

# Decay Properties of Spectral Projectors with Applications to Electronic Structure\*

Michele Benzi<sup>†</sup>  
Paola Boito<sup>‡</sup>  
Nader Razouk<sup>§</sup>

**Abstract.** Motivated by applications in quantum chemistry and solid state physics, we apply general results from approximation theory and matrix analysis to the study of the decay properties of spectral projectors associated with large and sparse Hermitian matrices. Our theory leads to a rigorous proof of the exponential off-diagonal decay (“nearsightedness”) for the density matrix of gapped systems at zero electronic temperature in both orthogonal and nonorthogonal representations, thus providing a firm theoretical basis for the possibility of linear scaling methods in electronic structure calculations for nonmetallic systems. We further discuss the case of density matrices for metallic systems at positive electronic temperature. A few other possible applications are also discussed.

**Key words.** electronic structure, localization, density functional theory, density matrix, spectral gap, matrix function, orthogonal projector

**AMS subject classifications.** Primary, 65F60, 65F50, 65N22; Secondary, 81Q05, 81Q10

**DOI.** 10.1137/100814019

**I. Introduction.** The physical and chemical properties of materials are largely determined by the electronic structure of the atoms and molecules found within them. In all but the simplest cases, the electronic structure can only be determined approximately, and since the late 1920s a huge amount of work has been devoted to finding suitable approximations and numerical methods for solving this fundamental problem. Traditional methods for electronic structure computations are based on the solution of generalized eigenvalue problems (“diagonalization”) for a sequence of large Hermitian matrices, known as one-particle Hamiltonians. The computational cost of this approach scales cubically in the size  $n$  of the problem, which is in turn determined by the number of electrons in the system. For large systems, the costs become prohibitive; this is often referred to as “the  $O(n^3)$  bottleneck” in the literature.

\*Received by the editors November 8, 2010; accepted for publication (in revised form) April 27, 2012; published electronically February 7, 2013. This work was supported by National Science Foundation grants DMS-0810862 and DMS-1115692 and by a grant of the University Research Committee of Emory University.

<http://www.siam.org/journals/sirev/55-1/81401.html>

<sup>†</sup>Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322 (benzi@mathcs.emory.edu).

<sup>‡</sup>Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322. Current address: DMI-XLIM, UMR 7252, Université de Limoges–CNRS, 123 avenue Albert Thomas, 87060 Limoges Cedex, France (paola.boito@unilim.fr).

<sup>§</sup>Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322. Current address: Ernst & Young GmbH Wirtschaftsprüfungsgesellschaft, Arnulfstraße 59, 80636 München, Germany (nrazouk@gmail.com).

In the last two decades, a number of researchers have developed approaches that are capable in many cases of achieving “optimal” computational complexity: the computational effort scales linearly in the number of electrons, leading to better performance for sufficiently large systems and making the electronic structure problem tractable for large-scale systems. These methods, often referred to as “ $O(n)$  methods,” apply mostly to insulators. They avoid diagonalization by computing instead the *density matrix*, a matrix which encodes all the important physical properties of the system. For insulators at zero temperature, this is the spectral projector onto the invariant subspace associated with the eigenvalues of the Hamiltonian falling below a certain value. For systems at positive temperatures, the density matrix can be expressed as a smooth function of the Hamiltonian.

The possibility of developing such methods rests on a deep property of electronic matter, called “nearsightedness” by W. Kohn [75]. Kohn’s “Nearsightedness Principle” expresses the fact that for a large class of systems the effects of disturbances, or perturbations, remain localized and thus do not propagate beyond a certain (finite) range; in other words, far away parts of the system do not “see” each other. Mathematically, this property translates into rapid off-diagonal decay in the density matrix. This fast fall-off in the density matrix entries has been often assumed without proof, or proved only in special cases. Moreover, the precise dependence of the rate of decay on properties of the system (such as the band gap in insulators or the temperature in metallic systems) has been the subject of much discussion.

The main goal of this paper is to provide a rigorous mathematical foundation for linear scaling methods in electronic structure computations. We do this by deriving estimates, in the form of decay bounds, for the entries of general density matrices for insulators and for metallic systems at positive electronic temperatures. We also address the question of the dependence of the rate of decay on the band gap and on the temperature. Although immediately susceptible to physical interpretation, our treatment is purely mathematical. By stripping the problem down to its essential features and working at the discrete level, we are able to develop an abstract theory covering nearly all types of systems and discretizations encountered in actual electronic structure problems.

Our results are based on a general theory of decay for the entries in analytic functions of sparse matrices, initially proposed in [12, 14, 106] and further developed here. The theory is based on classical approximation theory and matrix analysis. A bit of functional analysis is used when considering a simple model of “metallic behavior,” for which the decay in the density matrix is very slow.

The approach described in this paper has a number of potential applications beyond electronic structure computations, and can be applied to any problem involving functions of large matrices where “locality of interaction” plays a role. Toward the end of the paper we briefly review the possible use of decay bounds in the study of correlations in quantum statistical mechanics and information theory, in the analysis of complex networks, and in some classical problems in numerical linear algebra, like the computation of invariant subspaces of symmetric tridiagonal matrices. The discussion of these topics will be necessarily brief, but we hope it will stimulate further work in these areas.

In this paper we are mostly concerned with the theory behind  $O(n)$  methods rather than with specific algorithms. Readers who are interested in the computational aspects should consult any of the many recent surveys on algorithms for electronic structure computations; among them, [20, 97, 113, 116] are especially recommended.

The remainder of the paper is organized as follows. Section 2 provides some background on electronic structure theory. The formulation of the electronic structure problem in terms of spectral projectors is reviewed in section 3. A survey of previous, related work on decay estimates for density matrices is given in section 4. In section 5 we formulate our basic assumptions on the matrices (discrete Hamiltonians) considered in this paper, particularly their normalization and asymptotic behavior for increasing system size ( $n \rightarrow \infty$ ). The approximation (truncation) of matrices with decay properties is discussed in section 6. A few general properties of orthogonal projectors are established in section 7. The core of the paper is represented by section 8, where various types of decay bounds for spectral projectors are stated and proved. In section 9 we discuss the transformation to an orthonormal basis set. The case of vanishing gap is discussed in section 10. Other applications of our results and methods are mentioned in section 11. Finally, concluding remarks and some open problems are given in section 12.

**2. Background on Electronic Structure Theories.** In this section we briefly discuss the basic principles underlying electronic structure theory. For additional details the reader is referred to, e.g., [22, 77, 88, 89, 116, 125].

Consider a physical system formed by a number of nuclei and  $n_e$  electrons in three-dimensional (3D) space. The time-independent Schrödinger equation for the system is the eigenvalue problem

$$(2.1) \quad \mathcal{H}_{\text{tot}} \Psi_{\text{tot}} = E_{\text{tot}} \Psi_{\text{tot}},$$

where  $\mathcal{H}_{\text{tot}}$  is the many-body Hamiltonian operator,  $E_{\text{tot}}$  is the total energy, and the functions  $\Psi_{\text{tot}}$  are the eigenstates of the system.

The Born–Oppenheimer approximation allows us to separate the nuclear and electronic coordinates. As a consequence, we only seek to solve the quantum mechanical problem for the electrons, considering the nuclei as sources of external potential. Then the electronic part of (2.1) can be written as

$$(2.2) \quad \mathcal{H} \Psi = E \Psi,$$

where  $E$  is the electronic energy and the eigenstates  $\Psi$  are functions of  $3n_e$  spatial coordinates and  $n_e$  (discrete) spin coordinates.

We denote spatial coordinates as  $\mathbf{r}$  and the spin coordinate as  $\sigma$ ; each electron is then defined by 3 + 1 coordinates  $\mathbf{x}_i = \begin{pmatrix} \mathbf{r}_i \\ \sigma_i \end{pmatrix}$ , and wavefunctions are denoted as  $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_{n_e})$ . Then the electronic Hamiltonian operator in (2.2) can be written as

$$\mathcal{H} = T + V_{\text{ext}} + V_{\text{ee}},$$

where  $T = -\frac{1}{2} \nabla^2$  is the kinetic energy,  $V_{\text{ext}}$  is the external potential (i.e., the potential due to the nuclei), and  $V_{\text{ee}} = \frac{1}{2} \sum_{i \neq j}^{n_e} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$  is the potential due to the electron-electron repulsion.<sup>1</sup> Moreover, the ground-state energy is given by

$$E_0 = \min_{\Psi} \langle \mathcal{H} \Psi, \Psi \rangle,$$

where the minimum is taken over all the normalized antisymmetric wavefunctions (electrons being Fermions, their wavefunction is antisymmetric). The electronic den-

<sup>1</sup>As is customary in physics, we use here atomic units, that is,  $e^2 = \hbar = m = 1$ , with  $e =$  electronic charge,  $\hbar =$  reduced Planck's constant, and  $m =$  electronic mass.

sity is defined as

$$\rho(\mathbf{r}) = n_e \sum_{\sigma} \int d\mathbf{x}_2 \cdots \int d\mathbf{x}_{n_e} |\Psi(\mathbf{r}, \sigma, \mathbf{x}_2, \dots, \mathbf{x}_{n_e})|^2.$$

In this expression, the sum over  $\sigma$  is the sum over the spin values of the first electron, while integration with respect to  $\mathbf{x}_i$ , with  $2 \leq i \leq n_e$ , denotes the integral over  $\mathbb{R}^3$  and sums over both possible spin values for the  $i$ th electron.

Observe that (2.2) is a many-particle equation that cannot be separated into several one-particle equations because of the term  $V_{ee}$ . Of course, being able to turn (2.2) into a separable equation would simplify the problem considerably, since the number of unknowns per equation would drop from  $3n_e + n_e$  to  $3 + 1$ . This is the motivation for one-electron methods.

For noninteracting particles, the many-body eigenstates  $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_{n_e})$  can be written as Slater determinants of occupied orbitals  $\phi_1(\mathbf{x}_1), \dots, \phi_{n_e}(\mathbf{x}_{n_e})$ ,

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_{n_e}) = \frac{1}{\sqrt{n_e!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \cdots & \phi_{n_e}(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_{n_e}) & \cdots & \phi_{n_e}(\mathbf{x}_{n_e}) \end{vmatrix},$$

where each orbital satisfies a single-particle eigenstate equation  $\mathcal{H}_i \phi_i = E_i \phi_i$ . In general, the name “one-particle method” is used also when self-consistent terms (e.g., involving the density) are present in  $\mathcal{H}_i$ ; in this case, the equations are solved iteratively, computing at each step the solution to a single-particle problem and then filling the lowest eigenstates with one electron each, to form a Slater determinant. However, some of the properties of a true noninteracting system (such as the fact that the energy is the sum of the eigenvalues of occupied states) are lost.

A fundamental example of the one-particle method is density functional theory (DFT). The main idea behind DFT consists in rewriting the ground-state energy as a density functional rather than a wavefunction functional. Indeed, the first Hohenberg–Kohn theorem [66] states that the potential is uniquely (up to a constant) determined by the ground-state density  $\rho(\mathbf{r})$ . In other words, the system can be seen as characterized by the density rather than by the potential. Moreover, the ground-state density of a system with given external potential can be computed by minimizing a suitable energy functional of  $\rho$  (second Hohenberg–Kohn theorem).

While of crucial theoretical importance, though, these results do not give a recipe for computing electronic structures. The next important step comes with the Kohn–Sham construction [76]: roughly speaking, one replaces the original, nonseparable system with a fictitious system of noninteracting electrons that have exactly the same density as the original system. The single-particle equations for the Kohn–Sham system are (neglecting spin)

$$\left( -\frac{1}{2} \nabla^2 + V(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}),$$

where the  $\psi_i$ 's are the Kohn–Sham orbitals and  $V(\mathbf{r})$  is the single-electron potential. The associated density is

$$\rho(\mathbf{r}) = \sum_{i=1}^{n_e} |\psi_i(\mathbf{r})|^2.$$

The single-particle potential  $V(\mathbf{r})$  can be written as

$$V(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + \int_{\mathbb{R}^3} \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + V_{xc}[\rho](\mathbf{r}),$$

where the term  $V_{xc}[\rho](\mathbf{r})$  is called the *exchange-correlation potential* and depends on the density. It is important to point out that the Kohn–Sham construction is not an approximation, in that the Kohn–Sham equations are exact and yield the exact density.

On the other hand, the exchange-correlation energy is not known in practice and needs to be approximated. In the local density approximation (LDA) framework, for instance, the exchange energy is based on the energy of a uniform electron gas. Introducing spin allows for a more refined approximation (LSDA, or local spin-density approximation). One may also include gradient corrections, thus obtaining the so-called generalized gradient approximation (GGA).

The solution of the Kohn–Sham equations is usually computed via self-consistent iterations. The iterative process begins with an approximation of the density; the associated approximate exchange-correlation potential is injected into the Kohn–Sham equations. The output density is then used to form a new approximation of the potential. The process continues until the update term for the density or the potential becomes negligible. Observe that the basic building block of this computational technique is the solution of an eigenvalue problem for noninteracting particles.

Electrons at the lowest atomic-like levels (“core” electrons) do not change their state much within chemical processes. For this reason, many computational techniques do not consider them explicitly, and instead replace the Coulomb attraction of the nucleus with a potential (called pseudopotential) that includes the effect of the core electrons on the valence electrons. This approach is always employed when using plane waves as a basis for wavefunctions, since the number of plane waves required to represent core electrons is prohibitive.

**3. Density Matrices.** As mentioned earlier, conventional methods for electronic structure calculations require the repeated solution of linear eigenvalue problems for a one-electron Hamiltonian operator of the form  $\mathcal{H} = -\frac{1}{2}\nabla^2 + V(\mathbf{r})$ . In practice, operators are discretized by grid methods or via Galerkin projection onto the finite-dimensional subspace spanned by a set of basis functions  $\{\phi_i\}_{i=1}^n$ . When linear combinations of atom-centered Slater- or Gaussian-type functions (see below) are employed, the total number of basis functions is  $n \approx n_b \cdot n_e$ , where  $n_e$  is the number of (valence) electrons in the system and  $n_b$  is a small or moderate integer related to the number of basis functions per atom. Traditional electronic structure algorithms diagonalize the discrete Hamiltonian, resulting in algorithms with  $O(n_e^3)$  (equivalently,  $O(n^3)$ ) operation count [77, 89, 116]. In these approaches, a sequence of generalized eigenproblems of the form

$$(3.1) \quad H\psi_i = \varepsilon_i S\psi_i, \quad 1 \leq i \leq n_e,$$

is solved, where  $H$  and  $S$  are, respectively, the discrete Hamiltonian and the overlap matrix relative to the basis set  $\{\phi_i\}_{i=1}^n$ . The eigenvectors  $\psi_i$  in (3.1) are known as the *occupied states* and correspond to the  $n_e$  lowest generalized eigenvalues  $\varepsilon_1 \leq \dots \leq \varepsilon_{n_e}$ , the *occupied levels*. The overlap matrix  $S$  is just the Gram matrix associated with the basis set,  $S_{ij} = \langle \phi_j, \phi_i \rangle$  for all  $i, j$ , where  $\langle \cdot, \cdot \rangle$  denotes the standard  $L^2$ -inner product. In Dirac’s *bra-ket* notation, which is the preferred one in the physics and chemistry

literature, one writes  $S_{ij} = \langle \phi_i | \phi_j \rangle$ . For an orthonormal basis set,  $S = I_n$  (the  $n \times n$  identity matrix) and the eigenvalue problem (3.1) is a standard one.

Instead of explicitly diagonalizing the discretized Hamiltonian  $H$ , one may reformulate the problem in terms of the density operator  $P$ , which is the  $S$ -orthogonal projector<sup>2</sup> onto the  $H$ -invariant subspace corresponding to the occupied states, that is, the subspace spanned by the  $n_e$  eigenvectors  $\psi_i$  in (3.1). Virtually all quantities of interest in electronic structure theory can be computed as functionals of the density matrix  $P$ ; see, e.g., [24, 95, 97]. It is this reformulation of the problem that allows for the development of potentially more efficient algorithms for electronic structure, including algorithms that asymptotically require only  $O(n_e)$  (equivalently,  $O(n)$ ) arithmetic operations and storage. Most current methodologies, including Hartree–Fock, DFT (e.g., Kohn–Sham), and hybrid schemes (like BLYP), involve self-consistent field (SCF) iterations, in which the density matrix  $P$  must be computed at each SCF step, typically with increasing accuracy as the outer iteration converges; see, e.g., [77, 137].

As stated in section 1, in this paper we use some classical results from polynomial approximation theory and matrix analysis to provide a mathematical foundation for linear scaling electronic structure calculations for a very broad class of systems. We assume that the basis functions  $\phi_i$  are localized, i.e., decay rapidly outside of a small region. Many of the most popular basis sets used in quantum chemistry, such as Gaussian-type orbitals, which are functions of the form

$$\phi(x, y, z) = C x^{n_x} y^{n_y} z^{n_z} e^{-\alpha r^2},$$

where  $C$  is a normalization constant, satisfy this requirement [77]. For systems with sufficient separation between atoms, this property implies a fast off-diagonal decay of the entries of the Hamiltonian matrix; moreover, a larger distance between atoms corresponds to a faster decay of matrix entries [77, p. 381]. If the entries that fall below a given (small) truncation tolerance are set to zero, the Hamiltonian turns out to be a sparse matrix.

Decay results are especially easy to state in the banded case,<sup>3</sup> but more general sparsity patterns will be taken into account as well.

We can also assume from the outset that the basis functions form an orthonormal set. If this is not the case, we perform a congruence transformation to an orthogonal basis and replace the original Hamiltonian  $H$  with  $\tilde{H} = Z^T H Z$ , where  $S^{-1} = Z Z^T$  is either the Löwdin ( $Z = S^{-1/2}$ , [85]) or the inverse Cholesky ( $Z = L^{-T}$ , with  $S = L L^T$ ) factor of the overlap matrix  $S$ ; see, e.g., [24]. Here  $Z^T$  denotes the transpose of  $Z$ ; for the Löwdin factorization,  $Z$  is symmetric ( $Z = Z^T$ ). Up to truncation, the transformed matrix  $\tilde{H}$  is still a banded (or sparse) matrix, albeit not as sparse as  $H$ . Hence, in our decay results we can replace  $H$  with  $\tilde{H}$ . The entries in  $S^{-1}$ , and therefore those in  $Z$ , decay at a rate which depends on the conditioning of  $S$ . This, in turn, will depend on the particular basis set used, on the total number of basis functions, and on the interatomic distances, with larger separations leading to faster decay. This is discussed further in section 9. We note that the case of tight-binding Hamiltonians is covered by our theory. Indeed, the tight-binding method consists in expanding the states of the physical system (e.g., a crystal) in linear combinations of atomic orbitals of the composing atoms; such an approximation is successful if the

<sup>2</sup>That is, orthogonal with respect to the inner product associated with  $S$ .

<sup>3</sup>A square matrix  $A = (A_{ij})$  is said to be  $m$ -banded if  $A_{ij} = 0$  whenever  $|i - j| > m$ ; for instance, a tridiagonal matrix is 1-banded according to this definition.

atomic orbitals have little overlap, which translates to a sparse Hamiltonian. The same applies to “real space” finite difference (or finite element) approximations [116].

For a given sparse discrete Hamiltonian  $H$  in an orthonormal basis, we consider the problem of approximating the zero-temperature density matrix associated with  $H$ , that is, the spectral projector  $P$  onto the occupied subspace spanned by the eigenvectors corresponding to the smallest  $n_e$  eigenvalues of  $H$ :

$$P = \psi_1 \otimes \psi_1 + \cdots + \psi_{n_e} \otimes \psi_{n_e} \equiv |\psi_1\rangle\langle\psi_1| + \cdots + |\psi_{n_e}\rangle\langle\psi_{n_e}|,$$

where  $H\psi_i = \varepsilon_i\psi_i$  for  $i = 1, \dots, n_e$ . Clearly,  $P$  is Hermitian and idempotent:  $P = P^* = P^2$ . Consider now the Heaviside (step) function

$$h(x) = \begin{cases} 1 & \text{if } x < \mu, \\ \frac{1}{2} & \text{if } x = \mu, \\ 0 & \text{if } x > \mu, \end{cases}$$

where the number  $\mu$  (sometimes called the *Fermi level* or *chemical potential* [53]) is such that  $\varepsilon_{n_e} < \mu < \varepsilon_{n_e+1}$ . If the *spectral gap*  $\gamma = \varepsilon_{n_e+1} - \varepsilon_{n_e}$ , also known as the HOMO-LUMO gap,<sup>4</sup> is not too small, the step function  $h$  is well approximated by the Fermi–Dirac function<sup>5</sup>  $f_{FD}(x) = 1/(1 + e^{\beta(x-\mu)})$  for suitable values of  $\beta > 0$ :

$$P = h(H) \approx f_{FD}(H) = [I_n + \exp(\beta(H - \mu I_n))]^{-1}.$$

The smaller  $\gamma$ , the larger  $\beta$  must be in order to have a good approximation: see Figure 8.8. The parameter  $\beta$  can be interpreted as an (artificial) inverse temperature; the zero-temperature limit is quickly approached as  $\beta \rightarrow \infty$ . A major advantage of the Fermi–Dirac function is that it is analytic; hence, we can replace  $h$  with  $f_{FD}$  and apply to it a wealth of results from approximation theory for analytic functions.

We emphasize that the study of the zero-temperature limit—that is, the ground state of the system—is of fundamental importance in electronic structure theory. In the words of [89, Chapter 2, pp. 11–12],

... the lowest energy ground state of the electrons determines the structure and low-energy motions of the nuclei. The vast array of forms of matter—from the hardest material known, diamond carbon, to the soft lubricant, graphite carbon, to the many complex crystals and molecules formed by the elements of the periodic table—are largely manifestations of the ground state of the electrons.

The Fermi–Dirac distribution is also used when dealing with systems at positive electronic temperatures ( $T > 0$ ) with a small or null gap (e.g., metallic systems); in this case  $\beta = (k_B T)^{-1}$ , where  $k_B$  is Boltzmann’s constant. In particular, use of the Fermi–Dirac function allows one to compute thermodynamical properties (such as the specific heat) and the  $T$ -dependence of quantities from first principles. In this case,

<sup>4</sup>HOMO = highest occupied molecular orbital; LUMO = lowest unoccupied molecular orbital.

<sup>5</sup>Several other analytic approximations to the step function are known, some of which are preferable to the Fermi–Dirac function from the computational point of view; see, e.g., [80] for a comparative study. For theoretical analysis, however, we find it convenient to work with the Fermi–Dirac function.

of course, the matrix  $P = f_{FD}(H)$  is no longer an orthogonal projector, not even approximately.

We mention in passing that it is sometimes advantageous to impose the normalization condition  $\text{Tr}(P) = 1$  on the density matrix; indeed, such a condition is standard and part of the definition of a density matrix in the quantum mechanics literature, beginning with von Neumann [131, 133]. At zero temperature we have  $\text{Tr}(P) = \text{rank}(P) = n_e$ , and  $P$  is replaced by  $\frac{1}{n_e}P$ . With this normalization  $P$  is no longer idempotent, except when  $n_e = 1$ . In this paper we do not make use of such normalization.

The localization (“pseudosparsity”) of the density matrix for insulators has been long known to physicists and chemists; see the literature review in the following section. A number of authors have exploited this property to develop a host of linear scaling algorithms for electronic structure computations; see, e.g., [4, 5, 20, 23, 24, 53, 54, 75, 79, 80, 89, 97, 98, 99, 113, 123, 135]. In this paper we derive explicitly computable decay bounds which can be used, at least in principle, to determine a priori the bandwidth or sparsity pattern of the truncation of the density matrix corresponding to a prescribed error. As we shall see, however, our decay estimates tend to be conservative and may be pessimistic in practice. Hence, we regard our results primarily as a theoretical contribution, providing a rigorous (yet elementary) mathematical justification for some important localization phenomena observed by physicists. An important aspect of our work is that our bounds are universal, in the sense that they only depend on the bandwidth (or sparsity pattern) of the discrete Hamiltonian  $H$ , on the smallest and largest eigenvalues of  $H$ , on the gap  $\gamma$ , and, when relevant, on the temperature  $T$ . In particular, our results are valid for a wide range of basis sets and indeed for different discretizations and representations of the Hamiltonian.

**4. Related Work.** The localization properties of spectral projectors (more generally, density matrices) associated with electronic structure computations in quantum chemistry and solid state physics have been the subject of a large number of papers. Roughly speaking, the results found in the literature fall into three broad categories:

1. Fully rigorous mathematical results for model systems (some quite general).
2. “Semirigorous” results for specific systems; these results are often characterized as “exact” or “analytical” by the authors (usually physicists), but would not be recognized as mathematically rigorous by mathematicians.
3. Nonrigorous results based on a mixture of heuristics, physical reasoning, and numerics.

Contributions in the first group are typically due to researchers working in solid state and mathematical physics. These include the pioneering works of Kohn [74] and des Cloizeaux [36], and the more recent papers by Nenciu [96], Brouder et al. [21], and a group of papers by Prodan, Kohn, and collaborators [103, 104, 105].

Before summarizing the content of these contributions, we should mention that nearly all the results found in the literature are expressed at the continuous level, that is, in terms of decay in *functions* rather than decay in *matrices*. The functions are typically functions of (real) space; results are often formulated in terms of the *density kernel*, but sometimes in terms of the *Wannier functions*. The latter form an orthonormal basis set associated with a broad class of Hamiltonians and are widely used in solid state physics. Since the Wannier functions span the occupied subspace, localization results for the Wannier functions immediately imply similar localization results for the corresponding spectral projector. Note, however, that the spectral projector may be exponentially localized even when the Wannier functions are not.



At the continuous level, the density matrix  $\rho : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$  is the kernel of the density operator  $\mathcal{P}$  defined by

$$(\mathcal{P}\psi)(\mathbf{r}) = \int_{\mathbb{R}^d} \rho(\mathbf{r}, \mathbf{r}')\psi(\mathbf{r}')d\mathbf{r}' ,$$

regarded as an integral operator on  $L^2(\mathbb{R}^d)$ ; here  $d = 1, 2, 3$ . The vectors  $\mathbf{r}$  and  $\mathbf{r}'$  represent any two points in  $\mathbb{R}^d$ , and  $|\mathbf{r} - \mathbf{r}'|$  is their (Euclidean) distance. The density kernel can be expressed as

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{i=1}^{n_e} \psi_i(\mathbf{r})\psi_i(\mathbf{r}')^* ,$$

where now  $\psi_i$  is the (normalized) eigenfunction of the Hamiltonian operator  $\mathcal{H}$  corresponding to the  $i$ th lowest eigenvalue,  $i = 1, \dots, n_e$ , and the asterisk denotes complex conjugation; see, e.g., [88]. The density operator  $\mathcal{P}$  admits the Dunford integral representation

$$(4.1) \quad \mathcal{P} = \frac{1}{2\pi i} \int_{\Gamma} (zI - \mathcal{H})^{-1} dz ,$$

where  $\Gamma$  is a simple closed contour in  $\mathbb{C}$  surrounding the eigenvalues of  $\mathcal{H}$  corresponding to the occupied states, with the remaining eigenvalues on the outside.

In [74], Kohn proved the rapid decay of the Wannier functions for one-dimensional (1D), one-particle Schrödinger operators with periodic and symmetric potentials with nonintersecting energy bands. This type of Hamiltonian describes 1D, centrosymmetric crystals. Kohn’s main result takes the form

$$(4.2) \quad \lim_{x \rightarrow \infty} w(x) e^{qx} = 0 ,$$

where  $w(x)$  denotes a Wannier function (here  $x$  is the distance from the center of symmetry) and  $q$  is a suitable positive constant. In the same paper (p. 820) Kohn also points out that for free electrons (not covered by his theory, which deals only with insulators) the decay is very slow, like  $x^{-1}$ .

A few observations are in order: First, the decay result (4.2) is asymptotic, that is, it implies fast decay at sufficiently large distances  $|x|$  only. Second, (4.2) is consistent not only with strict exponential decay, but also with decay of the form  $x^p e^{-q'x}$ , where  $p$  is arbitrary (positive or negative) and  $q' > q$ . Hence, the actual decay could be faster, but also slower, than exponential. Since the result in (4.2) provides only an estimate (rather than an upper bound) for the density matrix in real space, it is not easy to use in actual calculations. To be fair, such practical aspects were not discussed by Kohn until much later (see, e.g., [75]). Also, later work showed that the asymptotic regime is already achieved for distances of the order of 1–2 lattice constants, and it helped clarify the form of the power-law prefactor, as discussed below.

The techniques used by Kohn, mostly the theory of analytic functions in one complex variable and some classical asymptotics for linear second-order differential operators with variable coefficients, did not lend themselves naturally to the treatment of higher-dimensional cases or more complicated potentials. The problem of the validity of Kohn’s results in two and three dimensions has remained open for a very long time, and has been long regarded as one of the last outstanding problems of one-particle condensed-matter physics. Partial results were obtained by des Cloizeaux

[36] and much later by Nenciu [96]. Des Cloizeaux, who studied both the decay of the Wannier functions and that of the associated spectral projectors, extended Kohn's localization results to 3D insulators with a center of inversion (a specific symmetry requirement) in the special case of simple, isolated (i.e., nondegenerate) energy bands; he also treated the tight-binding limit for arbitrary crystals. Nenciu further generalized Kohn's results to arbitrary  $d$ -dimensional insulators, again limited to the case of simple bands.

The next breakthrough came much more recently, when Brouder et al. [21] managed to prove localization of the Wannier functions for a broad class of insulators in arbitrary dimensions. The potentials considered by these authors are sufficiently general for the results to be directly applicable to DFT, within both the LDA and the GGA frameworks. The results in [21], however, also prove that for *Chern insulators* (i.e., insulators for which the *Chern invariants*, which characterize the band structure, are nonvanishing) the Wannier functions do *not* decay exponentially, therefore leaving open the question of proving the decay of the density matrix in this case [129]. It should be mentioned that the mathematics in [21] is fairly sophisticated and requires some knowledge of modern differential geometry and topology.

Further papers of interest include the work by Prodan, Kohn, and collaborators [103, 104, 105]. From the mathematical standpoint, the most satisfactory results are perhaps those presented in [104]. In this paper, the authors use norm estimates for complex symmetric operators in Hilbert space to obtain sharp exponential decay estimates for the resolvents of rather general Hamiltonians with spectral gap. Using the contour integral representation formula (4.1), these estimates yield (for sufficiently large separations) exponential spatial decay bounds of the form

$$(4.3) \quad |\rho(\mathbf{r}, \mathbf{r}')| \leq C e^{-\alpha|\mathbf{r}-\mathbf{r}'|} \quad (C > 0, \alpha > 0, \text{ const.})$$

for a broad class of insulators. A lower bound on the decay rate  $\alpha$  (also known as the *decay length* or *inverse correlation length*) is derived, and the behavior of  $\alpha$  as a function of the spectral gap  $\gamma$  is examined.

Among the papers in the second group, we mention [52, 64, 70, 73, 90, 127, 128]. These papers provide quantitative decay estimates for the density matrix, based on either fairly rigorous analyses of special cases or not fully rigorous discussions of general situations. Significant use is made of approximations, asymptotics, heuristics, and physically motivated assumptions, and the results are often validated by numerical calculations. Also, it is occasionally stated that while the results were derived in the case of simplified models, the conclusions should be valid in general. Several of these authors emphasize the difficulty of obtaining rigorous results for general systems in arbitrary dimension. Despite not being fully rigorous from a mathematical point of view, these contributions are certainly very valuable and seem to have been broadly accepted by physicists and chemists. We note, however, that the results in these papers usually take the form of order-of-magnitude estimates for the density matrix  $\rho(\mathbf{r}, \mathbf{r}')$  in real space, valid for sufficiently large separations  $|\mathbf{r} - \mathbf{r}'|$ , rather than strict upper bounds. As said before of Kohn's results, this type of estimate may be difficult to use for computational purposes.

In the case of insulators, the asymptotic decay estimates in these papers take the form

$$(4.4) \quad \rho(\mathbf{r}, \mathbf{r}') = C \frac{e^{-\alpha|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r} - \mathbf{r}'|^\sigma}, \quad |\mathbf{r} - \mathbf{r}'| \rightarrow \infty \quad (\alpha > 0, \sigma > 0, \text{ const.}),$$

where higher-order terms have been neglected. Many of these papers concern the precise form of the power-law factor (i.e., the value of  $\sigma$ ) in both insulators and metallic systems. The actual functional dependence of  $\alpha$  on the gap and of  $\sigma$  on the dimensionality of the problem have been the subject of intense discussion, with some authors claiming that  $\alpha$  is proportional to  $\gamma$ , and others finding it to be proportional to  $\sqrt{\gamma}$ ; see, e.g., [53, 70, 73, 90, 127, 128] and section 8.6. It appears that both types of behavior can occur in practice. For instance, in [73] the authors provide a tight-binding model of an insulator for which the density falls off exponentially with decay length  $\alpha = O(\gamma)$  in the diagonal direction of the lattice, and  $\alpha = O(\sqrt{\gamma})$  in nondiagonal directions, as  $\gamma \rightarrow 0+$ . We also note that in [73], the decay behavior of the density matrix for an insulator is found to be given (up to higher-order terms) by

$$\rho(\mathbf{r}, \mathbf{r}') = C \frac{e^{-\alpha|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|^{d/2}}, \quad |\mathbf{r}-\mathbf{r}'| \rightarrow \infty,$$

where  $d$  is the dimensionality of the problem. In practice, the power-law factor in the denominator is often ignored, since the exponential decay dominates.

In [52], Goedecker argued that the density matrix for  $d$ -dimensional ( $d = 1, 2, 3$ ) metallic systems at electronic temperature  $T > 0$  behaves to leading order like

$$(4.5) \quad \rho(\mathbf{r}, \mathbf{r}') = C \frac{\cos(|\mathbf{r}-\mathbf{r}'|)}{|\mathbf{r}-\mathbf{r}'|^{(d+1)/2}} e^{-k_B T |\mathbf{r}-\mathbf{r}'|}, \quad |\mathbf{r}-\mathbf{r}'| \rightarrow \infty.$$

Note that in the zero-temperature limit, a power-law decay (with oscillations) is observed. An analogous result was also obtained in [70]. Note that the decay length in the exponential goes to zero like the temperature  $T$  rather than like  $\sqrt{T}$ , as claimed, for instance, in [3]. We will return on this topic in section 8.7.

Finally, as representatives of the third group of papers we select [3] and [140]. The authors of [3] use the Fermi–Dirac approximation of the density matrix and consider its expansion in the Chebyshev basis. From an estimate of the rate of decay of the coefficients of the Chebyshev expansion of  $f_{FD}(x)$ , they obtain estimates for the number of terms needed to satisfy a prescribed error in the approximation of the density matrix. In turn, this yields estimates for the rate of decay as a function of the extreme eigenvalues and spectral gap of the discrete Hamiltonian. Because of some ad hoc assumptions and the many approximations used, the arguments in this paper cannot be considered mathematically rigorous, and the estimates thus obtained are not always accurate. Nevertheless, the idea of using a polynomial approximation for the Fermi–Dirac function and the observation that exponential decay of the expansion coefficients implies exponential decay in the (approximate) density matrix is quite valuable and, as we show in this paper, can be made fully rigorous.

Finally, in [140] the authors present the results of numerical calculations for various insulators in order to gain some insight into the dependence of the decay length on the gap. Their experiments confirm that the decay behavior of  $\rho(\mathbf{r}, \mathbf{r}')$  can be strongly anisotropic and that different rates of decay may occur in different directions; this is consistent with the analytical results in [73].

Despite this considerable body of work, the localization question for density matrices cannot be regarded as completely settled from the mathematical standpoint. We are not aware of any completely general and rigorous mathematical treatment of the decay properties in density matrices associated with general (localized) Hamiltonians, covering all systems with gap as well as metallic systems at positive temperature.

Moreover, rather than order-of-magnitude estimates, actual upper bounds would be more satisfactory.

Almost all the abovementioned results concern the continuous, infinite-dimensional case. In practice, of course, calculations are performed on discrete,  $n$ -dimensional approximations  $H$  and  $P$  to the operators  $\mathcal{H}$  and  $\mathcal{P}$ . The replacement of density operators with finite density matrices can be obtained via the introduction of a system of  $n$  basis functions  $\{\phi_i\}_{i=1}^n$ , leading to the density matrix  $P = (P_{ij})$  with

$$(4.6) \quad P_{ij} = \langle \phi_j, \mathcal{P} \phi_i \rangle = \langle \phi_i | \mathcal{P} | \phi_j \rangle = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \rho(\mathbf{r}, \mathbf{r}') \phi_i(\mathbf{r})^* \phi_j(\mathbf{r}') d\mathbf{r} d\mathbf{r}'.$$

As long as the basis functions are localized in space, the decay behavior of the density function  $\rho(\mathbf{r}, \mathbf{r}')$  for increasing spatial separation  $|\mathbf{r} - \mathbf{r}'|$  is reflected in the decay behavior of the matrix elements  $P_{ij}$  away from the main diagonal (i.e., for  $|i - j|$  increasing) or, more generally, for increasing distance  $d(i, j)$  in the graph associated with the discrete Hamiltonian; see section 6 for details.

In developing and analyzing  $O(n)$  methods for electronic structure computations, it is important to rigorously establish decay bounds for the entries of the density matrices that take into account properties of the discrete Hamiltonians. It is in principle possible to obtain decay estimates for finite-dimensional approximations using localized basis functions from the spatial decay estimates for the density kernel. Note, however, that any estimates obtained inserting (4.3) or (4.5) into (4.6) would depend on the particular set of basis functions used.

In this paper we take a different approach. Instead of starting with the continuous problem and discretizing it, we establish our estimates directly for sequences of matrices of finite, but increasing order. We believe that this approach is closer to the practice of electronic structure calculations, where matrices are the primary computational objects.

We impose a minimal set of assumptions on our matrix sequences so as to reproduce the main features of problems encountered in actual electronic structure computations, while at the same time ensuring a high degree of generality. Since our aim is to provide a rigorous and general mathematical justification to the possibility of  $O(n)$  methods, this approach seems to be quite natural.<sup>6</sup>

To put our work further into perspective, we quote from two prominent researchers in the field of electronic structure, one a mathematician, the other a physicist. In his excellent survey [77], Claude Le Bris, discussing the basis for linear scaling algorithms, i.e., the assumed sparsity of the density matrix, wrote (pp. 402 and 404):

The latter assumption is in some sense an *a posteriori* assumption, and not easy to analyse . . . It is to be emphasized that the numerical analysis of the linear scaling methods overviewed above that would account for cut-off rules and locality assumptions, is not yet available.

It is interesting to compare these statements with two earlier ones made by Stefan Goedecker. In [51] he wrote (p. 261):

To obtain a linear scaling, the extended orbitals [i.e., the eigenfunctions of the one-particle Hamiltonian corresponding to occupied states] have to be

<sup>6</sup>We refer the historically-minded reader to the interesting discussion given by John von Neumann in [132] on the benefits that can be expected from a study of the asymptotic properties of large matrices, in contrast to the study of the infinite-dimensional (Hilbert space) case.

replaced by the density matrix, whose physical behavior can be exploited to obtain a fast algorithm. This last point is essential. Mathematical and numerical analyses alone are not sufficient to construct a linear algorithm. They have to be combined with physical intuition.

A similar statement can be found in [53, p. 1086]:

Even though  $O(N)$  algorithms contain many aspects of mathematics and computer science they have, nevertheless, deep roots in physics. Linear scaling is not obtainable by purely mathematical tricks, but it is based on an understanding of the concept of locality in quantum mechanics.

In the following we provide a general treatment of the question of decay in spectral projectors that is as a priori as possible, in the sense that it relies on a minimal set of assumptions on the discrete Hamiltonians; furthermore, our theory is purely mathematical and therefore completely independent of any physical interpretation. Nevertheless, our theory allows us to shed light on questions like the dependence of the decay length on the temperature in the density matrix for metals at  $T > 0$ ; see section 8.7. We do this using for the most part fairly simple mathematical tools from classical approximation theory and linear algebra.

Of course, in the development of practical linear scaling algorithms a deep knowledge of the physics involved is extremely important; we think, however, that locality is as much a *mathematical phenomenon* as a physical one.

We hope that the increased level of generality attained in this paper (relative to previous treatments in the physics literature) will also help in the development of  $O(n)$  methods for other types of problems where spectral projectors and related matrix functions play a central role. A few examples are discussed in section 11.

**5. Normalizations and Scalings.** We will be dealing with sequences of matrices  $\{H_n\}$  of increasing size. We assume that each matrix  $H_n$  is an Hermitian  $n \times n$  matrix, where  $n = n_b \cdot n_e$ ; here  $n_b$  is fixed, while  $n_e$  is increasing. As explained in section 3, the motivation for this assumption is that in most electronic structure codes, once a basis set has been selected the number  $n_b$  of basis functions per particle is fixed, and one is interested in the scaling as  $n_e$ , the number of particles, increases. Hence, the parameter that controls the system size is  $n_e$ . We also assume that the system is contained in a  $d$ -dimensional box of volume  $V = L^d$  and that  $L \rightarrow \infty$  as  $n_e \rightarrow \infty$  in such a way that the average density  $n_e/L^d$  remains constant (*thermodynamic limit*). This is very different from the case of finite element or finite difference approximations to partial differential equations (PDEs), where the system (or domain) size is considered fixed while the number of basis functions increases or, equivalently, the mesh size  $h$  goes to zero.

Our scaling assumption has very important consequences on the structural and spectral properties of the matrix sequence  $\{H_n\}$ ; namely, the following properties hold:

1. The bandwidth of  $H_n$ , which reflects the *interaction range* of the discrete Hamiltonians, remains bounded as the system size increases [89, p. 454]. More generally, the entries of  $H_n$  decay away from the main diagonal at a rate independent of  $n_e$  (hence, of  $n$ ). See section 6 for precise definitions and generalizations.
2. The eigenvalue spectra  $\sigma(H_n)$  are also uniformly bounded as  $n_e \rightarrow \infty$ . In view of the previous property, this is equivalent to saying that the entries in  $H_n$  are uniformly bounded in magnitude; this is just a consequence of Gersgorin's theorem (see, e.g., [67, p. 344]).

3. For the case of Hamiltonians modeling insulators or semiconductors, the spectral (HOMO-LUMO) gap does not vanish as  $n_e \rightarrow \infty$ . More precisely, if  $\varepsilon_i^{(n)}$  denotes the  $i$ th eigenvalue of  $H_n$  and  $\gamma_n := \varepsilon_{n_e+1}^{(n)} - \varepsilon_{n_e}^{(n)}$ , then  $\inf_n \gamma_n > 0$ . This assumption does not hold for Hamiltonians modeling metallic systems; in this case,  $\inf_n \gamma_n = 0$ , i.e., the spectral gap goes to zero as  $n_e \rightarrow \infty$ .

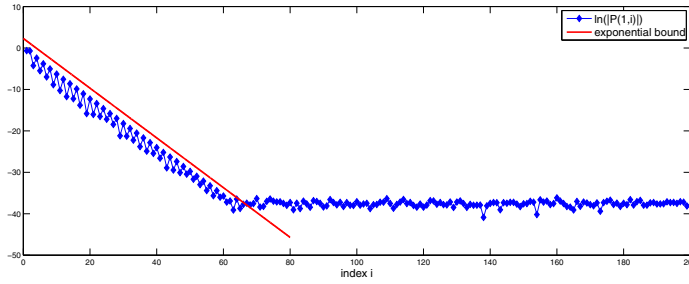
We emphasize that these properties hold for very general classes of physical systems and discretization methods for electronic structure, with few exceptions (i.e., nonlocalized basis functions, such as plane waves). It is instructive to contrast these properties with those of matrix sequences arising in finite element or finite difference approximations of PDEs, where the matrix size increases as  $h \rightarrow 0$ , with  $h$  a discretization parameter. Considering the case of a scalar, second-order elliptic PDE, we see that the first property only holds in the 1D case, or in higher-dimensional cases when the discretization is refined in only one dimension. (As we will see, this condition is rather restrictive and can be relaxed.) Furthermore, it is generally impossible to satisfy the second assumption and that on the nonvanishing gap ( $\inf_n \gamma_n > 0$ ) simultaneously. Indeed, normalizing the matrices so that their spectra remain uniformly bounded will generally cause the eigenvalues to completely fill the spectral interval as  $n \rightarrow \infty$ . That is, in general, given any two points inside this interval and for  $n$  large enough, at least one eigenvalue of the corresponding  $n \times n$  matrix falls between these two points.

Our assumptions allow us to refer to the spectral gap of the matrix sequence  $\{H_n\}$  without having to specify whether we are talking about an absolute or a relative gap. As we shall see, it is convenient to assume that all the matrices in the sequence  $\{H_n\}$  have spectrum contained in the interval  $[-1, 1]$ ; therefore, the absolute gap and the relative gap of any matrix  $H_n$  are the same, up to the factor 2. The spectral gap (more precisely, its reciprocal) is a natural measure of the conditioning of the problem of computing the spectral projector onto the occupied subspace, i.e., the subspace spanned by the eigenvectors of  $H_n$  corresponding to eigenvalues  $\varepsilon_i^{(n)} < \mu$ ; see, e.g., [109, p. B4] for a recent discussion. The assumption  $\inf_n \gamma_n > 0$  then simply means that the electronic structure problem is *uniformly well-conditioned*; note that this assumption is also very important for the convergence of the outer SCF iteration [77, 137]. This hypothesis is satisfied for insulators and semiconductors, but not in the case of metals.

**6. Approximation of Matrices by Numerical Truncation.** Discretization of  $\mathcal{H}$ , the Hamiltonian operator, by means of basis sets consisting of linear combinations of Slater- or Gaussian-type orbitals leads to matrix representations that are, strictly speaking, full. Indeed, since these basis functions are globally supported, almost all matrix elements  $H_{ij} = \langle \phi_j, \mathcal{H} \phi_i \rangle \equiv \langle \phi_i | \mathcal{H} | \phi_j \rangle$  are nonzero. The same is true for the entries of the overlap matrix  $S_{ij} = \langle \phi_j, \phi_i \rangle$ . However, owing to the rapid decay of the basis functions outside of a localized region, and due to the local nature of the interactions encoded by the Hamiltonian operator, the entries of  $H$  decay exponentially fast with the spatial separation of the basis functions. (For the overlap matrix corresponding to Gaussian-type orbitals, the decay is actually even faster than exponential.)

More formally, we say that a sequence of  $n \times n$  matrices  $A_n = ([A_n]_{ij})$  has the *exponential off-diagonal decay property* if there are constants  $c > 0$  and  $\alpha > 0$  independent of  $n$  such that

$$(6.1) \quad |[A_n]_{ij}| \leq c e^{-\alpha|i-j|} \quad \text{for all } i, j = 1, \dots, n.$$



**Fig. 6.1** Logarithmic plot of the first row of a density matrix and an exponential bound.

Corresponding to each matrix  $A_n$  we then define for a nonnegative integer  $m$  the matrix  $A_n^{(m)} = ([A_n^{(m)}]_{ij})$  as follows:

$$[A_n^{(m)}]_{ij} = \begin{cases} [A_n]_{ij} & \text{if } |i - j| \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, each matrix  $A_n^{(m)}$  is  $m$ -banded and can be thought of as an approximation, or truncation, of  $A_n$ . Note that the set of  $m$ -banded matrices forms a vector subspace  $\mathcal{V}_m \subseteq \mathbb{C}^{n \times n}$  and that  $A_n^{(m)}$  is just the orthogonal projection of  $A_n$  onto  $\mathcal{V}_m$  with respect to the Frobenius inner product  $\langle A, B \rangle_F := \text{Tr}(B^* A)$ . Hence,  $A_n^{(m)}$  is the best approximation of  $A_n$  in  $\mathcal{V}_m$  with respect to the Frobenius norm.

Note that we do not require the matrices to be Hermitian or symmetric here; we only assume (for simplicity) that the same pattern of nonzero off-diagonals is present on either side of the main diagonal. The following simple result from [14] provides an estimate of the rate at which the truncation error decreases as the bandwidth  $m$  of the approximation increases. In addition, it establishes  $n$ -independence of the truncation error for  $n \rightarrow \infty$  for matrix sequences satisfying (6.1).

**PROPOSITION 6.1** (see [14]). *Let  $A$  be a matrix with entries  $A_{ij}$  satisfying (6.1) and let  $A^{(m)}$  be the corresponding  $m$ -banded approximation. Then for any  $\epsilon > 0$  there is an  $\bar{m}$  such that  $\|A - A^{(m)}\|_1 \leq \epsilon$  for  $m \geq \bar{m}$ .*

The integer  $\bar{m}$  in the foregoing proposition is easily found to be given by

$$\bar{m} = \left\lceil \frac{1}{\alpha} \ln \left( \frac{2c}{1 - e^{-\alpha}} \epsilon^{-1} \right) \right\rceil.$$

Clearly, this result is of interest only for  $\bar{m} < n$  (in fact, for  $\bar{m} \ll n$ ).

**EXAMPLE 6.2.** *Let us consider a tridiagonal matrix  $H$  of size  $200 \times 200$ , with eigenvalues randomly chosen in  $[-1, -0.5] \cup [0.5, 1]$ , and let  $P$  be the associated density matrix with  $\mu = 0$ . Numerical computation shows that  $P$  satisfies the bound (6.1) with  $\alpha = 0.6$  and  $c = 10$  (as long as its entries are larger than the machine precision). Figure 6.1 depicts the absolute value of the entries in the first row of  $P$  and the bound (6.1), in a logarithmic scale. Choose, for instance, a tolerance  $\epsilon = 10^{-6}$ ; then it follows from the previous formula that the truncated matrix  $P^{(m)}$  satisfies  $\|P - P^{(m)}\|_1 \leq \epsilon$  for any bandwidth  $m \geq 29$ .*

What is important about this simple result is that when applied to a sequence  $\{A_n\} = ([A_n]_{ij})$  of  $n \times n$  matrices having the off-diagonal decay property (6.1) with  $c$  and  $\alpha$  independent of  $n$ , the bandwidth  $\bar{m}$  is itself independent of  $n$ . For convenience,

we have stated Proposition 6.1 in the 1-norm; when  $A = A^*$  the same conclusion holds for the 2-norm, owing to the inequality

$$(6.2) \quad \|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$$

(see [57, Corollary 2.3.2]). Moreover, a similar result also applies to other types of decay, such as algebraic (power-law) decay of the form

$$|[A_n]_{ij}| \leq \frac{c}{|i-j|^{p+1}} \quad \text{for all } i, j = 1, \dots, n,$$

with  $c$  and  $p$  independent of  $n$ , as long as  $p > 1$ .

REMARK 6.3. *It is worth emphasizing that the above considerations do not require that the matrix entries  $[A_n]_{ij}$  themselves actually decay exponentially away from the main diagonal, but only that they are bounded above in an exponentially decaying manner. In particular, the decay behavior of the matrix entries need not be monotonic.*

Although we have limited ourselves to absolute approximation errors in various norms, it is easy to accommodate relative errors by normalizing the matrices. Indeed, upon normalization all the Hamiltonians satisfy  $\|H_n\|_2 = 1$ ; furthermore, for density matrices this property is automatically satisfied, since they are orthogonal projectors. In the next section we also consider using the Frobenius norm for projectors.

The foregoing considerations can be extended to matrices with more general decay patterns, i.e., with exponential decay away from a subset of selected positions  $(i, j)$  in the matrix; see, e.g., [14] as well as [31]. In order to formalize this notion, we first recall the definition of *geodesic distance*  $d(i, j)$  in a graph [37]: it is the number of edges in the shortest path connecting two nodes  $i$  and  $j$ , possibly infinite if there is no such path. Next, given a (sparse) matrix sequence  $\{A_n\}$ , we associate with each matrix  $A_n$  a graph  $G_n$  with  $n$  nodes and  $m = O(n)$  edges. In order to obtain meaningful results, however, we need to impose some restrictions on the types of sparsity allowed. Recall that the *degree* of node  $i$  in a graph is just the number of neighbors of  $i$ , i.e., the number of nodes at distance 1 from  $i$ . We denote by  $\deg_n(i)$  the degree of node  $i$  in the graph  $G_n$ . We shall assume that the maximum degree of any node in  $G_n$  remains bounded as  $n \rightarrow \infty$ ; that is, there exists a positive integer  $D$  independent of  $n$  such that  $\max_{1 \leq i \leq n} \deg_n(i) \leq D$  for all  $n$ . Note that when  $A_n = H_n$  (discretized Hamiltonian), this property is a mathematical restatement of the physical notion of locality, or finite range, of interactions.

Now let us assume that we have a sequence of  $n \times n$  matrices  $A_n = ([A_n]_{ij})$  with associated graphs  $G_n$  and graph distances  $d_n(i, j)$ . We will say that  $A_n$  has the *exponential decay property relative to the graph  $G_n$*  if there are constants  $c > 0$  and  $\alpha > 0$  independent of  $n$  such that

$$(6.3) \quad |[A_n]_{ij}| \leq c e^{-\alpha d_n(i,j)} \quad \text{for all } i, j = 1, \dots, n.$$

PROPOSITION 6.4. *Let  $\{A_n\}$  be a sequence of  $n \times n$  matrices satisfying the exponential decay property (6.3) relative to a sequence of graphs  $\{G_n\}$  having uniformly bounded maximal degree. Then, for any given  $0 < \epsilon < c$ , each  $A_n$  contains at most  $O(n)$  entries greater than  $\epsilon$  in magnitude.*

*Proof.* For a fixed node  $i$ , the condition  $|[A_n]_{ij}| > \epsilon$  together with (6.3) immediately implies

$$(6.4) \quad d_n(i, j) < \frac{1}{\alpha} \ln \left( \frac{c}{\epsilon} \right).$$



Since  $c$  and  $\alpha$  are independent of  $n$ , inequality (6.4), together with the assumption that the graphs  $G_n$  have bounded maximal degree, implies that for any row of the matrix (indexed by  $i$ ), there is at most a constant number of entries that have magnitude greater than  $\epsilon$ . Hence, only  $O(n)$  entries in  $A_n$  can satisfy  $|[A_n]_{ij}| > \epsilon$ .  $\square$

REMARK 6.5. *Note that the hypothesis of uniformly bounded maximal degrees is certainly satisfied if the graphs  $G_n$  have uniformly bounded bandwidths (recall that the bandwidth of a graph is just the bandwidth of the corresponding adjacency matrix). This special case corresponds to the matrix sequence  $\{A_n\}$  having the off-diagonal exponential decay property.*

Under the same assumptions as Proposition 6.4, we can show that it is possible to approximate each  $A_n$  to within an arbitrarily small error  $\epsilon > 0$  in norm with a sparse matrix  $A_n^{(m)}$  (i.e., a matrix containing only  $O(n)$  nonzero entries).

PROPOSITION 6.6. *Assume the hypotheses of Proposition 6.4 are satisfied. Define the matrix  $A_n^{(m)} = ([A_n^{(m)}]_{ij})$ , where*

$$[A_n^{(m)}]_{ij} = \begin{cases} [A_n]_{ij} & \text{if } d_n(i, j) \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

*Then, for any given  $\epsilon > 0$ , there exists  $\bar{m}$  independent of  $n$  such that  $\|A_n - A_n^{(m)}\|_1 < \epsilon$  for all  $m \geq \bar{m}$ . Moreover, if  $A = A^*$ , then also  $\|A_n - A_n^{(m)}\|_2 < \epsilon$  for all  $m \geq \bar{m}$ . Furthermore, each  $A_n^{(m)}$  contains only  $O(n)$  nonzeros.*

*Proof.* For each  $n$  and  $m$  and for  $1 \leq j \leq n$ , let

$$K_n^m(j) := \{i \mid 1 \leq i \leq n \text{ and } d_n(i, j) > m\}.$$

We have

$$\|A_n - A_n^{(m)}\|_1 = \max_{1 \leq j \leq n} \sum_{i \in K_n^m(j)} |[A_n]_{ij}| \leq c \max_{1 \leq j \leq n} \sum_{i \in K_n^m(j)} e^{-\alpha d_n(i, j)}.$$

Letting  $\lambda = e^{-\alpha}$ , we obtain

$$\|A_n - A_n^{(m)}\|_1 \leq c \max_{1 \leq j \leq n} \sum_{i \in K_n^m(j)} \lambda^{d_n(i, j)} \leq c \sum_{k=m+1}^n \lambda^k < c \sum_{k=m+1}^{\infty} \lambda^k = c \frac{\lambda^{m+1}}{1 - \lambda}.$$

Since  $0 < \lambda < 1$ , for any given  $\epsilon > 0$  we can always find  $\bar{m}$  such that

$$c \frac{\lambda^{m+1}}{1 - \lambda} \leq \epsilon \quad \text{for all } m \geq \bar{m}.$$

If  $A_n = A_n^*$ , then  $\|A_n - A_n^{(m)}\|_2 \leq \|A_n - A_n^{(m)}\|_1 < \epsilon$  for all  $m \geq \bar{m}$ . The last assertion follows from the bounded maximal degree assumption.  $\square$

Hence, when forming the overlap matrices and discrete Hamiltonians, only matrix elements corresponding to “sufficiently nearby” basis functions (i.e., basis functions having sufficient overlap) need to be computed, the others being negligibly small. The resulting matrices are therefore sparse, and indeed banded for 1D problems, with a number of nonzeros that grows linearly in the matrix dimension. The actual bandwidth, or sparsity pattern, may depend on the choice and numbering (ordering) of basis functions and (for the discrete Hamiltonians) on the strength of the interactions, i.e., on the form of the potential function  $V$  in the Hamiltonian operator.

It should be kept in mind that while the number of nonzeros in the Hamiltonians discretized using (say) Gaussian-type orbitals is  $O(n)$ , the actual number of nonzeros per row can be quite high, indeed much higher than when finite differences or finite elements are used to discretize the same operators. It is not unusual to have hundreds or even thousands of nonzeros per row. On the other hand, the matrices are very often not huge in size. As already mentioned, the size  $n$  of the matrix is the total number of basis functions, which is a small or moderate multiple (between 2 and 25, say) of the number  $n_e$  of electrons. For example, if  $n_b \approx 10$  and  $n_e \approx 2000$ , the size of  $H$  will be  $n \approx 20,000$  and  $H$  could easily contain several millions of nonzeros. This should be compared with “real space” discretizations based on finite elements or high-order finite difference schemes [116]. The resulting Hamiltonians are usually very sparse, with a number of nonzero entries per row averaging a few tens at most [7]. However, these matrices are of much larger dimension than the matrices obtained using basis sets consisting of atom-centered orbitals. In this case, methodologies based on approximating the density matrix are currently not feasible, except for 1D problems. The same remark applies to discretizations based on plane waves, which tend to produce matrices of an intermediate size between those obtained using localized basis sets and those resulting from the use of real space discretizations. These matrices are actually dense and are never formed explicitly. Instead, they are only used in the form of matrix-vector products, which can be implemented efficiently by means of FFTs; see, e.g., [116].

The possibility of developing linear scaling methods for electronic structure largely depends on the localization properties of the density matrix  $P$ . It is therefore critical to understand the decay behavior of the density matrix. Since at zero temperature the density matrix is just a particular spectral projector, we consider next some general properties of such projectors.

**7. General Properties of Orthogonal Projectors.** While our main goal in this paper is to study decay properties in orthogonal projectors associated with certain sequences of sparse matrices of increasing size, it is useful to first establish some a priori estimates for the entries of general projectors. Indeed, the intrinsic properties of a projector, like idempotency, positive semidefiniteness, and the relations between their trace, rank, and Frobenius norm, tend to impose rather severe constraints on the magnitude of its entries, particularly for increasing dimension and rank.

We begin by observing that in an orthogonal projector  $P$ , all entries  $P_{ij}$  satisfy  $|P_{ij}| \leq 1$  and, since  $P$  is positive semidefinite, its largest entry is on the main diagonal. Also, the trace and rank coincide:  $\text{Tr}(P) = \text{rank}(P)$ . Moreover,  $\|P\|_2 = 1$  and  $\|P\|_F = \sqrt{\text{Tr}(P)}$ .

In the context of electronic structure computations, we deal with a sequence of  $n \times n$  orthogonal projectors  $\{P_n\}$  of rank  $n_e$ , where  $n = n_b \cdot n_e$  with  $n_e$  increasing and  $n_b$  fixed. Hence,

$$(7.1) \quad \text{Tr}(P_n) = \text{rank}(P_n) = n_e \quad \text{and} \quad \|P_n\|_F = \sqrt{n_e}.$$

For convenience, we will call a sequence of orthogonal projectors  $\{P_n\}$  satisfying (7.1) a *density matrix sequence*; the entries of  $P_n$  will be denoted by  $[P_n]_{ij}$ . We have the following lemma.

LEMMA 7.1. *Let  $\{P_n\}$  be a density matrix sequence. Then*

$$\frac{\sum_{i \neq j} |[P_n]_{ij}|^2}{\|P_n\|_F^2} \leq 1 - \frac{1}{n_b}.$$

*Proof.* Just observe that  $\text{Tr}(P_n) = \sum_{i=1}^n [P_n]_{ii} = n_e$  together with  $|[P_n]_{ii}| \leq 1$  for all  $i$  imply that the minimum of the sum  $\sum_{i=1}^n |[P_n]_{ii}|^2$  is achieved when  $[P_n]_{ii} = \frac{n_e}{n} = \frac{1}{n_b}$  for all  $i$ . Hence,  $\sum_{i=1}^n |[P_n]_{ii}|^2 \geq \frac{n_e^2}{n^2} = \frac{n_e}{n_b}$ . Therefore,

$$(7.2) \quad \sum_{i \neq j} |[P_n]_{ij}|^2 = \|P_n\|_F^2 - \sum_{i=1}^n |[P_n]_{ii}|^2 \leq \left(1 - \frac{1}{n_b}\right) n_e$$

and the result follows dividing through by  $\|P_n\|_F^2 = n_e$ .  $\square$

REMARK 7.2. From the proof one can trivially see that the bound (7.2) is sharp. In section 10 we give a nontrivial example where the bound is attained.

THEOREM 7.3. Let  $\{P_n\}$  be a density matrix sequence. Then, for any  $\epsilon > 0$ , the number of entries of  $P_n$  greater than or equal to  $\epsilon$  in magnitude grows at most linearly with  $n$ .

*Proof.* Clearly, it suffices to show that the number of off-diagonal entries  $[P_n]_{ij}$  with  $|[P_n]_{ij}| \geq \epsilon$  can grow at most linearly with  $n$ . Let

$$\mathcal{I} = \{(i, j) \mid 1 \leq i, j \leq n \text{ and } i \neq j\} \quad \text{and} \quad \mathcal{I}_\epsilon = \{(i, j) \in \mathcal{I} \mid |[P_n]_{ij}| \geq \epsilon\}.$$

Then obviously

$$\sum_{i \neq j} |[P_n]_{ij}|^2 = \sum_{(i,j) \in \mathcal{I}_\epsilon} |[P_n]_{ij}|^2 + \sum_{(i,j) \in \mathcal{I} \setminus \mathcal{I}_\epsilon} |[P_n]_{ij}|^2$$

and, if  $|\mathcal{I}_\epsilon| = K$ , then

$$\sum_{i \neq j} |[P_n]_{ij}|^2 \geq K\epsilon^2 \quad \Rightarrow \quad \frac{\sum_{i \neq j} |[P_n]_{ij}|^2}{\|P_n\|_F^2} \geq \frac{K\epsilon^2}{n_e} = \frac{K\epsilon^2 n_b}{n}.$$

Hence, by Lemma 7.1,

$$\frac{K\epsilon^2 n_b}{n} \leq \frac{\sum_{i \neq j} |[P_n]_{ij}|^2}{\|P_n\|_F^2} \leq 1 - \frac{1}{n_b},$$

from which we obtain the bound

$$(7.3) \quad K \leq \frac{n}{\epsilon^2 n_b} \left(1 - \frac{1}{n_b}\right),$$

which shows that the number  $K$  of entries of  $P_n$  with  $|[P_n]_{ij}| \geq \epsilon$  can grow at most as  $O(n)$  for  $n \rightarrow \infty$ .  $\square$

REMARK 7.4. Due to the presence of the factor  $\epsilon^2$  in the denominator of the bound (7.3), for small  $\epsilon$  the proportion of entries of  $P_n$  that are not smaller than  $\epsilon$  can actually be quite large unless  $n$  is huge. Nevertheless, the result is interesting because it shows that in any density matrix sequence, the proportion of entries larger than a prescribed threshold must vanish as  $n \rightarrow \infty$ . In practice, for density matrices corresponding to sparse Hamiltonians with gap, localization occurs already for moderate values of  $n$ .

We pointed out in the previous section that if the entries in a matrix sequence  $\{A_n\}$  decay at least algebraically with exponent  $p > 1$  away from the main diagonal, with rates independent of  $n$ , then for any prescribed  $\epsilon > 0$  it is possible to find a

sequence of approximants  $\{A_n^{(m)}\}$  with a fixed bandwidth  $m$  (or sparsity pattern) such that  $\|A_n - A_n^{(m)}\| < \epsilon$ . This applies in particular to density matrix sequences. The next result shows that, in principle, a *linear* rate of decay is enough to allow for banded (or sparse) approximation to within any prescribed *relative* error in the Frobenius norm.

**THEOREM 7.5.** *Let  $\{P_n\}$  be a density matrix sequence and assume that there exists  $c > 0$  independent of  $n$  such that  $|[P_n]_{ij}| \leq c/(|i-j|+1)$  for all  $i, j = 1, \dots, n$ . Then, for all  $\epsilon > 0$ , there exists a positive integer  $\bar{m}$  independent of  $n$  such that*

$$\frac{\|P_n - P_n^{(m)}\|_F}{\|P_n\|_F} \leq \epsilon \quad \text{for all } m \geq \bar{m},$$

where  $P_n^{(m)}$  is the  $m$ -banded approximation obtained by setting to zero all the entries of  $P_n$  outside the band.

*Proof.* We subtract  $P_n^{(m)}$  from  $P_n$  and compute  $\|P_n - P_n^{(m)}\|_F^2$  by adding the squares of the nonzero entries in the upper triangular part of  $P_n - P_n^{(m)}$  diagonal by diagonal and multiplying the result by 2 (since the matrices are Hermitian). Using the decay assumption we obtain

$$\|P_n - P_n^{(m)}\|_F^2 \leq 2c^2 \sum_{k=1}^{n-m-1} \frac{k}{(n-k+1)^2} = 2c^2 \sum_{k=1}^{n-m-1} \frac{k}{[k-(n+1)]^2}.$$

To obtain an upper bound for the right-hand side, we observe that the function

$$f(x) = \frac{x}{(x-a)^2}, \quad a = n+1,$$

is strictly increasing and convex on the interval  $[1, n-m]$ . Hence, the sum can be bounded above by the integral of  $f(x)$  taken over the same interval:

$$\sum_{k=1}^{n-m-1} \frac{k}{(n-k+1)^2} < \int_1^{n-m} \frac{x}{(x-a)^2} dx, \quad a = n+1.$$

Evaluating the integral and substituting  $a = n+1$  in the result, we obtain

$$\|P_n - P_n^{(m)}\|_F^2 < 2c^2 \left[ \ln \left( \frac{m+1}{n} \right) + (n+1) \left( \frac{1}{m+1} - \frac{1}{n} \right) \right].$$

Dividing by  $\|P_n\|_F^2 = n_e$ , we find

$$\frac{\|P_n - P_n^{(m)}\|_F^2}{\|P_n\|_F^2} < \frac{2c^2}{n_e} \left[ \ln \left( \frac{m+1}{n} \right) + (n+1) \left( \frac{1}{m+1} - \frac{1}{n} \right) \right] < \frac{2c^2}{n_e} \frac{n+1}{m+1}.$$

Recalling that  $n = n_b \cdot n_e$ , we can rewrite the last inequality as

$$\frac{\|P_n - P_n^{(m)}\|_F^2}{\|P_n\|_F^2} < \frac{2c^2}{m+1} \frac{n+1}{n_e} = \frac{2c^2}{m+1} \left( n_b + \frac{1}{n_e} \right) \leq \frac{2c^2}{m+1} (n_b + 1),$$

a quantity which can be made arbitrarily small by taking  $m$  sufficiently large.  $\square$

**REMARK 7.6.** *In practice, linear decay (or even algebraic decay with a small exponent  $p \geq 1$ ) is too slow to be useful in the development of practical  $O(n)$  algorithms.*

For example, from the above estimates we obtain  $\bar{m} = O(\epsilon^{-2})$ , which is clearly not a very encouraging result, even allowing for the fact that the above bound may be pessimistic in general. To date, practical linear scaling algorithms have been developed only for density matrix sequences exhibiting exponential off-diagonal decay.

In the case of exponential decay, one can prove the following result.

**THEOREM 7.7.** *Let  $\{P_n\}$  be a density matrix sequence with  $|(P_n)_{ij}| \leq c e^{-\alpha|i-j|}$ , where  $c > 0$  and  $\alpha > 0$  are independent of  $n$ . Let  $\{P_n^{(m)}\}$  be the corresponding sequence of  $m$ -banded approximations. Then there exists  $k_0 > 0$  independent of  $n$  and  $m$  such that*

$$\frac{\|P_n - P_n^{(m)}\|_F^2}{\|P_n\|_F^2} \leq k_0 e^{-2\alpha m}.$$

*Proof.* The proof is similar to that of Theorem 7.5, except that it is now easy to evaluate the upper bound and the constants exactly. We omit the details.  $\square$

**REMARK 7.8.** *It is immediate to see that the foregoing bound implies the much more favorable estimate  $\bar{m} = O(\ln \epsilon^{-1})$ .*

Again, similar results hold for arbitrary sparsity patterns, replacing  $|i - j|$  with the graph distance. More precisely, the following result holds.

**THEOREM 7.9.** *Let  $\{P_n\}$  be a density matrix sequence with the exponential decay property with respect to a sequence of graphs  $\{G_n\}$  having uniformly bounded maximal degree. Then, for all  $\epsilon > 0$ , there exists a positive integer  $\bar{m}$  independent of  $n$  such that*

$$\frac{\|P_n - P_n^{(m)}\|_F}{\|P_n\|_F} \leq \epsilon \quad \text{for all } m \geq \bar{m},$$

where  $P_n^{(m)}$  is sparse, i.e., it contains only  $O(n)$  nonzeros.

We consider now some of the consequences of approximating full, but localized matrices with sparse ones. The following quantity plays an important role in many electronic structure codes:

$$\langle E \rangle = \text{Tr}(PH) = \epsilon_1 + \epsilon_2 + \cdots + \epsilon_{n_e},$$

where  $\epsilon_i$  denotes the  $i$ th eigenvalue of the discrete Hamiltonian  $H$ . Minimization of  $\text{Tr}(PH)$ , subject to the constraints  $P = P^* = P^2$  and  $\text{Tr}(P) = n_e$ , is the basis of several linear scaling algorithms; see, e.g., [24, 53, 77, 79, 93, 95, 97]. Note that in the tight-binding model, and also within the independent electron approximation, the quantity  $\langle E \rangle$  represents the single-particle energy [6, 53, 97, 128]. Now, assume that  $\hat{H} \approx H$  and  $\hat{P} \approx P$  and define the corresponding approximation of  $\langle E \rangle$  as  $\langle \hat{E} \rangle = \text{Tr}(\hat{P}\hat{H})$ . (We note in passing that in order to compute  $\langle \hat{E} \rangle = \text{Tr}(\hat{P}\hat{H})$ , only the entries of  $\hat{P}$  corresponding to nonzero entries in  $\hat{H}$  need to be computed.) Let  $\Delta_P = \hat{P} - P$  and  $\Delta_H = \hat{H} - H$ . We have

$$\langle \hat{E} \rangle = \text{Tr}[(P + \Delta_P)(H + \Delta_H)] = \text{Tr}(PH) + \text{Tr}(P\Delta_H) + \text{Tr}(\Delta_P H) + \text{Tr}(\Delta_P \Delta_H).$$

Neglecting the last term, we obtain for  $\delta_E = |\langle E \rangle - \langle \hat{E} \rangle|$  the bound

$$\delta_E \leq |\text{Tr}(P\Delta_H)| + |\text{Tr}(\Delta_P H)|.$$

Recalling that the Frobenius norm is the matrix norm induced by the inner product  $\langle A, B \rangle = \text{Tr}(B^*A)$ , using the Cauchy-Schwarz inequality and  $\|P\|_F = \sqrt{n_e}$  we find

$$\delta_E \leq \sqrt{n_e} \|\Delta_H\|_F + \|\Delta_P\|_F \|H\|_F.$$

Now, since the orthogonal projector  $P$  is invariant with respect to scalings of the Hamiltonian, we can assume  $\|H\|_F = 1$ , so that  $\delta_E \leq \sqrt{n_e} \|\Delta_H\|_F + \|\Delta_P\|_F$  holds. In practice, a bound on the *relative* error would be more meaningful. Unfortunately, it is not easy to obtain a rigorous bound in terms of the relative error in the approximate projector  $\hat{P}$ . If, however, we replace the relative error in  $\langle \hat{E} \rangle$  with the normalized error obtained by dividing the absolute error by the number  $n_e$  of electrons, we obtain

$$\frac{\delta_E}{n_e} \leq \frac{\|\Delta_H\|_F}{\sqrt{n_e}} + \frac{\|\Delta_P\|_F}{n_e}.$$

A similar bound for  $\delta_E/n_e$  that involves matrix 2-norms can be obtained as follows. Recall that  $n = n_b \cdot n_e$ , and that  $\|A\|_F \leq \sqrt{n} \|A\|_2$  for any  $n \times n$  matrix  $A$ . Observing that the von Neumann trace inequality [68, pp. 182–183] implies  $|\text{Tr}(P\Delta_H)| \leq \text{Tr}(P) \|\Delta_H\|_2 = n_e \|\Delta_H\|_2$ , we obtain

$$(7.4) \quad \frac{\delta_E}{n_e} \leq \|\Delta_H\|_2 + \sqrt{\frac{n_b}{n_e}} \|\Delta_P\|_2.$$

Since  $n_b$  is constant, an interesting consequence of (7.4) is that for large system sizes (i.e., in the limit as  $n_e \rightarrow \infty$ ), the bound on the normalized error in  $\langle E \rangle$  is essentially determined by the truncation error in the Hamiltonian  $H$  rather than by the error in the density matrix  $P$ .

On the other hand, scaling  $H$  so that  $\|H\|_F = 1$  may not be advisable in practice. Indeed, since the Frobenius norm of the Hamiltonian grows unboundedly for  $n_e \rightarrow \infty$ , rescaling  $H$  so that  $\|H\|_F = 1$  would lead to a loss of significant information when truncation is applied in the case of large systems. A more sensible scaling, which is often used in algorithms for electronic structure computations, is to divide  $\|H\|$  by its largest eigenvalue in magnitude, so that  $\|H\|_2 = 1$ . This is consistent with the assumption, usually satisfied in practice, that the spectra of the Hamiltonians remain bounded as  $n_e \rightarrow \infty$ . (Note that this is the same normalization used to establish the decay bounds in section 8.) With this scaling we readily obtain, to first order, the bound

$$(7.5) \quad \frac{\delta_E}{n_e} \leq \|\Delta_H\|_2 + n_b \|\Delta_P\|_2,$$

showing that errors in  $\Delta_H$  and  $\Delta_P$  enter the estimate for the normalized error in the objective function  $\text{Tr}(PH)$  with approximately the same weight, since  $n_b$  is a moderate constant. We also note that since both error matrices  $\Delta_H$  and  $\Delta_P$  are Hermitian, (6.2) implies that the bounds (7.4) and (7.5) remain true if the 2-norm is replaced by the 1-norm. We mention that the problem of the choice of norm in the measurement of truncation errors has been discussed in [111, 114]. These authors emphasize the use of the 2-norm, which is related to the distance between the exact and inexact (perturbed) occupied subspaces  $\mathcal{X} := \text{Range}(P)$  and  $\hat{\mathcal{X}} := \text{Range}(\hat{P})$  as measured by the sine of the principal angle between  $\mathcal{X}$  and  $\hat{\mathcal{X}}$ ; see [111].

One important practical aspect, which we do not address here, is that in many quantum chemistry codes the matrices have a natural block structure (where each block corresponds, for instance, to the basis functions centered at a given atom); hence, dropping is usually applied to submatrices rather than to individual entries. Exploitation of the block structure is also desirable in order to achieve high performance in matrix-matrix products and other operations; see, e.g., [24, 25, 112].

We conclude this section with a few remarks on the infinite-dimensional case. Recall that any separable, complex Hilbert space  $\mathcal{H}$  is isometrically isomorphic to

the sequence space

$$\ell^2 := \left\{ (\xi_n) \mid \xi_n \in \mathbb{C} \ \forall n \in \mathbb{N} \text{ and } \sum_{n=1}^{\infty} |\xi_n|^2 < \infty \right\}.$$

Moreover, if  $\{e_n\}$  is an orthonormal basis in  $\mathcal{H}$ , to any bounded linear operator  $\mathcal{A}$  on  $\mathcal{H}$  there corresponds the infinite matrix  $A = (A_{ij})$  acting on  $\ell^2$ , uniquely defined by  $A_{ij} = \langle e_j, \mathcal{A}e_i \rangle$ . Note that each column of  $A$  must be in  $\ell^2$ , hence the entries  $A_{ij}$  in each column of  $A$  must go to zero for  $i \rightarrow \infty$ . The same is true for the entries in each row (for  $j \rightarrow \infty$ ) since  $A^* = (A_{ji}^*)$ , the adjoint of  $A$ , is also a (bounded) operator defined everywhere on  $\ell^2$ . More precisely, for any bounded linear operator  $A = (A_{ij})$  on  $\ell^2$  the following bounds hold:

$$(7.6) \quad \sum_{j=1}^{\infty} |A_{ij}|^2 \leq \|A\|_2^2 \quad \text{for all } i \quad \text{and} \quad \sum_{i=1}^{\infty} |A_{ij}|^2 \leq \|A\|_2^2 \quad \text{for all } j,$$

since  $\|A\|_2 = \|A^*\|_2$ .

An orthogonal projector  $\mathcal{P}$  on  $\mathcal{H}$  is a self-adjoint ( $\mathcal{P} = \mathcal{P}^*$ ), idempotent ( $\mathcal{P} = \mathcal{P}^2$ ) linear operator. Such an operator is necessarily bounded, with norm  $\|\mathcal{P}\| = 1$ . Hence, (7.6) implies

$$(7.7) \quad \sum_{j=1}^{\infty} |P_{ij}|^2 \leq 1,$$

where  $P = (P_{ij})$  denotes the matrix representation of  $\mathcal{P}$ . The idempotency condition implies

$$P_{ij} = \sum_{k=1}^{\infty} P_{ik}P_{kj} \quad \text{for all } i, j = 1, 2, \dots$$

In particular, for  $i = j$  we get, using the hermiticity property  $P_{ij} = P_{ji}^*$ ,

$$(7.8) \quad P_{ii} = \sum_{k=1}^{\infty} P_{ik}P_{ki} = \sum_{k=1}^{\infty} |P_{ik}|^2 \quad \text{for all } i = 1, 2, \dots$$

Now, since  $P$  is a projector its entries satisfy  $|P_{ij}| \leq 1$ ; therefore, (7.8) is a strengthening of inequality (7.7). Note in particular that the off-diagonal entries in the first row (or column) of  $P$  must satisfy

$$\sum_{j>1} |P_{1j}|^2 \leq 1 - |P_{11}|^2,$$

those in the second row (or column) must satisfy

$$\sum_{j>2} |P_{2j}|^2 \leq 1 - |P_{22}|^2 - |P_{12}|^2,$$

and in general the entries  $P_{ij}$  with  $j > i$  must satisfy

$$(7.9) \quad \sum_{j>i} |P_{ij}|^2 \leq 1 - \sum_{k=1}^i |P_{ki}|^2 \quad \text{for all } i = 1, 2, \dots$$

Hence, decay in the off-diagonal entries in the  $i$ th row of  $P$  must be fast enough for the bounds (7.9) to hold. In general, however, it is not easy to quantify the asymptotic rate of decay to zero of the off-diagonal entries in an arbitrary orthogonal projector on  $\ell^2$ . In general, the rate of decay can be rather slow. In section 10 we will see an example of a spectral projector associated with a very simple tridiagonal Hamiltonian for which the off-diagonal entries decay linearly to zero.

**8. Decay Results.** In this section we present and discuss some results on the decay of entries for the Fermi–Dirac function applied to Hamiltonians and for the density matrix (spectral projector corresponding to occupied states). We consider both the banded case and the case of more general sparsity patterns. The proofs, which require some basic tools from polynomial approximation theory, will be given in section 8.3.

**8.1. Bounds for the Fermi–Dirac Function.** We begin with the following result for the banded case. As usual in this paper, in the following one should think of the positive integer  $n$  as being of the form  $n = n_b \cdot n_e$ , with  $n_b$  constant and  $n_e \rightarrow \infty$ .

**THEOREM 8.1.** *Let  $m$  be a fixed positive integer and consider a sequence of matrices  $\{H_n\}$  such that*

- (i)  $H_n$  is an  $n \times n$  Hermitian,  $m$ -banded matrix for all  $n$ ;
- (ii) for every  $n$ , all the eigenvalues of  $H_n$  lie in the interval  $[-1, 1]$ .

For a given Fermi level  $\mu$  and inverse temperature  $\beta$ , define for each  $n$  the  $n \times n$  Hermitian matrix  $F_n := f_{FD}(H_n) = [I_n + e^{\beta(H_n - \mu I_n)}]^{-1}$ . Then there exist constants  $c > 0$  and  $\alpha > 0$ , independent of  $n$ , such that the following decay bound holds:

$$(8.1) \quad |[F_n]_{ij}| \leq c e^{-\alpha|i-j|}, \quad i \neq j.$$

The constants  $c$  and  $\alpha$  can be chosen as

$$(8.2) \quad c = \frac{2\chi M(\chi)}{\chi - 1}, \quad M(\chi) = \max_{z \in \mathcal{E}_\chi} |f_{FD}(z)|,$$

$$(8.3) \quad \alpha = \frac{1}{m} \ln \chi$$

for any  $1 < \chi < \bar{\chi}$ , where

$$(8.4) \quad \bar{\chi} = \frac{\sqrt{\sqrt{(\beta^2(1-\mu^2) - \pi^2)^2 + 4\pi^2\beta^2} - \beta^2(1-\mu^2) + \pi^2}}{\sqrt{2}\beta} + \frac{\sqrt{\sqrt{(\beta^2(1-\mu^2) - \pi^2)^2 + 4\pi^2\beta^2} + \beta^2(1+\mu^2) + \pi^2}}{\sqrt{2}\beta}$$

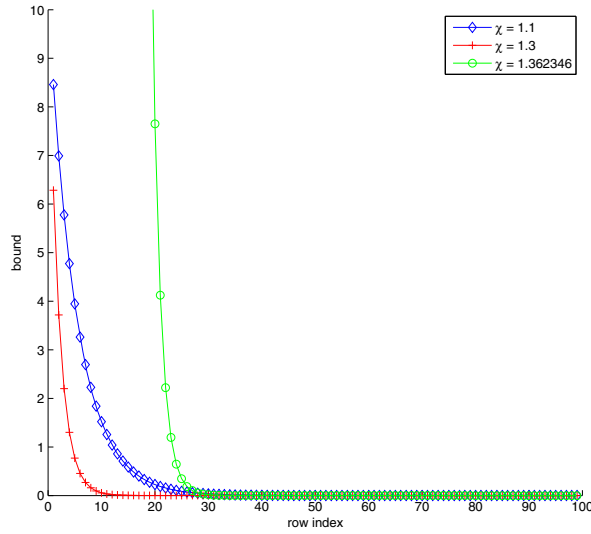
and  $\mathcal{E}_\chi$  is the unique ellipse with foci in  $-1$  and  $1$ , with semiaxes  $\kappa_1 > 1$  and  $\kappa_2 > 0$ , and  $\chi = \kappa_1 + \kappa_2$ .

**REMARK 8.2.** *The ellipse  $\mathcal{E}_\chi$  in the previous theorem is unique because the identity  $\sqrt{\kappa_1^2 - \kappa_2^2} = 1$ , valid for any ellipse with foci in  $1$  and  $-1$ , implies  $\kappa_1 - \kappa_2 = 1/(\kappa_1 + \kappa_2)$ , hence the parameter  $\chi = \kappa_1 + \kappa_2$  alone completely characterizes the ellipse.*

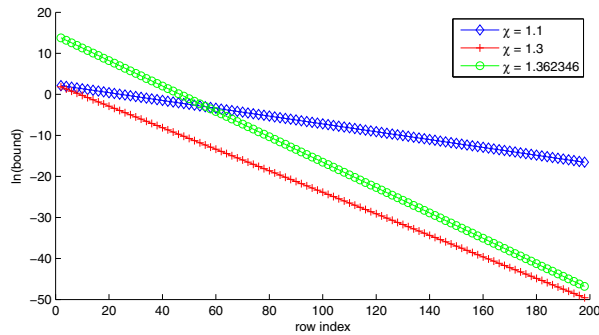
**REMARK 8.3.** *Theorem 8.1 can be immediately generalized to the case where the spectra of the sequence  $\{H_n\}$  are contained in an interval  $[a, b]$  for any  $a < b \in \mathbb{R}$ . It suffices to shift and scale each Hamiltonian,*

$$\hat{H}_n = \frac{2}{b-a} H_n - \frac{a+b}{b-a} I_n,$$





**Fig. 8.1** Bounds (8.1) with  $\mu = 0$  and  $\beta = 10$  for three different values of  $\chi$ .



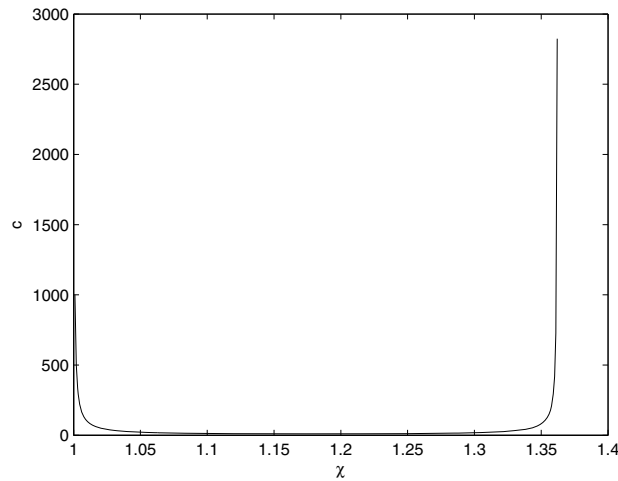
**Fig. 8.2** Logarithmic plot of the bounds (8.1) with  $\mu = 0$  and  $\beta = 10$ .

so that  $\widehat{H}_n$  has spectrum in  $[-1, 1]$ . For the decay bounds to be independent of  $n$ , however,  $a$  and  $b$  must be independent of  $n$ .

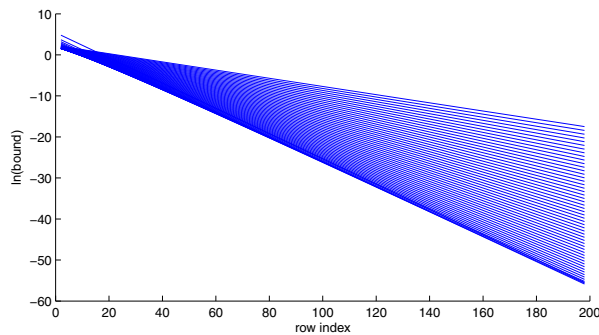
It is important to note that there is a certain amount of arbitrariness in the choice of  $\chi$ , and therefore of  $c$  and  $\alpha$ . If one is mainly interested in a fast asymptotic decay behavior (i.e., for sufficiently large  $|i - j|$ ), it is desirable to choose  $\chi$  as large as possible. On the other hand, if  $\chi$  is very close to  $\bar{\chi}$ , then the constant  $c$  is likely to be quite large and the bounds might be too pessimistic. Let us look at an example. Take  $\mu = 0$ ; in this case we have

$$\bar{\chi} = \left( \pi + \sqrt{\beta^2 + \pi^2} \right) / \beta \quad \text{and} \quad M(\chi) = |1 / (1 + e^{\beta\zeta})|, \quad \text{where} \quad \zeta = i \frac{\chi^2 - 1}{2\chi}.$$

Note that, in agreement with experience, decay is faster for smaller  $\beta$  (i.e., higher electronic temperatures); see sections 8.3 and 8.7 for additional details and discussion. Figures 8.1 and 8.2 show the behavior of the bound given by (8.1) on the first row



**Fig. 8.3** Plot of  $c$  as a function of  $\chi$  with  $\mu = 0$  and  $\beta = 10$ .



**Fig. 8.4** Logarithmic plot of the bounds (8.1) with  $\mu = 0$  and  $\beta = 10$  for several values of  $\chi$ .

of a  $200 \times 200$  tridiagonal matrix ( $m = 1$ ) for  $\beta = 10$  and for three values of  $\chi$ . It is easy to see from the plots that the asymptotic behavior of the bounds improves as  $\chi$  increases; however, the bound given by  $\chi = 1.362346$  is less useful than the bound given by  $\chi = 1.3$ . Figure 8.3 is a plot of  $c$  as a function of  $\chi$  and it shows that  $c$  grows very large when  $\chi$  is close to  $\bar{\chi}$ . This is expected, since  $f_{FD}(z)$  has two poles given by  $z = \pm i\pi/\beta$  on the regularity ellipse  $\mathcal{E}_{\bar{\chi}}$ . It is clear from Figures 8.1 and 8.2 that  $\chi = 1.3$  is the best choice among the three proposed values if one is interested in determining a bandwidth outside of which the entries of  $F_n$  can be safely neglected. As already observed in [14, 106], improved bounds can be obtained by adaptively choosing different (typically increasing) values of  $\chi$  as  $|i - j|$  grows, and by using as a bound the (lower) envelope of the curves plotted in Figure 8.4, which shows the behavior of the decay bounds for several values of  $\chi \in (1.1, \bar{\chi})$ , with  $\bar{\chi} \approx 1.3623463$ .

The results of Theorem 8.1 can be generalized to the case of Hamiltonians with rather general sparsity patterns; see [14, 31, 106]. To this end, we make use of the notion of geodesic distance in a graph already used in section 6. The following result holds.

THEOREM 8.4. Consider a sequence of matrices  $\{H_n\}$  such that

- (i)  $H_n$  is an  $n \times n$  Hermitian matrix for all  $n$ ;
- (ii) the spectra  $\sigma(H_n)$  are uniformly bounded and contained in  $[-1, 1]$  for all  $n$ .

Let  $d_n(i, j)$  be the graph distance associated with  $H_n$ . Then the following decay bound holds:

$$(8.5) \quad |[F_n]_{ij}| \leq c e^{-\theta d_n(i,j)}, \quad i \neq j,$$

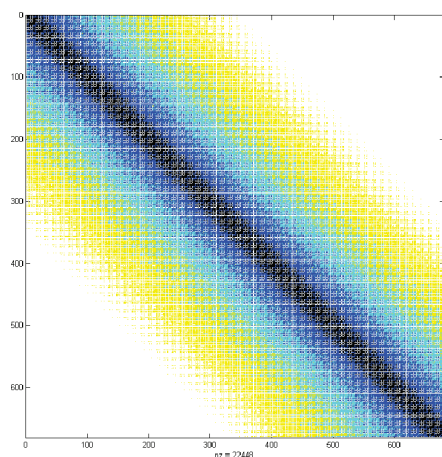
where  $\theta = \ln \chi$  and the remaining notation and choice of constants are as in Theorem 8.1.

We remark that in order for the bound (8.5) to be meaningful from the point of view of linear scaling, we need to impose some restrictions on the asymptotic sparsity of the graph sequence  $\{G_n\}$ . As discussed in section 6,  $O(n)$  approximations of  $F_n$  are possible if the graphs  $G_n$  have maximum degree uniformly bounded with respect to  $n$ . This guarantees that the distance  $d_n(i, j)$  grows unboundedly with  $|i - j|$ , at a rate independent of  $n$  for  $n \rightarrow \infty$ .

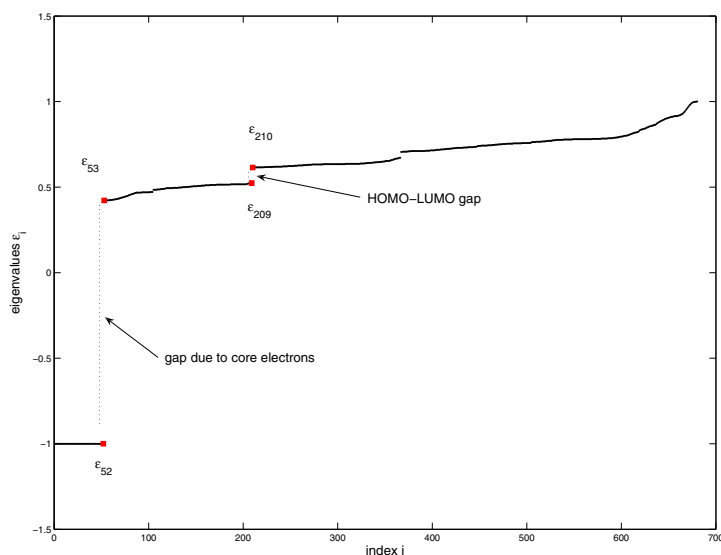
**8.2. Density Matrix Decay for Systems with Gap.** The previous results establish exponential decay bounds for the Fermi–Dirac function of general localized Hamiltonians and thus for density matrices of arbitrary systems at positive electronic temperature. In this subsection we consider the case of gapped systems (like insulators) at zero temperature. In this case, as we know, the density matrix is the spectral projector onto the occupied subspace. As an example, we consider the density matrix corresponding to the linear alkane n-Dopentacontane  $C_{52}H_{106}$  composed of 52 carbon and 106 hydrogen atoms, discretized in a Gaussian-type orbital basis. The number of occupied states is 209, or half the total number of electrons in the system.<sup>7</sup> The corresponding Hamiltonian in the original nonorthogonal basis is displayed in Figure 9.1 (top) and the “orthogonalized” Hamiltonian  $\tilde{H}$  is shown in Figure 9.1 (bottom). Figure 8.5 displays the zero-temperature density matrix, which is seen to decay exponentially away from the main diagonal. Comparing Figure 8.5 and Figure 9.1, we can see that for a truncation level of  $10^{-8}$ , the bandwidth of the density matrix is only slightly larger than that of the Hamiltonian. The eigenvalue spectrum of the Hamiltonian, scaled and shifted so that its spectrum is contained in the interval  $[-1, 1]$ , is shown in Figure 8.6. One can clearly see a large gap ( $\approx 1.4$ ) between the 52 low-lying eigenvalues corresponding to the core electrons in the system, as well as the smaller HOMO-LUMO gap ( $\approx 0.1$ ) separating the 209 occupied states from the virtual (unoccupied) ones. It is worth emphasizing that the exponential decay of the density matrix is independent of the size of the system; that is, if the alkane chain was made arbitrarily long by adding C and H atoms to it, the density matrix would be of course much larger in size, but its bandwidth would remain virtually unchanged for the same truncation level, due to the fact that the bandwidth and the HOMO-LUMO gap of the Hamiltonian do not appreciably change as the number of particles increases. It is precisely this independence of the rate of decay (hence, of the bandwidth) to system size that makes  $O(n)$  approximations possible (and competitive) for large  $n$ .

Let us now see how Theorem 8.1 can be used to prove decay bounds on the entries of density matrices. Let  $H$  be the discrete Hamiltonian associated with a certain physical system and let  $\mu$  be the Fermi level of interest for this system. We assume that the spectrum of  $H$  has a gap  $\gamma$  around  $\mu$ , that is, we have  $\gamma = \varepsilon^+ - \varepsilon^- > 0$ , where

<sup>7</sup>Here spin is being taken into account, so that the density kernel is given by  $\rho(\mathbf{r}, \mathbf{r}') = 2 \sum_{i=1}^{n_e/2} \psi_i(\mathbf{r}) \psi_i(\mathbf{r}')^*$ ; see, e.g., [88, p. 10].



**Fig. 8.5** Magnitude of the entries in the density matrix for the linear alkane  $C_{52}H_{106}$  chain, with 209 occupied states. White:  $< 10^{-8}$ ; yellow:  $10^{-8} - 10^{-6}$ ; green:  $10^{-6} - 10^{-4}$ ; blue:  $10^{-4} - 10^{-2}$ ; black:  $> 10^{-2}$ . Note:  $nz$  refers to the number of “black” entries.



**Fig. 8.6** Spectrum of the Hamiltonian for  $C_{52}H_{106}$ .

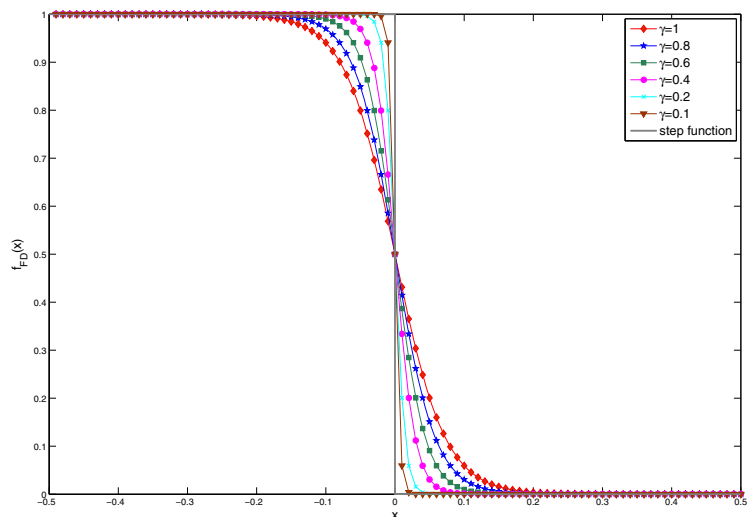
$\varepsilon^+$  is the smallest eigenvalue of  $H$  to the right of  $\mu$  and  $\varepsilon^-$  is the largest eigenvalue of  $H$  to the left of  $\mu$ . In the particular case of the HOMO-LUMO gap, we have  $\varepsilon^- = \varepsilon_{n_e}$  and  $\varepsilon^+ = \varepsilon_{n_e+1}$ .

The Fermi–Dirac function can be used to approximate the Heaviside function; the larger  $\beta$ , the better the approximation. More precisely, the following result is easy to prove (see [106]).

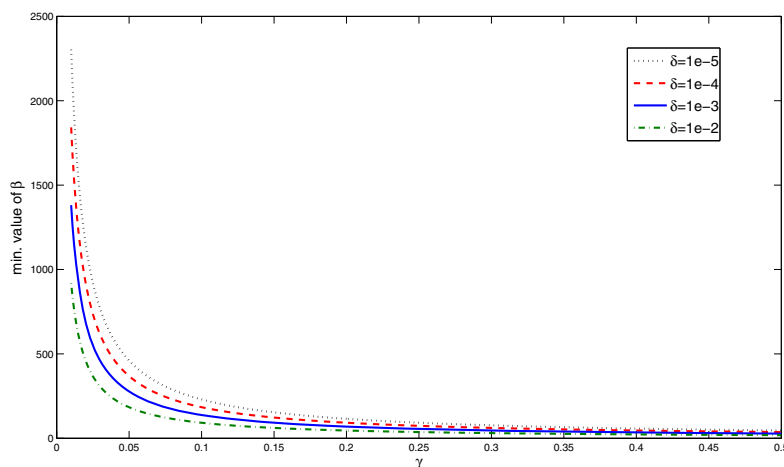
PROPOSITION 8.5. Let  $\delta > 0$  be given. If  $\beta$  is such that

$$(8.6) \quad \beta \geq \frac{2}{\gamma} \ln \left( \frac{1-\delta}{\delta} \right),$$

then  $1 - f_{FD}(\varepsilon^-) \leq \delta$  and  $f_{FD}(\varepsilon^+) \leq \delta$ .



**Fig. 8.7** Approximations of the Heaviside function by the Fermi-Dirac function ( $\mu = 0$ ) for different values of  $\gamma$  and  $\delta = 10^{-6}$ .



**Fig. 8.8** Behavior of the minimum acceptable value of  $\beta$  as a function of  $\gamma$  for different values of  $\delta$ .

In Figure 8.7 we show Fermi-Dirac approximations to the Heaviside function (with a jump at  $\mu = 0$ ) for different values of  $\gamma$  between 0.1 and 1, where  $\beta$  has been chosen so as to reduce the error in Proposition 8.5 above the value  $\delta = 10^{-6}$ . The behavior of  $\beta$  as a function of  $\gamma$  according to (8.6) is plotted in Figure 8.8.

As a consequence of Theorem 8.1 and Proposition 8.5 we have the following corollary.

**COROLLARY 8.6.** *Let  $n_b$  be a fixed positive integer and  $n = n_b \cdot n_e$ , where the integers  $n_e$  form a monotonically increasing sequence. Let  $\{H_n\}$  be a sequence of Hermitian  $n \times n$  matrices with the following properties:*

1. Each  $H_n$  has bandwidth  $m$  independent of  $n$ .

2. There exist two fixed intervals  $I_1 = [-1, a]$ ,  $I_2 = [b, 1] \subset \mathbb{R}$ , with  $\gamma = b - a > 0$ , such that for all  $n = n_b \cdot n_e$ ,  $I_1$  contains the smallest  $n_e$  eigenvalues of  $H_n$  (counted with their multiplicities) and  $I_2$  contains the remaining  $n - n_e$  eigenvalues.

Let  $P_n$  denote the  $n \times n$  spectral projector onto the subspace spanned by the eigenvectors associated with the  $n_e$  smallest eigenvalues of  $H_n$  for each  $n$ . Let  $\delta > 0$  be arbitrary. Then there exist constants  $c > 0, \alpha > 0$  independent of  $n$  such that

$$(8.7) \quad |[P_n]_{ij}| \leq \min \left\{ 1, c e^{-\alpha|i-j|} \right\} + \delta \quad \text{for all } i \neq j.$$

The constants  $c$  and  $\alpha$  can be computed from (8.2) and (8.3), where  $\chi$  is chosen in the interval  $(1, \bar{\chi})$ , with  $\bar{\chi}$  given by (8.4) and  $\beta$  such that (8.6) holds.

Corollary 8.6 allows us to determine a priori a bandwidth  $\bar{m}$  independent of  $n$  outside of which the entries of  $P_n$  are smaller than a prescribed tolerance  $\tau > 0$ . Observe that it is not possible to incorporate  $\delta$  in the exponential bound, but, at least in principle, one may always choose  $\delta$  smaller than a certain threshold. For instance, one may take  $\delta < \tau/2$  and define  $\bar{m}$  as the smallest integer value of  $m$  such that the relation  $c e^{-\alpha m} \leq \tau/2$  holds.

In the case of Hamiltonians with a general sparsity pattern one may apply Theorem 8.4 to obtain a more general version of Corollary 8.6. If the fixed bandwidth hypothesis is removed, the following bound holds:

$$(8.8) \quad |[P_n]_{ij}| \leq \min \left\{ 1, c e^{-\theta d_n(i,j)} \right\} + \delta \quad \text{for all } i \neq j,$$

with  $\theta = \ln \chi$ . Once again, for the result to be meaningful some restriction on the sparsity patterns, like the uniformly bounded maximum degree assumption already discussed, must be imposed.

**8.3. Proof of Decay Bounds.** Theorem 8.1 is a consequence of results proved in [12, Theorem 2.2] and [106, Theorem 2.2]; its proof relies on a fundamental result in polynomial approximation theory known as Bernstein's theorem [92]. Given a function  $f$  continuous on  $[-1, 1]$  and a positive integer  $k$ , the  $k$ th best approximation error for  $f$  is the quantity

$$E_k(f) = \inf \left\{ \max_{-1 \leq x \leq 1} |f(x) - p(x)| : p \in P_k \right\},$$

where  $P_k$  is the set of all polynomials with real coefficients and degree less than or equal to  $k$ . Bernstein's theorem describes the asymptotic behavior of the best approximation error for a function  $f$  analytic on a domain containing the interval  $[-1, 1]$ .

Consider the family of ellipses in the complex plane with foci in  $-1$  and  $1$ . As already mentioned, an ellipse in this family is completely determined by the sum  $\chi > 1$  of its half-axes and will be denoted as  $\mathcal{E}_\chi$ .

**THEOREM 8.7 (Bernstein).** *Let the function  $f$  be analytic in the interior of the ellipse  $\mathcal{E}_\chi$  and continuous on  $\mathcal{E}_\chi$ . Moreover, assume that  $f(z)$  is real for real  $z$ . Then*

$$E_k(f) \leq \frac{2M(\chi)}{\chi^k(\chi - 1)},$$

where  $M(\chi) = \max_{z \in \mathcal{E}_\chi} |f(z)|$ .

Let us now consider the special case where  $f(z) := f_{FD}(z) = 1/(1 + e^{\beta(z-\mu)})$  is the Fermi–Dirac function of parameters  $\beta$  and  $\mu$ . Observe that  $f_{FD}(z)$  has poles in  $\mu \pm i\frac{\pi}{\beta}$ , so the admissible values for  $\chi$  with respect to  $f_{FD}(z)$  are given by  $1 < \chi < \bar{\chi}$ , where the parameter  $\bar{\chi}$  is such that  $\mu \pm i\frac{\pi}{\beta} \in \mathcal{E}_{\bar{\chi}}$  (the *regularity ellipse* for  $f = f_{FD}$ ). Also observe that smaller values of  $\beta$  correspond to a greater distance between the poles of  $f_{FD}(z)$  and the real axis, which in turn yields a larger value of  $\bar{\chi}$ . In other words, the smaller  $\beta$ , the faster the decay in Theorem 8.1. Explicit computation of  $\bar{\chi}$  yields (8.4).

Now, let  $H_n$  be as in Theorem 8.1. We have

$$\|f_{FD}(H_n) - p_k(H_n)\|_2 = \max_{x \in \sigma(H_n)} |f_{FD}(x) - p_k(x)| \leq E_k(f_{FD}) \leq cq^{k+1},$$

where  $c = 2\chi M(\chi)/(\chi - 1)$  and  $q = 1/\chi$ . The Bernstein approximation of degree  $k$  gives a bound on  $|[f_{FD}(H_n)]_{ij}|$  when  $[p_k(H_n)]_{ij} = 0$ , that is, when  $|i - j| > mk$ . We may also assume  $|i - j| \leq m(k + 1)$ . Therefore, we have

$$|[f_{FD}(H_n)]_{ij}| \leq ce^{m(k+1)\ln(q^{1/m})} = ce^{-\alpha m(k+1)} \leq ce^{-\alpha|i-j|}.$$

As for Theorem 8.4, note that for a general sparsity pattern we have  $[(H_n)^k]_{ij} = 0$ , and therefore  $[p_k(H_n)]_{ij} = 0$ , whenever  $d_n(i, j) > k$ . Writing  $d_n(i, j) = k + 1$  we obtain

$$|[f_{FD}(H_n)]_{ij}| \leq c(1/\chi)^{k+1} = ce^{-\theta d_n(ij)}.$$

Let us now prove Corollary 8.6. Assume that  $\beta$  satisfies the inequality (8.6) for given values of  $\delta$  and  $\gamma$ . If we approximate the Heaviside function with step at  $\mu$  by means of the Fermi–Dirac function  $f_{FD}(x) = 1/(1 + e^{\beta(x-\mu)})$ , the pointwise approximation error is given by  $g(x) = e^{\beta(x-\mu)}/(1 + e^{\beta(x-\mu)})$  for  $x < \mu$  and by  $f_{FD}(x)$  for  $x > \mu$ . It is easily seen that  $g(x)$  is a monotonically increasing function, whereas  $f_{FD}$  is monotonically decreasing. As a consequence, for each Hamiltonian  $H_n$  we have that  $1 - f_{FD}(\lambda) \leq \delta$  for all eigenvalues  $\lambda \in I_1$  and  $f_{FD}(\lambda) \leq \delta$  for all  $\lambda \in I_2$ . In other words, the pointwise approximation error on the spectrum of  $H_n$  is always bounded by  $\delta$ . Therefore, we have

$$|[P_n - f_{FD}(H_n)]_{ij}| \leq \|P_n - f_{FD}(H_n)\|_2 \leq \delta.$$

We may then conclude using Theorem 8.1 that

$$|[P_n]_{ij}| \leq |[f_{FD}(H_n)]_{ij}| + \delta \leq ce^{-\alpha|i-j|} + \delta.$$

Finally, recall that in an orthogonal projector no entry can exceed unity in absolute value. With this in mind, (8.7) and (8.8) readily follow.

**8.4. Additional Bounds.** Theorems 8.1 and 8.4 rely on Bernstein’s result on best polynomial approximation. Following the same argument, one may derive decay bounds for the density matrix from any other estimate on the best polynomial approximation error for classes of functions that include the Fermi–Dirac function. For instance, consider the following result of Achieser (see [92, Theorem 78] and [1]).

**THEOREM 8.8.** *Let the function  $f$  be analytic in the interior of the ellipse  $\mathcal{E}_\chi$ . Suppose that  $|\operatorname{Re} f(z)| < 1$  holds in  $\mathcal{E}_\chi$  and that  $f(z)$  is real for real  $z$ . Then the following bound holds:*

$$(8.9) \quad E_k(f) \leq \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{(2\nu + 1) \cosh((2\nu + 1)(k + 1) \ln \chi)}.$$

The series in (8.9) converges quite fast; therefore, it suffices to compute a few terms explicitly to obtain a good approximation of the bound. A rough estimate shows that, in order to approximate the right-hand side of (8.9) within a tolerance  $\tau$ , one may truncate the series after  $\nu_0$  terms, where  $r^{\nu_0} < \tau(1-r)$  and  $r = \chi^{-\frac{k+1}{2}}$ .

Observe that, as in Bernstein's results, there is again a degree of arbitrariness in the choice of  $\chi$ . However, the admissible range for  $\chi$  is smaller here because of the hypothesis  $|\operatorname{Re} f(z)| < 1$ .

The resulting matrix decay bounds have the form

$$(8.10) \quad |[f_{FD}(H_n)]_{ij}| \leq \frac{4}{\pi} \sum_{\nu=0}^{\infty} \frac{(-1)^\nu}{(2\nu+1) \cosh((2\nu+1)(d(i,j)+1) \ln \chi)}$$

for the case of general sparsity patterns. While these bounds are less transparent than those derived from Bernstein's theorem, they are computable. We have found that the bounds (8.10) improve on (8.1) for entries close to the main diagonal, but do not seem to have a better asymptotic behavior. A possibility would be to combine the two bounds by taking the smaller of the two values.

So far we have only considered bounds based on best approximation of analytic functions defined on a single interval. In [61], Hasson obtained an interesting result on polynomial approximation of a step function defined on the union of two symmetric intervals. Let  $a, b \in \mathbb{R}$  with  $0 < a < b$  and let  $\operatorname{sgn}(x)$  be the sign function defined on  $[-b, -a] \cup [a, b]$ , i.e.,  $\operatorname{sgn}(x) = -1$  on  $[-b, -a]$  and  $\operatorname{sgn}(x) = 1$  on  $[a, b]$ . Notice that the sign function is closely related to the Heaviside function  $h(x)$ , since we have  $h(x) = \frac{1}{2}(1 + \operatorname{sgn}(x))$ .

PROPOSITION 8.9. *There exists a positive constant  $K$  such that*

$$(8.11) \quad E_k(\operatorname{sgn}; [-b, -a] \cup [a, b]) \leq K \frac{\left(\sqrt{\frac{b-a}{b+a}}\right)^k}{\sqrt{k}}.$$

Given a sequence of Hamiltonians  $\{H_n\}$  with gapped spectra, one may choose  $a$  and  $b$  and shift  $H_n$ , if necessary, so that the spectrum of each  $H_n$  is contained in  $[-b, -a] \cup [a, b]$  and the eigenvalues corresponding to occupied states belong to  $[-b, -a]$ . Then we obtain the following decay bound for the density matrix:

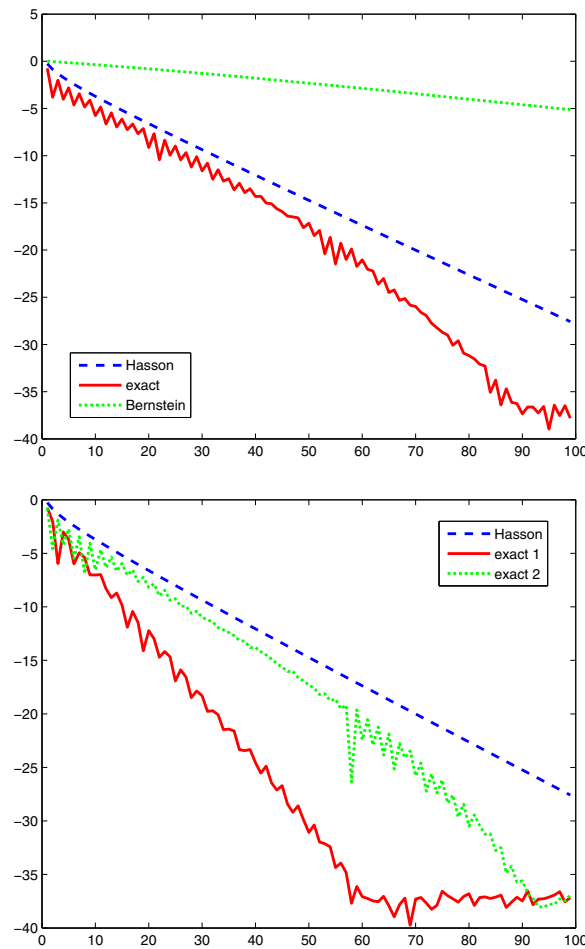
$$(8.12) \quad |[P_n]_{ij}| \leq K \frac{e^{-\xi d(i,j)}}{2\sqrt{d(i,j)}}, \quad \text{where } \xi = \frac{1}{2} \ln \frac{b+a}{b-a}.$$

Under the bounded maximal degree condition, the rate of decay is independent of  $n$ .

A few remarks on (8.12) are in order:

- Since (8.12) relies directly on a polynomial approximation of the step function, we do not need here the extra term  $\delta$  found in (8.8).
- Unfortunately, it is not possible to assess whether (8.12) may be useful in practice without an explicit formula—or at least an estimate—for the constant  $K$ . The asymptotic decay rate, however, is faster than exponential and indeed faster than for other bounds; a comparison is shown in Figure 8.9 (top). Notice that this logarithmic plot is only meant to show the slope of the bound (which is computed for  $K = 1$ ).
- A disadvantage of (8.12) is the requirement that the intervals containing the spectra  $\sigma(H_n)$  should be symmetric with respect to 0. Of course one may always choose  $a$  and  $b$  so that this hypothesis is satisfied, but the quality of





**Fig. 8.9** *Top: Logarithmic plot of Hasson (dashed line) and Bernstein-type (dotted line) decay bounds for a  $100 \times 100$  tridiagonal matrix with spectrum in  $[-1, -0.25] \cup [0.25, 1]$ . The solid line plots the first row of the “exact” density matrix. Bottom: Logarithmic plot of Hasson decay bounds (dashed line) and first rows of density matrices associated with matrices with different eigenvalue distributions (solid and dotted lines).*

the decay bound deteriorates if  $b$  (or  $-b$ ) is not close to the maximum (resp., minimum) eigenvalue; see Figure 8.9 (bottom). The dashed line shows the slope of the decay bound for  $a = 0.25$  and  $b = 1$ , in a logarithmic scale. The dotted line displays the behavior of the first row of the density matrix associated with a tridiagonal  $100 \times 100$  matrix with spectrum in  $[-1, -0.25] \cup [0.25, 1]$ . The solid line refers to the first row of the density matrix associated with a matrix with spectrum in  $[-0.4375, -0.25] \cup [0.25, 1]$ . The first matrix is clearly better approximated by the decay bound than the second one.

As one can see from the two plots in Figure 8.9, even for  $c = K = 1$  both types of decay bounds are rather conservative, and estimating the truncation bandwidth  $\bar{m}$  needed to achieve a prescribed error from these bounds would lead to an overly large band. Hence, the bounds may not be very useful in practice. For further discussion of these issues, see section 8.9.

**8.5. Further Results.** Let us assume again that we have a sequence  $\{H_n\}$  of Hermitian  $n \times n$  Hamiltonians (with  $n = n_b \cdot n_e$ ,  $n_b$  fixed,  $n_e \rightarrow \infty$ ) such that

- the matrices  $H_n$  are banded with uniformly bounded bandwidth, or sparse with graphs having uniformly bounded maximum degree;
- the spectra  $\sigma(H_n)$  are uniformly bounded;
- the sequence  $\{H_n\}$  has a “stable” spectral gap; i.e., there exist real numbers  $g_1 < g_2$  such that  $[g_1, g_2] \cap \sigma(H_n) = \emptyset$  for sufficiently large  $n$ .

In this subsection we let

- $\mu := (g_2 + g_1)/2$  (Fermi level);
- $\gamma := g_2 - \mu = \mu - g_1$  (absolute spectral gap).

Note that because of the uniformly bounded spectra assumption, the absolute spectral gap is within a constant of the relative gap previously defined.

Chui and Hasson study in [29] the asymptotic behavior of the error of best polynomial approximation for a sufficiently smooth function  $f$  defined on the set  $I = [-b, -a] \cup [a, b]$ , with  $0 < a < b$ . Denote as  $\mathcal{C}(I)$  the space of real-valued continuous functions on  $I$ , with the uniform convergence norm. Then we have (see [29, Theorem 1] and [84]) the following theorem.

**THEOREM 8.10.** *Let  $f \in \mathcal{C}(I)$  be such that  $f|_{[-b, -a]}$  is the restriction of a function  $f_1$  analytic on the left half-plane  $\operatorname{Re} z < 0$  and  $f|_{[a, b]}$  is the restriction of a function  $f_2$  analytic on the right half-plane  $\operatorname{Re} z > 0$ . Then*

$$\limsup_{k \rightarrow \infty} [E_k(f, I)]^{1/k} \leq \sqrt{\frac{b-a}{b+a}},$$

where  $E_k(f, I)$  is the error of best polynomial approximation for  $f$  on  $I$ .

The authors of [29] observe that the above result cannot be obtained by extending  $f(x)$  to a continuous function on  $[-b, b]$  and applying known bounds for polynomial approximation over a single interval. Theorem 8.10 looks potentially useful for our purposes, except that it provides an asymptotic result, rather than an explicit bound for each value of  $k$ . Therefore, we need to reformulate the argument in [29]. To this end, we prove a variant of Bernstein’s theorem (cf. Theorem 8.7) adapted to our goals. Instead of working on the interval  $[-1, 1]$ , we want to bound the approximation error on the interval  $[a^2, b^2]$ .

**THEOREM 8.11.** *Let  $f \in \mathcal{C}([a^2, b^2])$  be the restriction of a function  $f$  analytic in the interior of the ellipse  $\mathcal{E}_{a^2, b^2}$  with foci in  $a^2, b^2$  and a vertex at the origin. Then, for all  $\xi$  with*

$$1 < \xi < \bar{\xi} := \frac{a+b}{a-b},$$

there exists a constant  $K$  such that

$$E_k(f, [a^2, b^2]) \leq K \left( \frac{1}{\xi} \right)^k.$$

*Proof.* The proof closely parallels the argument given in [92] for the proof of Theorem 8.7. First of all, observe that the ellipse  $\mathcal{E}_\chi$  in Bernstein’s theorem has foci in  $\pm 1$  and vertices in  $\pm(\chi + 1/\chi)/2$  and  $\pm(\chi - 1/\chi)/2$ . The parameter  $\chi$  is the sum of the lengths of the semiaxes. Similarly, the ellipse  $\mathcal{E}_{a^2, b^2}$  has foci in  $a^2, b^2$  and vertices in  $0, a^2 + b^2$ , and  $(a^2 + b^2)/2 \pm iab$ . Also observe that  $\bar{\xi}$  is the sum of the lengths of the semiaxes of  $\mathcal{E}_{a^2, b^2}$ , normalized with respect to the semifocal length, so that it plays

exactly the same role as  $\chi$  for  $\mathcal{E}_\chi$ . Now we look for a conformal map that sends an annulus in the complex plane to the ellipse where  $f$  is analytic. When this ellipse is  $\mathcal{E}_\chi$ , a suitable map is  $u = c(v) = (v + 1/v)/2$ , which sends the annulus  $\chi^{-1} < |v| < \chi$  to  $\mathcal{E}_\chi$ . When the desired ellipse has foci in  $a^2, b^2$ , we compose  $c(v)$  with the change of variable

$$x = \psi(u) = \left( u + \frac{a^2 + b^2}{b^2 - a^2} \right) \frac{b^2 - a^2}{2},$$

thus obtaining a function that maps the annulus  $\mathcal{A} = \{\xi^{-1} < |v| < \xi\}$  to an ellipse. Denote this ellipse as  $\mathcal{E}_{a^2, b^2, \xi}$  and observe that it is contained in the interior of  $\mathcal{E}_{a^2, b^2}$ . Therefore we have that the function

$$f(\psi(c(v))) = f\left(\left[\frac{1}{2}\left(v + \frac{1}{v}\right) + \frac{a^2 + b^2}{b^2 - a^2}\right] \frac{b^2 - a^2}{2}\right)$$

is analytic on  $\mathcal{A}$  and continuous on  $|v| = \xi$ . The proof now proceeds as in the original Bernstein theorem. The Laurent expansion

$$f(\psi(c(v))) = \sum_{\nu=-\infty}^{\infty} \alpha_\nu v^\nu$$

converges in  $\mathcal{A}$  with  $\alpha_{-\nu} = \alpha_\nu$ . Moreover, we have the bound

$$|\alpha_\nu| = \left| \frac{1}{2\pi i} \int_{|v|=\xi} \frac{f(\psi(c(v)))}{v^{\nu+1}} dv \right| \leq \frac{M(\xi)}{\xi^\nu},$$

where  $M(\xi)$  is the maximum value (in modulus) taken by  $f$  on the ellipse  $\mathcal{E}_{a^2, b^2, \xi}$ .

Now observe that  $u = c(v)$  describes the real interval  $[-1, 1]$  for  $|v| = 1$ , so for  $u \in [-1, 1]$  we have

$$f(\psi(u)) = \alpha_0 + 2 \sum_{\nu=1}^{\infty} \alpha_\nu T_\nu(u),$$

where  $T_\nu(u)$  is the  $\nu$ th Chebyshev polynomial. Since  $\psi(u)$  is a linear transformation, we have  $E_k(f(z), [a^2, b^2]) = E_k(f(u), [-1, 1])$ , so from the theory of Chebyshev approximation [92] we obtain

$$E_k(f, [a^2, b^2]) = E_k(f(u), [-1, 1]) \leq 2M(\xi) \sum_{\nu=k+1}^{\infty} \xi^{-\nu} = \frac{2M(\xi)}{\xi - 1} \xi^{-k},$$

hence the thesis. Note that the explicit value of  $K$  is computable. □

The following result is based on [29, Theorem 1].

**THEOREM 8.12.** *Let  $f \in \mathcal{C}(I)$  be as in Theorem 8.10. Then, for all  $\xi$  with*

$$1 < \xi < \bar{\xi} := \frac{a + b}{a - b},$$

*there exists  $C > 0$  independent of  $k$  such that*

$$E_k(f, I) \leq C \xi^{-\frac{k}{2}}.$$

*Proof.* Let  $P_k$  and  $Q_k$  be polynomials of best uniform approximation of degree  $k$  on the interval  $[a^2, b^2]$  for the functions  $f_2(\sqrt{x})$  and  $f_2(\sqrt{x})/\sqrt{x}$ , respectively. Then by Theorem 8.11 there are constants  $K_1$  and  $K_2$  such that

$$(8.13) \quad \max_{x \in [a^2, b^2]} |P_k(x) - f_2(\sqrt{x})| \leq K_1 \xi^{-k}$$

and

$$(8.14) \quad \max_{x \in [a^2, b^2]} |Q_k(x) - f_2(\sqrt{x})/\sqrt{x}| \leq K_2 \xi^{-k}.$$

We use the polynomials  $P_k$  and  $Q_k$  to define a third polynomial  $R_{2k+1}(x) := [P_k(x^2) + xQ_k(x^2)]/2$ , of degree  $\leq 2k+1$ , which approximates  $f(x)$  on  $[a, b]$  and has small norm on  $[-b, -a]$ . Indeed, from (8.13) and (8.14) we have

$$(8.15) \quad \begin{aligned} \max_{x \in [a, b]} |R_{2k+1}(x) - f(x)| &\leq \frac{1}{2} \max_{x \in [a, b]} |P_k(x^2) - f(x)| + \frac{1}{2} \max_{x \in [a, b]} |xQ_k(x^2) - f(x)| \\ &\leq \frac{1}{2} K_1 \xi^{-k} + \frac{1}{2} b K_2 \xi^{-k} = \frac{K_1 + bK_2}{2} \xi^{-k} \end{aligned}$$

and

$$(8.16) \quad \max_{x \in [-b, -a]} |R_{2k+1}(x)| \leq \frac{1}{2} \max_{x \in [a, b]} |P_k(x^2) - f(x) + f(x) - xQ_k(x^2)|$$

$$(8.17) \quad \leq \frac{1}{2} \max_{x \in [a, b]} |P_k(x^2) - f(x)| + \frac{1}{2} \max_{x \in [a, b]} |xQ_k(x^2) - f(x)| \leq \frac{K_1 + bK_2}{2} \xi^{-k}.$$

Similarly, we can find another polynomial  $S_{2k+1}(x)$  such that

$$(8.18) \quad \max_{x \in [-b, -a]} |S_{2k+1}(x) - f(x)| \leq \frac{K_3 + bK_4}{2} \xi^{-k}$$

and

$$(8.19) \quad \max_{x \in [a, b]} |S_{2k+1}(x)| \leq \frac{K_3 + bK_4}{2} \xi^{-k}.$$

Then from the inequalities (8.15)–(8.19) we have

$$\begin{aligned} \max_{x \in I} |R_{2k+1}(x) + S_{2k+1}(x) - f(x)| &\leq \max_{x \in [a, b]} |R_{2k+1}(x) - f(x)| + \max_{x \in [a, b]} |S_{2k+1}(x)| \\ &\quad + \max_{x \in [-b, -a]} |S_{2k+1}(x) - f(x)| + \max_{x \in [-b, -a]} |R_{2k+1}(x)| \\ &\leq (K_1 + K_3 + b(K_2 + K_4)) \xi^{-k}, \end{aligned}$$

and therefore

$$E_k(f, I) \leq \sqrt{\xi} (K_1 + K_3 + b(K_2 + K_4)) \xi^{-\frac{k}{2}}$$

for odd values of  $k$  and

$$E_k(f, I) \leq \xi (K_1 + K_3 + b(K_2 + K_4)) \xi^{-\frac{k}{2}}$$

for even values of  $k$ . This completes the proof.  $\square$

In the following we assume, without loss of generality, that  $k$  is odd. In order to obtain bounds on the density matrix, we apply Theorem 8.12 to the step function  $f$  defined on  $I$  as

$$f(x) = \begin{cases} 1 & \text{for } -b \leq x \leq -a, \\ 0 & \text{for } a \leq x \leq b; \end{cases}$$

i.e.,  $f$  is the restriction of  $f_1(z) \equiv 1$  on  $[-b, -a]$  and the restriction of  $f_2(z) \equiv 0$  on  $[a, b]$ . Here the polynomial approximation of  $f_2(\sqrt{x})$ ,  $f_2(\sqrt{x})/\sqrt{x}$ , and  $f_1(\sqrt{-x})$  is exact, so we have  $K_1 = K_2 = K_3 = 0$ . As for  $K_4$ , observe that  $|1/\sqrt{z}|$  achieves its maximum on the vertex of  $\mathcal{E}_{a^2, b^2, \xi}$  with smallest abscissa; therefore we have

$$K_4 = \frac{2M(\xi)}{\xi - 1},$$

where

$$M(\xi) = \frac{1}{\sqrt{z_0}} \quad \text{with} \quad z_0 = \left[ -\frac{1}{2} \left( \xi + \frac{1}{\xi} \right) + \frac{a^2 + b^2}{b^2 - a^2} \right] \frac{b^2 - a^2}{2}.$$

Moreover, we find  $R_{2k+1}(x) \equiv 0$  and  $S_{2k+1}(x) = (1 + xV_k(x^2))/2$ , where  $V_k(x)$  is the polynomial of best uniform approximation for  $1/\sqrt{x}$  on  $[a^2, b^2]$ . Thus, we obtain the bound

$$E_k(f, I) \leq C\xi^{-\frac{k}{2}},$$

where  $C$  is given by

$$C = \sqrt{\xi} K_4 b.$$

Let us now apply this result to our sequence of Hamiltonians. We will assume that the matrices are shifted so that  $\mu = 0$ , that is, we replace each  $H_n$  by  $H_n - \mu I_n$ . Under this hypothesis, the natural choice for  $a$  is  $a = \gamma$ , whereas  $b$  is the smallest number such that  $\sigma(H_n) \subset [-b, -a] \cup [a, b]$  for every  $n$ .

Using the same argument used in section 8.3 for the derivation of matrix decay bounds (see also [12] and [14]), we can obtain bounds on the off-diagonal entries of  $f(H_n)$ . If  $H_n$  is banded with bandwidth  $m$  independent of  $n$ , we have

$$(8.20) \quad |[P_n]_{ij}| = |[f(H_n)]_{ij}| \leq \sqrt{\xi} \frac{2M(\xi)}{\xi - 1} b \xi^{-\frac{|i-j|}{2m}},$$

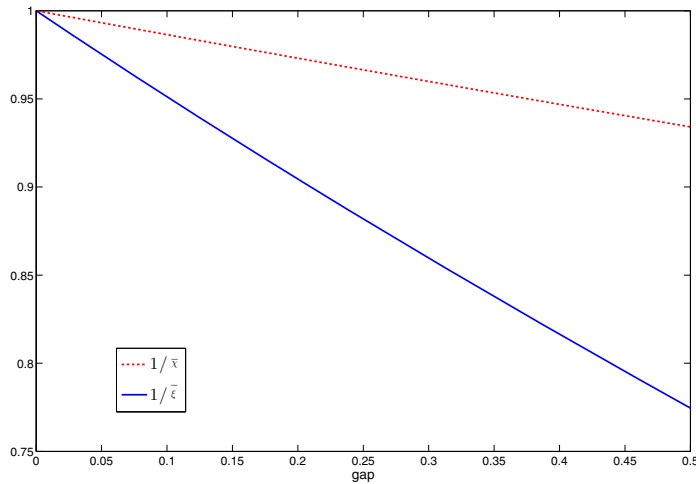
whereas if  $H_n$  has a more general sparsity pattern, we obtain

$$(8.21) \quad |[P_n]_{ij}| = |[f(H_n)]_{ij}| \leq \sqrt{\xi} \frac{2M(\xi)}{\xi - 1} b \xi^{-\frac{d_n(i,j)}{2}},$$

where  $d_n(i, j)$  is the distance between nodes  $i$  and  $j$  in the graph  $G_n$  associated with  $H_n$ .

Next, we compare the bounds derived in this section with those for the Fermi–Dirac approximation of the step function obtained in section 8.1, using a suitable choice of the inverse temperature  $\beta$ . Recall that if  $\mathcal{E}_\chi$  denotes the regularity ellipse for the Fermi–Dirac function, the earlier bounds for the banded case are

$$(8.22) \quad |[P_n]_{ij}| \leq \frac{2M(\chi)}{\chi - 1} \left( \frac{1}{\chi} \right)^{\frac{|i-j|}{m}}.$$



**Fig. 8.10** Comparison of parameters  $1/\bar{\xi}$  and  $1/\bar{\chi}$  for several values of the spectral gap. Here  $\delta = 10^{-5}$ .

For ease of computation, we assume in this section that  $\mu = 0$  and that the spectrum of each matrix  $H_n$  is contained in  $[-1, 1]$ . As explained in section 8.1, once  $\gamma$  is known, we pick a tolerance  $\delta$  and compute  $\beta$  so that the Fermi–Dirac function provides a uniform approximation of the step function with error  $\leq \delta$  outside the gap:

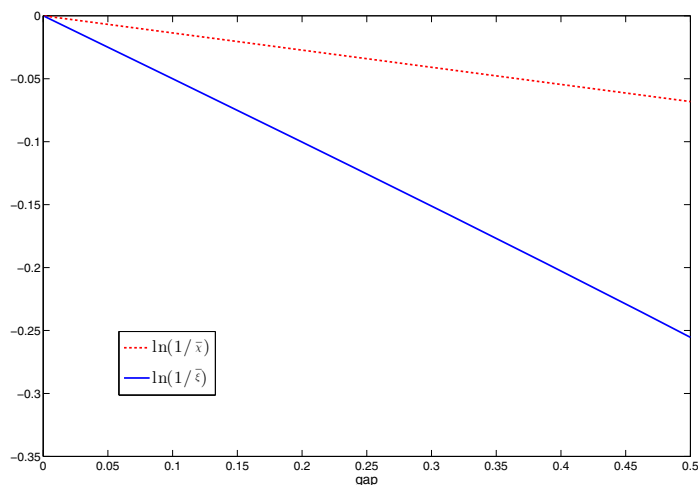
$$\beta \geq \frac{2}{\gamma} \ln \left( \frac{1 - \delta}{\delta} \right).$$

Then the supremum of the set of admissible values of  $\chi$ , which ensures optimal asymptotic decay in this framework, is

$$\bar{\chi} = \left( \pi + \sqrt{\beta^2 + \pi^2} \right) / \beta.$$

Figures 8.10 and 8.11 compare the values of  $1/\bar{\xi}$  and  $1/\bar{\chi}$  (which characterize the behavior of the bounds (8.20) and (8.22), respectively). Note that in general we find  $1/\bar{\xi} < 1/\bar{\chi}$ ; this means that the asymptotic decay rate is higher for the bound based on disjoint interval approximation. Moreover, the disjoint interval method directly approximates the step function and therefore does not require one to choose a tolerance for “intermediate” approximation. As a result, the bounds based on disjoint interval approximation prescribe a smaller truncation bandwidth  $\bar{m}$  in the approximation to the spectral projector in order to achieve a given level of error. For instance, in the tridiagonal case ( $m = 1$ ) we observed a factor of three reduction in  $\bar{m}$  compared to the previous bounds, independent of the size of the gap.

**8.6. Dependence of the Rate of Decay on the Spectral Gap.** As already mentioned in section 4, the functional dependence of the decay length (governing the rate of decay in the density matrix) on the spectral gap has been the subject of some discussion; see, for instance, [3, 70, 73, 104, 127, 140]. Some of these authors have argued that the decay length decreases like the square root of the gap if the Fermi level is located near one of the gap edges (i.e., close to either  $\varepsilon_{n_e}$  or  $\varepsilon_{n_e+1}$ ) and like the gap itself if the Fermi level falls in the middle of the gap. These estimates hold for the small gap limit.



**Fig. 8.11** Logarithmic plot of parameters  $1/\bar{\xi}$  and  $1/\bar{\chi}$  with respect to several values of the spectral gap. Here  $\delta = 10^{-5}$ .

In this section we address this problem by studying how the decay described by the bounds (8.20) and (8.21) behaves asymptotically with respect to  $\gamma$  or, equivalently, with respect to  $a$  (see the notation introduced in the previous section). Note that we are assuming here that the Fermi level falls exactly in the middle of the gap.

Let us rewrite (8.20) in the form

$$|[P_n]_{ij}| \leq C e^{-\alpha|i-j|/m},$$

where

$$\alpha = \frac{1}{2} \ln \xi = \frac{1}{2} \ln \left( \frac{a+b}{b-a} \right).$$

For a fixed  $m$ , the decay behavior is essentially described by the parameter  $\alpha$ . Let us assume for simplicity of notation that  $b = 1$ , so that the spectral gap is normalized and the expression for  $\alpha$  becomes

$$\alpha = \frac{1}{2} \ln \left( \frac{1+a}{1-a} \right).$$

The Taylor expansion of  $\alpha$  for  $a$  small yields

$$\alpha = a + \frac{a^3}{3} + o(a^3).$$

Therefore, for small values of  $\gamma$ , the decay behavior is described at first order by the gap itself, rather than by a more complicated function of  $\gamma$ . This result is consistent with similar ones found in the literature [70, 73, 140]. The fact that some systems exhibit density matrix decay lengths proportional to the square root of the gap (see, e.g., [73]) does not contradict our result: since we are dealing here with upper bounds, a square root dependence, which corresponds to faster decay for small  $a$ , is still consistent with our bounds. Given that our bounds are completely general, it does not come as a surprise that we obtain the more conservative estimate among the alternatives discussed in the literature.

**8.7. Dependence of the Rate of Decay on the Temperature.** Another issue that has stirred up some controversy in the literature concerns the precise rate of decay in the density matrix for metals at positive temperature; see, e.g., the results and discussion in [3, 52, 70]. Recall that in metals at positive temperatures  $T$ , the density matrix  $F_n = f_{FD}(H_n)$  decays exponentially. The question is whether the decay length is proportional to  $T$  or to  $\sqrt{T}$  for small  $T$ . Our approach shows that the decay length is proportional to  $T$ .

Indeed, from the analysis in section 8.1, in particular Theorems 8.1 and 8.4, we find that the decay length  $\alpha$  in the exponential decay bound (8.1) (or, more generally, the decay length  $\theta$  in the bound (8.5)) behaves like  $\ln \chi$ , where—assuming for simplicity that  $\mu = 0$ , as before—the parameter  $\chi$  is any number satisfying

$$1 < \chi < \bar{\chi}, \quad \bar{\chi} = \left( \pi + \sqrt{\beta^2 + \pi^2} \right) / \beta.$$

Letting  $x = \pi/\beta = \pi k_B T$  and observing that for small  $x$

$$\ln \left( x + \sqrt{1 + x^2} \right) = x + o(x^2),$$

we conclude that, at low temperatures, the decay length is proportional to  $k_B T$ . This conclusion is in complete agreement with the results in [52, 70]. To the best of our knowledge, this is the first time this result has been established in a fully rigorous and completely general manner.

**8.8. Other Approaches.** Decay bounds on the entries of spectral projectors can also be obtained from the contour integral representation

$$(8.23) \quad P_n = \frac{1}{2\pi i} \int_{\Gamma} (zI_n - H_n)^{-1} dz,$$

where  $\Gamma$  is a simple closed curve (counterclockwise oriented) in  $\mathbb{C}$  surrounding a portion of the real axis containing those eigenvalues of  $H_n$  that correspond to the occupied states, and only those. Componentwise, (8.23) becomes

$$[P_n]_{ij} = \frac{1}{2\pi i} \int_{\Gamma} [(zI_n - H_n)^{-1}]_{ij} dz, \quad 1 \leq i, j \leq n,$$

from which we obtain

$$|[P_n]_{ij}| \leq \frac{1}{2\pi} \int_{\Gamma} |[(zI_n - H_n)^{-1}]_{ij}| dz, \quad 1 \leq i, j \leq n.$$

Assume the matrices  $H_n$  are banded, with uniformly bounded spectra and bandwidths as  $n \rightarrow \infty$ . By [34, Proposition 2.3] there exist, for all  $z \in \Gamma$ , explicitly computable constants  $c(z) \geq 0$  and  $0 < \lambda(z) < 1$  (independent of  $n$ ) such that

$$(8.24) \quad |[(zI_n - H_n)^{-1}]_{ij}| \leq c(z)[\lambda(z)]^{|i-j|}$$

for all  $i, j = 1, \dots, n$ . Moreover,  $c$  and  $\lambda$  depend continuously on  $z \in \Gamma$ . Since  $\Gamma$  is compact we can set

$$(8.25) \quad c = \max_{z \in \Gamma} c(z) \quad \text{and} \quad \lambda = \max_{z \in \Gamma} \lambda(z).$$



Now let us assume that the matrices  $H_n$  have spectral gaps  $\gamma_n$  satisfying  $\inf_n \gamma_n > 0$ . It is then clear that  $c$  is finite and that  $\lambda \in (0, 1)$ . Hence, we obtain the following bound:

$$(8.26) \quad |[P_n]_{ij}| \leq \left( c \cdot \frac{\ell(\Gamma)}{2\pi} \right) \lambda^{|i-j|}$$

for all  $i, j = 1, \dots, n$ , where  $\ell(\Gamma)$  denotes the length of  $\Gamma$ . Finally, letting  $C = c \cdot \frac{\ell(\Gamma)}{2\pi}$  and  $\alpha = -\ln \lambda$ , we obtain the exponential decay bounds

$$(8.27) \quad |[P_n]_{ij}| \leq C \cdot e^{-\alpha|i-j|}, \quad 1 \leq i, j \leq n,$$

with both  $C > 0$  and  $\alpha > 0$  independent of  $n$ . As usual, the bounds can be easily extended to the case of general sparsity patterns. One disadvantage of this approach is that explicit evaluation of the constants  $C$  and  $\alpha$  is rather complicated.

The integral representation (8.23) is useful not only as a theoretical tool, but also increasingly as a computational tool. Indeed, quadrature rules with suitably chosen nodes  $z_1, \dots, z_k \in \Gamma$  can be used to approximate the integral in (8.23), leading to

$$(8.28) \quad P_n \approx \sum_{i=1}^k w_i (z_i I_n - H_n)^{-1}$$

for suitable quadrature weights  $w_1, \dots, w_k$ . Note that this amounts to a rational approximation of  $P_n = h(H_n)$ . In practice, using the trapezoidal rule with a small number of nodes suffices to achieve high accuracy, due to the exponential convergence of this quadrature rule for analytic functions [33]. Note that if  $P_n$  is real, then it is sufficient to use just the  $z_i$  in the upper half-plane and then take the real part of the result [65, p. 307]. If the spectral gap  $\gamma_n$  for  $H_n$  is not too small, all the resolvents  $(z_i I_n - H_n)^{-1}$  decay rapidly away from the main diagonal, with exponential rate independent of  $n_e$ . Hence,  $O(n)$  approximation is possible, at least in principle. Rational approximations of the type (8.28) are especially useful in those situations where only selected entries of  $P_n$  are required. Then only the corresponding entries of the resolvents  $(z_i I_n - H_n)^{-1}$  need to be computed. For instance, in some cases only the diagonal entries of  $P_n$  are needed [116]. In others, only entries in positions corresponding to the nonzero entries in the Hamiltonian  $H_n$  must be computed; this is the case, for instance, when computing the objective function  $\langle E \rangle = \text{Tr}(P_n H_n)$  in density matrix minimization algorithms. Computing selected entries of a resolvent is not an easy problem. However, progress has been made on this front in several recent papers; see, e.g., [78, 82, 83, 124, 126].

**8.9. Computational Considerations.** In the preceding sections we have rigorously established exponential decay bounds for zero-temperature density matrices corresponding to finite-range Hamiltonians with nonvanishing spectral gap (“insulators”), as well as for density matrices corresponding to arbitrary finite-range Hamiltonians at positive electronic temperatures. Our results are very general and apply to a wide variety of physical systems and discretizations. Hence, a mathematical justification of the physical phenomenon of “nearsightedness” has been obtained, and the possibility of  $O(n)$  methods firmly established.<sup>8</sup>

<sup>8</sup>Heuristics relating the “nearsightedness range of electronic matter” and the linear complexity of the divide-and-conquer method of Yang [138], essentially a domain decomposition approach to DFT, were given by Kohn himself; see, e.g., [75, 105].

Having thus achieved our main purpose, the question remains whether our estimates can be of practical use in the design of  $O(n)$  algorithms. As shown in section 6, having estimated the rate of decay in the density matrix  $P$  allows one to prescribe a priori a sparsity pattern for the computed approximation  $\tilde{P}$  to  $P$ . Estimating an “envelope” for the nonnegligible entries in  $P$  means that one can estimate beforehand the storage requirements and set up static data structures for the computation of the approximate density matrix  $\tilde{P}$ . An added advantage is the possibility of using the prescribed sparsity pattern to develop efficient parallel algorithms; it is well known that adaptive computations, in which the sparsity pattern is determined “on the fly,” may lead to load imbalances and loss of parallel efficiency due to the need for large amounts of communication and unpredictable memory accesses. This is completely analogous to prescribing a sparsity pattern vs. using an adaptive one when computing sparse approximate inverses for use as preconditioners when solving linear systems; see [10].

Most of the  $O(n)$  algorithms currently in use consist of iterative schemes producing increasingly accurate approximations to the density matrix. These approximations may correspond to successive terms in an expansion of  $P$  with respect to a prescribed basis [54, 80, 81], or they may be the result of a gradient or descent method in density matrix minimization approaches [23, 24, 79, 93]. Closely related methods include *purification* and algorithms based on approximating the sign function [95]; we refer the reader again to [20, 97, 113] for recent surveys on state-of-the-art linear scaling methods for electronic structure. Most of these algorithms construct a sequence of approximations

$$P^{(0)}, P^{(1)}, \dots, P^{(k)}, \dots,$$

which, under appropriate conditions, converge to  $P$ . Each iterate is obtained from the preceding one by some matrix-matrix multiplication, or powering, scheme; each step introduces new nonzeros (fill-in), and the matrices  $P^{(k)}$  become increasingly dense. The exponential decay property, however, implies that most of these nonzeros will be negligible, with only  $O(n)$  of them being above any prescribed threshold  $\delta > 0$ . Clearly, knowing a priori the location of the nonnegligible entries in  $P$  can be used to drastically reduce the computational burden and to achieve linear scaling, since only those entries need to be computed. Negligible entries that fall *within* the prescribed sparsity pattern may be removed using a drop tolerance; this strategy further decreases storage and arithmetic complexity, but its implementation demands the use of dynamic data structures.

An illustration of this use of the decay estimates can be found, for instance, in [14], where a Chebyshev expansion of the Fermi–Dirac function  $f_{FD}(H)$  was used to approximate the density matrix at finite temperatures. Given a prescribed error tolerance, exponential decay bounds were applied to the Fermi–Dirac function to determine the truncation bandwidth needed to satisfy the required approximation error. When computing the polynomial  $p_k(H) \approx f_{FD}(H)$  using the Chebyshev expansion, only entries within the prescribed bandwidth were retained. Combined with an estimate of the approximation error obtained by monitoring the magnitude of the coefficients in the Chebyshev expansion, this approach worked well for some simple 1D model problems resulting in linear scaling computations. A related approach, based on qualitative decay estimates for the density matrix, was used in [4], whose authors present computational results for a variety of 1D and 2D systems including insulators at zero temperature and metals at finite temperature; see further [80].

Unfortunately, the practical usefulness of our bounds for more realistic calculations is limited. The bounds are generally pessimistic and tend to be overly conservative, especially for the case of zero or low temperatures. This is to be expected, since the bounds were obtained by estimating the degree of a polynomial approximation to the Fermi–Dirac matrix function needed to satisfy a prescribed error tolerance. These bounds tend to be rather pessimistic because they do not take into account the possibility of numerical cancellation when evaluating the matrix polynomial. For instance, the bounds must apply in the worst-case scenario where the Hamiltonian has nonnegative entries and the approximating polynomial has nonnegative coefficients. Moreover, the bounds do not take into account the size of the entries in the Hamiltonian, particularly the fact that the nonzeros within the band (or sparsity pattern) are not of uniform size but may be spread out over several orders of magnitude. It should be emphasized that the presence of a gap is only a sufficient condition for localization of the density matrix, not a necessary one: it has been pointed out, for example, in [90], that disordered systems may exhibit strong localization even in the absence of a well-defined gap. This is the case, for instance, of the *Anderson model of localization* in condensed matter physics [2]. Obviously, our approach is unable to account for such phenomena in the zero-temperature case. The theory reviewed in this paper is primarily a qualitative one; nevertheless, it captures many of the features of actual physical systems, like the asymptotic dependence of the decay rate on the gap size or on the electronic temperature.

A natural question is whether the bounds can be improved to the point where they can be used to obtain practical estimates of the entries in the density matrix. In order to achieve this, additional assumptions on the Hamiltonians would be needed, making the theory less general. In other words, the price we pay for the generality of our theory is that we get pessimistic bounds. Recall that for a given sparsity pattern in the normalized Hamiltonians  $H_n$  our decay bounds depend on just one essential parameter, the gap  $\gamma$ . Our bounds are the same no matter what the eigenvalue distribution is to the left of the highest occupied level,  $\varepsilon_{n_e}$ , and to the right of the lowest unoccupied one,  $\varepsilon_{n_e+1}$ . If more spectral information were at hand, the bounds could be improved. The situation is very similar to that arising in the derivation of error bounds for the convergence of Krylov methods, such as the conjugate gradient (CG) method for solving symmetric positive definite (SPD) linear systems  $Ax = b$ ; see, e.g., [57, Theorem 10.2.6]. Bounds based on the spectral condition number  $\kappa_2(A)$  alone, while sharp, do not in general capture the actual convergence behavior of CG. They represent the worst-case behavior, which is rarely observed in practice. Much more accurate bounds can be obtained by making assumptions on the distribution of the eigenvalues of  $A$ . For instance, if  $A$  has only  $k$  distinct eigenvalues, then the CG method converges (in exact arithmetic) to the solution  $x_* = A^{-1}b$  in at most  $k$  steps. Similarly, suppose the Hamiltonian  $H_n$  has only  $k < n$  distinct eigenvalues (with  $\mu$  not one of them), and that the multiplicities of the eigenvalues to the left of  $\mu$  add up to  $n_e$ , the number of electrons. Then there is a polynomial  $p_k(\lambda)$  of degree at most  $k - 1$  such that  $p_k(H_n) = P_n$ , the density matrix. This is just the interpolation polynomial that takes the value 1 on the eigenvalues to the left of  $\mu$ , and zero on the eigenvalues to the right of  $\mu$ . This polynomial “approximation” is actually exact. If  $k \ll n$  and is independent of  $n$ , then  $P_n$  will be a matrix with  $O(n)$  nonzero entries; moreover, the sparsity pattern of  $P_n$  can be determined a priori from the graph structure of  $H_n$ . Another situation is that in which the eigenvalues of  $H_n$  fall in a small number  $k$  of narrow bands, or tight clusters, with the rightmost band to the left of  $\mu$  well separated from the leftmost band to the right of  $\mu$ . In this case

we can find again a low-degree polynomial  $p_k(\lambda)$  with  $p_k(H_n) \approx P_n$ , and improved bounds can be obtained.

The problem, of course, is that these are rather special eigenvalue distributions, and it is difficult to know a priori whether or not such conditions hold.

Another practical issue that should be at least briefly mentioned is the fact that our bounds assume knowledge of lower and upper bounds on the spectra of the Hamiltonians  $H_n$ , as well as estimates for the size and location of the spectral gap (this is also needed in order to determine the Fermi level  $\mu$ ). These issues have received a great deal of attention in the literature, and here we limit ourselves to observing that  $O(n)$  procedures exist to obtain sufficiently accurate estimates of these quantities; see, e.g., [53].

**9. Transformation to an Orthonormal Basis.** In this section we discuss the transformation of a Hamiltonian from a nonorthogonal to an orthogonal basis. The main point is that while this transformation results in matrices with less sparsity, the transformed matrices retain the decay properties of the original matrices, only with (possibly) different constants. What is important, from the point of view of asymptotic complexity, is that the rate of decay remains independent of system size.

We begin with a discussion of decay in the inverse of the overlap matrix. To this end, consider a sequence  $\{S_n\}$  of overlap matrices of size  $n = n_b \cdot n_e$ , with  $n_b$  constant and  $n_e$  increasing to infinity. We make the following assumptions:

1. Each  $S_n$  is a banded SPD matrix with unit diagonal entries and with bandwidth uniformly bounded with respect to  $n$ .
2. The spectral condition number (ratio of the largest to the smallest eigenvalue) of each  $S_n$ ,  $\kappa_2(S_n)$ , is uniformly bounded with respect to  $n$ . Because of assumption 1, this is equivalent to requiring that the smallest eigenvalue of  $S_n$  remain bounded away from zero for all  $n$ .

As always in this paper, the bandedness assumption in item 1 is not essential and can be replaced by the weaker hypothesis that each  $S_n$  is sparse and that the corresponding graphs  $\{G_n\}$  have bounded maximal degree with respect to  $n$ . Actually, it would be enough to require that the sequence  $\{S_n\}$  have the exponential decay property relative to a sequence of graphs  $\{G_n\}$  of bounded maximal degree. In order to simplify the discussion, and also in view of the fact that overlap matrices usually exhibit exponential or even superexponential decay, we assume from the outset that each  $S_n$  has already been truncated to a sparse (or banded) matrix. Again, this is for notational convenience only, and it is straightforward to modify the following arguments to account for the more general case. On the other hand, the assumption on condition numbers in item 2 is essential and cannot be weakened.

**REMARK 9.1.** *We note that assumption 2 above is analogous to the condition that the sequence of Hamiltonians  $\{H_n\}$  have spectral gap bounded below uniformly in  $n$ ; while this condition ensures (as we have shown) the exponential decay property in the associated spectral projectors  $P_n$ , assumption 2 above ensures exponential decay in the inverses (or inverse factors) of the overlap matrices. Both conditions amount to requiring that the corresponding problems be uniformly well-conditioned in  $n$ . The difference is that the decay on the spectral projectors depends on the spectral gap of the Hamiltonians and therefore on the nature of the system under study (i.e., insulator vs. metallic system), whereas the sparsity and spectral properties of the overlap matrices depend on other features of the system, mainly the interatomic distances.*

In the following we shall need some basic results on the decay of the inverses [34], inverse Cholesky factors [15], and inverse square roots (Löwdin factors) [12] of banded SPD matrices; see also [71].

Let  $A$  be SPD and  $m$ -banded, and let  $a$  and  $b$  denote the smallest and largest eigenvalues of  $A$ , respectively. Write  $\kappa$  for the spectral condition number  $\kappa_2(A)$  of  $A$  (hence,  $\kappa = b/a$ ). Define

$$q := \frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} \quad \text{and} \quad \lambda := q^{1/m}.$$

Furthermore, let  $K_0 := (1 + \sqrt{\kappa})^2/(2b)$ . In [34], Demko, Moss, and Smith obtained the following bound on the entries of  $A^{-1}$ :

$$(9.1) \quad |[A^{-1}]_{ij}| \leq K \lambda^{|i-j|}, \quad 1 \leq i, j \leq n,$$

where  $K := \max\{a^{-1}, K_0\}$ . Note that the bound (9.1) “blows up” as  $\kappa \rightarrow \infty$ , as one would expect.

As shown in [15], the decay bound (9.1) and the bandedness assumption on  $A$  imply a similar decay bound on the inverse Cholesky factor  $Z = R^{-1} = L^{-T}$ , where  $A = R^T R = L L^T$  with  $R$  upper triangular ( $L$  lower triangular). Assuming that  $A$  has been scaled so that  $\max_{1 \leq i \leq n} A_{ii} = 1$  (which is automatically true if  $A$  is an overlap matrix corresponding to a set of normalized basis functions), we have

$$(9.2) \quad |Z_{ij}| \leq K_1 \lambda^{j-i}, \quad j \geq i,$$

with  $K_1 = K \frac{1-\lambda^m}{1-\lambda}$ ; here  $K, \lambda$  are the same as before. We further note that while  $K_1 > K$ , for some classes of matrices it is possible to show that the actual magnitude of the  $(i, j)$  entry of  $Z$  (as opposed to the bound (9.2)) is actually less than the magnitude of the corresponding entry of  $A^{-1}$ . This is true, for instance, for an irreducible  $M$ -matrix; see [15].

Finally, let us consider the inverse square root,  $A^{-1/2}$ . In [12] the following bound is established:

$$(9.3) \quad |[A^{-1/2}]_{ij}| \leq K_2 \lambda^{|i-j|}, \quad 1 \leq i, j \leq n.$$

Here  $K_2$  depends again on the extreme eigenvalues  $a$  and  $b$  of  $A$ , whereas  $\lambda = q^{1/m}$ , where now  $q$  is any number satisfying the inequalities

$$\frac{\sqrt{\kappa} - 1}{\sqrt{\kappa} + 1} < q < 1.$$

As before, the bound (9.3) blows up as  $\kappa \rightarrow \infty$ , as one would expect.

Introducing the positive scalar  $\alpha = -\ln \lambda$ , we can rewrite all these bounds in the form

$$|B_{ij}| \leq K e^{-\alpha|i-j|}, \quad 1 \leq i, j \leq n,$$

for the appropriate matrix  $B$  and suitable constants  $K$  and  $\alpha > 0$ .

Now let  $\{S_n\}$  be a sequence of  $n \times n$  overlap matrices, where  $n = n_b \cdot n_e$  with  $n_b$  fixed and  $n_e \rightarrow \infty$ . Assuming that the matrices  $S_n$  satisfy assumptions 1–2 above, then their inverses satisfy the uniform exponential decay bounds (9.1), with  $K$  and  $\lambda$  constant and independent of  $n$ . Hence, as discussed in section 6, for any given  $\epsilon > 0$  there exists an integer  $\bar{m}$  independent of  $n$  such that each matrix  $S_n$  in the sequence can be approximated, in norm, by an  $\bar{m}$ -banded matrix with an error less than  $\epsilon$ . As usual, this result can be extended from the banded case to the sparse case,

assuming that the corresponding graphs  $G_n$  have bounded maximal degree as  $n \rightarrow \infty$ . Moreover, given assumptions 1–2 above, the inverse Cholesky factors  $Z_n$  satisfy a uniform (in  $n$ ) exponential decay bound of the type (9.2), and therefore uniform approximation with banded triangular matrices is possible. Again, generalization to more general sparsity patterns is possible, provided the usual assumption on the maximum degree of the corresponding graphs  $G_n$  holds. Similarly, under the same conditions we obtain a uniform rate of exponential decay for the entries of the inverse square roots  $S_n^{-1/2}$ , with a corresponding result on the existence of a banded (or sparse) approximation.

Let us now consider the sequence of transformed Hamiltonians,  $\tilde{H}_n = Z_n^T H_n Z_n$ . Here  $Z_n$  denotes either the inverse Cholesky factor or the inverse square root of the corresponding overlap matrix  $S_n$ . Assuming that the sequence  $\{H_n\}$  satisfies the off-diagonal exponential decay property and that  $\{S_n\}$  satisfies assumptions 1–2 above, it follows from the decay properties of the matrix sequence  $\{Z_n\}$  that the sequence  $\{\tilde{H}_n\}$  also enjoys off-diagonal exponential decay. This is a straightforward consequence of the following result, which is adapted from a similar one for infinite matrices due to Jaffard [71, Proposition 1].

**THEOREM 9.2.** *Consider two sequences  $\{A_n\}$  and  $\{B_n\}$  of  $n \times n$  matrices (where  $n \rightarrow \infty$ ) whose entries satisfy*

$$|[A_n]_{ij}| \leq c_1 e^{-\alpha|i-j|} \quad \text{and} \quad |[B_n]_{ij}| \leq c_2 e^{-\alpha|i-j|}, \quad 1 \leq i, j \leq n,$$

where  $c_1$ ,  $c_2$ , and  $\alpha > 0$  are independent of  $n$ . Then the sequence  $\{C_n\}$ , where  $C_n = A_n B_n$ , satisfies a similar bound:

$$(9.4) \quad |[C_n]_{ij}| \leq c e^{-\alpha'|i-j|}, \quad 1 \leq i, j \leq n,$$

for any  $0 < \alpha' < \alpha$ , with  $c$  independent of  $n$ .

*Proof.* First note that the entries of each  $A_n$  clearly satisfy

$$|[A_n]_{ij}| \leq c_1 e^{-\alpha'|i-j|} \quad \text{for any } \alpha' < \alpha.$$

Let  $\omega = \alpha - \alpha'$ . Then  $\omega > 0$  and the entries  $[C_n]_{ij}$  of  $C_n = A_n B_n$  satisfy

$$|[C_n]_{ij}| \leq \sum_{k=1}^n |[A_n]_{ik}| |[B_n]_{kj}| \leq c_1 c_2 \left( \sum_{k=1}^n e^{-\omega|k-j|} \right) e^{-\alpha'|i-j|}.$$

To complete the proof just observe that for any  $j$ ,

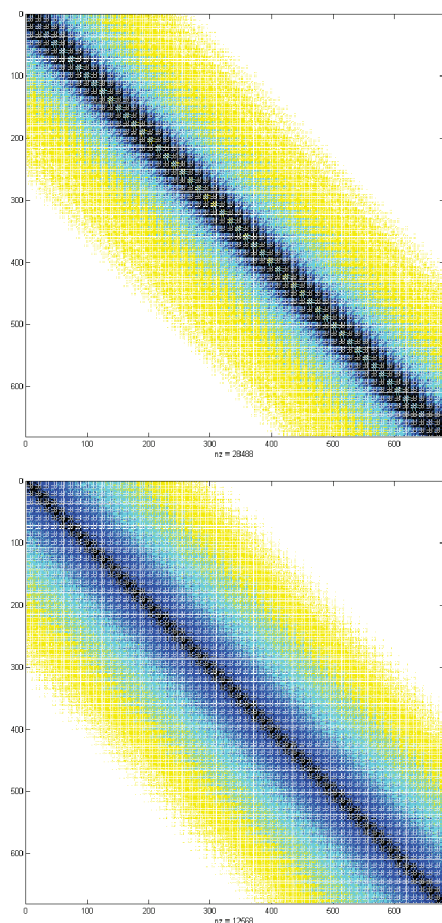
$$\sum_{k=1}^n e^{-\omega|k-j|} = \sum_{k=0}^{j-1} e^{-\omega k} + \sum_{k=1}^{n-j} e^{-\omega k} < \sum_{k=0}^{\infty} e^{-\omega k} + \sum_{k=1}^{\infty} e^{-\omega k} = \frac{1 + e^{-\omega}}{1 - e^{-\omega}}.$$

Since the last term is independent of  $n$ , the entries of  $C_n$  satisfy (9.4) with a constant  $c$  that is also independent of  $n$ .  $\square$

The foregoing result can obviously be extended to the product of three matrices. Thus, the entries of the matrix sequence  $\{\tilde{H}_n\}$ , where  $\tilde{H}_n = Z_n^T H_n Z_n$ , enjoy the exponential off-diagonal decay property

$$|[\tilde{H}_n]_{ij}| \leq c e^{-\alpha|i-j|}, \quad 1 \leq i, j \leq n,$$

for suitable constants  $c$  and  $\alpha > 0$ .



**Fig. 9.1** Magnitude of the entries in the Hamiltonian for the  $C_{52}H_{106}$  linear alkane. Top: Nonorthogonal (Gaussian-type orbital) basis. Bottom: Orthogonal basis. White:  $< 10^{-8}$ ; yellow:  $10^{-8} - 10^{-6}$ ; green:  $10^{-6} - 10^{-4}$ ; blue:  $10^{-4} - 10^{-2}$ ; black:  $> 10^{-2}$ . Note:  $nz$  refers to the number of “black” entries.

Alternatively, one could first approximate  $H_n$  and  $Z_n$  with banded matrices  $\bar{H}_n$  and  $\bar{Z}_n$  and then define the (approximate) transformed Hamiltonian as  $\tilde{H}_n := \bar{Z}_n^T \bar{H}_n \bar{Z}_n$ , possibly subject to further truncation. Using the fact that both  $H_n$  and  $Z_n$  have 2-norm bounded independently of  $n$ , it is easy to show that the final approximation error can be reduced below any prescribed tolerance by reducing the error in  $\bar{H}_n$  and  $\bar{Z}_n$ . Hence, with either approach, the transformed Hamiltonians  $\tilde{H}_n$  can be approximated uniformly in  $n$  within a prescribed error by banded matrices of constant bandwidth, just like the original (“nonorthogonal”) Hamiltonians. While the bandwidth of the approximations will be larger than for the original Hamiltonians, the truncated matrices retain a good deal of sparsity and asymptotically contain  $O(n)$  nonzeros. Hence, we have a justification of the statement (see section 1) that in our theory we can assume from the outset that the basis set  $\{\phi_i\}_{i=1}^n$  is orthonormal.

In Figure 9.1 we show the Hamiltonian  $H$  for the already mentioned linear alkane  $C_{52}H_{106}$  (see section 8.2) discretized in a Gaussian-type orbital basis (top) and the “orthogonalized” Hamiltonian  $\tilde{H} = \bar{Z}^T \bar{H} \bar{Z}$  (bottom). This figure shows that while

the transformation to the orthogonal basis alters the magnitude of the entries in the Hamiltonian, the bandwidth of  $\tilde{H}$  (truncated to a tolerance of  $10^{-8}$ ) is only slightly wider than that of  $H$ . In this case the overlap matrix  $S$  is well-conditioned, hence the entries of  $Z$  exhibit fast decay. An ill-conditioned overlap matrix would lead to a less sparse transformed Hamiltonian  $\tilde{H}$ .

As usual, the bandedness assumption was made for simplicity of exposition only; similar bounds can be obtained for more general sparsity patterns, assuming the matrices  $H_n$  and  $S_n$  have the exponential decay property relative to a sequence  $\{G_n\}$  of graphs having maximal degree uniformly bounded with respect to  $n$ .

It is important to emphasize that, in practice, the explicit formation of  $\tilde{H}_n$  from  $H_n$  and  $Z_n$  is not needed and is never carried out. Indeed, in all algorithms for electronic structure computation the basic matrix operations are matrix-matrix and matrix-vector products, which can be performed without explicit transformation of the Hamiltonian to an orthonormal basis. On the other hand, for the study of the decay properties it is convenient to assume that all the relevant matrices are explicitly given in an orthogonal representation.

One last issue to be addressed is whether the transformation to an orthonormal basis should be effected via the inverse Cholesky factor or via the Löwdin (inverse square root) factor of the overlap matrix. Comparing the decay bounds for the two factors suggests that the inverse Cholesky factor should be preferred (smaller  $\alpha$ ). Also note that the inverse Cholesky factor is triangular, and its sparsity can be increased by suitable reorderings of the overlap matrix. The choice of ordering may also be influenced by the computer architecture used. We refer the reader to [30] for the use of bandwidth-reducing orderings like reverse Cuthill–McKee, and to [25] for the use of space-filling curve orderings like the 3D Hilbert curve to improve load balancing and data locality on parallel architectures. In contrast, the Löwdin factor is a full symmetric matrix, regardless of the ordering. On the other hand, the multiplicative constant  $c$  is generally smaller for the Löwdin factor. Closer examination of a few examples suggests that in practice there is no great difference in the actual decay behavior of these two factors. However, approximating  $S_n^{-1/2}$  is generally more expensive and considerably more involved than approximating the inverse Cholesky factor. For the latter, the AINV algorithm [13] and its variants [24, 110, 136] are quite efficient and have been successfully used in various quantum chemistry codes. For other  $O(n)$  algorithms for transformation to an orthonormal basis, see [72, 100, 122]. In all these algorithms, sparsity is preserved by dropping small entries in the course of the computation. Explicit decay bounds for the  $Z_n$  factors could be used, in principle, to establish a priori which matrix elements *not* to compute, thus reducing the amount of overhead. Notice, however, that even if asymptotically bounded, the condition numbers  $\kappa_2(S_n)$  can be fairly large, leading to rather pessimistic decay estimates. This is again perfectly analogous to the situation with the condition-number-based error bounds for the CG method applied to a linear system  $Ax = b$ . Indeed, both the CG error bounds and the estimates (9.1) are obtained using Chebyshev polynomial approximation for the function  $f(\lambda) = \lambda^{-1}$ .

**10. The Vanishing Gap Case.** In this section we discuss the case of a sequence  $\{H_n\}$  of bounded, finite-range Hamiltonians for which the spectral gap around the Fermi level  $\mu$  vanishes as  $n \rightarrow \infty$ . Recall that this means that  $\inf_n \gamma_n = 0$ , where  $\gamma_n := \varepsilon_{n_e+1}^{(n)} - \varepsilon_{n_e}^{(n)}$  is the HOMO-LUMO gap for the  $n$ th Hamiltonian; it is assumed here that  $\varepsilon_{n_e}^{(n)} < \mu < \varepsilon_{n_e+1}^{(n)}$  for all  $n = n_b \cdot n_e$ . The reciprocal  $\gamma_n^{-1}$  of the gap can be interpreted as the condition number of the problem [109], so a vanishing spectral



gap means that the conditioning deteriorates as  $n_e \rightarrow \infty$  and the problem becomes increasingly difficult.

As already mentioned, in the zero-temperature limit our decay bounds blow up and therefore lose all meaning as  $\gamma_n \rightarrow 0$ . On the other hand, we know a priori that some type of decay should be present, in view of the results in section 7. A general treatment of the vanishing gap case appears to be rather difficult, for the main reason that in the limit as  $\beta \rightarrow \infty$  the Fermi–Dirac approximation to the Heaviside function becomes discontinuous, and therefore we can no longer make use of tools from classical approximation theory for analytic functions. Similarly, in the vanishing gap case the decay bounds (8.27) based on the resolvent estimates (8.24) break down since  $c \rightarrow \infty$  and  $\lambda \rightarrow 1$  in (8.26).

Rather than attacking the problem in general, in this section we give a complete analysis of what is perhaps the simplest nontrivial example of a sequence  $\{H_n\}$  with vanishing gap. While this is only a special case, this example captures some of the essential features of the “metallic” case, such as the rather slow off-diagonal decay of the entries of the density matrix. The simple model studied in this section may appear at first sight to be too simple and unrealistic to yield any useful information about actual physical systems. However, calculation of the density matrix at zero temperature on a system composed of 500 Al atoms reported in [140] reveals a decay behavior which is essentially identical to that obtained analytically for a free electron gas, a model very close to ours (which is essentially a discrete variant of the one in [140]). We believe that our analysis will shed some light on more general situations in which a slowly decaying density matrix occurs.

We begin by considering the infinite tridiagonal Toeplitz matrix

$$(10.1) \quad H = \begin{pmatrix} 0 & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & \ddots & \ddots & \ddots \end{pmatrix},$$

which defines a bounded, banded, self-adjoint operator on  $\ell^2$ . The graph of this matrix is just a (semi-infinite) path. The operator can be interpreted as an averaging operator or as a centered second-difference operator with a zero Dirichlet condition at one end, shifted and scaled so as to have spectrum contained in  $[-1, 1]$ . From a physical standpoint,  $H$  is the shifted and scaled discrete one-electron Hamiltonian, where the electron is constrained to the half-line  $[0, \infty)$ .

For  $n$  even ( $n = 2 \cdot n_e$ , with  $n_e \in \mathbb{N}$ ) consider the  $n$ -dimensional approximation

$$(10.2) \quad H_n = \begin{pmatrix} 0 & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & \frac{1}{2} & 0 & \frac{1}{2} \\ & & & \frac{1}{2} & 0 \end{pmatrix}.$$

This corresponds to truncating the semi-infinite path and imposing zero Dirichlet conditions at both ends. Now let  $\{e_1, e_2, \dots\}$  denote the standard basis of  $\ell^2$ , and let  $\hat{I}$  denote the identity operator restricted to the subspace of  $\ell^2$  spanned by  $e_{n+1}, e_{n+2}, \dots$

Letting

$$H_{(n)} := \begin{pmatrix} H_n & 0 \\ 0 & \hat{I} \end{pmatrix},$$

the sequence  $\{H_{(n)}\}$  is now a sequence of bounded self-adjoint linear operators on  $\ell^2$  that converges strongly to  $H$ . Note that  $\sigma(H_n) \subset [-1, 1]$  for all  $n$ ; also,  $0 \notin \sigma(H_n)$  for all even  $n$ . It is easy to see that half of the eigenvalues of  $H_n$  lie in  $[-1, 0)$  and the other half in  $(0, 1]$ . We set  $\mu = 0$  and we label as ‘‘occupied’’ the states corresponding to negative eigenvalues. The spectral gap of each  $H_n$  is then  $\varepsilon_{n/2+1}^{(n)} - \varepsilon_{n/2}^{(n)}$ .

The eigenvalues and eigenvectors of  $H_n$  are known explicitly [35, Lemma 6.1]. Indeed, the eigenvalues, in descending order, are given by  $\varepsilon_k^{(n)} = \cos\left(\frac{k\pi}{n+1}\right)$  (with  $1 \leq k \leq n$ ) and the corresponding normalized eigenvectors are given by  $v_k^{(n)} = (v_k^{(n)}(j))$  with entries

$$v_k^{(n)}(j) = \sqrt{\frac{2}{n+1}} \sin\left(\frac{jk\pi}{n+1}\right), \quad 1 \leq j \leq n.$$

Note that the eigenvalues are symmetric with respect to the origin, and that the spectral gap at 0 vanishes, since  $\varepsilon_{n/2+1}^{(n)} = -\varepsilon_{n/2}^{(n)} \rightarrow 0$  as  $n \rightarrow \infty$ . We also point to the well-known fact that the eigenvectors of this operator are strongly delocalized. Nevertheless, as we will see, some localization (decay) is present in the density matrix, owing to cancellation (i.e., destructive interference).

Now let  $P_n$  be the zero-temperature density matrix associated with  $H_n$ , i.e., the spectral projector onto the subspace of  $\mathbb{C}^n$  spanned by the eigenvectors of  $H_n$  associated with the lowest  $n_e$  eigenvalues (the occupied subspace). We extend  $P_n$  to a projector acting on  $\ell^2$  by embedding  $P_n$  into an infinite matrix  $P_{(n)}$  as follows:

$$P_{(n)} := \begin{pmatrix} P_n & 0 \\ 0 & 0 \end{pmatrix}.$$

Note that  $P_{(n)}$  is just the orthogonal projector onto the subspace of  $\ell^2$  spanned by the eigenvectors of  $H_{(n)}$  associated with eigenvalues in the interval  $[-1, 0)$ . Moreover,  $\text{Tr}(P_{(n)}) = \text{Tr}(P_n) = \text{rank}(P_n) = \frac{n}{2} = n_e$ . The limiting behavior of the sequence  $\{P_{(n)}\}$  (hence, of  $\{P_n\}$ ) is completely described by the following result.

**THEOREM 10.1.** *Let  $H$ ,  $H_n$ , and  $P_{(n)}$  be as described above. Then the following hold:*

- (i)  $H$  has purely absolutely continuous spectrum,<sup>9</sup> given by the interval  $[-1, 1]$ . In particular,  $H$  has no eigenvalues.
- (ii) The union of the spectra of the  $n$ -dimensional sections  $H_n$  of  $H$  is everywhere dense in  $\sigma(H) = [-1, 1]$ . In other words, every point in  $[-1, 1]$  is the limit of a sequence of the form  $\{\varepsilon_k^{(n)}\}$  for  $n \rightarrow \infty$ , where  $\varepsilon_k^{(n)} \in \sigma(H_n)$  and  $k = k(n)$ .
- (iii) The sequence  $\{H_n\}$  has vanishing gap:  $\inf_n \gamma_n = 0$ .
- (iv) The spectral projectors  $P_{(n)}$  converge strongly to  $P = h(H)$ , where  $h(x) = \chi_{[-1, 0)}(x)$ , the characteristic function of the interval  $[-1, 0)$ .
- (v)  $P$  is the orthogonal projector onto an infinite-dimensional subspace of  $\ell^2$ .

<sup>9</sup>The absolutely continuous spectrum of a self-adjoint linear operator  $H$  on a Hilbert space  $\mathcal{H}$  is the spectrum of the restriction of  $H$  to the subspace  $\mathcal{H}_{\text{ac}} \subseteq \mathcal{H}$  of vectors  $\psi$  whose spectral measures  $\mu_\psi$  are absolutely continuous with respect to the Lebesgue measure. For details, see [107, pp. 224–231].

*Proof.* Statements (i)–(ii) are straightforward consequences of classical results on the asymptotic eigenvalue distribution of Toeplitz matrices, while (iv)–(v) follow from general results in spectral theory. Statement (iii) was already noted (the eigenvalues of  $H_n$  are explicitly known) and it also follows from (i)–(ii). In more detail, statement (i) is a special case of Rosenblum’s theorem on the spectra of banded infinite Toeplitz matrices; see [108] or [17, Theorem 1.31]. For the facts that the spectrum of  $H$  coincides with the interval  $[-1, 1]$  and that the finite section eigenvalues  $\varepsilon_k^{(n)}$  are dense in  $\sigma(H) = [-1, 1]$  (statement (ii)), see the paper by Hartman and Wintner [60] or the book by Grenander and Szegö [58, Chapter 5]. Statement (iv) can be proved as follows. For a linear operator  $A$  on  $\ell^2$ , write  $R_\lambda(A) = (A - \lambda I)^{-1}$ , with  $\lambda \notin \sigma(A)$ . A sequence  $\{A_n\}$  of self-adjoint (Hermitian) operators is said to converge in the *strong resolvent sense* to  $A$  if  $R_\lambda(A_n) \rightarrow R_\lambda(A)$  strongly for all  $\lambda \in \mathbb{C}$  with  $\text{Re } \lambda \neq 0$ , that is,

$$\lim_{n \rightarrow \infty} \|R_\lambda(A_n)x - R_\lambda(A)x\| = 0 \quad \text{for all } x \in \ell^2.$$

It is easy to check, using, for instance, the results in [18, Chapter 2], that the sequence  $\{H_n\}$  converges in the strong resolvent sense to  $H$ . Statement (iv) (as well as (ii)) now follows from [107, Theorem VIII.24]. The fact (v) that  $P = h(H)$  is an orthogonal projector onto an infinite-dimensional subspace of  $\ell^2$  follows from the fact that  $\mu = 0$  is not an eigenvalue of  $H$  (because of (i)) and from the spectral theorem for self-adjoint operators in Hilbert space; see, e.g., [107, Chapter VII] or [115, Chapter 12].  $\square$

The foregoing result implies that the Toeplitz matrix sequence  $\{H_n\}$  given by (10.2) exhibits some of the key features of the discrete Hamiltonians describing metallic systems, in particular, the vanishing gap property and the fact that the eigenvalues tend to fill the entire energy spectrum. The sequence  $\{H_n\}$  can be thought of as a 1D “toy model” that can be solved analytically to gain some insight into the decay properties of the density matrix of such systems. Indeed, from the knowledge of the eigenvectors of  $H_n$  we can write down the spectral projector corresponding to the lowest  $n_e = n/2$  eigenvalues explicitly. Recalling that the eigenvalues  $\varepsilon_k^{(n)}$  are given in descending order, it is convenient to compute  $P_n$  as the projector onto the orthogonal complement of the subspace spanned by the eigenvectors corresponding to the  $n/2$  largest eigenvalues:

$$P_n = I_n - \sum_{k=1}^{n_e} v_k^{(n)}(v_k^{(n)})^T.$$

The  $(i, j)$  entry of  $P_n$  is therefore given by

$$[P_n]_{ij} = e_i^T P_n e_j = \delta_{ij} - \frac{2}{n+1} \sum_{k=1}^{n_e} \sin\left(\frac{ik\pi}{n+1}\right) \sin\left(\frac{jk\pi}{n+1}\right).$$

For  $i = j$ , we find

$$(10.3) \quad [P_n]_{ii} = 1 - \frac{2}{n+1} \sum_{k=1}^{n_e} \sin^2\left(\frac{ik\pi}{n+1}\right) = \frac{1}{2} \quad \text{for all } i = 1, \dots, n \text{ and for all } n.$$

Hence, for this system the charge density  $P_{ii}$  is constant and the system essentially behaves like a noninteracting electron gas; see, for example, [50]. We note in passing that this example confirms that the bound (7.2) is sharp, since equality is attained for this particular projector. Moreover, the trigonometric identity

$$(10.4) \quad \sin \theta \sin \phi = -\frac{1}{2} [\cos(\theta + \phi) - \cos(\theta - \phi)]$$

implies for all  $i, j = 1, \dots, n$  that

$$(10.5) \quad [P_n]_{ij} = \frac{1}{n+1} \sum_{k=1}^{n_e} \left[ \cos\left(\frac{(i+j)k\pi}{n+1}\right) - \cos\left(\frac{(i-j)k\pi}{n+1}\right) \right].$$

From (10.5) it immediately follows, for all  $i$  and for all  $n$ , that

$$(10.6) \quad [P_n]_{i,i+2l} = 0, \quad l = 1, 2, \dots$$

Since (10.3) and (10.6) hold for all  $n$ , they also hold in the limit as  $n \rightarrow \infty$ . Hence, the strong limit  $P$  of the sequence of projectors  $\{P_{(n)}\}$  satisfies  $P_{ii} = 1/2$  and  $P_{i,j} = 0$  for all  $j = i + 2l$ , where  $i, l = 1, 2, \dots$ . To determine the remaining off-diagonal entries  $P_{ij}$  (with  $j \neq i$  and  $j \neq i + 2l$ ) we directly compute the limit of  $[P_n]_{ij}$  as  $n \rightarrow \infty$ , as follows. Observe that using the substitution  $x = k/(n+1)$  and taking the limit as  $n \rightarrow \infty$  in (10.5), we obtain for all  $i \geq 1$  and for all  $j \neq i + 2l$  ( $l = 0, 1, \dots$ )

$$(10.7) \quad \begin{aligned} P_{ij} &= \int_0^{\frac{1}{2}} \cos[(i+j)\pi x] dx - \int_0^{\frac{1}{2}} \cos[(i-j)\pi x] dx \\ &= \frac{1}{\pi} \left[ \frac{(-1)^{\frac{i+j-1}{2}}}{i+j} + \frac{(-1)^{\frac{i-j+1}{2}}}{i-j} \right]. \end{aligned}$$

It follows from (10.7) that  $|P_{ij}|$  is bounded by a quantity that decays only linearly in the distance from the main diagonal. As a result,  $O(n)$  approximation of  $P_n$  for large  $n$  involves a huge prefactor. Therefore, from this very simple example we can gain some insight into the vanishing gap case. The analytical results obtained show that the density matrix can exhibit rather slow decay, confirming the well-known fact that  $O(n)$  approximations pose a formidable challenge in the vanishing gap case.

The 2D case is easily handled as follows. We consider for simplicity the case of a square lattice consisting of  $n^2$  points in the plane. The 2D Hamiltonian is given by

$$H_{n^2} = \frac{1}{2}(H_n \otimes I_n + I_n \otimes H_n),$$

where the scaling factor  $\frac{1}{2}$  is needed so as to have  $\sigma(H_{n^2}) \subset [-1, 1]$ . The eigenvalues and eigenvectors of  $H_{n^2}$  can be explicitly written in terms of those of  $H_n$ ; see, e.g., [35]. Assuming again that  $n$  is even, exactly half of the  $n^2$  eigenvalues of  $H_{n^2}$  are negative (counting multiplicities), the other half positive. As before, we are interested in finding the spectral projector associated with the eigenvectors corresponding to negative eigenvalues. Note again that the spectral gap tends to zero as  $n \rightarrow \infty$ . If  $P_{n^2}$  denotes the spectral projector onto the occupied states, it is not difficult to show that

$$(10.8) \quad P_{n^2} = P_n \otimes (I_n - P_n) + (I_n - P_n) \otimes P_n.$$

It follows from (10.8) that the spectral projector  $P_{n^2}$  has a natural  $n \times n$  block structure, where the following hold:

- Each diagonal block is equal to  $\frac{1}{2}I_n$ ; note that this gives the correct trace,  $\text{Tr}(P_{n^2}) = \frac{n^2}{2}$ .
- The  $(k, l)$  off-diagonal block  $\Pi_{kl}$  is given by  $\Pi_{kl} = [P_n]_{kl}(I_n - 2P_n)$ . Hence, each off-diagonal block has a “striped” structure, with the main diagonal as well as the third, fifth, etc., off-diagonals identically zero. Moreover, every block  $\Pi_{kl}$  with  $l = k + 2m$  ( $m \geq 1$ ) is zero.

This shows that in the 2D case, the rate of decay in the spectral projector is essentially the same as in the 1D case. The 3D case can be handled in a similar manner, leading to the same conclusion.

For this simple example we can also compute the entries of the density matrix at positive electronic temperature  $T > 0$ . Recalling that the density matrix in this case is given by the Fermi–Dirac function with parameter  $\beta = 1/(k_B T)$ , we have in the 1D case (assuming  $\mu = 0$ )

$$(10.9) \quad P_{ij} = \frac{2}{n+1} \sum_{k=1}^n \frac{\sin\left(\frac{ik\pi}{n+1}\right) \sin\left(\frac{jk\pi}{n+1}\right)}{1 + \exp\left[\beta \cos\left(\frac{k\pi}{n+1}\right)\right]}.$$

Making use again of the trigonometric identity (10.4) and using the same substitution  $x = k/(n+1)$ , we can reduce the computation of the density matrix element  $P_{ij}$  for  $n \rightarrow \infty$  to the evaluation of the integral

$$(10.10) \quad P_{ij} = \int_0^1 \frac{\cos[(i-j)\pi x] - \cos[(i+j)\pi x]}{1 + \exp(\beta \cos \pi x)} dx.$$

Unfortunately, this integral cannot be evaluated explicitly in terms of elementary functions. Note, however, that the integral

$$I_k = \int_0^1 \frac{\cos(k\pi x)}{1 + \exp(\beta \cos \pi x)} dx$$

(where  $k$  is an integer) becomes, under the change of variable  $\pi x = \arccos t$ ,

$$I_k = \frac{1}{\pi} \int_{-1}^1 \frac{\cos(k \arccos t)}{1 + e^{\beta t}} \frac{dt}{\sqrt{1-t^2}}.$$

Hence, up to a constant factor,  $I_k$  is just the  $k$ th coefficient in the Chebyshev expansion of the Fermi–Dirac function  $1/(1+e^{\beta t})$ . Since the Fermi–Dirac function is analytic on the interior of an ellipse containing the interval  $[-1, 1]$  and continuous on the boundary of such an ellipse, it follows from the general theory of Chebyshev approximation that the coefficients  $I_k$  decay at least exponentially fast as  $k \rightarrow \infty$ ; see, e.g., [92]. This in turn implies that the entries  $P_{ij}$  given by (10.10) decay at least exponentially fast away from the main diagonal, the faster the larger the temperature is, as already discussed in section 8.7. Hence, for this special case we have established in a more direct way the exponential decay behavior already proved in general in section 8.1. In the present case, however, for any value of  $\beta$  the decay rate of the entries  $P_{ij}$  given by (10.10) can be determined to arbitrary accuracy by numerically computing the Chebyshev coefficients of the Fermi–Dirac function.

We mention that a simple, 1D model of a system with arbitrarily small gap was described in [49]. The (continuous) Hamiltonian in [49] consists of the kinetic term plus a potential given by a sum of Gaussian wells located at the nuclei sites  $X_i$ :

$$\mathcal{H} = -\frac{1}{2} \frac{d^2}{dx^2} + V(x), \quad V(x) = - \sum_{i=-\infty}^{\infty} \frac{a}{\sqrt{2\pi\sigma^2}} \exp(-(x - X_i)^2/2\sigma^2),$$

with  $a > 0$  and  $\sigma > 0$  tunable parameters. The spectra of this family of Hamiltonians present a band structure with band gap proportional to  $\sqrt{a}/\sigma$ . Note that the model

essentially reduces to ours for  $a \rightarrow 0$  and/or for  $\sigma \rightarrow \infty$ . On the other hand, while the gap can be made arbitrarily small by tuning the parameters in the model, for any choice of  $a > 0$  and  $\sigma > 0$  the gap does not vanish; therefore, no approximation of the infinite-size system with a sequence of finite-size ones can lead to a vanishing gap in the thermodynamic limit. This means that our bounds, when applied to this model, will yield exponential decay, albeit very slow (since the correlation lengths will be quite large for small  $a \rightarrow 0$  and/or for large  $\sigma$ ). The model in [49], on the other hand, can be useful for testing purposes when developing algorithms for metal-like systems with slowly decaying density matrices.

**11. Other Applications.** In this section we sketch a few possible applications of our decay results to areas other than electronic structure computations.

**11.1. Density Matrices for Thermal States.** In quantum statistical mechanics, the equilibrium density matrix for a system of particles subject to a heat bath at absolute temperature  $T$  is defined as

$$(11.1) \quad P = \frac{e^{-\beta H}}{Z}, \quad \text{where } Z = \text{Tr}(e^{-\beta H}).$$

As usual,  $\beta = (k_B T)^{-1}$ , where  $k_B$  denotes the Boltzmann constant; see [102]. The matrix  $P$  is the quantum analogue of the canonical Gibbs state. The Hamiltonian  $H$  is usually assumed to have been shifted so that the smallest eigenvalue is zero [87, p. 112]. Note that  $P$  as defined in (11.1) is not an orthogonal projector. It is, however, Hermitian and positive semidefinite. Normalization by the *partition function*  $Z$  ensures that  $\sigma(P) \subset [0, 1]$  and that  $\text{Tr}(P) = 1$ .

It is clear that for increasing temperature, i.e., for  $T \rightarrow \infty$  (equivalently, for  $\beta \rightarrow 0$ ), the canonical density matrix  $P$  approaches the identity matrix, normalized by the matrix size  $n$ . In particular, the off-diagonal entries tend to zero. The physical interpretation of this is that in the limit of large temperatures the system states become totally uncorrelated. For temperatures approaching absolute zero, on the other hand, the canonical matrix  $P$  tends to the orthogonal projector associated with the zero eigenvalue (ground state). In this limit, the correlation between state  $i$  and state  $j$  is given by the  $(i, j)$  entry of the orthogonal projector onto the eigenspace corresponding to the zero eigenvalue, normalized by  $n$ .

For finite, positive values of  $T$ , the canonical density matrix  $P$  is full but decays away from the main diagonal (or, more generally, away from the sparsity pattern of  $H$ ). The rate of decay depends on  $\beta$ : the smaller it is, the faster the decay. Application of the bounds developed in section 8 to the matrix exponential is straightforward. For instance, the bounds based on Bernstein's theorem take the form

$$(11.2) \quad |[e^{-\beta H}]_{ij}| \leq C(\beta) e^{-\alpha d(i,j)}, \quad i \neq j,$$

where

$$C(\beta) = \frac{2\chi}{\chi - 1} e^{\beta \|H\|_2 (\kappa_1 - 1)/2} \quad \text{and} \quad \alpha = 2 \ln \chi.$$

In these expressions,  $\chi > 1$  and  $\kappa_1 > 1$  are the parameters associated with the Bernstein ellipse with foci in  $-1$  and  $1$  and major semiaxis  $\kappa_1$ , as described in section 8. Choosing  $\chi$  large makes the exponential term decay  $e^{-\alpha d(i,j)}$  very fast, but causes  $C(\beta)$  to grow larger. Clearly, a smaller  $\beta$  makes the upper bound (11.2) smaller. Bounds on the entries of the canonical density matrix  $P$  can be obtained by dividing

through the upper bounds by  $Z$ . Techniques for estimating  $Z$  can be developed using the techniques described in [56]; see also [11].

Although the bound (11.2) is an exponentially decaying one, it can be shown that the decay in the entries of a banded or sparse matrix is actually *superexponential*. This can be shown by expanding the exponential in a series of Chebyshev polynomials and using the fact that the coefficients in the expansion, which can be expressed in terms of Bessel functions, decay to zero superexponentially; see [92] and also [69]. The decay bounds obtained in this way are, however, less transparent and more complicated to evaluate than (11.2).

Finally, exponential decay bounds for spectral projectors and other matrix functions might provide a rigorous justification for  $O(n)$  algorithms recently developed for disordered systems; see, e.g., [117, 118].

**11.2. Quantum Information Theory.** A related area of research where our decay bounds for matrix functions have proven useful is the study of quantum many-body systems in information theory; see, e.g., [31, 32, 38, 120, 121]. In particular, relationships between spectral gaps and rates of decay for functions of finite-range Hamiltonians have been established in [31] using the techniques introduced in [12]. The exponential decay of correlations and its relation to the spectral gap have also been studied in [62, 63].

As shown in [32], exponential decay bounds for matrix functions play a crucial role in establishing so-called *area laws* for the *entanglement entropy* of ground states associated with bosonic systems. These area laws essentially state that the entanglement entropy associated with a 3D bosonic lattice is proportional to the surface area, rather than to the volume, of the lattice. Intriguingly, such area laws are analogous to those governing the Beckenstein–Hawking black hole entropy. We refer the interested reader to the recent, comprehensive survey paper [38] for additional information.

**11.3. Complex Networks.** The study of complex networks is an emerging field of science currently undergoing vigorous development. Researchers in this highly interdisciplinary field include mathematicians, computer scientists, physicists, chemists, engineers, neuroscientists, biologists, social scientists, etc. Among the mathematical tools used in this field, linear algebra and graph theory, in particular spectral graph theory, play a major role. Also, statistical mechanics concepts and techniques have been found to be ideally suited to the study of large-scale networks.

In recent years, quantitative methods of network analysis have increasingly made use of matrix functions. This approach has been spearheaded in the works of Estrada, Rodríguez-Velázquez, D. Higham, and Hatano; see, e.g., [39, 40, 41, 42, 43, 46], as well as the recent surveys [45, 44] and the references therein. Functions naturally arising in the context of network analysis include the exponential, the resolvent, and hyperbolic functions, among others. Physics-based justifications for the use of these matrix functions in the analysis of complex networks are thoroughly discussed in [44].

For example, the exponential of the adjacency matrix  $A$  associated with a simple, undirected graph  $G = (V, E)$  can be used to give natural definitions of important measures associated with nodes in  $G$ , such as the *subgraph centrality* associated with node  $i$ , defined as  $C(i) = [e^A]_{ii}$ , and the *communicability* associated with two distinct nodes  $i$  and  $j$ , defined as  $C(i, j) = [e^A]_{ij}$ . Other network quantities that can be expressed in terms of the entries in appropriate matrix functions of  $A$  include betweenness, returnability, vulnerability, and so forth. The graph Laplacian  $L = D - A$ , where  $D = \text{diag}(d_1, \dots, d_n)$  with  $d_i$  denoting the degree of node  $i$ , is sometimes used instead of the adjacency matrix, as well as weighted analogues of both  $A$  and  $L$ .

Most networks arising in real-world applications are sparse, often with degree distributions closely approximated by power laws. Because the maximum degree in such “scale-free” networks increases as the number of nodes tends to infinity, one cannot expect uniform exponential decay rates to hold asymptotically for the matrix functions associated with such graphs unless additional structure is imposed, for instance, in the form of weights. Nevertheless, our bounds for the entries of functions of sparse matrices can be used to obtain estimates on quantities such as the communicability between two nodes. A discussion of locality (or the lack thereof) in matrix functions used in the analysis of complex networks can be found in [44]. We also refer the reader to [11] for a description of quadrature-rule-based bounds for the entries of matrix functions associated with complex networks.

**11.4. Tridiagonal Eigensolvers.** The solution of symmetric tridiagonal eigenvalue problems plays an important role in many fields of computational science. As noted, for example, in [130], solving such problems is key for most dense real symmetric (and complex Hermitian) eigenvalue computations and therefore plays a central role in standard linear algebra libraries such as LAPACK and ScaLAPACK. Even in the sparse case, the symmetric tridiagonal eigenvalue problem appears as a step in the Lanczos algorithm.

The efficiency of symmetric tridiagonal eigensolvers can be significantly increased by exploiting localization in the eigenvectors (more generally, invariant subspaces) associated with an isolated cluster of eigenvalues. It would be highly desirable to identify beforehand any localization in the eigenspace in a cost-effective manner, as this would lead to reduced computational costs [101, 130]. It is clear that this problem is essentially the same as the one considered in this paper, with the additional assumption that the matrix  $H$  is tridiagonal. Given estimates on the location of the cluster of eigenvalues and on the size of the gaps separating it from the remainder of the spectrum, the techniques described in this paper can be used to bound the entries in the spectral projector associated to the cluster of interest; in turn, the bounds can be used to identify banded approximations to the spectral projectors with guaranteed prescribed error. Whether the estimates obtained in this manner are accurate enough to lead to practical algorithms with run times and storage demands substantially improved over current ones remains an open question for further research.

Finally, in the recent paper [139] the exponential decay results in [12] are used to derive error bounds and stopping criteria for the Lanczos method applied to the computation of  $e^{-tA}v$ , where  $A$  is a large SPD matrix,  $v$  is a vector, and  $t > 0$ . The bounds are applied to the exponential of the tridiagonal matrix  $T_k$  generated after  $k$  steps by the Lanczos process in order to obtain the approximation error after  $k$  steps.

**11.5. Non-Hermitian Extensions.** Although the main focus of the paper has been the study of functions of sparse Hermitian matrices, many of our results can be extended, under appropriate conditions, to non-Hermitian matrices. The generalizations of our decay bounds to *normal* matrices, including, for example, skew-Hermitian matrices, is relatively straightforward; see, e.g., the results in [14] and [106]. Further generalizations to diagonalizable matrices were given in [14], although the bounds now contain additional terms taking into account the departure from normality. These bounds may be difficult to use in practice, as knowledge of the eigenvectors or of the field of values of the matrix is needed. Bounds for functions of general sparse matrices can also be obtained using contour integration; see, e.g., [106] and [91]. It is quite possible that these bounds will prove useful in applications involving functions of sparse, nonnormal matrices. Examples include functions of digraphs in network analysis,



like returnability, or functions of the Hamiltonians occurring in the emerging field of non-Hermitian quantum mechanics; see, respectively, [42] and [8, 9, 94].

**12. Conclusions and Open Problems.** In this paper we have described a general theory of localization for the density matrices associated with certain sequences of banded or sparse discrete Hamiltonians of increasing size. We have obtained, under very general conditions, exponential decay bounds for the off-diagonal entries of zero-temperature density matrices for gapped systems (“insulators”) and for density matrices associated with systems at positive electronic temperature. The theory, while purely mathematical, recovers well-known physical phenomena such as the fact that the rate of decay is faster at higher temperatures and for larger gaps, and even captures the correct asymptotics for small gaps and low temperatures. Thus, we have provided a theoretical justification for the development of  $O(n)$  methods for electronic structure computations. As an integral part of this theory, we have also surveyed the approximation of rapidly decaying matrices by banded or sparse ones, the effects of transforming a Hamiltonian from a nonorthogonal to an orthogonal basis, and some general properties of orthogonal projectors.

In the case of zero-temperature and vanishing gaps, our bounds deteriorate for increasing  $n$ . In the limit as  $n \rightarrow \infty$  we no longer have exponentially decaying bounds, which is entirely consistent with the physics. For metallic systems at zero temperature the decay in the spectral projector follows a power law, and we have exhibited a simple model Hamiltonian for which the decay in the corresponding density matrix is only linear in the distance from the main diagonal.

Because of the slow decay, the development of  $O(n)$  methods in the metallic case at zero temperature is problematic. We refer the reader to [5, 19, 81, 134] for some attempts in this direction, but the problem remains essentially open. In the metallic case it may be preferable to keep  $P$  in the factorized form  $P = XX^*$ , where  $X \in \mathbb{C}^{n \times n_e}$  is any matrix whose columns span the occupied subspace, and to seek a maximally localized  $X$ . Note that

$$P = XX^* = (XU)(XU)^*$$

for any unitary  $n_e \times n_e$  matrix  $U$ , so the question is whether the occupied subspace admits a set of basis vectors that can be rotated so as to become as localized as possible. Another possibility is to research the use of rank-structured approximations (such as hierarchical matrix techniques [59]) to the spectral projector. Combinations of tensor product approximations and wavelets appear to be promising. We refer here to [55] for a study of the decay properties of density matrices in a wavelet basis (see also [119]), and to [16] for an early attempt to exploit near low-rank properties of spectral projectors. See also the more recent works by W. Hackbusch and collaborators [26, 27, 28, 47, 48, 86].

Besides the motivating application of electronic structure, our theory is also applicable to other problems where localization plays a prominent role. We hope that this paper will stimulate further research in this fascinating and important area at the crossroads of mathematics, physics, and computing.

**Acknowledgments.** We are indebted to three anonymous referees for carefully reading the original manuscript and for suggesting a number of corrections and improvements. Thanks also to the handling editor, Fadil Santosa, for useful feedback and for his patience and understanding during the several months it took us to revise the paper. We would also like to acknowledge useful discussions with David Borthwick, Matt Challacombe, Jean-Luc Fattebert, Roberto Grena, Daniel Kressner, and

Maxim Olshanskii. Finally, we are grateful to Jacek Jakowski for providing the data for the linear alkane.

## REFERENCES

- [1] N. I. ACHESER, *Theory of Approximation*, Frederick Ungar, New York, 1956.
- [2] P. W. ANDERSON, *Absence of diffusion in certain random lattices*, Phys. Rev., 109 (1958), pp. 1492–1505.
- [3] R. BAER AND M. HEAD-GORDON, *Sparsity of the density matrix in Kohn–Sham density functional theory and an assessment of linear system-size scaling methods*, Phys. Rev. Lett., 79 (1997), pp. 3962–3965.
- [4] R. BAER AND M. HEAD-GORDON, *Chebyshev expansion methods for electronic structure calculations on large molecular systems*, J. Chem. Phys., 107 (1997), pp. 10003–10013.
- [5] R. BAER AND M. HEAD-GORDON, *Energy renormalization-group method for electronic structure of large systems*, Phys. Rev. B, 58 (1998), pp. 15296–15299.
- [6] K. R. BATES, A. D. DANIELS, AND G. E. SCUSERIA, *Comparison of conjugate gradient density matrix search and Chebyshev expansion methods for avoiding diagonalization in large-scale electronic structure calculations*, J. Chem. Phys., 109 (1998), pp. 3308–3312.
- [7] C. BEKAS, E. KOKIOPOULOU, AND Y. SAAD, *Computation of large invariant subspaces using polynomial filtered Lanczos iterations with applications in density functional theory*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 397–418.
- [8] C. M. BENDER, S. BOETTCHER, AND P. N. MEISINGER, *PT-symmetric quantum mechanics*, J. Math. Phys., 40 (1999), pp. 2201–2229.
- [9] C. M. BENDER, D. C. BRODY, AND H. F. JONES, *Must a Hamiltonian be Hermitian?*, Amer. J. Phys., 71 (2003), pp. 1095–1102.
- [10] M. BENZI, *Preconditioning techniques for large linear systems: A survey*, J. Comput. Phys., 182 (2002), pp. 418–477.
- [11] M. BENZI AND P. BOITO, *Quadrature rule-based bounds for functions of adjacency matrices*, Linear Algebra Appl., 433 (2010), pp. 637–652.
- [12] M. BENZI AND G. H. GOLUB, *Bounds for the entries of matrix functions with applications to preconditioning*, BIT, 39 (1999), pp. 417–438.
- [13] M. BENZI, C. D. MEYER, AND M. TÜMA, *A sparse approximate inverse preconditioner for the conjugate gradient method*, SIAM J. Sci. Comput., 17 (1996), pp. 1135–1149.
- [14] M. BENZI AND N. RAZOUK, *Decay bounds and  $O(n)$  algorithms for approximating functions of sparse matrices*, Electr. Trans. Numer. Anal., 28 (2007), pp. 16–39.
- [15] M. BENZI AND M. TÜMA, *Orderings for factorized sparse approximate inverse preconditioners*, SIAM J. Sci. Comput., 21 (2000), pp. 1851–1868.
- [16] G. BEYLKIN, N. COULT, AND M. J. MOHLENKAMP, *Fast spectral projection algorithms for density-matrix computations*, J. Comput. Phys., 152 (1999), pp. 32–54.
- [17] A. BÖTTCHER AND S. M. GRUDSKY, *Spectral Properties of Banded Toeplitz Matrices*, SIAM, Philadelphia, PA, 2005.
- [18] A. BÖTTCHER AND B. SILBERMANN, *Introduction to Large Truncated Toeplitz Matrices*, Springer, New York, 1998.
- [19] D. R. BOWLER, J.-L. FATTEBERT, M. J. GILLAN, P.-D. HAYNES, AND C.-K. SKYLARIS, *Introductory remarks: Linear scaling methods*, J. Phys. Condensed Matter, 20 (2008), 290301.
- [20] D. R. BOWLER AND T. MIYAZAKI,  *$O(N)$  methods in electronic structure calculations*, Rep. Progr. Phys., 75 (2012), 036503.
- [21] C. BROUDER, G. PANATI, M. CALANDRA, C. MOURougANE, AND N. MARZARI, *Exponential localization of Wannier functions in insulators*, Phys. Rev. Lett., 98 (2007), 046402.
- [22] K. BURKE ET AL., *The ABC of DFT*, available online from <http://chem.uci.edu/~kieron/dftold2/literature.php>, accessed October 2012.
- [23] M. CERIOTTI, T. D. KÜHNE, AND M. PARRINELLO, *An efficient and accurate decomposition of the Fermi operator*, J. Chem. Phys., 129 (2008), 024707.
- [24] M. CHALLACOMBE, *A simplified density matrix minimization for linear scaling self-consistent field theory*, J. Chem. Phys., 110 (1999), pp. 2332–2342.
- [25] M. CHALLACOMBE, *A general parallel sparse-blocked matrix multiply for linear scaling SCF theory*, Comput. Phys. Commun., 128 (2000), pp. 93–107.
- [26] S. R. CHINNAMSETTY, *Wavelet Tensor Product Approximation in Electronic Structure Calculations*, Ph.D. thesis, Universität Leipzig, 2008.
- [27] S. R. CHINNAMSETTY, M. ESPIG, H.-J. FLAD, AND W. HACKBUSCH, *Canonical tensor products as a generalization of Gaussian-type orbitals*, Z. Phys. Chem., 224 (2010), pp. 681–694.

- [28] S. R. CHINNAMSETTY, M. ESPIG, B. N. KHOROMSKIJ, W. HACKBUSCH, AND H.-J. FLAD, *Tensor product approximation with optimal rank in quantum chemistry*, J. Chem. Phys., 127 (2007), 084110.
- [29] C. K. CHUI AND M. HASSON, *Degree of uniform approximation on disjoint intervals*, Pacific J. Math., 105 (1983), pp. 291–297.
- [30] L. COLOMBO AND W. SAWYER, *A parallel implementation of tight-binding molecular dynamics*, Mater. Sci. Engrg. B, 37 (1996), pp. 228–231.
- [31] M. CRAMER AND J. EISERT, *Correlations, spectral gap and entanglement in harmonic quantum systems on generic lattices*, New J. Phys., 8 (2006), 71.
- [32] M. CRAMER, J. EISERT, M. B. PLENIO, AND J. DREISSIG, *Entanglement-area law for general bosonic harmonic lattice systems*, Phys. Rev. A, 73 (2006), 012309.
- [33] P. J. DAVIS AND P. RABINOWITZ, *Methods of Numerical Integration*, 2nd ed., Academic Press, London, 1984.
- [34] S. DEMKO, W. F. MOSS, AND P. W. SMITH, *Decay rates for inverses of band matrices*, Math. Comp., 43 (1984), pp. 491–499.
- [35] J. W. DEMMEL, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, PA, 1997.
- [36] J. DES CLOIZEAUX, *Energy bands and projection operators in a crystal: Analytic and asymptotic properties*, Phys. Rev., 135 (1964), pp. A685–A697.
- [37] R. DIESTEL, *Graph Theory*, Springer, Berlin, 2000.
- [38] J. EISERT, M. CRAMER, AND M. B. PLENIO, *Colloquium: Area laws for the entanglement entropy*, Rev. Modern Phys., 82 (2010), pp. 277–306.
- [39] E. ESTRADA, *Generalized walks-based centrality measures for complex biological networks*, J. Theoret. Biol., 263 (2010), pp. 556–565.
- [40] E. ESTRADA, *The Structure of Complex Networks: Theory and Applications*, Oxford University Press, Oxford, UK, 2012.
- [41] E. ESTRADA AND N. HATANO, *Communicability in complex networks*, Phys. Rev. E, 77 (2008), 036111.
- [42] E. ESTRADA AND N. HATANO, *Returnability in complex directed networks (digraphs)*, Linear Algebra Appl., 430 (2009), pp. 1886–1896.
- [43] E. ESTRADA AND N. HATANO, *A vibrational approach to node centrality and vulnerability in complex networks*, Phys. A, 389 (2010), pp. 3648–3660.
- [44] E. ESTRADA, N. HATANO, AND M. BENZI, *The physics of communicability in complex networks*, Phys. Rep., 514 (2012), pp. 89–119.
- [45] E. ESTRADA AND D. J. HIGHAM, *Network properties revealed through matrix functions*, SIAM Rev., 52 (2010), pp. 696–714.
- [46] E. ESTRADA AND J. A. RODRÍGUEZ-VELÁZQUEZ, *Subgraph centrality in complex networks*, Phys. Rev. E, 71 (2005), 056103.
- [47] H.-J. FLAD, W. HACKBUSCH, D. KOLB, AND R. SCHNEIDER, *Wavelet approximation of correlated wavefunctions. I. Basics*, J. Chem. Phys., 116 (2002), pp. 9641–9657.
- [48] H.-J. FLAD, W. HACKBUSCH, B. N. KHOROMSKIJ, AND R. SCHNEIDER, *Concepts of data-sparse tensor-product approximation in many-particle modelling*, in Matrix Methods: Theory, Algorithms and Applications: Dedicated to the Memory of Gene Golub, V. Olshevsky and E. Tyrtshnikov, eds., World Scientific, Hackensack, NJ, 2010, pp. 313–343.
- [49] C. J. GARCÍA-CERVERA, J. LU, Y. XUAN, AND W. E, *A linear scaling subspace iteration algorithm with optimally localized non-orthogonal wave functions for Kohn–Sham density functional theory*, Phys. Rev. B, 79 (2009), 115110.
- [50] G. GIULIANI AND G. VIGNALE, *Quantum Theory of the Electron Liquid*, Cambridge University Press, Cambridge, UK, 2005.
- [51] S. GOEDECKER, *Low complexity algorithms for electronic structure calculations*, J. Comput. Phys., 118 (1995), pp. 261–268.
- [52] S. GOEDECKER, *Decay properties of the finite-temperature density matrix in metals*, Phys. Rev. B, 58 (1998), pp. 3501–3502.
- [53] S. GOEDECKER, *Linear scaling electronic structure methods*, Rev. Modern Phys., 71 (1999), pp. 1085–1123.
- [54] S. GOEDECKER AND L. COLOMBO, *Efficient linear scaling algorithm for tight-binding molecular dynamics*, Phys. Rev. Lett., 73 (1994), pp. 122–125.
- [55] S. GOEDECKER AND O. V. IVANOV, *Frequency localization properties of the density matrix and its resulting hypersparsity in a wavelet representation*, Phys. Rev. B, 59 (1999), pp. 7270–7273.
- [56] G. H. GOLUB AND G. MEURANT, *Matrices, Moments and Quadrature with Applications*, Princeton University Press, Princeton, NJ, 2010.
- [57] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.

- [58] U. GRENANDER AND G. SZEGÖ, *Toeplitz Forms and Their Applications*, Chelsea, New York, 1958.
- [59] W. HACKBUSCH, *Hierarchische Matrizen: Algorithmen und Analysis*, Springer, Berlin, 2009.
- [60] P. HARTMAN AND A. WINTNER, *The spectra of Toeplitz's matrices*, Amer. J. Math., 76 (1954), pp. 867–882.
- [61] M. HASSON, *The degree of approximation by polynomials on some disjoint intervals in the complex plane*, J. Approx. Theory, 144 (2007), pp. 119–132.
- [62] M. B. HASTINGS, *Locality in quantum and Markov dynamics on lattices and networks*, Phys. Rev. Lett., 93 (2004), 140402.
- [63] M. B. HASTINGS AND T. KOMA, *Spectral gap and exponential decay of correlations*, Comm. Math. Phys., 265 (2006), pp. 781–804.
- [64] L. HE AND D. VANDERBILT, *Exponential decay properties of Wannier functions and related quantities*, Phys. Rev. Lett., 86 (2001), pp. 5341–5344.
- [65] N. J. HIGHAM, *Functions of Matrices: Theory and Computation*, SIAM, Philadelphia, PA, 2008.
- [66] P. HOHENBERG AND W. KOHN, *Inhomogeneous electron gas*, Phys. Rev., 136 (1964), pp. B864–871.
- [67] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1991.
- [68] R. A. HORN AND C. R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1994.
- [69] A. ISERLES, *How large is the exponential of a banded matrix?*, New Zealand J. Math., 29 (2000), pp. 177–192.
- [70] S. ISMAIL-BEIGI AND T. A. ARIAS, *Locality of the density matrix in metals, semiconductors, and insulators*, Phys. Rev. Lett., 82 (1999), pp. 2127–2130.
- [71] S. JAFFARD, *Propriétés des matrices “bien localisées” près de leur diagonale et quelques applications*, Ann. Inst. Henri Poincaré, 7 (1990), pp. 461–476.
- [72] B. JANSIK, S. HØST, P. JØRGENSEN, AND J. OLSEN, *Linear-scaling symmetric square-root decomposition of the overlap matrix*, J. Chem. Phys., 126 (2007), 124104.
- [73] J. JĘDRZEJEWSKI AND T. KROKHMALSKII, *Exact results for spatial decay of the one-body density matrix in low-dimensional insulators*, Phys. Rev. B, 70 (2004), 153102.
- [74] W. KOHN, *Analytic properties of Bloch waves and Wannier functions*, Phys. Rev., 115 (1959), pp. 809–821.
- [75] W. KOHN, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett., 76 (1996), pp. 3168–3171.
- [76] W. KOHN AND L. J. SHAM, *Self-consistent equations including exchange and correlation effects*, Phys. Rev. (2), 140 (1965), pp. A1133–A1138.
- [77] C. LE BRIS, *Computational chemistry from the perspective of numerical analysis*, Acta Numer., 14 (2005), pp. 363–444.
- [78] S. LI, S. AHMED, G. GLIMIECK, AND E. DARVE, *Computing entries of the inverse of a sparse matrix using the FIND algorithm*, J. Comput. Phys., 227 (2008), pp. 9408–9427.
- [79] X.-P. LI, R. W. NUNES, AND D. VANDERBILT, *Density-matrix electronic structure method with linear system-size scaling*, Phys. Rev. B, 47 (1993), pp. 10891–10894.
- [80] W. LIANG, C. SARAVANAN, Y. SHAO, R. BAER, A. T. BELL, AND M. HEAD-GORDON, *Improved Fermi operator expansion methods for fast electronic structure calculations*, J. Chem. Phys., 119 (2003), pp. 4117–4124.
- [81] L. LIN, J. LU, L. YING, R. CAR, AND W. E, *Multipole representation of the Fermi operator with application to the electronic structure analysis of metallic systems*, Phys. Rev. B, 79 (2009), 115133.
- [82] L. LIN, C. YANG, J. LU, L. YING, AND W. E, *A fast parallel algorithm for selected inversion of structured sparse matrices with application to 2D electronic structure calculations*, SIAM J. Sci. Comput., 33 (2011), pp. 1329–1351.
- [83] L. LIN, C. YANG, J. C. MEZA, J. LU, AND L. YING, *SelInv—An algorithm for selected inversion of a sparse symmetric matrix*, ACM Trans. Math. Software, 37 (2011), pp. 1–19.
- [84] G. G. LORENTZ, *Approximation of Functions*, Holt, Rinehart and Winston, New York, 1966.
- [85] P.-O. LÖWDIN, *Linear Algebra for Quantum Theory*, John Wiley and Sons, New York, 1998.
- [86] H. LUO, D. KOLB, H.-J. FLAD, W. HACKBUSCH, AND T. KOPRUCKI, *Wavelet approximation of correlated wavefunctions. II. Hyperbolic wavelets and adaptive approximation schemes*, J. Chem. Phys., 117 (2002), pp. 3625–3638.
- [87] G. MACKEY, *Mathematical Foundations of Quantum Mechanics*, Dover, New York, 2004.
- [88] N. H. MARCH, W. H. YOUNG, AND S. SAMPANTHAR, *The Many-Body Problem in Quantum Mechanics*, Cambridge University Press, Cambridge, UK, 1967.

- [89] R. M. MARTIN, *Electronic Structure. Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, UK, 2004.
- [90] P. E. MASLEN, C. OCHSENFELD, C. A. WHITE, M. S. LEE, AND M. HEAD-GORDON, *Locality and sparsity of ab initio one-particle density matrices and localized orbitals*, J. Phys. Chem. A, 102 (1998), pp. 2215–2222.
- [91] N. MASTRONARDI, M. NG, AND E. E. TYRTYSHNIKOV, *Decay in functions of multiband matrices*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2721–2737.
- [92] G. MEINARDUS, *Approximation of Functions: Theory and Numerical Methods*, Springer Tracts in Natural Philosophy 13, Springer, New York, 1967.
- [93] J. M. MILLAM AND G. SCUSERIA, *Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations*, J. Chem. Phys., 106 (1997), pp. 5569–5577.
- [94] N. MOISEYEV, *Non-Hermitian Quantum Mechanics*, Cambridge University Press, Cambridge, UK, 2011.
- [95] K. NEMETH AND G. SCUSERIA, *Linear scaling density matrix search based on sign matrices*, J. Chem. Phys., 113 (2000), pp. 6035–6041.
- [96] G. NENCIU, *Existence of the exponentially localised Wannier functions*, Comm. Math. Phys., 91 (1983), pp. 81–85.
- [97] A. M. N. NIKLASSON, *Density matrix methods in linear scaling electronic structure theory*, in Linear-Scaling Techniques in Computational Chemistry and Physics, R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski, eds., Springer, New York, 2011, pp. 439–473.
- [98] P. ORDEJÓN, *Order-N tight-binding methods for electronic-structure and molecular dynamics*, Comput. Materials Sci., 12 (1998), pp. 157–191.
- [99] P. ORDEJÓN, D. A. DRABOLD, R. M. MARTIN, AND M. P. GRUMBACH, *Linear system-size scaling methods for electronic-structure calculations*, Phys. Rev. B, 51 (1995), pp. 1456–1476.
- [100] T. OZAKI, *Efficient recursion method for inverting an overlap matrix*, Phys. Rev. B, 64 (2001), 195110.
- [101] B. PARLETT, *Invariant subspaces for tightly clustered eigenvalues of tridiagonals*, BIT, 36 (1996), pp. 542–562.
- [102] R. K. PATHRIA, *Statistical Mechanics*, Internat. Ser. Natural Philos. 45, Pergamon Press, Oxford, UK, 1986.
- [103] E. PRODAN, *Nearsightedness of electronic matter in one dimension*, Phys. Rev. B, 73 (2006), 085108.
- [104] E. PRODAN, S. R. GARCIA, AND M. PUTINAR, *Norm estimates of complex symmetric operators applied to quantum systems*, J. Phys. A Math. Gen., 39 (2006), pp. 389–400.
- [105] E. PRODAN AND W. KOHN, *Nearsightedness of electronic matter*, Proc. Natl. Acad. Sci. USA, 102 (2005), pp. 11635–11638.
- [106] N. RAZOUK, *Localization Phenomena in Matrix Functions: Theory and Algorithms*, Ph.D. thesis, Emory University, Atlanta, GA, 2008.
- [107] M. REED AND B. SIMON, *Methods of Modern Mathematical Physics. Volume I: Functional Analysis*, Academic Press, New York, London, 1972.
- [108] M. ROSENBLUM, *The absolute continuity of Toeplitz’s matrices*, Pacific J. Math., 10 (1960), pp. 987–996.
- [109] E. H. RUBENSSON, *Controlling errors in recursive Fermi–Dirac operator expansions with applications in electronic structure theory*, SIAM J. Sci. Comput., 34 (2012), pp. B1–B23.
- [110] E. H. RUBENSSON, N. BOCK, E. HOLMSTRÖM, AND A. M. N. NIKLASSON, *Recursive inverse factorization*, J. Chem. Phys., 128 (2008), 104105.
- [111] E. H. RUBENSSON, E. RUDBERG, AND P. SALEK, *Rotations of occupied invariant subspaces in self-consistent field calculations*, J. Math. Phys., 49 (2008), 032103.
- [112] E. H. RUBENSSON, E. RUDBERG, AND P. SALEK, *Truncation of small matrix elements based on the Euclidean norm for blocked data structures*, J. Comput. Chem., 30 (2009), pp. 974–977.
- [113] E. H. RUBENSSON, E. RUDBERG, AND P. SALEK, *Methods for Hartree–Fock and density functional theory electronic structure calculations with linearly scaling processor time and memory usage*, in Linear-Scaling Techniques in Computational Chemistry and Physics, R. Zalesny, M. G. Papadopoulos, P. G. Mezey, and J. Leszczynski, eds., Springer, New York, 2011, pp. 269–300.
- [114] E. H. RUBENSSON AND P. SALEK, *Systematic sparse matrix error control for linear scaling electronic structure calculations*, J. Comput. Chem., 26 (2005), pp. 1628–1637.
- [115] W. RUDIN, *Functional Analysis*, McGraw–Hill, New York, 1973.

- [116] Y. SAAD, J. R. CHELIKOWSKY, AND S. M. SHONTZ, *Numerical methods for electronic structure calculations of materials*, SIAM Rev., 52 (2010), pp. 3–54.
- [117] V. E. SACKSTEDER, *Linear Algebra with Disordered Sparse Matrices That Have Spatial Structure: Theory and Computation*, Ph.D. thesis, Department of Physics, Università degli Studi di Roma “La Sapienza,” Rome, Italy, 2004.
- [118] V. E. SACKSTEDER,  *$O(N)$  algorithms for disordered systems*, Numer. Linear Algebra Appl., 12 (2005), pp. 827–838.
- [119] R. SCHNEIDER AND T. WEBER, *Wavelets for density matrix computation in electronic structure calculations*, Appl. Numer. Math., 56 (2006), pp. 1383–1396.
- [120] N. SCHUCH, *Quantum Entanglement: Theory and Applications*, Ph.D. thesis, Department of Physics, Technischen Universität München, Munich, Germany, 2007.
- [121] N. SCHUCH, J. I. CIRAC, AND M. M. WOLF, *Quantum states on harmonic lattices*, Comm. Math. Phys., 267 (2006), pp. 65–92.
- [122] S. SCHWEIZER, J. KUSSMANN, B. DOSER, AND C. HOCHSENFELD, *Linear-scaling Cholesky decomposition*, J. Comput. Chem., 29 (2008), pp. 1004–1010.
- [123] G. SCUSERIA, *Linear scaling density functional calculations with Gaussian orbitals*, J. Phys. Chem. A, 25 (1999), pp. 4782–4790.
- [124] R. B. SIDJE AND Y. SAAD, *Rational approximation to the Fermi–Dirac function with applications in density functional theory*, Numer. Algorithms, 56 (2011), pp. 455–479.
- [125] A. SZABO AND N. OSTLUND, *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory*, Dover, New York, 1996.
- [126] J. TANG AND Y. SAAD, *A probing method for computing the diagonal of a matrix inverse*, Numer. Linear Algebra Appl., 19 (2012), pp. 485–501.
- [127] S. N. TARASKIN, D. A. DRABOLD, AND S. R. ELLIOTT, *Spatial decay of the single-particle density matrix in insulators: Analytic results in two and three dimensions*, Phys. Rev. Lett., 88 (2002), 196495.
- [128] S. N. TARASKIN, P. A. FRY, X. ZHANG, D. A. DRABOLD, AND S. R. ELLIOTT, *Spatial decay of the single-particle density matrix in tight-binding metals: Analytic results in two dimensions*, Phys. Rev. B, 66 (2002), 233101.
- [129] T. THONHAUSER AND D. VANDERBILT, *Insulator/Chern-insulator transition in the Haldane model*, Phys. Rev. B, 74 (2006), 23511.
- [130] C. VÖMEL AND B. N. PARLETT, *Detecting localization in an invariant subspace*, SIAM J. Sci. Comput., 33 (2011), pp. 3447–3467.
- [131] J. VON NEUMANN, *Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik*, Gött. Nach., 1927. See also Collected Works, Volume I, A. H. Taub, ed., Pergamon Press, New York, Oxford, London, Paris, 1961, pp. 208–235.
- [132] J. VON NEUMANN, *Approximative properties of matrices of high finite order*, Port. Math., 3 (1942), pp. 1–62. See also Collected Works, Volume I, A. H. Taub, ed., Pergamon Press, New York, Oxford, London, Paris, 1961, pp. 270–331.
- [133] J. VON NEUMANN, *Mathematical Foundations of Quantum Mechanics*, Princeton University Press, Princeton, NJ, 1955.
- [134] S. C. WATSON AND E. CARTER, *Linear-scaling parallel algorithms for the first principles treatment of metals*, Comput. Phys. Commun., 128 (2000), pp. 67–92.
- [135] S. Y. WU AND C. S. JAYANTHI, *Order- $N$  methodologies and their applications*, Phys. Rep., 358 (2002), pp. 1–74.
- [136] H. J. XIANG, J. YANG, J. G. HOU, AND Q. ZHU, *Linear scaling calculation of band edge states and doped semiconductors*, J. Chem. Phys., 126 (2007), 244707.
- [137] C. YANG, W. GAO, AND J. C. MEZA, *On the convergence of the self-consistent field iteration for a class of nonlinear eigenvalue problems*, SIAM J. Matrix Anal. Appl., 30 (2009), pp. 1773–1788.
- [138] W. YANG, *Direct calculation of electron density in density-functional theory*, Phys. Rev. Lett., 66 (1991), pp. 1438–1441.
- [139] Q. YE, *Error bounds for the Lanczos method for approximating matrix exponentials*, SIAM J. Numer. Anal., 51 (2013), pp. 68–87.
- [140] X. ZHANG AND D. A. DRABOLD, *Properties of the density matrix from realistic calculations*, Phys. Rev. B, 63 (2001), 233109.