, which is equal to the number of bits in which two bitstrings differ [44]. If we set $\epsilon = 1$, this reduces to a single-bit flip stability criterion. Only domains which are maximal in their probability within this distance are retained as MPDs.

If we suppose that the computational cost for diagonalising the Hubbard model for L sites is of the order of the dimension of the Hilbert space \mathcal{H} cubed,

$$C_{\text{Hubbard}} = [\dim \mathcal{H}]^3,\quad (51)$$

and the cost for the MPD procedure is of the order of the product of the number of coefficients, the number of domains and the number of single-bit flips for a given domain,

$$C_{\text{MPD}} = [\dim \mathcal{H}] 2^L (L - 1),\quad (52)$$

then for Hubbard models at half-filling, imposing S_z symmetry, $C_{\text{Hubbard}} > C_{\text{MPD}}$. Hence, for Hubbard models at half-filling, this analysis procedure will always take less time than the determination of the eigenvector to be analysed. This allows for the use of MPDs as an efficient interpretational tool, since they can be calculated at a fraction of the computational cost of the determination of the eigenvector to be analysed.

4. Applications: 1D Hubbard with six sites at half-filling

As discussed in Section 1, we analyse the 1D Hubbard model with six sites at half-filling (denoted as $6s, 6e$ in the accompanying plots), with and without periodic boundary conditions, for a U/t range from 0 to 100 in unit steps. For every value of U/t , the corresponding Hamiltonian is diagonalised through an FCI calculation and the resulting ground-state wave function is used as input for the domain probability computation. The domains are then tested for stability against single-bit flips, resulting in a set of (point group) symmetric classes of MPDs. In all cases,

only one representative structure of a symmetry equivalent class of MPDs will be plotted.

In the following we will focus on the set of MPDs for two electrons regardless of their spin (i.e. $\{\text{MPD}(\nu)\} = \{\Omega^* | P(\nu, \Omega^*) = \max_{\Omega} P(\nu, \Omega)\}$ with $\nu = 2$) as well as the set of MPDs for spin-resolved electrons (i.e. $\{\text{MPD}(\nu_{\uparrow}, \nu_{\downarrow})\} = \{\Omega^* | P(\nu_{\uparrow}, \uparrow, \nu_{\downarrow}, \downarrow, \Omega^*) = \max_{\Omega} P(\nu_{\uparrow}, \uparrow, \nu_{\downarrow}, \downarrow, \Omega)\}$ with $\nu_{\uparrow} + \nu_{\downarrow} = 2$).

4.1. With periodic boundary conditions: Hubbard benzene

The 1D Hubbard model with six sites at half-filling and periodic boundary conditions can be used as a minimal model for benzene (Hubbard benzene). As shown in Figure 1, $\{\text{MPD}(2)\}$ always consists of the same three classes of MPDs, irrespective of the U/t value. Note that only 2-site domains emerge from the algorithm, even though nothing limits the size of the domains. The fact that no larger domains emerge is therefore solely due to the structure of the wave function itself. Note that an MPD is found for every possible 2-site domain.

In the large U/t regime, the probabilities for each of the three 2-site MPD classes become equal. At lower U/t values, the probabilities of the classes differ. This is in line with what we expect from the underlying physics: in the large U/t regime, we expect static correlation to become very important, in other words, we expect large on-site repulsion to force one electron per site, leading to

the anti-ferromagnetic character of the wave function. At lower U/t values, a mean field solution of the Hamiltonian dominates the FCI expansion. This solution should render the probabilities of the classes different.

When turning to spin-resolved MPDs we note that the theory itself does not require that $\{\text{MPD}(2)\} = \{\text{MPD}(1_{\uparrow}, 1_{\downarrow})\} = \{\text{MPD}(2_{\uparrow}, 0_{\downarrow})\}$, nor that the ranking of the associated probability values should remain the same between groups. The fact that in benzene for almost all U/t values all three 2-site MPD classes have a significant probability is noteworthy (Figures 2 and 3). This allows us to expand the spin-unresolved MPDs in terms of their spin-resolved counterparts. If the sets of MPDs would not have been the same, one could still calculate the probabilities associated with the non-maximal domains, but they should not be interpreted as maximum probability domains.

Although the classes of MPDs remain the same, we can observe that in general the ranking of the probability values between the three groups does not remain the same. Figure 2 shows that MPD structure ‘1’ remains the most probable when choosing one up-spin and one down-spin electron. This indicates that, for a typical electron pair consisting of one spin-up and one spin-down electron, the highest probability is found for an MPD containing two adjacent sites. However, this situation is reversed when choosing two up-spin and zero down-spin electrons, where structure ‘3’ is the most probable (Figure 3). Same-spin electrons therefore tend to remain

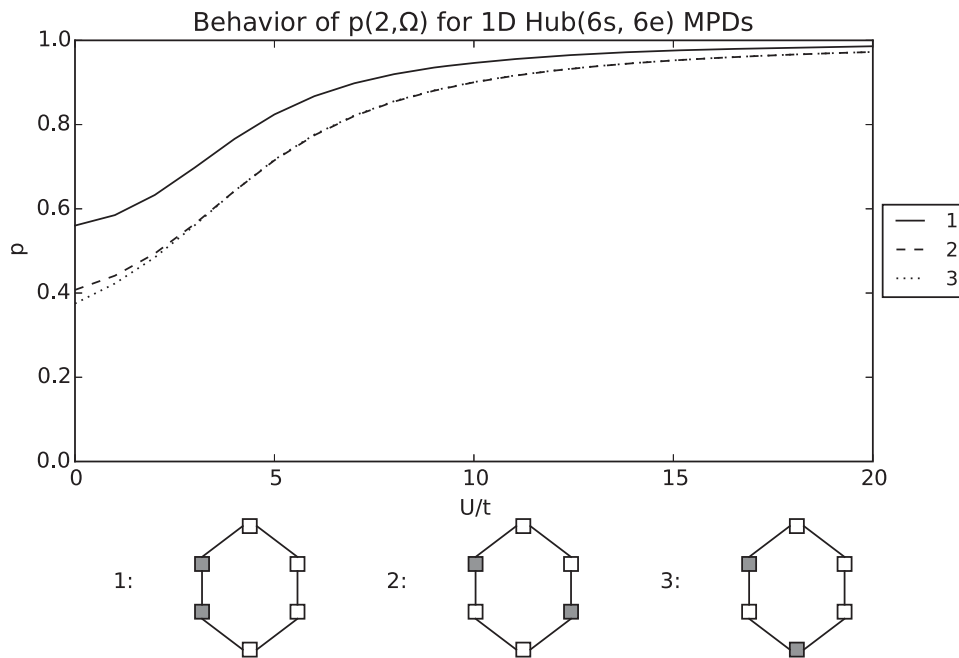


Figure 1. MPDs for choosing two spin-unresolved electrons for 1D Hubbard with periodic boundary conditions for six sites at half-filling for different U/t values.

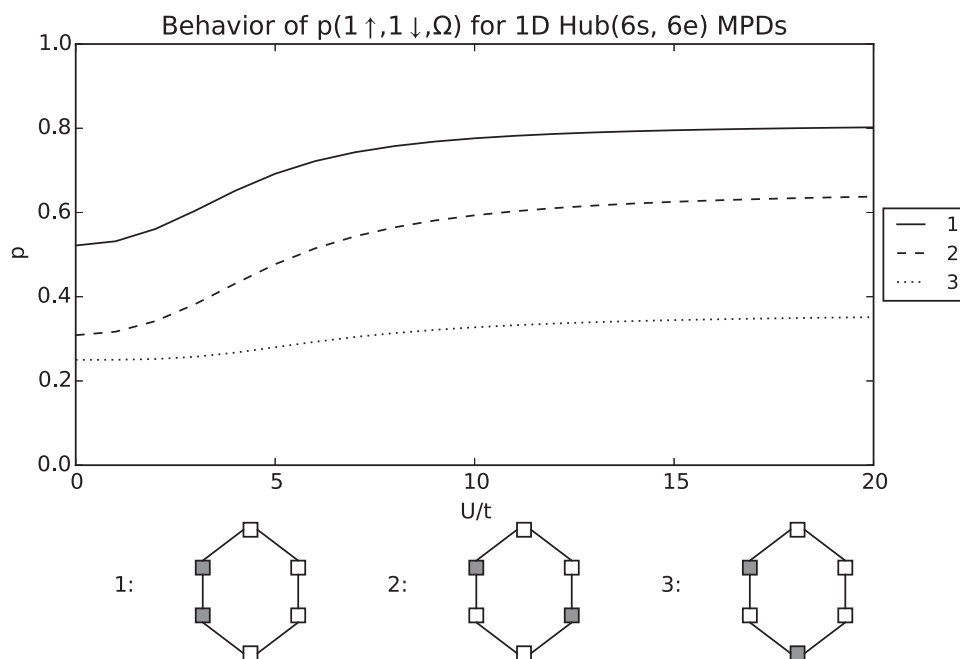


Figure 2. MPDs for choosing one up-spin and one down-spin electron for 1D Hubbard with periodic boundary conditions for six sites at half-filling for different U/t values.

in non-adjacent sites. In both spin configurations, the second highest probability is associated with what could be called the para-delocalisation of class '2'.

Note that the associated probabilities for the set of MPDs are quite small when choosing two up-spin and zero down-spin electrons compared to the other two

choices. Also, structure '1' is not stable against the single-bit flip criterion in the low U/t regime, where a higher probability can be obtained by adding another site to the domain. However, the resulting 3-site domain is again 'unstable' and can flip to a genuine MPD, which is why 3-site MPDs do not appear.

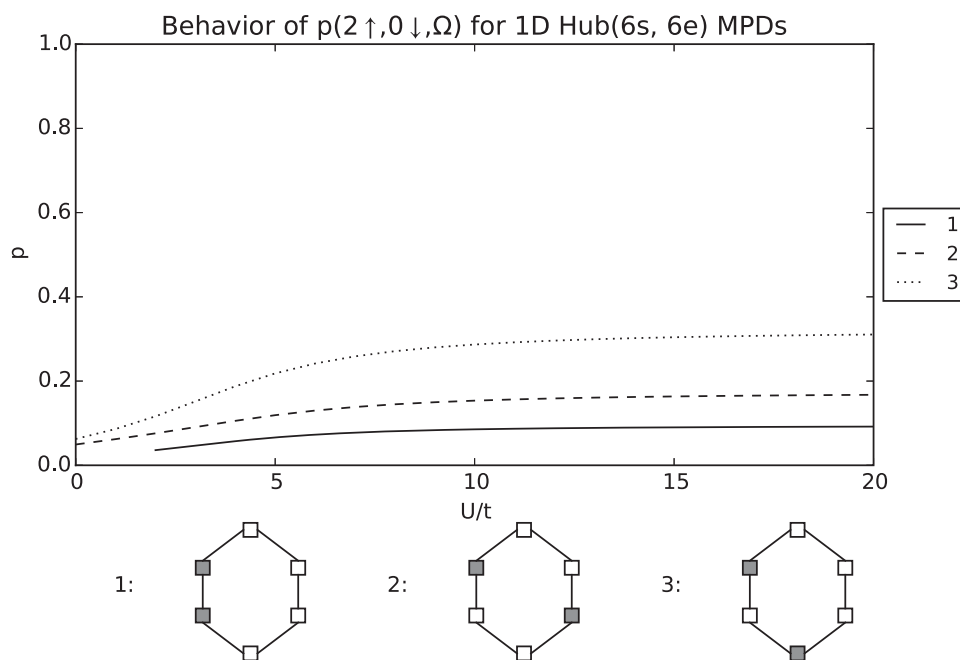


Figure 3. MPDs for choosing two up-spin and zero down-spin electrons for 1D Hubbard with periodic boundary conditions for six sites at half-filling for different U/t values.

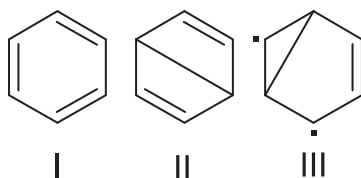


Figure 4. Kékulé, Dewar and bi-radical structures of Hubbard benzene.

It is tempting to relate these MPD findings to the well-known resonance structures of benzene, where the two Kekulé structures are the most important ones followed by the three Dewar structures shown as structure classes I and II, respectively, in Figure 4. The third type of MPD could be interpreted as related to class III in the same figure. It is indeed gratifying that, as was the case for delocalisation indices [45,46], the MPD results also seem to be in line with Valence Bond (VB) Theory results. However, our FCI wave functions are expressed in terms of orthogonal orbitals and as such, the individual VB wave functions for the different structures do not appear immediately in the FCI expansion. In future work, a CASVB-[47,48] type approach will be used to re-express – without altering the energy – the FCI wave function in terms of VB structures using a non-unitary transformation. As this lies outside the scope of the present work the apparent agreement momentarily remains to be treated with caution. However, from a global perspective, the MPDs do behave as expected for the different correlation regimes, from the mean field regime at zero U to the high static

correlation regime, and appear to be in line with what is expected from VB Theory-based knowledge.

4.2. Without periodic boundary conditions: Hubbard 1,3,5-hexatriene

The 1D Hubbard model with six sites at half-filling and no periodic boundary conditions can be used as a model for 1,3,5-hexatriene (Hubbard 1,3,5-hexatriene). If we determine the MPDs for two electrons (regardless of spin), we again find that only domains consisting of two sites occur even though the algorithm allows other domains to be found. In total, fifteen 2-site domains are possible, which are gathered in nine symmetry classes. We find the same nine symmetry classes to have significant probabilities over the entire U/t range (Figure 5).

The MPDs consisting of terminal sites are more probable than the MPDs containing the interior sites, with a clear decrease in probability for other domain configurations. This is again in line with what we expect from the underlying physics: the on-site repulsion ‘pushes’ the electrons towards the extremal sites. All probabilities again converge for high U/t values due to the anti-ferromagnetic properties of the wave function.

If we determine the spin-resolved MPDs (Figures 6 and 7), we again obtain the same classes of MPDs as in the spin-unresolved case, although this is not dictated by the theory as such. For typical electron pairs, the highest probability is again found for MPDs consisting of two adjacent sites; the relative low probability of structure ‘3’

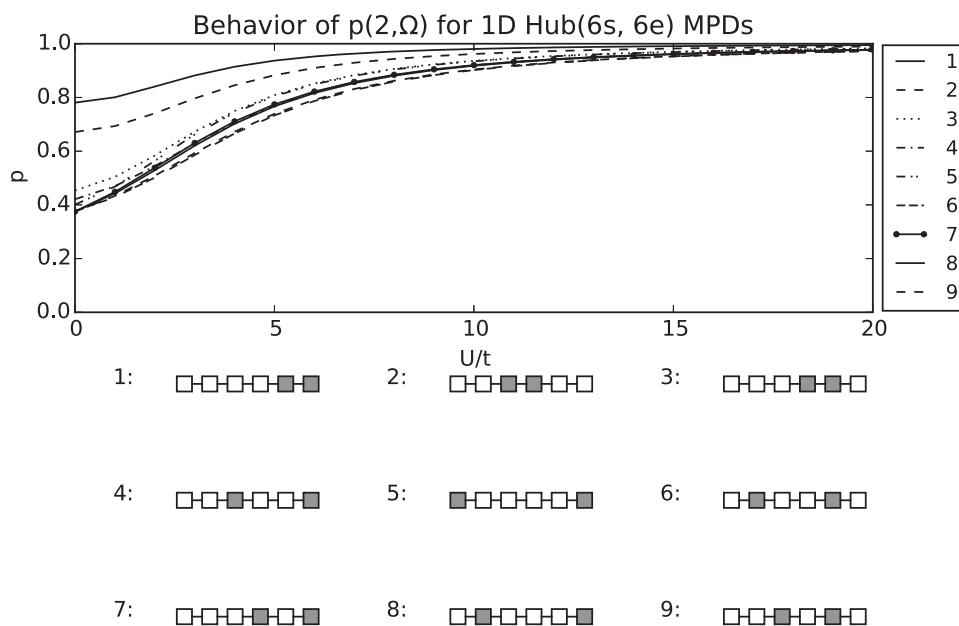


Figure 5. MPDs for choosing two spin-unresolved electrons for 1D Hubbard without periodic boundary conditions for six sites at half-filling for different U/t values.

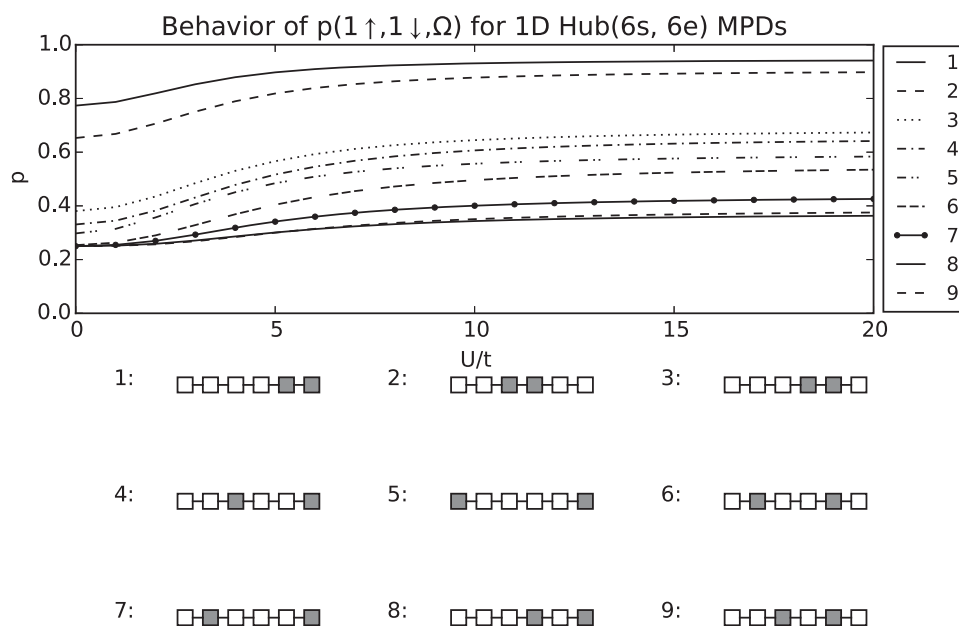


Figure 6. MPDs for choosing one up-spin and one down-spin electron for 1D Hubbard without periodic boundary conditions for six sites at half-filling for different U/t values.

for $\{\text{MPD}(1\uparrow, 1\downarrow)\}$ can be linked to the required nodal structure of the underlying wave function. The probabilities associated with $\{\text{MPD}(2\uparrow, 0\downarrow)\}$ are again lower compared to the two other choices.

Again, the MPDs behave as can be expected for different correlation regimes. Furthermore, they appear to be in line with what is expected from VB Theory (see for instance the discussion of 1,3,5-hexatriene in [49]).

5. Concluding remarks and perspectives

We have shown that the framework of the Hubbard model can be used to highlight the essential theoretical structure behind MPDs, both for FCI and HF wave functions. We note that the resulting framework is applicable to all models for which Equation (13) is valid. Furthermore, we have implemented a fast analysis procedure which can be used to analyse numerical results, and we have illustrated this

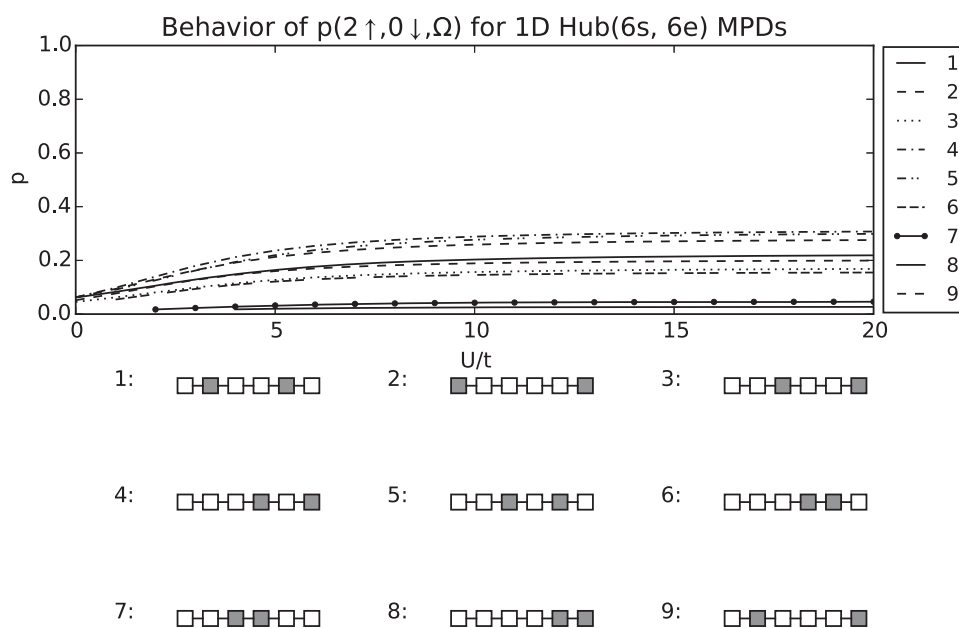


Figure 7. MPDs for choosing two up-spin and zero down-spin electrons for 1D Hubbard without periodic boundary conditions for six sites at half-filling for different U/t values.

with two applications at the FCI level. We have shown that the physics behind the considered Hubbard model can be captured using MPDs.

It is important to note that the theory of MPDs allows the user to choose a number of electrons which is different than two for determining MPDs. For the small systems under scrutiny, these types of analyses typically provide less information. However, we are currently investigating larger systems, where we might be able to link MPDs to other chemical concepts.

We also note that the problem of multi-domain optimisation, where several domains are optimised simultaneously for a given partition of electrons, becomes feasible in the Hubbard model using the method of projection operators. We are currently investigating algorithms along these lines.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Research Foundation Flanders (FWO-Vlaanderen); Special Research Fund of Ghent University; the Flemish Government – department EWI and Ghent University.

References

- [1] P.L.A. Popelier in *Chemical Concepts from Quantum Mechanics*, Vol. 135 of Faraday Discussions, edited by P.L.A. Popelier (The Royal Society of Chemistry, Cambridge, 2007), pp. 1–516.
- [2] G.N. Lewis, *J. Am. Chem. Soc.* **38**(4), 762–785 (1916).
- [3] E. Schrödinger, *Ann. Phys.* **385**(13), 437–490 (1926).
- [4] R.J. Gillespie and R.S. Nyholm, *Q. Rev. Chem. Soc.* **11**(4), 339–380 (1957).
- [5] R.J. Gillespie, *J. Chem. Educ.* **47**(1), 18–23 (1970).
- [6] R.S. Mulliken, *J. Chem. Phys.* **23**(10), 1833–1840 (1955).
- [7] F.L. Hirshfeld, *Theor. Chim. Acta* **44**(2), 129–138 (1977).
- [8] R.F.W. Bader, *Chem. Rev.* **91**(5), 893–928 (1991).
- [9] P. Bultinck and P. Popelier, in *Chemical Reactivity Theory, A Density Functional View*, edited by P.K. Chattaraj (CRC Press, Boca Raton, FL, 2009), pp. 215–227.
- [10] R. Ponec and M. Strnad, *Int. J. Quant. Chem.* **50**(1), 43–53 (1994).
- [11] K.B. Wiberg, *Tetrahedron* **24**(3), 1083–1096 (1968).
- [12] X. Fradera, M.A. Austen, and R.F.W. Bader, *J. Phys. Chem. A* **103**(2), 304–314 (1999).
- [13] R. Ponec and D.L. Cooper, *J. Mol. Struct.: THEOCHEM* **727**(1), 133–138 (2005).
- [14] M. Giambiagi, M.S. de Giambiagi, and K.C. Mundim, *Struct. Chem.* **1**(5), 423–427 (1990).
- [15] P. Bultinck, R. Ponec, and S. van Damme, *J. Phys. Org. Chem.* **18**(8), 706–718 (2005).
- [16] R. Ponec, *J. Math. Chem.* **21**(3), 323–333 (1997).
- [17] E. Francisco, Á.M. Pendás, and M.A. Blanco, *J. Chem. Phys.* **126**(9), 094102 (2007).
- [18] Á.M. Pendás, E. Francisco, and M.A. Blanco, *Phys. Chem. Chem. Phys.* **9**(9), 1087–1092 (2007).
- [19] Á.M. Pendás, E. Francisco, and M.A. Blanco, *J. Chem. Phys.* **127**(14), 144103 (2007).
- [20] Á.M. Pendás, E. Francisco, and M.A. Blanco, *J. Phys. Chem. A* **111**(6), 1084–1090 (2007).
- [21] E. Francisco, Á.M. Pendás, and M.A. Blanco, *Comput. Phys. Commun.* **178**(8), 621–634 (2008).
- [22] E. Francisco and Á.M. Pendás, *Comput. Phys. Commun.* **185**, 2663 (2014).
- [23] A. Savin, in *Reviews of Modern Quantum Chemistry: A Celebration of the Contributions of Robert G. Parr*, edited by K.D. Sen (World Scientific, Singapore, 2002), Vol. 1, pp. 43–62.
- [24] R. Daudel, H. Brion, and S. Odier, *J. Chem. Phys.* **23**(11), 2080–2083 (1955).
- [25] P.W. Ayers, *J. Chem. Sci.* **117**(5), 441–454 (2005).
- [26] M. Kohout, *Int. J. Quant. Chem.* **97**(1), 651–658 (2004).
- [27] A. Savin, *J. Phys. Chem. Solid.* **65**(12), 2025–2029 (2004).
- [28] E. Cancès, R. Keriven, F. Lodier, and A. Savin, *Theor. Chem. Acc.* **111**(2–6), 373–380 (2004).
- [29] E. Chamorro, P. Fuentealba, and A. Savin, *J. Comput. Chem.* **24**(4), 496–504 (2003).
- [30] A. Gallegos, R. Carbó-Dorca, F. Lodier, E. Cancès, and A. Savin, *J. Comput. Chem.* **26**(5), 455–460 (2005).
- [31] A. Scemama, *J. Theor. Comput. Chem.* **4**(02), 397–409 (2005).
- [32] A. Scemama, M. Caffarel, and A. Savin, *J. Comput. Chem.* **28**(1), 442–454 (2007).
- [33] M. Causà and A. Savin, *J. Phys. Chem. A* **115**(45), 13139–13148 (2011).
- [34] M. Causà and A. Savin, *Z. Anorg. Allg. Chem.* **637**(7–8), 882–884 (2011).
- [35] M. Causà, M. D’Amore, C. Garzillo, F. Gentile, and A. Savin, in *Applications of Density Functional Theory to Biological and Bioinorganic Chemistry*, edited by M.V. Putz and D.M.P. Mingos (Springer-Verlag, Berlin, 2013), pp. 119–141.
- [36] M. Menéndez, Á.M. Pendás, B. Braïda, and A. Savin, *Comput. Theor. Chem.* **1053**, 142 (2015).
- [37] M. Causà, M. D’Amore, F. Gentile, M. Menendez, and M. Calatayud, *Comput. Theor. Chem.* **1053**, 315 (2015).
- [38] O.M. Lopes Jr., B. Braïda, M. Causà, and A. Savin, in *Advances in the Theory of Quantum Systems in Chemistry and Physics*, edited by P.E. Hoggan, E.J. Brändas, J. Maruani, P. Piecuch, and G. Delgado-Barrio (Springer, Dordrecht, 2012), pp. 173–184.
- [39] M. Menéndez and Á.M. Pendás, *Theor. Chem. Acc.* **133**(9), 1–8 (2014).
- [40] J. Hubbard, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (The Royal Society, London, 1963), Vol. 276, pp. 238–257.
- [41] E. Hückel, *Z. Phys.* **60**(7–8), 423–456 (1930).
- [42] N.C. Handy, *Chem. Phys. Lett.* **74**(2), 280–283 (1980).
- [43] J. Olsen, B.O. Roos, P. Jørgensen, and J.H. Jensen, *J. Chem. Phys.* **89**(4), 2185–2192 (1988).
- [44] R.W. Hamming, *Bell Syst. Tech. J.* **29**(2), 147–160 (1950).
- [45] R.F.W. Bader, A. Streitwieser, A. Neuhaus, K.E. Laidig, and P. Speers, *J. Am. Chem. Soc.* **118**(21), 4959–4965 (1996).
- [46] R.F.W. Bader and M.E. Stephens, *J. Am. Chem. Soc.* **97**(26), 7391–7399 (1975).

- [47] T. Thorsteinsson, D.L. Cooper, J. Gerratt, P.B. Karadakov, and M. Raimondi, *Theor. Chim. Acta* **93**(6), 343–366 (1996).
- [48] T. Thorsteinsson and D.L. Cooper, *Theor. Chim. Acta* **94**(4), 233–245 (1996).
- [49] W. Wu, D. Danovich, A. Shurki, and S. Shaik, *J. Phys. Chem. A* **104**(38), 8744–8758 (2000).

Appendix A: Generating function

A.1. Factorisation of the generating function

In the present appendix, we will make abstraction of the spin- σ variable. Before proving the factorisation of the generating function, it is instructive to consider a simple recurrence relation at the operator level. Assume that we construct a domain $\Omega \cup \{i\}$ by adding a single site i to an existing domain Ω , then one has

$$\hat{P}(v, \Omega \cup \{i\}) = \hat{P}(v, \Omega)\hat{P}(0, \{i\}) + \hat{P}(v-1, \Omega)\hat{P}(1, \{i\}). \quad (\text{A1})$$

The proof goes via the explicit definition of the probability domain operator (25). Indeed, using the permutation symmetry among occupied (O_i) and unoccupied (U_i) operators, one can decompose the sum over all permutations into permutations restricted within the smaller domain and the single site i . Note that for the special case of $v = 0$, the recurrence relation simplifies to

$$\hat{P}(0, \Omega \cup \{i\}) = \hat{P}(0, \Omega)\hat{P}(0, \{i\}). \quad (\text{A2})$$

The recurrence relation (A1) has a nice classical interpretation. The probability of finding exactly v particles within an extended domain $\Omega \cup \{i\}$ is indeed the probability of finding exactly v particles in Ω and zero particles at the site i , plus the probability of finding $v-1$ particles in Ω and the remaining particle at site i . It is worth stressing though that this interpretation is strictly classical and is not necessarily true in general at the quantum level

$$P(v, \Omega \cup \{i\}) = \langle \Psi | [\hat{P}(v, \Omega)\hat{P}(0, \{i\}) + \hat{P}(v-1, \Omega)\hat{P}(1, \{i\})] | \Psi \rangle, \quad (\text{A3})$$

$$\neq \langle \Psi | \hat{P}(v, \Omega) | \Psi \rangle \langle \Psi | \hat{P}(0, \{i\}) | \Psi \rangle + \langle \Psi | \hat{P}(v-1, \Omega) | \Psi \rangle \langle \Psi | \hat{P}(1, \{i\}) | \Psi \rangle, \quad (\text{A4})$$

$$= P(v, \Omega)P(0, \{i\}) + P(v-1, \Omega)P(1, \{i\}). \quad (\text{A5})$$

The equality only holds when the quantum state $|\Psi\rangle$ is an eigenstate of the projection operators (17), which means that the resolution of the identity is equivalent to taking the expectation values of the projection operators. This is explicitly the case for Slater determinant states in the site basis (9).

We are now at the position of proving the factorisation of the generating function at the operator level

$$\hat{G}(\Omega, t) = \sum_{v=0}^M \hat{P}(v, \Omega) t^v = \prod_{k=1}^M (\hat{U}_{i_k} + t \hat{O}_{i_k}). \quad (\text{A6})$$

The proof goes by induction. Assume that the factorisation is valid for domains Ω of size M . Adding one additional site i gives rise to a domain $\Omega \cup \{i\}$ of size $M+1$. The generating function for the latter domain is

$$\hat{G}(\Omega \cup \{i\}, t) = \sum_{v=0}^{M+1} \hat{P}(v, \Omega \cup \{i\}) t^v \quad (\text{A7})$$

$$= \sum_{v=0}^{M+1} [\hat{P}(v, \Omega)\hat{P}(0, \{i\}) + \hat{P}(v-1, \Omega)\hat{P}(1, \{i\})] t^v, \quad (\text{A8})$$

in which we have used the recurrence relation (A1). Careful reorganisation of both summation terms leads to

$$\hat{G}(\Omega \cup \{i\}, t) = \sum_{v=0}^M \hat{P}(v, \Omega) t^v (\hat{P}(0, \{i\}) + t \hat{P}(1, \{i\})) \quad (\text{A9})$$

$$= \hat{G}(\Omega, t) (\hat{U}_i + t \hat{O}_i). \quad (\text{A10})$$

If we label site i as i_{M+1} , the factorisation is proven.

It is possible to condense the generating function even further in notation using the Taylor expansion of the projection operator

$$e^{\alpha \hat{O}_i} = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} \hat{O}_i^k = 1 + \sum_{k=1}^{\infty} \frac{\alpha^k}{k!} \hat{O}_i = 1 + (e^\alpha - 1) \hat{O}_i = \hat{U}_i + e^\alpha \hat{O}_i. \quad (\text{A11})$$

Identifying $e^\alpha = t$, we can rewrite the generating function (A6) in exponential form

$$\hat{G}(\Omega, t) = \prod_{k=1}^M t^{\hat{O}_{i_k}} = t^{\sum_{k=1}^M \hat{O}_{i_k}} = t^{\hat{n}(\Omega)}, \quad (\text{A12})$$

with $\hat{n}(\Omega)$ the number operator counting the number of particles contained within the domain Ω .

A.2. Generating function for Slater determinants

We will again make abstraction of the spin- σ variable and consider a Slater determinant wave function with N occupied orthonormal single-particle states

$$|\Phi\rangle = \prod_{\alpha=1}^N \left(\sum_{i=1}^L q_{i\alpha} \hat{a}_i^\dagger \right) |\theta\rangle, \quad (\text{A13})$$

in which $q_{i\alpha}$ denotes the unitary transformation from the site basis to the Slater determinant basis, with $\sum_{i=1}^L q_{i\alpha}^* q_{i\beta} = \delta_{\alpha\beta}$. It is now possible to evaluate the generating function

$$\begin{aligned} \langle \Phi | t^{\hat{n}(\Omega)} | \Phi \rangle &= \sum_{\substack{j_1, \dots, j_N \\ k_1, \dots, k_N}}^L q_{j_1 \alpha_1}^* \dots q_{j_N \alpha_N}^* q_{k_1 \alpha_1} \dots q_{k_N \alpha_N} \\ &\quad \times \langle \theta | a_{j_N} \dots \hat{a}_{j_1} t^{\hat{n}(\Omega)} \hat{a}_{k_1}^\dagger \dots \hat{a}_{k_N}^\dagger | \theta \rangle \\ &= \sum_{j_1 \dots j_N}^L \sum_{\pi \in S_N} (-1)^\pi q_{j_1 \alpha_1}^* \dots q_{j_N \alpha_N}^* q_{\pi(j_1) \alpha_1} \dots q_{\pi(j_N) \alpha_N} t^{n_\Omega(j_1 \dots j_N)}, \end{aligned} \quad (\text{A14})$$

$$(\text{A15})$$

with S_N the group of permutations of $[j_1 \dots j_N]$, $n_\Omega(j_1 \dots j_N)$ denoting the number of sites of Ω in $[j_1 \dots j_N]$, and where we have used the fact that $t^{\hat{n}(\Omega)}$ is diagonal in the site basis, with the fermionic anticommutation relations resulting in a phase factor. This can be further simplified to

$$\begin{aligned} \langle \Phi | t^{\hat{n}(\Omega)} | \Phi \rangle &= \sum_{j_1 \dots j_N}^L \sum_{\pi \in S_N} (-1)^\pi q_{j_1 \alpha_1}^* \dots q_{j_N \alpha_N}^* \\ &\quad q_{j_1 \pi(\alpha_1)} \dots q_{j_N \pi(\alpha_N)} t^{n_\Omega(j_1 \dots j_N)} \quad (\text{A16}) \\ &= \sum_{j_1 \dots j_N}^L \det [Q(j_1 \dots j_N)] t^{n_\Omega(j_1 \dots j_N)}, \quad (\text{A17}) \end{aligned}$$

with $Q(j_1 \dots j_N)$ an $N \times N$ matrix with matrix elements $Q(j_1 \dots j_N)_{ik} = q_{j_i \alpha_i}^* q_{j_i \alpha_k}$. This can be seen as the expansion of a single determinant, where each column contains a summation over j_i and is given a weight t if $j_i \in \Omega$ and a weight 1 if $j_i \notin \Omega$. This results in

$$\langle \Phi | t^{\hat{n}(\Omega)} | \Phi \rangle = \det \{ S(\bar{\Omega}) + tS(\Omega) \}, \quad (\text{A18})$$

where we have introduced two $N \times N$ matrices defined as

$$S(\bar{\Omega})_{jk} = \sum_{i \notin \Omega} q_{i\alpha_j}^* q_{i\alpha_k}, \quad S(\Omega)_{jk} = \sum_{i \in \Omega} q_{i\alpha_j}^* q_{i\alpha_k}. \quad (\text{A19})$$

The orthonormality condition results in $S(\bar{\Omega}) + S(\Omega) = I_N$, which allows us to rewrite

$$\langle \Phi | t^{\hat{n}(\Omega)} | \Phi \rangle = \det \{ [I_N - S(\Omega)] + tS(\Omega) \}. \quad (\text{A20})$$

This is the first determinant expression to be proven, which can be further rewritten by noting that $S(\Omega) = q(\Omega)^\dagger q(\Omega)$, with $q(\Omega)$ an $M \times N$ matrix defined as $q(\Omega)_{i\alpha} = q_{i\alpha}$, with $i \in \Omega$. From Sylvester's determinant theorem, we have

$$\begin{aligned} &\det \{ I_N + (t-1)q(\Omega)^\dagger q(\Omega) \} \\ &= \det \{ I_M + (t-1)q(\Omega)q(\Omega)^\dagger \}, \end{aligned} \quad (\text{A21})$$

resulting in the second determinant expression for the generating function.