In contrast, for the mixed ensemble we have $\boldsymbol{\rho}^2 \neq \boldsymbol{\rho}$, and $\mathrm{Tr}\,(\boldsymbol{\rho}^2) < 1$. The decomposition $\boldsymbol{\rho}=\sum_i w_i \; |\psi^{(i)} >< \psi^{(i)}|$ is not unique, as we shall see in the examples.

EXP **Example 1**

We work out an illustrative example for spin 1/2. The linear space has dimension 2, and the density operator is represented by a (2×2) matrix. Due to hermiticity and normalization, we are left with three real independent parameters to determine $\boldsymbol{\rho}$. Those parameters can be identified with the three mean values for the average of the spin operator, $[S_x]\,, [S_y]\,, [S_z]$. This is a particular characteristic of spin 1/2. For convenience, we introduce the Pauli spin operator $\overrightarrow{\boldsymbol{\sigma}}$ by

$$\overrightarrow{\mathbf{S}}=\frac{\hbar}{2}\overrightarrow{\boldsymbol{\sigma}} \;,$$

with the standard representation:

$$\boldsymbol{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \;, \qquad \boldsymbol{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \;, \qquad \boldsymbol{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \;. \tag{1.12}$$

We note that the three Pauli matrices plus the identity form a basis of the linear space of the complex (2×2) matrices. So, in general we have

$$\boldsymbol{\rho} = \frac{1}{2}m_0\;\mathbf{1}+\frac{1}{2}m_x\boldsymbol{\sigma}_x + \frac{1}{2}m_y\boldsymbol{\sigma}_y + \frac{1}{2}m_z\boldsymbol{\sigma}_z = \frac{1}{2}m_0\;\mathbf{1}+\frac{1}{2}\overrightarrow{\mathbf{m}}\cdot\overrightarrow{\boldsymbol{\sigma}} \;, \tag{1.13}$$

with the vector $\overrightarrow{\mathbf{m}}= (m_x,m_y,m_z)$ being called *polarization*. Due to hermiticity of $\boldsymbol{\rho}$ and the Pauli matrices, all the coefficients in (1.13) are real. Also, remembering that Pauli matrices are traceless, the normalization condition reads:

$$\mathrm{Tr}\left(\boldsymbol{\rho}\right) = 1 = \frac{1}{2} m_0 \ \mathrm{Tr}\ \left(\mathbf{1}\right) = m_0 \ ,$$

and we are left with three real independent parameters given by $\overrightarrow{\mathbf{m}}= (m_x,m_y,m_z)$, with the density operator written in the general form:

$$\boldsymbol{\rho} = \frac{1}{2}\left(\mathbf{1} + \overrightarrow{\mathbf{m}} \cdot \overrightarrow{\boldsymbol{\sigma}}\right) \ .$$

Using the algebra associated with the Pauli matrices, one easily finds that

$$\boldsymbol{\rho}^2 = \frac{1}{4}\left[\mathbf{1} + \mathbf{2}\overrightarrow{\mathbf{m}} \cdot \overrightarrow{\boldsymbol{\sigma}} + \left(\overrightarrow{\mathbf{m}} \cdot \overrightarrow{\boldsymbol{\sigma}}\right)^2\right] = \frac{1}{2}\left[\left(\frac{1+m^2}{2}\right)\mathbf{1} + \overrightarrow{\mathbf{m}} \cdot \overrightarrow{\boldsymbol{\sigma}}\right] \ ,$$

with $m^2 = \left|\overrightarrow{\mathbf{m}}\right|^2 = m_x^2 + m_y^2 + m_z^2$. Taking the trace, we obtain

$$\mathrm{Tr}\left(\boldsymbol{\rho}^2\right) = \frac{1+m^2}{2} \leqq 1 \ ,$$

implying that $m^2 \leqq 1$, or $0 \leqq \left|\overrightarrow{\mathbf{m}}\right| \leqq 1$. We have a pure ensemble if and only if $m^2 = 1$, which means maximum polarization. For the mixed ensemble, $0 \leqq m^2 < 1$. The case $m = 0$ is called unpolarized or random ensemble. Using the anticommuting properties of Pauli matrices:

$$\boldsymbol{\sigma}_x\boldsymbol{\sigma}_y = -\boldsymbol{\sigma}_y\boldsymbol{\sigma}_x = i\boldsymbol{\sigma}_z \ ,$$

$$\boldsymbol{\sigma}_y\boldsymbol{\sigma}_z = -\boldsymbol{\sigma}_z\boldsymbol{\sigma}_y = i\boldsymbol{\sigma}_x \ ,$$

$$\boldsymbol{\sigma}_z\boldsymbol{\sigma}_x = -\boldsymbol{\sigma}_x\boldsymbol{\sigma}_z = i\boldsymbol{\sigma}_y \ ,$$

it follows

$$\begin{aligned}
[\mathbf{S}_x] &= \mathrm{Tr}\,(\boldsymbol{\rho}\mathbf{S}_x) = \frac{\hbar}{2}m_x \ , \\
[S_y] &= \mathrm{Tr}\left(\boldsymbol{\rho}\mathbf{S}_y\right) = \frac{\hbar}{2}m_y \ , \\
[S_z] &= \mathrm{Tr}\,(\boldsymbol{\rho}\mathbf{S}_z) = \frac{\hbar}{2}m_z \ .
\end{aligned}$$

For the random ensemble, $[S_x] = [S_y] = [S_z] = 0$, that is $m = 0$ and the density operator is written as:

$$\boldsymbol{\rho}_0 = \frac{1}{2}\mathbf{1} \doteq \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \ . \tag{1.14}$$

We will discover that this case corresponds to the maximal mixing. Suppose that we use the basis of states that diagonalize $\mathbf{S}_z$. We call them $|\hat{\mathbf{z}}; + >$ and $|\hat{\mathbf{z}}; - >$, corresponding to the eigenvalues $\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$ respectively. The density operator for the random case in (1.14) can be represented in the form

$$\begin{aligned}
\boldsymbol{\rho}_0 &= \frac{1}{2}|\hat{\mathbf{z}}; + >< \hat{\mathbf{z}}; +| + \frac{1}{2}|\hat{\mathbf{z}}; - > < \hat{\mathbf{z}}; - \,| \ = \\
&= w_+|\hat{\mathbf{z}}; + >< \hat{\mathbf{z}}; +| + w_-|\hat{\mathbf{z}}; - > < \hat{\mathbf{z}}; - \,| \ ,
\end{aligned}$$

with $w_+ = w_- = \frac{1}{2}$, *i.e.* $\boldsymbol{\rho}_0$ can be considered as a mixing of the states $|\hat{\mathbf{z}}; + >$ and $|\hat{\mathbf{z}}; - >$ with equal weighs. One important issue for the mixed ensemble, is that the decomposition in terms of pure ensembles is not unique. In the example above, we can consider a different basis of states, let's say the states $|\hat{\mathbf{x}}; + >$ and $|\hat{\mathbf{x}}; - >$ that diagonalize the component $\mathbf{S}_x$ of the spin operator. We have the unitary transformation:

$$
\begin{aligned}
|\hat{\mathbf{z}}; + > &= \frac{1}{\sqrt{2}}|\hat{\mathbf{x}}; + > + \frac{1}{\sqrt{2}}|\hat{\mathbf{x}}; - > \ , \\
|\hat{\mathbf{z}}; - > &= \frac{1}{\sqrt{2}}|\hat{\mathbf{x}}; + > - \frac{1}{\sqrt{2}}|\hat{\mathbf{x}}; - > \ ,
\end{aligned}
$$

that leads to

$$
\boldsymbol{\rho}_0 = \frac{1}{2}|\hat{\mathbf{x}}; + >< \hat{\mathbf{x}}; +| + \frac{1}{2}|\hat{\mathbf{x}}; - > < \hat{\mathbf{x}}; - \ | \ ,
$$

which means that the random ensemble (1.14) may be considered, at the same time, as a mixing of states $|\hat{\mathbf{x}}; + >$ and $|\hat{\mathbf{x}}; - >$ with equal weighs. Actually, there is an infinite number of possibilities, saying that the random mixed ensemble can be decomposed equally in terms of black and white, or red and green, or blue and yellow, and so on, at the same time. This fact is a manifestation of the quantum nature of the state, in spite of the maximal mixture. The pure ensemble can be parameterized as:

$$\overrightarrow{\mathbf{m}}= (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \ ,$$

with $|\tilde{\mathbf{m}}|^2 = 1$, with the angles (θ, φ) giving the direction of the polarization. In matrix form, the density operator reads:

$$\boldsymbol{\rho} = \begin{pmatrix} \frac{1+\cos\theta}{2} & \frac{e^{-i\varphi}\sin\theta}{2} \\ \frac{e^{i\varphi}\sin\theta}{2} & \frac{1-\cos\theta}{2} \end{pmatrix} \ .$$

For many other examples, see the exercise section. ■

A

1.3 Coupled systems and non-separability

Two quantum systems that interacted in the past, remain correlated forever and cannot be represented separately by ket states. In plain words, a subsystem can not in general, be described by a wave function. This feature is called the *Non-separability Property.* This is a limitation of the wave function formalism, since in Statistical Mechanics, one usually wants to refer to a part of the whole system, independently of the rest. Important cases are those when the system is coupled to *reservoirs*, and we want to eliminate the latter degrees of freedom from the physical description. Those are instances where the density operator furnishes a superior view of the state of the system, since the wave function representation has been ruled out. To fully understand the nonseparable case, we will firstly discuss the case of non-interacting systems, that is systems that are separable.

A diluted system, as for example an *ideal gas*, can be well approximated as a system of many non-interacting particles (molecules). Within this ideal limit, particles are not correlated, and the Statistical Mechanics can be reduced to a *one-particle* description. In a general way, let us consider two systems which span two different Hilbert spaces of ket states, $\mathcal{R}$ and $\mathcal{S}$, with basis $\{|N>\}$ and $\{|n>\}$, respectively. The combined space of the two systems, $\mathcal{R} \times \mathcal{S}$, is spanned by the direct product states, which in the Dirac's notation, are written as:

$$|Nn> \equiv |N> |n> \quad .$$

Direct products of operators are represented by direct products of the corresponding matrices.

Example. Direct product of two matrices A and B, where:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} , \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \quad .$$

One defines $A \times B$ as the matrix:

$$A \times B \equiv \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix} \quad .$$

The direct product results in a (6×6) matrix, since the terms $a_{ij}B$ are understood as (3×3) submatrices.

The following properties are easily demonstrated:

1. Matrix elements are understood as:

$$< Nn|\mathbf{A} \times \mathbf{B}|N'n' >=< N|\mathbf{A}|N' >< n|\mathbf{B}|n' > \ .$$

   In particular, scalar products are obtained as:

$$< Nn|N'n' >=< N|N' >< n|n' >= \delta_{NN'}\delta_{nn'} \ .$$

2. From the point above, we get:

$$\mathrm{Tr}\,(\mathbf{A} \times \mathbf{B}) = \mathrm{Tr}\,(\mathbf{A})\,\mathrm{Tr}(\mathbf{B}) \ .$$

3. Direct products of column (row) vectors are obtained using the same rule defined for matrices. Take for example the case below:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ 0 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ 0 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \ .$$

A general state of the system is written as:

$$|\psi >= \sum_{N,n} \mathfrak{C}(N,n)\ |N > |n > \ ,$$

but since the systems are uncorrelated, one must have $\mathfrak{C}(N,n) = C(N)c(n)$ and the state is separable:

$$|\psi >= \sum_{N,n} \mathfrak{C}(N,n)\ |N > |n >= \left(\sum_{N} C\left(N\right)|N > \right) \left(\sum_{n} c\left(n\right)|n > \right) = |\psi_R > |\psi_S > \ .$$

The same is true for the density matrix, which in general is written as

$$\boldsymbol{\rho} = \sum_{N,M,n,m} |Nn >< Nn|\rho|Mm >< Mm|\ ,$$

but for uncorrelated systems, one should have

$$< Nn|\rho|Mm >= A_{NM}B_{nm}\ ,$$

factorizing the density operator as:

$$\boldsymbol{\rho} = \left(\sum_{N} A_{NM}|N >< M| \right) \left(\sum_{n} B_{nm}|n >< m| \right) = \boldsymbol{\rho}_{\mathcal{R}} \times \boldsymbol{\rho}_{\mathcal{S}}\ , \qquad (1.15)$$

in the form of a direct product. Separability for the density operator has a broader sense in Quantum Mechanics, but we shall not pursue this discussion here [4]. A situation as the one depicted in (1.15), is called *simple separability.* In the case of an ideal gas, in the absent of interactions, all the particles are uncorrelated. If we use the particle coordinates as labels, separability leads to:

$$\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = \boldsymbol{\rho}_1(\mathbf{x}_1) \times \boldsymbol{\rho}_2(\mathbf{x}_2) \times ... \times \boldsymbol{\rho}_N(\mathbf{x}_N)\ ,$$

where N is the total number of particles and $\boldsymbol{\rho}_i(\mathbf{x}_i)$ is the density operator for one-particle states. If the particles are identical,

$$\boldsymbol{\rho}_j = \boldsymbol{\rho}_1,$$

for all $j = 2, 3, ..., N$, and the factorization is written as

$$\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N) = \bigotimes_{i=1}^{N} \ \boldsymbol{\rho}_1(\mathbf{x}_i) \ .$$

In particular, if $\boldsymbol{\rho}_1$ is normalized, we obtain:

$$\mathrm{Tr}\left[\boldsymbol{\rho}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)\right] = \prod_{i=1}^{N} \mathrm{Tr}\left[\boldsymbol{\rho}_1(\mathbf{x}_i)\right] = 1 \ .$$

In this case, the calculation of the density operator is reduced to the calculation of the one-particle operator $\boldsymbol{\rho}_1(\mathbf{x})$.

In the interacting case, correlations among the particles appear, and the states are not any more separable. Assume that at the initial time ($t_0 = 0$), the two subsystem are not interacting, and we prepare the initial state as separable:

$$|\psi, 0 >= |N > |n > \ .$$

After that, the interaction is turned on during a finite interval of time, and finally turned off again. The interaction made the systems to be correlated. The asymptotic state ($t \to \infty$) is of the type:

$$|\psi, \infty >= \sum_{M,m} a_\infty(M, m; N, n)|M > |m > \ , \tag{1.16}$$

where the coefficient $a_\infty(M, m; N, n)$ is the probability amplitude for the transition $|N>|n> \to |M>|m>$. Since the systems are correlated, we have in general that

$$a_\infty(M, m; N, n) \neq C(M, N)c(m, n) \ ,$$

for all pairs (M, m) and (N, n), and the state (1.16) is not separable. In this case, we cannot assign a wave function to a subsystem (either R or S). This result is called *nonseparability principle.*

A

1.4 Density matrix of a subsystem

We have seen in the previous subsection that the language of ket states does not allow us, in general, to describe an isolated system from the remainder of the universe. But this is possible when one describes the state of the system through the density operator. Consider two interacting quantal systems, whose states span the spaces R and S. We want to pay attention to subsystem S, leaving R undetected (R may be a reservoir, and we want to eliminate its degrees of freedom). The states of the total system span $R \times S$, but in general, physical states are not separable in the presence of interactions. Assume basis $\{|N>\}$ and $\{|n>\}$ for R and S, respectively, as in the subsection above. We want to calculate an average of an observable that refers only to S, which is written in the form

$$\mathbf{\Omega} = \mathbf{I}_R \times \mathbf{\Omega}_S \ ,$$

where I_R is the identity in R:

$$[\mathbf{\Omega}] = \mathrm{Tr}\left[\boldsymbol{\rho}\left(\mathbf{I}_R \times \mathbf{\Omega}_S\right)\right] = \sum_{N,M,n,m} < Nn|\rho|Mm >< Mm|\left(\mathbf{I}_R \times \mathbf{\Omega}_S\right)|Nn > \ . \tag{1.17}$$

Note that $< Mm|\left(\mathbf{I}_R \times \mathbf{\Omega}_S\right)|Nn >= \delta_{MN} < m|\mathbf{\Omega}_S|n >$, and substituting in (1.17) yields:

$$[\mathbf{\Omega}] = \sum_{n,m} < m|\mathbf{\Omega}_S|n > \left(\sum_N < Nn|\boldsymbol{\rho}|Nm >\right) \ .$$

The quantity $\sum_N < Nn|\boldsymbol{\rho}|Nm >$ is the partial trace of $\boldsymbol{\rho}$ in the R space. We then make the following definition:

Definition 5 *Reduced density operator, $\bar{\boldsymbol{\rho}}_S$, relative to S.*

Its matrix elements in S are given by:

$$< n|\bar{\boldsymbol{\rho}}_S|m > \equiv \sum_N < Nn|\boldsymbol{\rho}|Nm > \ . \tag{1.18}$$

We rewrite the above definition (1.18) in a formal fashion as:

$$\bar{\boldsymbol{\rho}}_S = \mathrm{Tr}_R\left(\boldsymbol{\rho}\right) \ ,$$

meaning that $\bar{\boldsymbol{\rho}}_S$ is obtained from $\boldsymbol{\rho}$ by taking the partial trace in R. The average value (1.17) is now referred to space S only:

$$[\mathbf{\Omega}] = \mathrm{Tr}\left[\boldsymbol{\rho}\left(\mathbf{I}_R \times \mathbf{\Omega}_S\right)\right] = \mathrm{Tr}_S\left[\bar{\boldsymbol{\rho}}_S \mathbf{\Omega}_S\right] \ .$$

By taking the partial trace, we lose the detailed information relative to subsystem R. It remains to prove that $\bar{\boldsymbol{\rho}}_S$ is a *bona fide* density operator. This is achieved, if we assume that $\boldsymbol{\rho}$ is a density operator for the whole universe $R \times S$:

i) hermiticity,

$$< n|\bar{\boldsymbol{\rho}}_S|m >\equiv \sum_N < Nn|\boldsymbol{\rho}|Nm >= \sum_N < Nm|\boldsymbol{\rho}|Nn >^* =< m|\bar{\boldsymbol{\rho}}_S|n >^* ,$$

because $\boldsymbol{\rho}$ is Hermitian;

ii) normalization,

$$\mathrm{Tr}_S\left(\bar{\boldsymbol{\rho}}_S\right) = \sum_n < n|\bar{\boldsymbol{\rho}}_S|n >\equiv \sum_n \sum_N < Nn|\boldsymbol{\rho}|Nn >= \mathrm{Tr}_S\left[\mathrm{Tr}_R\left(\boldsymbol{\rho}\right)\right] = \mathrm{Tr}\left(\boldsymbol{\rho}\right) = 1\ ;$$

iii) positivity,

$$< n|\bar{\boldsymbol{\rho}}_S|n >\equiv \sum_N < Nn|\boldsymbol{\rho}|Nn > \geqslant 0\ ,$$

since it is a sum of positive terms.

To measure properties of the S subsystem we do not need the complete density operator $\boldsymbol{\rho}$, but only the reduced operator $\bar{\boldsymbol{\rho}}_S$ relative to S. Detailed information of the other subsystem is lost, but some ‘average properties’ of R are still contained in $\bar{\boldsymbol{\rho}}_S$.

EXP **Example 2**

Consider two interacting particles of spin 1/2, which are coupled in a singlet state of the total spin:

$$|\psi_0 >= \frac{1}{\sqrt{2}}\,|+\rangle_R\,|-\rangle_S - \frac{1}{\sqrt{2}}\,|-\rangle_R\,|+\rangle_S\ , \tag{1.19a}$$

where we have used the labels R and S for the particles. The above state is said to be *entangled*, and clearly there is no ket state to represent either subsystem R or S. The state (1.19a) is a pure state, with density operator given by:

$$\begin{aligned}\boldsymbol{\rho} &= |\psi_0 >< \psi_0| = \\ &= \frac{1}{2}\{|+\rangle_R\,|-\rangle_S\ \ {}_S\langle -|_R\,\langle +|\ +\ |-\rangle_R\,|+\rangle_S\ \ {}_S\langle +|_R\,\langle -|\ - \\ &\quad -|+\rangle_R\,|-\rangle_S\ \ {}_S\langle +|_R\,\langle -|\ -\ \ |-\rangle_R\,|+\rangle_S\ \ {}_S\langle -|_R\,\langle +|\}\ .\end{aligned}$$

We now take the partial trace relative to R:

$$\begin{aligned}\bar{\boldsymbol{\rho}}_S &= \mathrm{Tr}_R\,(\boldsymbol{\rho})\ =\ {}_R\langle +|\,\boldsymbol{\rho}\,|+\rangle_R\ +\ {}_R\langle -|\,\boldsymbol{\rho}\,|-\rangle_R = \\ &= \frac{1}{2}\,|-\rangle_S\ \ {}_S\langle -|\ +\ \frac{1}{2}\,|+\rangle_S\ \ {}_S\langle +|\ \doteq \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}_S\ . \qquad (1.20)\end{aligned}$$

Note that $\bar{\boldsymbol{\rho}}_S$ represents a mixed ensemble (maximum mixture), while the original $\boldsymbol{\rho}$ was a pure state: by eliminating the degrees of freedom of R, some information is lost in a way that cannot be recovered lately. Just to stress the irreversible character of the process, we note that partially tracing the density matrix of the pure ensemble of the triplet state

$$|\psi_1 >= \frac{1}{\sqrt{2}}\,|+\rangle_R\,|-\rangle_S + \frac{1}{\sqrt{2}}\,|-\rangle_R\,|+\rangle_S$$

leads to the same reduced density matrix (1.20). ■

A

1.5 Representing the Density Operator: Density Matrix

We discuss in first place the case of discreet spectrum, with a complete and orthonormal basis $\{|n\rangle\}$. In many instances, $\{|n\rangle\}$ is the basis that makes diagonal a complete set of observables. We expand the density operator in terms of this basis :

$$\begin{aligned}\boldsymbol{\rho} \;&=\; \sum_i w_i \;|\psi^{(i)} >< \psi^{(i)}| \;=\sum_{n,m}\sum_i w_i |n >< n|\psi^{(i)} >< \psi^{(i)}|m >< m| \;= \\ &=\; \sum_{n,m} |n >< n|\boldsymbol{\rho}|m><m| \;=\sum_{n,m} |n >< m| \left(\sum_i w_i < n|\psi^{(i)} >< \psi^{(i)}|m > \right) \;.\end{aligned}$$

Defining the linear coefficients as $a_n^{(i)} \equiv < n|\psi^{(i)} >$, *i.e.* $|\psi^{(i)} >= \sum_n a_n^{(i)}|n >$, the matrix elements of $\boldsymbol{\rho}$ are written as:

$$< n|\boldsymbol{\rho}|m>=\sum_i w_i \; a_n^{(i)} a_m^{(i)*} \equiv \overline{a_n a_m^*} \;,$$

where the bar means average over the mixed ensemble and a^* is the complex conjugate of a. We remember that $a_n^{(i)} \equiv < n|\psi^{(i)} >$ is the probability amplitude that the state $|n >$ is contained in $|\psi^{(i)} >$. The corresponding probability is $P_n^{(i)} = \left|a_n^{(i)}\right|^2 = \left|< n|\psi^{(i)} >\right|^2$, which appears in the diagonal matrix elements of $\boldsymbol{\rho}$:

$$< n|\boldsymbol{\rho}|n>=\sum_i w_i \;\; \left|a_n^{(i)}\right|^2 = \sum_i w_i \; P_n^{(i)} \;\; \geqq 0 \;,$$

which in turn, can be interpreted as a probability, since:

$$\sum_n < n|\boldsymbol{\rho}|n>=\sum_i w_i \sum_n P_n^{(i)} = \sum_i w_i = 1 \;.$$