

# Lecture Notes on Quantum Mechanics & Quantum Annealing – Some Background for the D-Wave Quantum Computer and Adiabatic Quantum Computation

SUMMER 2016 – VERSION PRL-QMN-2016.v1.2b – Copyright © 2016 and 2017

Ken Kreutz–Delgado

Electrical & Computer Engineering, Jacobs School of Engineering

Qualcomm Institute Pattern Recognition Laboratory, Calit2

University of California, San Diego

August 10, 2017

## 1 Preface

These lecture notes were provided in the summer of 2016 to a “D-Wave Interest Group” (DWIG) predominantly comprised of first-year MS students in the Electrical and Computer Engineering (ECE) department at the University of California, San Diego. Participation was entirely voluntary and no course credit was given. The DWIG, which met weekly, was interested in participating in a course of directed reading aimed at understanding Adiabatic Quantum Computation (AQC) and Quantum Annealing (QA). The purpose of the lecture notes given below was to supplement and expand on some of the quantum mechanics (QM) discussion provided in the NASA/JPL/D-Wave/USC report [24], the D-Wave research memorandum [22], and the monograph [18]. Specifically, because the participants in the DWIG had little or no prior exposure to quantum mechanics, the notes attempt to “fill in the blanks” of some of the discussion given in references [24, 22, 18]. In particular, other than some aspects of its pedagogic content, there is no claim to originality of these notes over and above the material presented in standard references such as [24, 22, 18], among others.

The background of the approximately ten participants during Summer 2016 was typical for contemporary BS degree graduates of EE and CS departments who have had additional courses in machine learning and pattern recognition roughly corresponding to the level of [3]. In particular, the participants had a level of proficiency in probability and random variables roughly corresponding to the level of the textbook [21], and in linear algebra roughly corresponding to the level of the textbook [20]. Typical for EE/CS graduates, they were also knowledgeable about the theory of ordinary differential equations. Other than the cursory introduction given in standard lower division physics courses, the participants had had no prior exposure to quantum mechanics.

Fortunately, the prior exposure to linear algebra at the relatively sophisticated and geometrically-motivated level of [20] meant that the participants effectively had already been introduced to the theory of finite-dimensional Hilbert spaces (except for the concept of a dual space, which was straightforward to

present and motivate). With this background, and some required translation of notation and terminology from linear algebra to the Dirac notation-based finite-dimensional Hilbert space formalism, the DWIG could be quickly exposed to the (non-relativistic) finite-dimensional quantum mechanics of systems of distinguishable (non-identical) spin- $\frac{1}{2}$  particles (qubits) needed to understand AQC. This was accomplished by assigning as mandatory reading the first seven chapters of the very accessible textbook by Susskind and Friedman [26], which has a primary focus on spin- $\frac{1}{2}$  systems and provides an introduction to the density matrix. Suggested additional sources for the mathematics of qubit systems were [14] and [12], the former reference providing a detailed discussion of manipulating the Pauli spin matrices and the latter providing additional discussion of the density matrix.<sup>1</sup>

As relevant background to these notes, prior lectures were given on the classical (non-quantum mechanical) stochastic Ising spin-glass model used both for modeling neural networks and for optimization of boolean objective functions via simulated annealing. The discussion of stochastic neural networks was based on material drawn from the classic textbooks of Hertz, Krogh, and Palmer [9] and Amit [1], while the discussion of simulated annealing (SA) was drawn from the foundational paper [15] and the presentation in [9].<sup>2</sup> Because, as discussed in [15, 9], classical equilibrium statistical mechanics plays a fundamental role in simulated annealing-based optimization, prior lectures based on material drawn from [4, 19] were given to provide an overview of the classical canonical ensemble, Boltzmann-Gibbs distribution, and Helmholtz free energy. Further, As noted in the D-Wave memorandum [22], equilibrium statistical mechanics is equally important for understanding key aspects of quantum annealing, motivating the discussion of quantum and classical statistical mechanics given in the notes below.<sup>3</sup>

In the lectures given to the DWIG, it is emphasized that the D-Wave computer is an extraordinarily sophisticated and advanced applied physics, engineering, and mathematical system that can be understood and engaged with at several levels:

1. Most fundamental is the device physics and engineering level. This requires a deep understanding of experimental and applied low-temperature physics and devices, such as SQUIDS, which are based on low-temperature phenomena and structures such as the Josephson junction.
2. Closely tied to the device and engineering physics level are the detailed mathematical models that guide the design, construction, and assembly of the components of the D-Wave computer to ensure that the system behaves like a controllable quantum mechanical spin-glass system of interconnected distinguishable qubits (via the use of “manufactured spins” [13]). For a discussion of issues pertaining to Levels 1 and 2, the DWIG participants were referred to references [8, 13, 24].
3. The next higher level of abstraction is provided by the quantum mechanical spin-glass behavior that is enforced by Levels 1 and 2. This behavior is used to analyze, design, and implement quantum annealing algorithms capable of combinatorial optimization (specifically quadratic unconstrained binary optimization (QUBO) problems) subject to device imposed constraints.
4. This level of abstraction assumes that the D-Wave computer functions as a black box capable of solving a certain family of QUBO problems. The onus is on the user to translate a combinatorial optimization problem of interest to QUBO form.

---

<sup>1</sup>More advanced QM references used in the drafting of these notes are [16, 6, 2, 5, 23].

<sup>2</sup>In the discussions and notes below, following [18], we often refer to SA as “Classical Annealing” (CA) to distinguish SA from the “Quantum Annealing” (QA) described in [24, 18].

<sup>3</sup>Useful references for quantum statistical mechanics are [10, 17, 19, 11, 25]. The extension of the Ising spin-glass model to the quantum mechanical Heisenberg model of ferromagnetism is briefly discussed in [11].

5. The highest level of abstraction exists when a user merely specifies a combinatorial optimization problem and a “compiler” exists that is capable of translating it to QUBO form that can be solved by the D-Wave computer.

## 2 Finite Dimensional Hilbert Space and Hermitian Operators

### 2.1 General Background

Let  $\mathcal{H}$  be a *finite-dimensional* Hilbert space of dimension  $n$  over the field of complex numbers. *It is an axiom of quantum mechanics* that a complex Hilbert space can serve as the state-space of (the *pure states* of) a quantum mechanical system of interest.<sup>4</sup>

Let  $M$  denote an *observable quantity* in physical space that can take experimentally measured real values  $m_0, \dots, m_{n-1}$ . *It is a second axiom of quantum mechanics* that the physical observable  $M$  is represented by  $\hat{M}$ , a self-adjoint (*hermitian*) linear operator in the Hilbert space  $\mathcal{H}$  whose eigenvalues are precisely equal to the experimentally measurable values  $m_i$ . The situation is as follows, when a measurement of  $M$  is taken, one of the real eigenvalues of  $\hat{M}$ ,  $m = m_i$ , is observed and the quantum state “jumps” to the corresponding eigenvector,  $|\phi\rangle = |\phi_i\rangle$ , of the operator  $\hat{M}$ .<sup>5</sup>

The eigenvectors and eigenvalues of the self-adjoint operator  $\hat{M}$  are given by,<sup>6</sup>

$$\hat{M}|\phi_j\rangle = m_j|\phi_j\rangle, \quad j = 0, \dots, n-1$$

with  $m_j \in \mathbb{R}$ . The information provided by the eigenvectors and eigenvalues are encoded in the *spectral representation* (or *spectral decomposition*),<sup>7</sup>

$$\hat{M} = \sum_{j=0}^{n-1} m_j |\phi_j\rangle\langle\phi_j| = \sum_{j=0}^{n-1} m_j \hat{P}^{(\phi_j)}, \quad (1)$$

where

$$\hat{P}^{(\phi_j)} = |\phi_j\rangle\langle\phi_j|.$$

is the orthogonal projection operator onto the one-dimensional subspace spanned by  $|\phi_j\rangle$ .<sup>8</sup> The identity operator  $\hat{I}$  can be resolved as,

$$\hat{I} = \sum_{j=0}^{n-1} |\phi_j\rangle\langle\phi_j| = \sum_{j=0}^{n-1} \hat{P}^{(\phi_j)}.$$

By induction on the eigenvalue/vector equation,  $\hat{M}|\phi_j\rangle = m_j|\phi_j\rangle$ , we have

$$\alpha \hat{M}^k |\phi_j\rangle = \alpha m_j^k |\phi_j\rangle \quad \text{for } j = 0, \dots, n-1 \quad \text{and } k = 0, 1, \dots$$

<sup>4</sup>We use the standard “Copenhagen interpretation” axiomatic approach articulated by von Neumann; see the discussion in [6]. The derivations given in this note are (for the most part) valid and straightforward because we are working in a finite-dimensional Hilbert space; things are trickier in infinite-dimensional Hilbert spaces. For example, linear operators on a finite-dimensional Hilbert space  $\mathcal{H}$ ,  $n = \dim(\mathcal{H}) < \infty$ , are all bounded, have point spectra ( $n$  eigenvalues only in the spectrum), and a domain of definition equal to  $\mathcal{H}$  [7].

<sup>5</sup>This occurs probabilistically, as we discuss when we present “Axiom 3” in the next section.

<sup>6</sup>We assume that the (orthogonal) eigenvectors have been normalized so that  $\langle\phi_i|\phi_j\rangle = \delta_{ij}$ .

<sup>7</sup>All self-adjoint linear operators on a finite dimensional Hilbert space of dimension  $n$  have a complete set of  $n$  orthonormal eigenvectors that can be used as a basis for the space, and  $n$  associated real eigenvalues. This means that every self-adjoint operator on a finite dimensional Hilbert space has a *discrete spectrum*.

<sup>8</sup>It is straightforward to show that  $\hat{P}^{(\phi_j)}$  is self-adjoint and idempotent.

for any scalar  $\alpha \in \mathbb{C}$ . As a consequence,

$$\text{Poly}(\hat{M})|\phi_j\rangle = \text{Poly}(m_j)|\phi_j\rangle \text{ for } j = 0, \dots, n-1$$

for any polynomial function  $\text{Poly}(z)$  and, more generally,

$$F(\hat{M})|\phi_j\rangle = F(m_j)|\phi_j\rangle \text{ for } j = 0, \dots, n-1 \quad (2)$$

for any analytic function  $F(z)$ . Note that  $F(m_j)$  and  $|\phi_j\rangle$ ,  $j = 0, \dots, n-1$ , are the eigenvalues and orthonormal eigenvectors of  $F(\hat{M})$  and therefore the operator  $F(\hat{M})$  has a spectral decomposition,

$$F(\hat{M}) = \sum_{j=0}^{n-1} F(m_j) |\phi_j\rangle\langle\phi_j| = \sum_{j=0}^{n-1} F(m_j) \hat{P}^{(\phi_j)}. \quad (3)$$

## 2.2 Energy Eigenstates and Energy Representation

Consider the very important observable represented by the operator  $\hat{H}$ , the *Hamiltonian operator*. It represents the observable energy  $E$  which takes real values  $E_0, \dots, E_{n-1}$ . The possible energy values are assumed to be ordered as,

$$E_0 \leq E_1 \leq E_2 \leq \dots \leq E_{n-1},$$

where  $E^{(0)} = E_0$  is the *ground state* and  $E^{(1)} = \min\{E_j | E_j > E_0\}$  is the *first excited state*.<sup>9</sup> The eigenvectors and eigenvalues (energy values) of  $\hat{H}$  are, of course, related by,<sup>10</sup>

$$\hat{H}|\psi_j\rangle = E_j|\psi_j\rangle, \quad j = 0, \dots, n-1$$

$$\hat{H} = \sum_{j=0}^{n-1} E_j |\psi_j\rangle\langle\psi_j| = \sum_{j=0}^{n-1} E_j \hat{P}^{(\psi_j)},$$

and

$$\hat{I} = \sum_{j=0}^{n-1} |\psi_j\rangle\langle\psi_j| = \sum_{j=0}^{n-1} \hat{P}^{(\psi_j)}$$

with  $E_j \in \mathbb{R}$ .

When a general state vector  $|\phi\rangle$  is represented in terms of the eigenvectors of  $\hat{H}$ ,

$$|\phi\rangle = \hat{I}|\phi\rangle = \left( \sum_{j=0}^{n-1} |\psi_j\rangle\langle\psi_j| \right) |\phi\rangle = \sum_{j=0}^{n-1} c_j |\psi_j\rangle, \quad c_j = \langle\psi_j|\phi\rangle,$$

we call this the *energy representation* of  $|\phi\rangle$ .<sup>11</sup> As a preliminary application of the very useful Equations (2) and (3), note that they imply

$$e^{-\beta\hat{H}}|\psi_j\rangle = e^{-\beta E_j}|\psi_j\rangle.$$

and

$$e^{-\beta\hat{H}} = \sum_{j=0}^{n-1} e^{-\beta E_j} |\psi_j\rangle\langle\psi_j|.$$

<sup>9</sup>More generally we index distinct energy levels by  $E^{(\ell)}$  in order of increasing magnitude,  $E^{(\ell+1)} > E^{(\ell)}$ . Note that if the ground state is nondegenerate then  $E^{(1)} = E_1$ .

<sup>10</sup>Again assume that the eigenvectors have been normalized,  $\langle\psi_i|\psi_j\rangle = \delta_{ij}$ .

<sup>11</sup>Note that the energy representation is equivalent to the column vector representation  $\mathbf{c} = (c_0, \dots, c_{n-1})^T \in \mathbb{C}^n$ .

### 3 Probabilities and the Density Matrix

Given that the system is in state  $|\psi\rangle \in \mathcal{H}$ , it is a third axiom of quantum mechanics that the probability that a measurement will observe the system to be in state  $|\phi\rangle$  is given by,<sup>12</sup>

$$p_\psi(\phi) = |\langle\psi|\phi\rangle|^2.$$

Note that,<sup>13</sup>

$$|\langle\psi|\phi\rangle|^2 = \langle\psi|\phi\rangle\langle\phi|\psi\rangle = \text{tr} \left( |\psi\rangle\langle\psi| |\phi\rangle\langle\phi| \right) = \text{tr} \hat{P}^{(\psi)} \hat{P}^{(\phi)}.$$

Suppose, then, that the quantum mechanical system is (has been prepared to be) in state  $\psi$ . We define the *density matrix* (aka *density operator*) by

$$\hat{\rho} = \hat{P}^{(\psi)} = |\psi\rangle\langle\psi|.$$

Note that in this case  $\hat{\rho} \iff \psi$ , so that  $p_\psi(\phi) = p_{\hat{\rho}}(\phi)$ , which begins to explain why in quantum mechanics *the density matrix  $\hat{\rho}$  itself is also called the state of the system*. With this definition we have,

$$p_\psi(\phi) = \text{tr} \hat{\rho} \hat{P}^{(\phi)} \quad \text{with} \quad \hat{\rho} = \hat{P}^{(\psi)} = |\psi\rangle\langle\psi| \quad \text{and} \quad \hat{P}^{(\phi)} = |\phi\rangle\langle\phi|. \quad (4)$$

The *most general form* of the density matrix is

$$\hat{\rho} = w_1 \hat{P}^{(1)} + \dots + w_m \hat{P}^{(m)} \quad (5)$$

where  $\hat{P}^{(j)} = |\psi_j\rangle\langle\psi_j|$  are projection operators onto one-dimensional subspaces of  $\mathcal{H}$  spanned by unit vectors  $|\psi_j\rangle$ ,  $w_j > 0$ , and  $w_1 + \dots + w_m = 1$ . It is important to note that the subspaces spanned by  $|\psi_j\rangle$  do *not* have to be mutually orthogonal.<sup>14</sup> Note that the general density matrix  $\hat{\rho}$  is self-adjoint,  $\hat{\rho} = \hat{\rho}^*$ , and that  $\text{tr} \hat{\rho} = 1$ . One can utilize a decomposition of the form (5) to prepare the system to be in state  $\hat{\rho}$  by choosing an basis  $\{|\psi_0\rangle, \dots, |\psi_{n-1}\rangle\}$  and then selecting a state  $|\psi_j\rangle$  (which is equivalent to selecting  $\hat{P}^{(j)} = |\psi_j\rangle\langle\psi_j|$ ) according to a probability equal to  $w_j$ . Once the state  $\hat{\rho}$  has been prepared, however, it is an important fact that *the decomposition (5) generally is not unique*, which can be interpreted to mean that the system “forgets” (or “does not know”) how it was prepared to be in state  $\hat{\rho}$  since there are many possible ways to do this.<sup>15</sup> Nonetheless, given the mathematical expression (5) (which describes a *possible* preparation procedure), one can interpret  $w_j$  to be the “prior probability” that the system was prepared to be in state  $|\psi_j\rangle$  *if* in fact Eq. (5) did describe the actual preparation procedure (which it might not, given that other decompositions are possible).

<sup>12</sup>Some authors refer to  $p_\psi(\phi)$  as the transition probability because it is the probability that state  $|\psi\rangle$  transitions to state  $|\phi\rangle$ ,  $|\psi\rangle \rightarrow |\phi\rangle$ , as a consequence of taking a measurement. The third axiom is known as *Born’s Rule*, after Max Born who proposed it in 1926.

<sup>13</sup>For any bounded operators (e.g., for matrices)  $A$  and  $B$ ,

$$\text{tr} AB = \text{tr} BA.$$

This is a very useful result. Here we use the fact that given two column vectors  $a$  and  $b$ ,  $a^H b = \text{tr} a^H b = \text{tr} b a^H$ .

<sup>14</sup>Note that when the unit vectors  $|\psi_j\rangle$  are not orthogonal, then Eq. (5) is not a spectral representation of  $\hat{\rho}$ . However, when they are orthogonal, then (5) is a spectral representation.

<sup>15</sup>Another interpretation is that  $\hat{\rho}$  contains the simultaneous quantum superposition of all of the (potentially infinite) ways that it could have been prepared into existence, which is another example of “quantum parallelism.” Discussion of the nonuniqueness of the density matrix can be found in Chapter 2 of [2] and in Appendix A below.

If there exists a state  $|\phi\rangle \in \mathcal{H}$  such that the general density matrix can be written as  $\hat{\rho} = \hat{P}^{(\phi)} = |\phi\rangle\langle\phi|$ , then we say that  $\hat{\rho}$  is a *pure state*. If  $\hat{\rho}$  is pure, then the system is in a definite quantum mechanical state  $|\phi\rangle$ . If  $\hat{\rho}$  is not pure, and must be written in the general form (5), then it is a *mixed state*.<sup>16</sup>

Given that the quantum mechanical system has been prepared in a state  $\hat{\rho}$ , then the expected value of an observable  $\hat{M}$  is given by,

$$\langle M \rangle \triangleq \mathbb{E} \{M\} = \text{tr} \hat{\rho} \hat{M}. \quad (6)$$

Representing the density matrix  $\hat{\rho}$  by a decomposition (5), this is straightforward to show:

$$\begin{aligned} \text{tr} \hat{\rho} \hat{M} &= \sum_{j=0}^{n-1} w_j \underbrace{|\psi_j\rangle\langle\psi_j|}_{\hat{P}^{(\psi_j)}} \sum_{i=0}^{n-1} m_i \underbrace{|\phi_i\rangle\langle\phi_i|}_{\hat{P}^{(\phi_i)}} \\ &= \sum_{j=0}^{n-1} w_j \sum_{i=0}^{n-1} m_i |\langle\psi_i|\phi_j\rangle|^2 \\ &= \sum_{j=0}^{n-1} w_j \sum_{i=0}^{n-1} m_i p_\psi(\phi_i) \\ &= \sum_{j=0}^{n-1} w_j \mathbb{E}_\psi \{M\} \\ &= \mathbb{E} \{M\} = \langle M \rangle. \end{aligned}$$

## 4 The Boltzmann-Gibbs Distribution

### 4.1 Some Canonical Statistical Mechanics

For a mechanically isolated quantum mechanical system placed in a heat bath at constant temperature  $T$ , constant volume  $V$ , and constant particle number  $N$ , the density matrix is given by the *canonical density matrix*, or *Boltzmann-Gibbs density matrix*,<sup>17</sup>

$$\hat{\rho} = \frac{e^{-\beta\hat{H}}}{\text{tr} e^{-\beta\hat{H}}}, \quad (7)$$

with  $\beta = \frac{1}{kT}$ , where for physical systems  $k$  is taken to be Boltzmann's constant,  $k = k_B$ .<sup>18</sup> Note that,

$$e^{-\beta\hat{H}} = e^{-\beta\hat{H}} I = e^{-\beta\hat{H}} \sum_{j=0}^{n-1} |\psi_j\rangle\langle\psi_j| = \sum_{j=0}^{n-1} e^{-\beta\hat{H}} |\psi_j\rangle\langle\psi_j| = \sum_{j=0}^{n-1} e^{-\beta E_j} |\psi_j\rangle\langle\psi_j|.$$

With  $\text{tr} |\psi_j\rangle\langle\psi_j| = \langle\psi_j|\psi_j\rangle = 1$ , we have

$$\text{tr} e^{-\beta\hat{H}} = \sum_{j=0}^{n-1} e^{-\beta E_j} = Z_\beta, \quad (8)$$

<sup>16</sup>To reiterate: if  $\hat{\rho}$  is a mixed state, then it *cannot* be written as  $\hat{\rho} = P^{(\phi)}$  for some state  $|\phi\rangle \in \mathcal{H}$ . I.e., the system is *not* in a definite quantum mechanical state  $|\phi\rangle \in \mathcal{H}$ .

<sup>17</sup>We will also refer to this form of  $\hat{\rho}$  as the *canonical state* or the *Boltzmann-Gibbs state*.

<sup>18</sup>We will take  $k = 1$ .

where  $Z_\beta$  is the partition function.<sup>19</sup> Thus the Boltzmann-Gibbs density matrix takes the form,

$$\hat{\rho} = \sum_{j=0}^{n-1} \frac{e^{-\beta E_j}}{Z_\beta} |\psi_j\rangle\langle\psi_j| = \sum_{j=0}^{n-1} w_j |\psi_j\rangle\langle\psi_j|, \quad w_j = \frac{e^{-\beta E_j}}{Z_\beta}. \quad (9)$$

Note that  $w_j$  is the Boltzmann-Gibbs probability that a mechanically isolated, constant particle and constant volume statistical system in a heat bath is in the energy eigenstate  $|\psi_j\rangle$ . Indeed, consider the observable  $M$  represented by the operator,

$$\hat{M} = \hat{P}^{(j)} = |\psi_j\rangle\langle\psi_j| = 1 \cdot |\psi_j\rangle\langle\psi_j| + 0 \cdot \sum_{i \neq j} |\psi_i\rangle\langle\psi_i|.$$

The observable  $M$  corresponds to asking (and taking a measurement to answer) the question:<sup>20</sup>

*Is the system in the particular energy eigenstate  $|\psi_j\rangle$  with energy (eigenvalue)  $E_j$ ?*

Note that this question corresponds to the (random) indicator function,<sup>21</sup>

$$M = \mathbf{1}(E(\psi_j) = E_j) \iff \hat{\rho} = \hat{P}^{(j)} = |\psi_j\rangle\langle\psi_j|.$$

We have

$$p(\psi_j) = \mathbb{E} \{ \mathbf{1}(E(\psi_j) = E_j) \} = \langle M \rangle = \text{tr} \hat{\rho} \hat{M} = \underbrace{\text{tr} \hat{\rho} \hat{P}^{(j)}}_{\langle \hat{P}^{(j)} \rangle} = w_j = \frac{e^{-\beta E_j}}{Z_\beta},$$

which is the Boltzmann-Gibbs distribution,  $p_j = p(\psi_j) = \frac{e^{-\beta E_j}}{Z_\beta}$ , encountered in classical statistical mechanics.<sup>22</sup>

Having gone from the quantum mechanical statement of the canonical distribution to the classical statement, let's go in the reverse direction:

$$\begin{aligned} p_j &= \frac{e^{-\beta E_j}}{Z_\beta} = \sum_{i=0}^{n-1} \frac{e^{-\beta E_i}}{Z_\beta} \underbrace{|\langle\psi_i|\psi_j\rangle|^2}_{\delta_{ij}} \\ &= \text{tr} \sum_{i=0}^{n-1} \frac{e^{-\beta E_i}}{Z_\beta} |\psi_i\rangle\langle\psi_i| |\psi_j\rangle\langle\psi_j| \\ &= \text{tr} \left( \sum_{i=0}^{n-1} \frac{e^{-\beta \hat{H}}}{Z_\beta} |\psi_i\rangle\langle\psi_i| \right) |\psi_j\rangle\langle\psi_j| \end{aligned}$$

<sup>19</sup>For  $N \gg 1$ , the number of states,  $n$ , is a very large number and it is generally intractable to compute the partition function  $Z_\beta$ .

<sup>20</sup>Note that this question is more precise than asking if the system is in *any* eigenstate that has energy  $E_j$ . This is because of the possibility that the energy  $E_j$  could be degenerate (i.e., have more than one eigenstate with the same eigenvalue  $E_j$ ). Thus we are *not* asking for the probability that the system has energy  $E_j$ ; we are asking for the probability that the system is in state  $\psi_j$  which is only one of the energy  $E_j$  eigenstates if energy level  $E_j$  is degenerate.

<sup>21</sup>This is an important fact: A *pure state*,  $\hat{P}^j = |\psi_j\rangle\langle\psi_j|$  represents a measurement that corresponds to asking a Yes-No question, i.e., to an indicator function.

<sup>22</sup>In classical statistical mechanics,  $j$  indexes a classical phase space state, whereas in quantum statistical mechanics it indexes a (pure) quantum mechanical state.

$$\begin{aligned}
&= \text{tr} \underbrace{\frac{e^{-\beta \hat{H}}}{Z_\beta}}_{\hat{\rho}} \underbrace{\left( \sum_{i=0}^{n-1} |\psi_i\rangle \langle \psi_i| \right)}_{\hat{I}} \underbrace{|\psi_j\rangle \langle \psi_j|}_{P^{(j)} = \hat{M}} \\
&= \text{tr} \hat{\rho} \hat{M} = \langle M \rangle = \langle \hat{P}^{(j)} \rangle = w_j.
\end{aligned}$$

## 4.2 Free Energy for a Finite-Dimensional Quantum Mechanical System

The relevant Thermodynamic Potential for the canonical distribution is the (Helmoltz) free energy, which is classically given by

$$F_T(p) = U(p) - TS(p),$$

where  $p = (p_0, \dots, p_{n-1})^T$ ,  $\sum_i p_i = 1$ , denotes the classical canonical distribution.<sup>23</sup> In *classical statistical mechanics*, the quantity

$$S = S(p) = - \sum_{j=0}^{n-1} p_j \ln p_j = -\mathbb{E} \{ \ln p \}$$

is the (Gibbs) thermodynamic entropy (setting  $k = 1$ ) and

$$U = U(p) = \mathbb{E} \{ E \} = \langle E \rangle = \sum_{j=0}^{n-1} p_j E_j$$

is the average energy (internal energy) of the system. It is evident that the free energy is a function of the temperature  $T$  (or, equivalently, of  $\beta = 1/T$ ) and of the Boltzmann-Gibbs distribution  $p$ ,  $F = F_T(p)$ .

*It is a fundamental fact of equilibrium statistical mechanics that the free energy takes its minimum value at thermal equilibrium*, a fact which is true both for classical and quantum statistical mechanics.<sup>24</sup> Note that minimizing the free energy corresponds to a balance between reducing the internal energy  $U$  and increasing the entropy  $S$ . Classical simulated annealing (CA), draws MCMC samples from a system having the Boltzmann-Gibbs distribution as its stationary distribution while taking  $T \rightarrow 0$  very slowly. This ultimately results in samples being drawn from, or near, the  $T = 0$  energy ground state of the system.

Quantum mechanically, the free energy for the canonical state given in Equations  $\hat{\rho}$  (7) and (9) is

$$F_T(\hat{\rho}) = U(\hat{\rho}) - TS(\hat{\rho})$$

where

$$U(\hat{\rho}) = \langle E \rangle = \text{tr} \hat{\rho} \hat{H} \quad \text{and} \quad S(\hat{\rho}) = -\text{tr} \hat{\rho} \ln \hat{\rho}.$$

In this form,  $S = S(\hat{\rho})$  is known as the quantum mechanical, or von Neumann, entropy. Thus, quantum mechanically,

$$F_T(\hat{\rho}) = \text{tr} \hat{\rho} \hat{H} + T \text{tr} \hat{\rho} \ln \hat{\rho}. \quad (10)$$

<sup>23</sup>But again note the equivalence of the classical and quantum mechanical Boltzmann-Gibbs distribution,  $p_j = w_j = Z_\beta^{-1} e^{-\beta E_j}$ , and see the next paragraph.

<sup>24</sup>This has interesting consequences. For example, it allows for an alternative derivation of the classical Boltzmann-Gibbs distribution as the distribution that minimizes the free energy with respect to  $p$  subject to the constraint that  $\sum_j p_j = 1$ .

One can readily see that this yields the classical free energy as we can show that  $U(\hat{\rho}) = U(p)$  and  $S(\hat{\rho}) = S(p)$  as follows:

$$\begin{aligned}
 U(\hat{\rho}) &= \text{tr } \hat{\rho} \hat{H} = \text{tr } \frac{e^{-\beta \hat{H}}}{Z_\beta} \sum_{j=0}^{n-1} E_j |\psi_j\rangle \langle \psi_j| \\
 &= \text{tr } \frac{1}{Z_\beta} \sum_{j=0}^{n-1} E_j e^{-\beta \hat{H}} |\psi_j\rangle \langle \psi_j| = \text{tr } \frac{1}{Z_\beta} \sum_{j=0}^{n-1} E_j e^{-\beta E_j} |\psi_j\rangle \langle \psi_j| \\
 &= \sum_{j=0}^{n-1} \underbrace{\frac{e^{-\beta E_j}}{Z_\beta}}_{=w_j=p_j} E_j \underbrace{\text{tr } |\psi_j\rangle \langle \psi_j|}_{=\langle \psi_j | \psi_j \rangle = 1} = \sum_{j=0}^{n-1} p_j E_j = U(p).
 \end{aligned}$$

Note that the canonical state  $\hat{\rho}$  can be written in terms of the orthonormal energy eigenvectors  $|\psi_j\rangle$  as shown in Eq. (9). For this reason  $w_j = p_j$  and  $|\psi_j\rangle$  are eigenvalue/eigenvector pairs for the operator  $\hat{\rho}$ ,

$$\hat{\rho} |\psi_j\rangle = p_j |\psi_j\rangle,$$

and Eq. (9) gives the spectral representation of  $\hat{\rho}$ . Therefore (see Eq. (2))

$$(\hat{\rho} \ln \hat{\rho}) |\psi_j\rangle = (p_j \ln p_j) |\psi_j\rangle$$

and (see Eq. (3))

$$\hat{\rho} \ln \hat{\rho} = \sum_{j=0}^{n-1} (p_j \ln p_j) |\psi_j\rangle \langle \psi_j|.$$

We have,

$$\begin{aligned}
 S(\hat{\rho}) &= -\text{tr } \hat{\rho} \ln \hat{\rho} = -\text{tr } \sum_{j=0}^{n-1} (p_j \ln p_j) |\psi_j\rangle \langle \psi_j| \\
 &= -\sum_{j=0}^{n-1} (p_j \ln p_j) \underbrace{\text{tr } |\psi_j\rangle \langle \psi_j|}_{\langle \psi_j | \psi_j \rangle = 1} = -\sum_{j=0}^{n-1} p_j \ln p_j = S(p).
 \end{aligned}$$

## 5 Optimization via Classical and Quantum Annealing

As mentioned above, a canonical statistical mechanical system<sup>25</sup> has a free energy,

$$F_T(\hat{\rho}) = U(\hat{\rho}) - TS(\hat{\rho}) = \text{tr } \hat{\rho} \hat{H} + T \text{tr } \hat{\rho} \ln \hat{\rho},$$

which is a minimum. For low enough temperature,  $T \approx 0$ , minimum free energy effectively corresponds to minimum internal energy,  $U = \langle E \rangle$ . Further, at  $T \approx 0$ , energy fluctuations are negligible and the internal energy is approximately equal to the ground state energy  $U = \langle E \rangle \approx E_0 \approx \text{constant}$ . Exploiting these two facts, we can “anneal” the system into an equilibrium configuration that approximates a ground-state configuration. In the D-Wave computer, annealing is not only controlled by the temperature parameter  $T$

<sup>25</sup>I.e., a mechanically isolated, constant particle, constant volume, system in thermal equilibrium with a heat bath at temperature  $T$  in the state (9).

in an initial “*classical annealing*” phase used to initialize the system, but also by the subsequent use of magnetic coupling of spin- $\frac{1}{2}$  qubits to an external driving field and to each other via interaction fields in a second “*quantum annealing*” phase that adiabatically evolves the Hamiltonian into a desired final form, all the while maintaining the system in a ground state configuration.

Specifically, in the D-Wave computer we take,<sup>26</sup>

$$\hat{H} = \hat{H}(\Gamma_P, \Gamma_D) = \Gamma_P \hat{H}_P + \Gamma_D \hat{H}_D,$$

where  $\Gamma_D$  and  $\Gamma_P$  take positive real number values.  $\hat{H}_P$  is a “problem Hamiltonian operator”, which encodes an optimization loss-function that we wish to minimize, and  $\hat{H}_D$  is the initial, or *disorder*, Hamiltonian operator that is used to prepare the state of the D-Wave computer into a maximally non-committal (“maximally disordered”) initialization configuration. The real scalar  $\Gamma_P \in [0, 1]$  is a control parameter that allows one to “turn on” the problem Hamiltonian  $H_P$ . The real scalar  $\Gamma_D \in [0, 1]$  is a control parameter that allow one to “turn off” the disorder Hamiltonian  $H_D$ . With this form for  $\hat{H}$ , we have

$$F_T(\hat{\rho}; \Gamma_P, \Gamma_D) = \text{tr } \hat{\rho} \hat{H}(\Gamma_P, \Gamma_D) + T \text{tr } \hat{\rho} \ln \hat{\rho} = \Gamma_P \text{tr } \hat{\rho} \hat{H}_P + \Gamma_D \text{tr } \hat{\rho} \hat{H}_D + T \text{tr } \hat{\rho} \ln \hat{\rho}.$$

To perform a two-stage (classical plus quantum) annealing process, we start with  $\Gamma_P = 0$  and  $\Gamma_D = 1$ , and work with the system

$$F_T(\hat{\rho}; 0, 1) = \text{tr } \hat{\rho} \hat{H}(0, 1) + T \text{tr } \hat{\rho} \ln \hat{\rho} = \text{tr } \hat{\rho} \hat{H}_D + T \text{tr } \hat{\rho} \ln \hat{\rho}.$$

We then slowly cool (thermally anneal) the system to a very low temperature,  $T \rightarrow T_0 \approx 0$  (around 15 millikelvin). This essentially corresponds to a classical annealing (CA) step which gets us (close to) the ground state of  $H_D$ , which is designed to be a maximally disordered superposition of all of the accessible states of the system.

Now assuming that  $T = T_0 \approx 0$ , we next consider the system

$$F_{T_0}(\hat{\rho}; \Gamma_P, \Gamma_D) = \Gamma_P \text{tr } \hat{\rho} \hat{H}_P + \Gamma_D \text{tr } \hat{\rho} \hat{H}_D + T_0 \text{tr } \hat{\rho} \ln \hat{\rho}.$$

We very slowly (adiabatically) take  $\Gamma_D$  from 1 to 0 while simultaneously taking  $\Gamma_P$  from 0 to 1. This is the quantum annealing (QA) stage. This is done very slowly (“adiabatically”) so that the system all the while has a high probability of being maintained in a ground state configuration. This adiabatic transition from  $\hat{H}_D$  to  $\hat{H}_P$  is formally indicated by

$$\lim_{\substack{\Gamma_D \rightarrow 0 \\ \Gamma_P \rightarrow 1}} F_{T_0}(\hat{\rho}; \Gamma_P, \Gamma_D) = \text{tr } \hat{\rho} \hat{H}_P + T_0 \text{tr } \hat{\rho} \ln \hat{\rho}.$$

In this limit, the free energy being at a minimum for a system in thermal equilibrium, we should be in a (near) ground-state configuration of the system with (final) Hamiltonian  $\hat{H}_P$ , a configuration which can be read from the system registers.

---

<sup>26</sup>Here we follow the discussion in [22].

## 6 The D-Wave Initialization State and Hamiltonian Operator $\hat{H}_D$

### 6.1 The Completely Sideways State $|r\rangle$

The D-Wave computer is designed to behave like a coupled quantum mechanical spin-glass system comprised of  $N$  distinguishable spin- $\frac{1}{2}$  qubits [24, 22, 18].<sup>27</sup> A single qubit can be represented in the two dimensional complex state space  $\mathcal{H} \simeq \mathbb{C}^2$ .<sup>28</sup> The measurements of the qubits are taken along the vertical (or  $z$ ) direction, and for this reason it is natural take as a basis for a single-qubit system to be the two orthonormal vectors “spin up” =  $|u\rangle \simeq \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in \mathbb{C}^2$  and “spin down” =  $|d\rangle \simeq \begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \mathbb{C}^2$ .<sup>29</sup> It is convenient to represent a basis vector in  $\mathbb{C}^2$  by  $|z\rangle$  where  $z \in \{u, d\}$ . We can think of  $|z\rangle$  as representing a “vertical state” (up or down) for a single qubit system.

For an  $N$  qubit system, let  $\mathcal{H}_i \simeq \mathbb{C}^2$  be the *component state space* for qubit  $i$ .  $\mathcal{H}_i$  has basis vectors  $|z_i = u\rangle \simeq \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and  $|z_i = d\rangle \simeq \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  which we denote as  $|z_i\rangle$  with  $z_i \in \{u, d\}$ . For convenience we set<sup>30</sup>

$$\mathbf{z} = (z_1, \dots, z_N) \in \{u, d\}^N = \underbrace{\{u, d\} \times \dots \times \{u, d\}}_{(N-1)\text{-fold cartesian product}}.$$

Let the interacting compound  $N$ -qubit system have a *compound state-space* given by

$$\mathcal{H} = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_N.$$

The typical basis vector for the compound space  $\mathcal{H}$  when using the  $N$ -fold tensor product “ $\mathbf{z}$ -basis” is<sup>31</sup>

$$|\mathbf{z}\rangle = |z_1, \dots, z_N\rangle = |z_1\rangle \otimes \dots \otimes |z_N\rangle = \bigotimes_{j=1}^N |z_j\rangle, \quad \mathbf{z} \in \{u, d\}^N. \quad (11)$$

We call the compound basis vector  $|\mathbf{z}\rangle$  a “vertical state” for the compound system.<sup>32</sup> Note that  $\mathcal{H}$  is an  $n = 2^N$  dimensional Hilbert space.<sup>33</sup> A general compound system state vector  $|\phi\rangle \in \mathcal{H}$  has the  $\mathbf{z}$ -basis representation

$$|\phi\rangle = \hat{I}|\phi\rangle = \left( \sum_{\mathbf{z} \in \{u, d\}^N} |\mathbf{z}\rangle \langle \mathbf{z}| \right) |\phi\rangle = \underbrace{\sum_{\mathbf{z} \in \{u, d\}^N} \mathbf{c}_{\mathbf{z}} |\mathbf{z}\rangle}_{2^N \text{ terms in this sum}}, \quad \mathbf{c}_{\mathbf{z}} = \langle \mathbf{z} | \phi \rangle.$$

In general, we cannot write  $|\phi\rangle \in \mathcal{H}$  as a simple tensor product of vectors in the component subspaces  $\mathcal{H}_i, i = 1, \dots, N$ , in which case we say that the component qubits are *entangled*. Otherwise they are *unentangled*.<sup>34</sup>

<sup>27</sup>Where  $N = 1,000$  in the current generation.

<sup>28</sup>The symbol “ $\simeq$ ” is used to indicate that the right-hand-side is a convenient choice of a concrete representation of the abstract entity on the left-hand-side.

<sup>29</sup>Alternative common symbols used in place of  $\{|u\rangle, |d\rangle\}$  for the spin up/down basis vectors are  $\{|+1\rangle, |-1\rangle\}, \{|+\rangle, |-\rangle\}, \{|\uparrow\rangle, |\downarrow\rangle\}, \{|0\rangle, |1\rangle\}$ , and  $\{|1\rangle, |0\rangle\}$ .

<sup>30</sup>Note that  $\mathbf{z} \in \{u, d\}^N$  is just shorthand for  $|z_j\rangle \in \{u, d\}, j = 1, \dots, N$ .

<sup>31</sup>It is useful to note that the tensor product is linear in each position. I.e.,

$$|a\rangle \otimes (\beta_1 |b_1\rangle + \beta_2 |b_2\rangle) \otimes |c\rangle \otimes |d\rangle = \beta_1 |a\rangle \otimes |b_1\rangle \otimes |c\rangle + \beta_2 |a\rangle \otimes |b_2\rangle \otimes |c\rangle \otimes |d\rangle.$$

To describe this property, we say that the tensor product is *multilinear*.

<sup>32</sup>Again, “vertical” means that both, and only, “pointing up” and “pointing down” component qubits are allowed. Thus we can say that the bases are comprised of all of the vertical unit vectors

<sup>33</sup>I.e., there are  $2^N$  basis vectors.

<sup>34</sup>It is very important to clearly understand the distinction between entanglement and quantum superposition.

Assume that we can turn on a globally transverse magnetic field that is directed “to the right”, i.e., along the positive  $x$ -direction in physical space. We will denote a single qubit oriented along the positive  $x$ -direction (“to the right”) by the unit vector  $|r\rangle \in \mathcal{H}$ , which we denote by  $x = r$ , and along the negative  $x$ -direction (“to the left”) by the unit vector  $|l\rangle$ , which we denote by  $x = l$ . Note that these two vectors are orthogonal. In terms of the  $z$ -basis for a single qubit system, we have

$$|r\rangle = \frac{1}{\sqrt{2}}(|u\rangle + |d\rangle) \quad \text{and} \quad |l\rangle = \frac{1}{\sqrt{2}}(|u\rangle - |d\rangle).$$

We will say that  $|r\rangle$  and  $|l\rangle$  represent the “sideways states” for a single qubit system.

Now consider the compound system. We define a “sideways state” to be

$$|\mathbf{x}\rangle = |x_1\rangle \otimes \cdots \otimes |x_N\rangle \in \mathcal{H}$$

where  $x_j \in \{r, l\}$ .<sup>35</sup> If *all*  $N$  qubits are oriented “to the right” along the positive  $x$ -direction in physical space, we have the compound state,

$$|\mathbf{x} = \mathbf{r}\rangle = |x_1 = r\rangle \otimes \cdots \otimes |x_N = r\rangle \in \mathcal{H},$$

which, using a natural notation, we conveniently write as,

$$|\mathbf{r}\rangle = |r_1\rangle \otimes \cdots \otimes |r_N\rangle = \bigotimes_{j=1}^N |r_j\rangle \in \mathcal{H}. \quad (12)$$

Similarly, we have

$$|\mathbf{l}\rangle = \bigotimes_{j=1}^N |l_j\rangle \in \mathcal{H}.$$

We will call  $|\mathbf{r}\rangle$  and  $|\mathbf{l}\rangle$  the “completely sideways states” of the compound system.<sup>36</sup>

Note that the completely sideways state  $|\mathbf{r}\rangle$  is obviously a tensor product state over all of the component spaces, so that the component qubits are unentangled. We can rewrite  $|\mathbf{r}\rangle$  as

$$|\mathbf{r}\rangle = \left[ \frac{1}{\sqrt{2}}(|u_1\rangle + |d_1\rangle) \right] \otimes \cdots \otimes \left[ \frac{1}{\sqrt{2}}(|u_N\rangle + |d_N\rangle) \right], \quad (13)$$

or

$$|\mathbf{r}\rangle = 2^{-\frac{N}{2}} \left[ \underbrace{(|u_1\rangle + |d_1\rangle)}_{\in \mathcal{H}_1} \otimes \cdots \otimes \underbrace{(|u_N\rangle + |d_N\rangle)}_{\in \mathcal{H}_N} \right] = \bigotimes_{j=1}^N \underbrace{\left( \sum_{z_j \in \{u,d\}} |z_j\rangle \right)}_{\in \mathcal{H}_j} \in \mathcal{H}. \quad (14)$$

Again, the tensor product factorization into component spaces is evident. Utilizing the multilinearity property of the tensor product, we can rewrite the completely sideways state  $|\mathbf{r}\rangle$  as

$$|\mathbf{r}\rangle = \sum_{z_j \in \{u,d\}} \bigotimes_{j=1}^N |z_j\rangle = \sum_{\mathbf{z} \in \{u,d\}^N} |\mathbf{z}\rangle. \quad (15)$$

Note that when written in this manner, it is not immediately apparent that the completely sideways state  $|\mathbf{r}\rangle$  is an unentangled state.

<sup>35</sup>Note that there are  $2^N$  different sideways vectors.

<sup>36</sup>Note that a completely sideways state requires that the component qubit spins are either *all* pointing to the right, in which case we have the compound state  $|\mathbf{r}\rangle$ , or *all* pointing to the left, in which case we have the state  $|\mathbf{l}\rangle$ .

## 6.2 Statistical Properties of the Completely Sideways State $|\mathbf{r}\rangle$

Consider the  $j^{\text{th}}$  qubit in isolation. This system can be prepared to be in any state  $|v_j\rangle \in \mathcal{H}_j$ , but assume that it has been prepared to be in the sideways state  $|v_j\rangle = |r_j\rangle$ ,

$$|r_j\rangle = \frac{1}{\sqrt{2}}(|u_j\rangle + |d_j\rangle).$$

Then the probability that a vertical measurement detects the qubit in the vertical state  $|u_j\rangle$  is

$$p_{r_j}(u_j) = |\langle r_j, u_j \rangle|^2 = \frac{1}{2}.$$

Similarly,<sup>37</sup>

$$p_{r_j}(d_j) = |\langle r_j, d_j \rangle|^2 = \frac{1}{2}.$$

We say that state  $|r_j\rangle$  is *maximally noncommittal* in terms of providing a bias towards either of the vertical states  $|u_j\rangle$  and  $|d_j\rangle$ . Note that because  $p_{r_j}(u_j) = p_{r_j}(d_j)$ , preparing the system in state  $|r_j\rangle$  creates the *maximum entropy distribution* over all the possible probabilities  $p_{v_j}(u_j)$  and  $p_{v_j}(d_j)$  determined from preparing the system to be in a state  $|v_j\rangle \in \mathcal{H}_j$ . The fact that  $|r_j\rangle$  creates the maximum entropy distribution is another way of saying that it is maximally noncommittal in terms of biasing a vertical measurement in favor of  $|u_j\rangle$  or  $|d_j\rangle$ . Note that using the formalism of the density matrix, the above derivations correspond to,

$$\hat{\rho} = P^{(r_j)} = |r_j\rangle\langle r_j| = \text{pure state}, \quad P^{(u_j)} = |u_j\rangle\langle u_j|, \quad P^{(d_j)} = |d_j\rangle\langle d_j|.$$

$$p_{r_j}(u_j) = \text{tr } \hat{\rho} P^{(u_j)} = \frac{1}{2} \quad \text{and} \quad p_{r_j}(d_j) = \text{tr } \hat{\rho} P^{(d_j)} = \frac{1}{2}.$$

Let  $|z_j\rangle \in \mathcal{H}_i$  be a vertical state in  $\mathcal{H}_i$ ,  $z_j \in \{u_j, d_j\}$ . We have,

$$p_{r_j}(z_j) = \text{tr } \hat{\rho} P^{(z_j)} = |\langle r_j | z_j \rangle|^2 = \frac{1}{2}.$$

Now assume that the entire compound system has been prepared to be in the completely sideways state  $|\mathbf{r}\rangle \in \mathcal{H}$ . This corresponds to the pure state  $\hat{\rho} = |\mathbf{r}\rangle\langle \mathbf{r}|$ . Given the completely sideways state  $\hat{\rho}$  what is the probability of a vertical state  $|\mathbf{z}\rangle$ ? We have,

$$p_{\mathbf{r}}(\mathbf{z}) = \text{tr } \hat{\rho} P^{(\mathbf{z})} = \text{tr } (|\mathbf{r}\rangle\langle \mathbf{r}|) (|\mathbf{z}\rangle\langle \mathbf{z}|) = |\langle \mathbf{r} | \mathbf{z} \rangle|^2.$$

However (see Equations (11) and (12)),

$$|\langle \mathbf{r} | \mathbf{z} \rangle|^2 = \left| \left( \langle r_1 | \otimes \cdots \otimes \langle r_N | \right) \left( |z_1\rangle \otimes \cdots \otimes |z_N\rangle \right) \right|^2 = \prod_{j=1}^N |\langle r_j | z_j \rangle|^2 = \prod_{j=1}^N p_{r_j}(z_j) = \frac{1}{2^N}.$$

Note that it is reasonable to write,

$$p_{r_j}(z_j) = p_{\mathbf{r}}(z_j)$$

---

<sup>37</sup>Of course  $p_{r_j}(u_j) + p_{r_j}(d_j) = 1$ .

because  $\mathbf{r} \implies r_j$ .<sup>38</sup> Thus we have that for any vertical state  $|\mathbf{z}\rangle = |z_1, \dots, z_N\rangle$ ,<sup>39</sup>

$$p_{\mathbf{r}}(\mathbf{z}) = \prod_{j=1}^N p_{\mathbf{r}}(z_j) = \frac{1}{2^N} = \text{constant}. \quad (16)$$

Also note that,

$$\sum_{\mathbf{z} \in \{u,d\}^N} p_{\mathbf{r}}(\mathbf{z}) = 2^N \left( \frac{1}{2^N} \right) = 1,$$

so that the set of all possible vertical states forms a probability sample space (an exhaustive set of disjoint events).<sup>40</sup> Thus if the compound qubit system is prepared to be in the completely sideways state  $|\mathbf{r}\rangle$ , then 1) all vertical states are equally likely; 2) the measurements of the component qubits of a vertical state are completely independent from each other,<sup>41</sup> and 3) each component qubit is equally likely to be up or down. This means that the probability distribution on the vertical states induced by the completely sideways state  $|\mathbf{r}\rangle$  is a maximum entropy distribution.

It is evidently, then, that being in the completely sideways state  $|\mathbf{r}\rangle$  corresponds to being maximally noncommittal as to the possible vertical values of the component qubits. More to the point, the initialization state is in a complete quantum superposition of all possible vertical values (up or down) of the component qubits. We can view the state  $|\mathbf{r}\rangle$  as one that induces “maximum disorder” on the vertical states. Another way to think about this is that all possible up-down qubit states simultaneously exist as a consequence of quantum superposition allowing for a “quantum parallelism” and that initializing the D-Wave computer to this state leads to the system exploring all possible states simultaneously as the Hamiltonian operator is evolved to the problem Hamiltonian as discussed earlier.<sup>42</sup>

It is the job of the “disorder Hamiltonian operator” to ensure that at thermal equilibrium the system is in this state of “maximum disorder”. For this reason we call a Hamiltonian operator that has  $|\mathbf{r}\rangle$  as a *ground energy eigenvalue* a “disorder Hamiltonian”  $\hat{\mathbf{H}}_D$ . This maximally disordered state  $|\mathbf{r}\rangle$  is the initialization state that the D-Wave computer is placed into in the first stage of thermal annealing discussed earlier. In the next subsection we describe this in more detail.

### 6.3 Initialization using the Disorder Hamiltonian Operator $\hat{\mathbf{H}}_D$

We now describe the “disorder Hamiltonian” operator  $\hat{\mathbf{H}}_D$  for the compound  $N$ -qubit system that has the completely sideways state  $|\mathbf{r}\rangle$  as its lowest energy eigenstate,<sup>43</sup>

$$\hat{\mathbf{H}}_D |\mathbf{r}\rangle = \mathbf{E}_0 |\mathbf{r}\rangle.$$

<sup>38</sup>Alternatively, one can take the component state  $|z_j\rangle$  to be equivalent to the compound state  $0 \otimes \dots \otimes 0 \otimes |z_j\rangle \otimes 0 \otimes \dots \otimes 0$ , where each 0 is the zero vector in a Hilbert space  $\mathcal{H}_j$ .

<sup>39</sup>Recall that factorizability of a joint distribution into the product of its marginals is the definition of independence.

<sup>40</sup>This is because we are restricting ourselves to only measuring qubit events in the vertical direction.

<sup>41</sup>This is related to the fact that the state  $|\mathbf{r}\rangle$  is an unentangled state.

<sup>42</sup>In Section 3.E of reference [24], the state  $|\mathbf{r}\rangle$  is called the “symmetric state”  $|\Phi_S\rangle$ , presumably because it is “symmetric” (non-committal) with respect to a preference for up-down qubit states. However, this state is not fully symmetric (“unpolarized”) with respect to (hypothetical) measurements in other directions—see the comments on pages 207-208 of [25] and the discussion regarding the lack of uniqueness of the density matrix in [2]. Of course since D-Wave only provides measurements in the up-down directions this point is irrelevant in this context.

<sup>43</sup>I.e., as its ground state.

We assume that the compound qubit system is in thermal equilibrium with a heat bath, and we therefore work with the Boltzmann-Gibbs canonical density function shown in Equations (7) and (9). At thermal equilibrium, the free energy

$$F_T(\hat{\rho}; 0, 1) = \text{tr } \hat{\rho} \hat{\mathbf{H}}_D + T \text{tr } \hat{\rho} \ln \hat{\rho} = U(\hat{\rho}) - TS(\hat{\rho}) \quad (17)$$

takes its minimum value. We will give a condition for which minimum free energy corresponds to the compound system being in the completely sideways ground state  $|r\rangle$  with high probability when  $\hat{\mathbf{H}}_D$  is chosen as follows.

Since our goal is to drive all the component qubits “to the right”, we choose a Hamiltonian operator that gives the energy of each qubit in a transverse magnetic field pointing to the right. For the  $j^{\text{th}}$  qubit this corresponds to a transverse field interaction Hamiltonian of

$$\hat{H}_D^{(j)} = -b_j \sigma_x^{(j)} \quad (18)$$

where  $\sigma_x^{(j)}$  is the standard Pauli spin matrix expressed in the  $|u_j\rangle$ - $|d_j\rangle$  reference frame and  $b_j > 0$  is a local transverse magnetic field intensity affecting qubit  $j$ .<sup>44</sup> Because the sideways vectors  $|r_j\rangle$  and  $|l_j\rangle$  are the eigenvectors of  $\sigma_x^{(j)}$  with eigenvalues  $+1$  and  $-1$  respectively<sup>45</sup> we have,

$$\hat{H}_D^{(j)} |r_j\rangle = E_0 |r_j\rangle = -b_j |r_j\rangle \quad \text{and} \quad \hat{H}_D^{(j)} |l_j\rangle = E_1 |l_j\rangle = +b_j |l_j\rangle.$$

For the composite system comprised of  $N$  qubits, we set<sup>46</sup>

$$b_j = b = \text{constant} > 0 \quad (19)$$

for all component qubits and take,<sup>47</sup>

$$\hat{\mathbf{H}}_D = \hat{\mathbf{H}}_D^{(1)} + \cdots + \hat{\mathbf{H}}_D^{(N)} = \sum_{j=1}^N \hat{\mathbf{H}}_D^{(j)}, \quad (20)$$

where

$$\hat{\mathbf{H}}_D^{(j)} = \hat{I} \otimes \cdots \otimes \hat{I} \otimes \hat{H}_D^{(j)} \otimes \hat{I} \otimes \cdots \otimes \hat{I}. \quad (21)$$

If we set

$$\hat{\mathbf{X}}_j = \hat{I} \otimes \cdots \otimes \hat{I} \otimes \sigma_x^{(j)} \otimes \hat{I} \otimes \cdots \otimes \hat{I},$$

then

$$\hat{\mathbf{H}}_D^{(j)} = -b \hat{\mathbf{X}}_j$$

and

$$\hat{\mathbf{H}}_D = -b \sum_{j=1}^N \hat{\mathbf{X}}_j \quad (22)$$

<sup>44</sup>The scalar factor of the magnetic dipole moment  $\mu_x^{(j)} = \gamma \sigma_x^{(j)}$  has been taken to be one,  $\gamma = 1$ .

<sup>45</sup>See the appendix of [26].

<sup>46</sup>I.e., we turn on a uniform transverse magnetic field with uniform magnetic field intensity  $b$ .

<sup>47</sup>The energy Hamiltonian is additive because there are no spin-spin interaction terms. I.e., all qubit-qubit interactions are “turned off.”

which is Equation (3.34) of [24].<sup>48</sup>

Note that

$$\hat{\mathbf{H}}^{(j)}|\mathbf{x}\rangle = \pm b|\mathbf{x}\rangle, \quad j = 1, \dots, N, \quad (23)$$

for any sideways state  $|\mathbf{x}\rangle$ , and

$$\hat{\mathbf{H}}^{(j)}|\mathbf{r}\rangle = -b|\mathbf{r}\rangle, \quad \text{for all } j = 1, \dots, N, \quad (24)$$

for the completely sideways state  $|\mathbf{r}\rangle$ . It is evident from Equations (20) and (23) that the sideways states  $|\mathbf{x}\rangle$  are the eigenvectors of  $\hat{\mathbf{H}}_D$  and that the energy levels (eigenvalues) range from the ground state energy  $\mathbf{E}_0 = -Nb$  to the most energetic state  $\mathbf{E}_{n-1} = +Nb$ , with  $n = 2^N$ .<sup>49</sup> In particular the completely sideways vector  $|\mathbf{r}\rangle$  is the unique (i.e., *nondegenerate*) ground state eigenvector,

$$\hat{\mathbf{H}}_D|\mathbf{r}\rangle = \sum_{j=1}^N \hat{\mathbf{H}}_D^{(j)}|\mathbf{r}\rangle = -Nb|\mathbf{r}\rangle = \mathbf{E}_0|\mathbf{r}\rangle, \quad (25)$$

with ground state energy  $\mathbf{E}_0 = -Nb$ .<sup>50</sup> Note that the difference between adjacent distinct energy levels is

$$\Delta \mathbf{E} = \mathbf{E}^{(\ell+1)} - \mathbf{E}^{(\ell)} = 2b.$$

In particular, because the ground state is nondegenerate, we have,

$$\mathbf{E}_1 - \mathbf{E}_0 = 2b$$

and, more generally,

$$\mathbf{E}_j - \mathbf{E}_0 \geq 2b,$$

for  $j \neq 0$ .

Let us now proceed to compute the two terms on the right-hand-side of the free-energy shown in equation (17). From our discussion in Section 4.2 we know that

$$\hat{\rho} = \sum_{j=0}^{n-1} p_j \mathbf{P}^{(\mathbf{x}_j)} = \sum_{j=0}^{n-1} p_j |\mathbf{x}_j\rangle \langle \mathbf{x}_j|,$$

$$S(\hat{\rho}) = -\text{tr} \hat{\rho} \ln \hat{\rho} = -\sum_{j=0}^{n-1} p_j \ln p_j$$

and

$$U(\hat{\rho}) = \langle \mathbf{E} \rangle = \sum_{j=1}^{n-1} p_j \mathbf{E}_j$$

<sup>48</sup>Using the spectral representations of the hermitian operators  $\hat{\mathbf{X}}_j$  in (22) leads to Equation (3.32) of [24]. Note that the spectral representation of  $\hat{\mathbf{X}}_j$  easily follows from that of  $\sigma_x$ ,  $\sigma_x = |r\rangle\langle r| - |l\rangle\langle l|$ .

<sup>49</sup>We denote the energy of the sideways state  $|\mathbf{x}_j\rangle$  by  $\mathbf{E}_j$ . Because there are  $N + 1$  distinct energy levels (eigenvalues) and  $n = 2^N$  eigenvectors, it is evident that most of the energy levels are highly degenerate for  $N \gg 1$ .

<sup>50</sup>Similarly it is easy to show that the completely sideways vector  $|\ell\rangle$  is uniquely the most energetic state with energy  $\mathbf{E}_{n-1} = Nb$ .

with

$$p_j = \frac{e^{-\beta \mathbf{E}_j}}{Z_\beta} = \frac{e^{-\beta \mathbf{E}_j}}{e^{-\beta \mathbf{E}_0} + e^{-\beta \mathbf{E}_1} + \dots + e^{-\beta \mathbf{E}_{n-1}}} = \frac{e^{-(\mathbf{E}_j - \mathbf{E}_0)/T}}{1 + e^{-(\mathbf{E}_1 - \mathbf{E}_0)/T} + \dots + e^{-(\mathbf{E}_{n-1} - \mathbf{E}_0)/T}}. \quad (26)$$

Note that for  $j \neq 0$ ,

$$p_j \leq e^{-(\mathbf{E}_j - \mathbf{E}_0)/T} \leq e^{-2b/T} \rightarrow 0 \quad \text{as} \quad \frac{b}{T} \rightarrow \infty,$$

while

$$p_0 \rightarrow 1 \quad \text{as} \quad \frac{b}{T} \rightarrow \infty.$$

This yields,<sup>51</sup>

$$\lim_{\frac{b}{T} \rightarrow \infty} \hat{\rho} = \mathbf{P}^{(x_0)} = |\mathbf{x}_0\rangle\langle \mathbf{x}_0| = |\mathbf{r}\rangle\langle \mathbf{r}|, \quad (27)$$

$$\lim_{\frac{b}{T} \rightarrow \infty} U(\hat{\rho}) = E_0 \quad (28)$$

and

$$\lim_{\frac{b}{T} \rightarrow \infty} TS(\hat{\rho}) = 0. \quad (29)$$

Thus for small enough temperature,  $T = T_0 \approx 0$ , and large enough transverse magnetic field,  $b \gg 1$ , we have  $p_j \approx \delta_{j0}$ , yielding a (minimum) free energy value of,

$$F_{T_0}(\hat{\rho}; 0, 1) = \underbrace{U(\hat{\rho})}_{\approx E_0} - \underbrace{T_0 S(\hat{\rho})}_{\approx 0} \approx E_0 = -Nb. \quad (30)$$

Because the ground state  $|\mathbf{r}\rangle$  is nondegenerate, this means for a large enough value of  $b/T$  the system is in the completely sideways state  $|\mathbf{r}\rangle$  with high probability ( $p_0 \approx 1$ ) once it has reached thermal equilibrium. This result is quite reasonable: because the first excited state energy is  $2b$  higher than the ground state level, a large value of  $b$  ensures a wide *energy gap* between the first two energy levels, while a small value of the temperature  $T$ ,  $T = T_0 \approx 0$ , ensures that it is unlikely that a thermal fluctuation can push the ground state through the energy gap to the next energy level.

One can arrive at Eq. (30) in a slightly different, but equivalent way, by noting the important fact that that Eq. (17) is equivalent to (see below)

$$F_T(\hat{\rho}; 0, 1) = -T \ln Z_T \quad (31)$$

where (see Eq. (8))

$$\begin{aligned} Z_T &= \text{tr} e^{-\frac{1}{T} \hat{H}} = \sum_{j=0}^{n-1} e^{-\frac{1}{T} E_j} \\ &= e^{-\frac{1}{T} E_0} \sum_{j=0}^{n-1} e^{-\frac{1}{T} (E_j - E_0)} \\ &= e^{-\frac{1}{T} E_0} \left( 1 + \sum_{j=1}^{n-1} e^{-\frac{1}{T} (E_j - E_0)} \right). \end{aligned}$$

<sup>51</sup>For fixed  $b$ , Eq. (27) shows that, with probability one, the equilibrium state is the (assumed unique) ground state in the limit that  $T \rightarrow 0$ .

Therefore,

$$F_T(\hat{\rho}; 0, 1) = E_0 - T \ln \left( 1 + \sum_{j=1}^{n-1} e^{-\frac{1}{T}(E_j - E_0)} \right). \quad (32)$$

Thus<sup>52</sup>

$$\lim_{T \rightarrow 0} F_T(\hat{\rho}; 0, 1) = E_0. \quad (33)$$

Note that this result is consistent with Eq. (30).

It is straightforward to derive (31) from Eq. (17). Simplifying the notation slightly, we have

$$F_T = U - TS = \text{tr } \hat{\rho} \hat{H} + T \text{tr } \hat{\rho} \ln \hat{\rho} = T \left( \text{tr } \hat{\rho} \ln \hat{\rho} + \beta \text{tr } \hat{\rho} \hat{H} \right).$$

But

$$\ln \hat{\rho} = \ln \left[ (Z^{-1} I) e^{-\beta \hat{H}} \right] = -(\ln Z) I - \beta \hat{H},$$

so that

$$\text{tr } \hat{\rho} \ln \hat{\rho} = -(\ln Z) \underbrace{\text{tr } \hat{\rho}}_{=1} - \beta \text{tr } \hat{\rho} \hat{H}$$

proving Eq. (31).

## 7 Some Final Comments

Once the system has thermally relaxed to the initialization state  $|\mathbf{r}\rangle$  under the influence of the disorder Hamiltonian  $\hat{H}_D$ , the system can be adiabatically evolved to obtain a minimizing state to the problem Hamiltonian  $\hat{H}_P$  as sketched in Section 5 and described in much more detail in references [24, 22, 18].

A general quantum system has a time-dependent Hamiltonian. Unfortunately, it is usually difficult to analyze and shape the behavior of a multivariate, time-dependent linear system.<sup>53</sup> However, the situation is much more benign for time invariant systems. A quantum system that has a Hamiltonian that is independent of time, which we refer to as *stationary*, is an example of such a system. As is evident from the discussion on pages 119-124 of [26] and Section III.B of [24], a stationary quantum system has the property that if it is prepared to be in an energy eigenstate, then it remains in that eigenstate forever. In particular, this is true if the system is prepared in a ground energy eigenstate. Because of the difficulty of a general analysis of nonstationary quantum systems, it is natural to ask if there are “nice” circumstances for which this property is preserved. The *Adiabatic Theorem* says that 1) if a nonstationary quantum system is prepared in a ground state; 2) the instantaneous energy gap between the ground and first excited state is “large enough”; and 3) the time-dependent Hamiltonian varies slowly enough (“adiabatically” in the parlance of physicists) then the system remains in its ground state as it evolves in time. This is the rationale for the procedure described in Section 5 above, and much more thoroughly in references [24]

<sup>52</sup>Recall that we are assuming that the ground state is nondegenerate. Reference [7] refers to results (27) and (33) as the *Feynman-Kac Theorem* (see Theorem 9.4 of [7]); however a Google search will show that this term is more often used to refer to an alternative theorem by Feynman and Kac.

<sup>53</sup>This is a general fact that extends beyond the domain of quantum mechanics. For example, the theory of control of linear time invariant (LTI) state-space systems is quite mature, as is evidenced by the many LTI control theory textbooks which exist, whereas the theory of control for general time-varying linear systems is much less well-developed.

and [18] which cite the foundational literature in AQC and QA. Proofs and discussions of the Adiabatic Theorem can be found in Chapter 10 of [5] and in [23], pages 346-348.

Finally, many examples of optimization problems that can be solved using Quantum Annealing and the construction of associated problem Hamiltonians are given in the report [24] and the monograph [18], which the reader should carefully study, as well as the references cited therein.

## Appendix A. Nonuniqueness of Admissible Factorizations of the Density Matrix

We will show the nonuniqueness of factorizations of the density operator on a finite-dimensional Hilbert space of dimension  $n$  following the discussion given in [2]. In the discussion we will interchangeably represent vectors and their duals respectively by<sup>54</sup>

$$\phi \iff |\phi\rangle \quad \text{and} \quad \phi^* \iff \langle\phi|.$$

The general form of the density matrix is

$$\hat{\rho} = v_1 \hat{P}^{(1)} + \dots + v_r \hat{P}^{(m)} = \sum_{j=1}^m v_j P^{(j)} = \sum_{j=1}^m v_j \phi_j \phi_j^*, \quad (34)$$

with  $P^{(j)} = \phi_j \phi_j^*$  and, in general,  $m \neq n$ . The conditions are imposed that

$$\|\phi_j\|^2 = \phi_j^* \phi_j = \langle\phi_j|\phi_j\rangle = 1, \quad j = 1, \dots, m \quad (35)$$

and

$$v_j > 0, \quad j = 1, \dots, m \quad \text{with} \quad v_1 + \dots + v_m = 1. \quad (36)$$

Note that  $\phi_j, j = 1, \dots, m$  are not required to be orthogonal. If  $\phi_j$  is represented by a column vector, then  $\phi_j^* = \phi_j^H$  is a row vector and  $P^{(j)} = \phi_j \phi_j^H$  and  $\hat{\rho}$  are  $n \times n$  matrices. We will see below that  $m$  can be no smaller than the rank of  $\hat{\rho}$ .

A density matrix  $\hat{\rho}$  represents the state of a quantum mechanical system, and is a self-adjoint (hermitian), positive operator of trace one,

$$\hat{\rho} = \hat{\rho}^* \geq 0, \quad \text{tr} \hat{\rho} = 1,$$

where here  $\hat{\rho}^*$  denotes the adjoint of  $\hat{\rho}$ . The rank of  $\hat{\rho}$  is

$$r = \text{rank} \rho \leq n,$$

where, in general,  $r \neq n$  and  $r \neq m$ .<sup>55</sup> As mentioned previously, the factorization

$$\hat{\rho} = \sum_{j=1}^m v_j \phi_j \phi_j^* \quad (37)$$

<sup>54</sup>If  $|\phi\rangle = \phi$  is represented by a finite-dimensional column vector,  $\phi \in \mathcal{H} = \mathbb{C}^n$  then  $\langle\phi| = \phi^*$  is given by the hermitian transpose,  $\phi^* = \phi^H \in \mathcal{H}^* \sim \mathbb{C}^n$ , which is a finite-dimensional row vector. (Note that the finite dimensional Hilbert space is self-dual,  $\mathcal{H}^* \sim \mathcal{H}$ .) In this case, the dyad (outer product) is a rank one  $n \times n$  matrix,  $|\phi\rangle\langle\phi| = \phi\phi^* = \phi\phi^H \in \mathbb{C}^{n \times n}$ .

<sup>55</sup>Note that we are carefully distinguishing between the dimension of the Hilbert space,  $n$ , the number of terms used to prepare the state  $\hat{\rho}$ ,  $m$ , and the rank of  $\hat{\rho}$ ,  $r$ .

can be interpreted as a procedure for preparing the system into the state  $\hat{\rho}$ . We say “a procedure” rather than “the procedure” as the factorization shown on the right-hand-side is not unique. In fact, there are an infinity of such factorizations and hence an infinity of possible procedures for preparing the system into the same quantum mechanical state  $\hat{\rho}$ .

Because  $\hat{\rho}$  of (37) is positive and hermitian it has a spectral representation which yields an alternative *canonical decomposition*

$$\hat{\rho} = \sum_{i=1}^r w_i \psi_i \psi_i^*, \quad (38)$$

where, as defined earlier,  $r = \text{rank } \hat{\rho}$ . The eigenvectors of the spectral representation,  $\psi_i$ ,  $i = 1, \dots, r$ , are orthonormal,

$$\psi_i^* \psi_j = \langle \psi_i | \psi_j \rangle = \delta_{ij}.$$

and the corresponding eigenvalues  $w_i$  satisfy,

$$w_j > 0, \quad j = 1, \dots, r \quad \text{with} \quad w_1 + \dots + w_r = 1. \quad (39)$$

We proceed to show that given a density matrix<sup>56</sup>  $\hat{\rho}$ , which must have a canonical representation (38), then there are an infinity of alternative factorizations of the form (37).

## A.1 Special Case: Degeneracy of the Canonical Factorization

Before we deal with the general case, we describe an important special case when some of the eigenvalues of the right-hand-side of (38) are degenerate. If (say)

$$\underbrace{w_\ell = w_{\ell+1} = \dots = w_{\ell+g-1}}_{\text{these eigenvalues are } g\text{-fold degenerate}},$$

then the associated eigenvectors  $\psi_\ell, \dots, \psi_{\ell+g-1}$  span a  $g$ -dimensional eigenspace. We can use any alternative orthonormal set of vectors that span this eigenspace<sup>57</sup> in lieu of  $\psi_i$ ,  $i = \ell, \dots, \ell + g - 1$ . Let us call these alternative eigenvectors  $\psi'_i$ ,  $i = \ell, \dots, \ell + g - 1$ . Then,

$$\hat{\rho} = \sum_{i=1}^{\ell-1} w_i \psi_i \psi_i^* + w_\ell \sum_{i=\ell}^{\ell+g-1} \psi'_i \psi_i'^* + \sum_{i=\ell+g}^r w_i \psi_i \psi_i^*. \quad (40)$$

The most extreme case of this is when  $\ell = 0$  and  $g = r = n$ , in which case,

$$\hat{\rho} = \frac{1}{n} \sum_{i=1}^n \psi_i \psi_i^* = \frac{1}{n} I.$$

A brief discussion of this latter case is given on page 207 of [25].

<sup>56</sup>I.e., give a positive, hermitian operator  $\hat{\rho}$  with trace  $\hat{\rho} = 1$ .

<sup>57</sup>Note that there are an infinite number of possibilities

## A.2 The General Case

Given a density matrix  $\hat{\rho}$  we have,

$$\hat{\rho} = \underbrace{\sum_{i=1}^r w_i \psi_i \psi_i^*}_{F_1} = \underbrace{\sum_{j=1}^m v_j \phi_j \phi_j^*}_{F_2}, \quad (41)$$

where  $F_1$  is the canonical (spectral) factorization of  $\hat{\rho}$ , which we assume is at hand, and  $F_2$  is an alternative, general admissible factorization. We will show that with  $F_1$  at hand we can produce an infinity of factorizations of the form  $F_2$ .

It is evident from (41) that the range of  $\hat{\rho}$  satisfies,

$$\mathcal{R}(\hat{\rho}) = \text{span} \{\psi_1, \dots, \psi_r\} = \text{span} \{\phi_1, \dots, \phi_m\} \subset \mathcal{H}.$$

Furthermore, since  $r = \text{rank } \hat{\rho} = \dim \mathcal{R}(\hat{\rho})$ , we must have  $m \geq r$ , as claimed earlier, and that the orthogonal projection operator onto  $\mathcal{R}(\hat{\rho})$  is given by,

$$P = P^{\mathcal{R}(\hat{\rho})} = \sum_{i=1}^r \psi_i \psi_i^*.$$

Since  $\phi_j \in \mathcal{R}(\hat{\rho})$ , we have  $\phi_j = P\phi_j$ . Thus, for  $j = 1, \dots, m$ ,

$$\phi_j = P\phi_j = \sum_{i=1}^r \psi_i \psi_i^* \phi_j = \sum_{i=1}^r b_{ij} \psi_i = [\psi_1 \ \dots \ \psi_r] \begin{pmatrix} b_{1j} \\ \vdots \\ b_{rj} \end{pmatrix} = \Psi \mathbf{b}_j \quad (42)$$

where

$$b_{ij} = \psi_i^* \phi_j = \langle \psi_i | \phi_j \rangle, \\ \mathbf{b}_j^T = (b_{1j} \ \dots \ b_{rj}) \in \mathbb{C}^r$$

and

$$\Psi = [\psi_1 \ \dots \ \psi_r].$$

The column vector  $\mathbf{b}_j$  is the representation of  $\phi_j$  in the  $\psi_i$ -basis for  $\mathcal{R}(\hat{\rho})$ . Note that

$$\Psi^* \Psi = \begin{bmatrix} \psi_1^* \\ \vdots \\ \psi_r^* \end{bmatrix} [\psi_1 \ \dots \ \psi_r] = I_{r \times r} \quad (43)$$

because  $\psi_i$  are orthonormal. Equations (35), (42), and (43) give

$$1 = \|\phi_j\|^2 = \phi_j^* \phi_j = \mathbf{b}_j^H \Psi^* \Psi \mathbf{b}_j = \mathbf{b}_j^H \mathbf{b}_j$$

or

$$1 = \|\mathbf{b}_j\|^2 = \sum_{i=1}^r |b_{ij}|^2, \quad j = 1, \dots, m. \quad (44)$$

Also note that setting

$$\Phi = [\phi_1 \ \cdots \ \phi_m] = \Psi (\mathbf{b}_1 \ \cdots \ \mathbf{b}_m),$$

we have

$$\Phi = \Psi \mathbf{B}$$

where

$$\mathbf{B} = (\mathbf{b}_1 \ \cdots \ \mathbf{b}_m) \in \mathbb{C}^{r \times m} \quad \text{with} \quad (\mathbf{B})_{ij} = b_{ij}.$$

Note that  $\mathbf{B}$  is  $r \times m$  with normalized columns,  $m \geq r$ ,  $\text{rank } \mathbf{B} = r$ , and  $\mathbf{B}^* = \mathbf{B}^H$ .

Let  $W = \text{diag}(w_1 \ \cdots \ w_r)$  subject to (39) and  $V = \text{diag}(v_1 \ \cdots \ v_m)$  subject to (36). Then the consistency condition  $F_1 = F_2$  becomes,

$$\Psi W \Psi^* = \Phi V \Phi^*,$$

or

$$\Psi W \Psi^* = \Psi \mathbf{B} V \mathbf{B}^H \Psi^*.$$

Finally, using the fact that  $\Psi^* \Psi = I$ , we obtain our desired consistency condition as the matrix equation,

$$W = \mathbf{B} V \mathbf{B}^H. \quad (45)$$

With the above results in hand, we can create an infinity of consistent factorizations of the form  $F_2$  for a density matrix  $\hat{\rho}$  as follows:

1. Find  $r$  orthonormal eigenvectors and eigenvalues,  $w_i, \psi_i$ , of  $\hat{\rho}$ . I.e., determine the canonical factorization  $\hat{\rho} = F_1$  (see (41)). Construct  $W = \text{diag}(w_1 \ \cdots \ w_r)$ , which necessarily satisfies (39).
2. Find  $m$  unit vectors  $\phi_j, j = 1, \dots, m$ , that span  $\mathcal{R}(\hat{\rho})$ ,  $m \geq r = \text{rank } \hat{\rho}$ . This can done via any procedure that produces  $m$  spanning vectors  $\phi_i$  that satisfy  $P^{\mathcal{R}(\hat{\rho})} \phi_j = \phi_j$ . Our goal is to select these unit vectors such that Step 4 below is successful.
3. Compute the elements of the matrix  $\mathbf{B}$ ,  $b_{ij} = \psi_i^* \phi_j = \langle \psi_i | \phi_j \rangle$ . Note the requirement that conditions (44) hold.
4. Determine  $v_j, i = 1, \dots, m$ , by solving the matrix consistency condition (45) for a solution

$$V = \text{diag}(v_1 \ \cdots \ v_m).$$

*Note the requirement that the solution satisfies conditions (36). If this cannot be done, go to Step 2 and make adjustments to the unit vectors  $\phi_j$  (which will then produce adjustments to the values of the elements,  $b_{ij}$ , of  $\mathbf{B}$ ).*

Note that Steps (2), (3), and (4) are tightly coupled and must typically be dealt with simultaneously.

Once Step 4 has been successfully completed we will have produced a valid factorization,

$$\hat{\rho} = \sum_{j=1}^m v_j \phi_j \phi_j^*$$

which means that we will have effectively produced an admissible preparation procedure for placing the system into the state  $\hat{\rho}$  which is generally different from that provided by the canonical spectral factorization of the state.

Note that a solution  $V = \text{diag}(v_1 \cdots v_m)$  to the matrix consistency condition (45) must satisfy the diagonal conditions,

$$w_i = \sum_{k=1}^m v_k |b_{ik}|^2, \quad i = 1, \dots, r, \quad (46)$$

and the off-diagonal conditions,<sup>58</sup>

$$0 = \sum_{k=1}^m v_k b_{ik} \bar{b}_{\ell k}, \quad \ell \neq i, \quad i, \ell = 1, \dots, r. \quad (47)$$

An example of the solution procedure outlined above is given on page 10 of [2].<sup>59</sup>

## References

- [1] Daniel J Amit. *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press, 1992.
- [2] E.G. Beltrametti and G. Cassinelli. *The Logic of Quantum Mechanics*. Cambridge University Press, 1981/1984.
- [3] C. Bishop, editor. *Pattern Recognition and Machine Learning*. Springer, 2007.
- [4] Herbert B Callen. *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons, 2nd edition, 1985.
- [5] D.J. Griffiths. *Introduction to Quantum Mechanics*. Prentice-Hall, second edition, 2005.
- [6] S.P. Gudder. *Stochastic Methods in Quantum Mechanics*. Dover, 1979/2005.
- [7] S.J. Gustafson and I.M. Sigal. *Mathematical Concepts of Quantum Mechanics*. Springer, 2003.
- [8] R. Harris *et al.* “Experimental Demonstration of a Robust and Scalable Flux Qubit”. *Physical Review B*, 81, 2010.
- [9] John Hertz, Anders Krogh, and Richard G Palmer. *Introduction to the Theory of Neural Computation*. Addison-Wesley, 1991.
- [10] T.L. Hill. *Statistical Mechanics – Principles and Selected Applications*. Dover, 1956/1987.
- [11] Kerson Huang. *Statistical Mechanics*. Wiley, 1987.

---

<sup>58</sup>Recall that  $W = \text{diag}(w_1 \cdots w_r)$  so that its off-diagonal terms are zero. The diagonal and off-diagonal equations shown here are Equations (2.3.1) given in [2].

<sup>59</sup>In that solution the tight coupling between Steps (2), (3), and (4) of the solution procedure is reflected in the imposed condition that

$$\frac{|b_{12}|^2}{|b_{22}|^2} = \frac{|b_{13}|^2}{|b_{23}|^2} = \frac{|\lambda_1|^2}{|\lambda_2|^2} < \frac{w_1}{w_2}$$

for  $\lambda_1$  and  $\lambda_2$  as defined in [2]. This effectively puts a constrained on the choice of the unit vectors  $\phi_j$  chosen in Step 2 of the procedure as they depend on the values of  $\lambda_1$  and  $\lambda_2$ . Note that nonetheless there still are an infinite number of choices one can make for their values that satisfy the constraint condition.

- [12] R.I.G. Hughes. *The Structure and Interpretation of Quantum Mechanics*. Harvard University Press, 1989.
- [13] M.W. Johnson *et al.* “Quantum Annealing with Manufactured Spins” (plus Supplementary Information). *Nature*, 473, 2011.
- [14] T.F. Jordan. *Quantum Mechanics in Simple Matrix Form*. Dover, 1986/2005.
- [15] S. Kirkpatrick, C. D. Gelatt Jr., and M.P. Vecchi. “Optimization by Simulated Annealing”. *Science*, 220(4598):671–680, 1983.
- [16] L.D. Landau and E.M. Lifshitz. *Quantum Mechanics – Non-Relativistic Theory*. Pergamon Press, second revised edition, 1965.
- [17] L.D. Landau and E.M. Lifshitz. *Statistical Physics*. Pergamon Press, second revised and enlarged edition, 1969.
- [18] C.C. McGeoch. *Adiabatic Quantum Computing and Quantum Annealing: Theory and Practice*. Morgan & Claypool, 2014.
- [19] D.A. McQuarrie. *Statistical Mechanics*. Harper & Row, 1976.
- [20] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, 2001.
- [21] A. Papoulis and S.U. Pillai. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, Boston, 4th edition, 2002.
- [22] G. Rose and W.G. Mcready. “An Introduction to Quantum Annealing”. *DWAVE/Technical Document 0712*, 2007.
- [23] J.J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Addison-Wesley, second edition, 2011.
- [24] V.N. Smelyanskiy *et al.* “A Near-Term Quantum Computing Approach for Hard Computational Problems in Space Exploration”. *NASA/JPL/D-Wave/USC Report, arXiv:1204.2821v2*, 2012.
- [25] R.F. Streater. *Statistical Dynamics – A Stochastic Approach to Nonequilibrium Thermodynamics*. Imperial College Press, second edition, 2009.
- [26] L. Suskind and A. Friedman. *Quantum Mechanics: The Theoretical Minimum*. Basic Books, 2014.