Single Systems

Classical States and probability vectors. X will represent the system being considered, and Σ the possible states of X (so if X were a die then $\Sigma = \{1, 2, 3, 4, 5, 6\}$).
For probabilistic computations, there is a probability associated with each value. For a bit, for example, we could have $\Pr(X = 0) = 3/4$, $\Pr(X = 1) = 1/4$. This is represented by the column vector $\begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix}$. Any probabilistic state can be represented with these column vectors, called probability vectors.

Measuring probabilistic states. If we know with certainty that X is in state $a \in \Sigma$, we denote the probability vector of it (which is 0 everywhere except for 1 at a) as $|a\rangle$. For a bit, any of the column vectors can be represented as a linear combination of $|0\rangle$ and $|1\rangle$, i.e. $\begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \end{bmatrix} = 3/4|0\rangle + 1/4|1\rangle$ (basic linear algebra).
Observing a probabilistic state can convert it to a $|a\rangle$ state, for example if we flip a coin and cover the result it has a 50% chance to be heads or tails, but once we observe it, it must be one or the other.

Classical operations. Deterministic operations are functions $f : \Sigma \to \Sigma$. These can be represented by matrix-vector multiplication (since we are taking vectors to vectors this is just a linear transformation). With shorthand, we can write these matrices more easily: $\langle a|$ is the row version of $|a\rangle$, so $\langle 0| = \begin{bmatrix} 1 & 0 \end{bmatrix}$ and similarly for $\langle 1|$. Then $|0\rangle\langle 1| = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Another nice property is that $\langle a||b\rangle$ (which is written as $\langle a|b\rangle = 1 \iff a = b$ and 0 otherwise.

Probabilistic operations. We also have probabilistic operations, taking determined vectors to ambiguous probability states. For example $\begin{bmatrix} 1 & 1/2 \\ 0 & 1/2 \end{bmatrix}$ sends $|0\rangle$ to $|0\rangle$ but sends $|1\rangle$ to $\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$.
The set of all probabilistic operations is the set of stochastic matrices, which contain only nonnegative real number entries and the entries in every column sum to 1. These are the matrices which always send probability vectors to probability vectors.
We can also compose probabilistic operations. If $M_1, ..., M_n$ are stochastic matrices on the system X, then for u representing a probability vector over X, $M_2(M_1 u) = (M_2 M_1)u$ by associativity.

Quantum information. A quantum state of a system is represented by a column vector. Vectors representing quantum states have the following two properties:

1. The entries of a quantum state vector are complex numbers
2. The sum of the squares of the absolute values of these entries is 1

The Euclidean norm of $v = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$ is denoted by $||v|| = \sqrt{\sum_{k=1}^{n} |\alpha_k^2|}$. Quantum state vectors are thus unit vectors withrespect to the Euclidean norm.
A qubit is a quantum system whose classical state set is $\{0, 1\}$. One more unusual example of a qubit is $\begin{bmatrix} (1+2i)/3 \\ -2/3 \end{bmatrix}$ (the math for the norm checks out here). Certainly the traditional bit basis vectors are also qubits. These vectors are all linear combinations of the standard basis states, and these are superpositions of the states 0 and 1.
The notation $|+\rangle = 1/\sqrt{2}|0\rangle + 1/\sqrt{2}|1\rangle$, and $|-\rangle = 1/\sqrt{2}|0\rangle - 1/\sqrt{2}|1\rangle$. We can use the $|\rangle$ notation on any vectors, not just standard basis ones, usually written as $|\psi\rangle$, but in this case $\langle\psi|$ refers to taking the *conjugate* transpose of $|\psi\rangle$. We can further extend Dirac notation onto systems which have arbitrary state sets, writing something like $\frac{1}{\sqrt{385}} \sum_{k=0}^{9} (k+1)|k\rangle$, representing the 10 element column vector from 1 to 10 descending. We can also express a vector like $\frac{1}{\sqrt{\Sigma}} \sum_{a\in\Sigma} |a\rangle$, the uniform superposition over Σ.

Measuring Quantum States. When a quantum state is measured, similar to the probabilistic setting, we will see a classical state rather than a state of probabilities. If a quantum state is measured, each classical state of teh system appears with probability equal to the absolute value *squared* of the entry in the quantum state vector corresponding to that classical state. This is called the Born Rule. For example, measuring the plus state results in a 50% chance of seeing 0 and a 50% chance of seeing 1, and actually the minus state has the same result when measured (their differences are theoretical and can emerge when multiple state operations are applied on them).

Unitary operations. Quantum information has a different set of allowed operations. Rather than being represented by stochastic matrices, quantum operations are represented by unitary matrices. A matrix is unitary if $UU^{\tau} = U^{\tau}U = I$, where U^{τ} is the conjugate transpose of U (so $\overline{U^T}$. The condition that U is unitary means that multiplying by U does not change the Euclidean norm of a vector (so it will keep the second property). There are some examples of common unitary operations:

1. The Pauli operations, the four Pauli matrices are the following: $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \sigma_x \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$, and $\sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.
2. The Hadmard operation, the Hadmard operation is the following: $H = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{bmatrix}$.
3. The Phase operations, a phase operations is described by the matrix $P_\theta = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{bmatrix}$. $P_{\pi/2} = S$ and $P_{\pi/4} = T$ are quite common.

All of these matrices are unitary. The Hadmard operation is quite related to the vectors we have seen so far: $H|0\rangle = |+\rangle, H|+\rangle = |0\rangle, H|1\rangle = |-\rangle, H|-\rangle = |1\rangle$.
Compositions of these qubit unitary operations are represented by matrix multiplication as well. On larger state systems, we can have $n \times n$ matrices, where n is the number of states an element can be in, which are unitary matrices that represent transformations. One example is $U = 1/2 \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$, known as the quantum Fourier transform.

Multiple Systems

Although we can view multiple systems together as a larger single system, in which case the previous section's work all applies, we may want to separate things into multiple systems to make many computations more easily, such as measuring some parts of a system.

Classical states of multiple systems. Let (X, Σ) and (Y, Γ) be two system-state pairs. If we place systems X and Y side-by-side, we can view these two systems as the system (X, Y). This has the state set $\Sigma \times \Gamma$. In general, the classical state set of an n-tuple is just their Cartesian product. Qiskit will typically name multiple qubits in the pattern $(a_{n-1}, ..., a_0)$.
We often represent the state $(a_{n-1}, ..., a_0)$ as the string $a_{n-1} \dots a_0$ for shorthand, especially when Σ_i are alphabets. For example, if $X_0, ..., X_9$ are bits, then written as a string their states are like 0000000000 or 0000110010.

Probabilistic states. A probabilistic state of multiple systems associates a probability with each element of the Cartesian product of the classical state sets of the individual systems. An example is we could have $\Pr((X, Y) = (0, 0)) = 1/2, \Pr((X, Y) = (1, 1)) = 1/2$, and the others obviously 0. Then both X and Y have a 50% chance to be 0 and a 50% chance to be 1, but they are also always the same.
When representing probabilistic states, we used vectors in the single-system example. Now, we have to order the Cartesian products of sets in a way that is consistent. We typically do this alphabetically, so we order on the first argument of the product, then the second, and so on. With this ordering, our prior example has

probability vector $\begin{bmatrix} 1/2 \\ 0 \\ 0 \\ 1/2 \end{bmatrix}$.

Two systems can also be independent if learning the classical state of either system has no effect on the probabilities of the other (normal probablity definition of independence). With probability vectors of multiple states, the Dirac notation is written as $\sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|ab\rangle$. Correlation is therefore defined as lack of independence. Independence can be calculated in the normal probability manner.

Tensor products of vectors. Given two vectors $|\phi\rangle = \sum_{a\in\Sigma}\alpha_a|a\rangle$ and $|\psi\rangle = \sum_{b\in\Gamma}\beta_b|b\rangle$, their tensor product $|\phi\rangle\otimes|\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma}\alpha_a\beta_b|ab\rangle$. The entries of this new vector correspond to elements of the Cartesian product. Equivlaently, $|\pi\rangle = |\phi\rangle\otimes|\psi\rangle$ is defined by the equation $\langle ab|\pi\rangle = \langle a|\phi\rangle\langle b|\psi\rangle$ being true for all $a\in\Sigma, b\in\Gamma$.

Using the tensor product, we can redefine independence as the claim that if a joint system (X,Y) in a probabilistic state is reprsented by $|\pi\rangle$, then X and Y are independent if $|\pi\rangle = |\phi\rangle\otimes|\psi\rangle$. We generally omit the $\otimes$. For standard basis bectors, $|a\rangle|b\rangle = |ab\rangle$. We tend to write ab as a string rather than (a,b) over simple alphabets.

Generalizing Independece and Tensor Products. Tensor products also can be generalized along with independence to three or more systems in a relatively simple way. In the general case, the vector $|\psi\rangle = |\phi_{n-1}\rangle\otimes...\otimes|\phi_0\rangle$ is defined by the equation $\langle a_{n-1}\dots a_0|\psi\rangle = \langle a_{n-1}|\phi_{n-1}\rangle\dots\langle a_0|\phi_0\rangle$. It can also be defined equivalently as a recursive tensor product of two vectors (common induction W). The tensor product is multilinear (linear in each of its arguments).

On independnece, we don't consider much of pairwise independence, but instead only consider mutual independence. In addition, for the standard basis vectors, $|a_{n-1}\rangle\otimes...\otimes|a_0\rangle = |a_{n-1}\dots a_0\rangle$.

Measurements of probabilistic states of multiple systems. We can take both total and partial measurements. With partial measurements, we learn information that changes the overall system, but does not give us the whole picture.

Let (X,Σ) and (Y,Γ) be the classical state-set pairs. Assume we measure X and ignore Y (the opposite is symmetrical).

We know that $\Pr(X=a) = \sum_{b\in\Gamma}\Pr((X,Y)=(a,b))$. There is also the usual conditional probability formula: $\Pr(Y=b|X=a) = \frac{\Pr((X,Y)=(a,b))}{\Pr(X=a)}$. In terms of probability vectors, we can also express these formulas. Let $|\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} p_{ab}|ab\rangle$. Measuring X alone yields for each $a\in\Sigma$ the equation $\Pr(X=a) = \sum_{c\in\Gamma} p_{ac}$. The probabilistic state of X alone is given by $\sum_{a\in\Sigma}(\sum_{c\in\Gamma} p_{ac})|a\rangle$. The probabilistc state of Y is updated according to the conditional probability formula, so that it is represented by $|\pi_a\rangle = \frac{\sum_{b\in\Gamma} p_{ab}|b\rangle}{\sum_{c\in\Gamma} p_{ac}}$. If our measurment of X resulted in state a, we update our description of the system (X,Y) to $|a\rangle\otimes|\pi_a\rangle$.

$|\pi_a\rangle$ can be viewed as the normalization of the vector $\sum_{b\in\Gamma} p_{ab}|b\rangle$.

Operations on probabilistic states. We can view multiple systems as single compound systems again to derive operations on multiple systems. Any operation on two systems X and Y is represented by a stochastic matrix whose rows and columns are indexed by $\Sigma\times\Gamma$. Suppose that X and Y are bits, and consider the operation which performs a NOT on Y if X = 1, and otherwise does nothing. Then the matrix representation of this operation is $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$. Exchanging the roles of X and Y rotates the bottom 3 rows.

Another example is to perform the operations of either setting Y to be X or X to be Y, each with probability 1/2. Then the matrix representation of this operation is $\begin{bmatrix} 1 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 1 \end{bmatrix}$. This can be extended to any number of systems. If we have three bits and increment them modulo 8, we can express this operation as

$\sum_{k=0}^{7} |(k+1) \mod 8\rangle\langle k|$. (basically the thing in the left part is the thing after the transformation and the thing in the right part is the thing before).

Independent Operations. Independent operations can be defined with the tensor product of matrices.
The tensor product $M \otimes N$ of the matrices $M = \sum_{a,b\in\Sigma} \alpha_{ab}|a\rangle\langle b|$ and $N = \sum_{c,d\in\Gamma} \beta_{cd}|c\rangle\langle d|$ is the matrix $M\otimes N = \sum_{a,b\in\Sigma}\sum_{c,d\in\Gamma} \alpha_{ab}\beta_{cd}|ac\rangle\langle bd|$. Equivalently, it is defined by $\langle ab|M\otimes N|bd\rangle = \langle a|M|b\rangle\langle c|N|d\rangle$ being true for every a, b, c, d. It can also equivalently be described as the unique matrix satisfying the equation $(M \otimes N)(|\phi\rangle \otimes |\psi\rangle) = (M|\phi\rangle) \otimes (N|\psi\rangle)$ for every $|\phi\rangle, |\psi\rangle$. There is also a matrix representation that looks worse than almost any other matrix ever so yeah.
Tensor products of 3 or more matrices are defined as $\langle a_{n-1}...a_0|M_{n-1}\otimes...\otimes M_0|b_{n-1}...b_0\rangle = \langle a_{n-1}|M_{n-1}|b_{n-1}\rangle...\langle a_0|M_0|b_0\rangle$ for every choice of classical states.
For both probabilistc states and operations, tensor products respect independnece. If we have two systems X and Y that are independently in states $|\phi\rangle$ and $|\psi\rangle$, then the compound system is in probabilistic state $|\phi\rangle \otimes |\psi\rangle$, and if we apply probabilistic operatiosn M and N to the two systems independently, the resulting action is described by $M \otimes N$.

Quantum Information. Quantum states of multiple systems are represented by column vectors having complex number entries and Euclidean norm 1, and the entries of these vectors are placed in correspondence with the Cartesian product of the classical state sets associated with the individual systems. For example, if X and Y are qubits, thenn the state set of the pair of qubits is $\{0, 1\} \times \{0, 1\}$. We associate this set with $\{00, 01, 10, 11\}$. One example of a quantum state vector is $1/\sqrt{2}|00\rangle - 1/\sqrt{6}|01\rangle + i/\sqrt{6}|10\rangle 1/\sqrt{6}|11\rangle$. We may use the fact that $|ab\rangle = |a\rangle|b\rangle$ to instead write $1/\sqrt{2}|0\rangle|0\rangle$ and so on, we may explicitly write the tensor product with $1/\sqrt{2}|0\rangle \otimes |0\rangle$ and so on, we may subscript the kets with systems like $1/\sqrt{2}|0\rangle_X|0\rangle_Y$ and so on, and certainly we may also write them as column vectors $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{6} \\ i/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}$.

Tensor Products of Quantum Vectors. Suppose that $|\phi\rangle$ is a q.s.v of X and $|\psi\rangle$ is a q.s.v for Y. Then their tensor product $|\phi\rangle \otimes |\psi\rangle$ is a quantum state vector of (X,Y). The tensor product is a q.s.v because the Euclidean norm is multiplicative, so it also has Euclidean norm 1 since its factors are also of norm 1.

Entangled States. Not all q.s.v's are product states. For example, the q.s.v $1/\sqrt{2}|00\rangle + 1/\sqrt{2}|11\rangle$ is not a product state. This holds because if it were a product state, there would exists qsv's $|\phi\rangle, |\psi\rangle$ for which $|\phi\rangle \otimes |\psi\rangle = 1/\sqrt{2}|00\rangle + 1/\sqrt{2}|11\rangle$, but then $\langle 0|\phi\rangle\langle 1|\psi\rangle = \langle 01|\phi \otimes \psi\rangle = 0$, so either $\langle 0|\phi\rangle = 0$ or $\langle 1|\psi\rangle = 0$, contradicting the fact that $\langle 0|\phi\rangle\langle 0|\psi\rangle = \langle 00|\phi \otimes \psi\rangle = 1/\sqrt{2}$ and $\langle 1|\phi\rangle\langle 1|\psi\rangle = \langle 11|\phi \otimes \psi\rangle = 1/\sqrt{2}$ are both nonzero. Thus, this qsv represents a correlation between two systems, and specifically we say that the systems are entangled.
Some important multi-qubit quantum states are:

- Bell states
  $|\phi^+\rangle = 1/\sqrt{2}|00\rangle + 1/\sqrt{2}|11\rangle$, $|\phi^-\rangle = 1/\sqrt{2}|00\rangle - 1/\sqrt{2}|11\rangle$,
  $|\psi^+\rangle = 1/\sqrt{2}|01\rangle - 1/\sqrt{2}|10\rangle$, $|\psi^-\rangle = 1/\sqrt{2}|01\rangle - 1/\sqrt{2}|10\rangle$. The collection of the four Bell states is called the Bell basis
- GHZ and W states
  GHZ $= 1/\sqrt{2}|000\rangle + 1/\sqrt{2}|111\rangle$
  W $= 1/\sqrt{3}|001\rangle + 1/\sqrt{3}|010\rangle + 1/\sqrt{3}|100\rangle$. Neither of these states is a product state.

Measurements of quantum states. Suppose $X_0, ..., X_{n-1}$ are states. We can view $(X_{n-1}, ..., X_0)$ collectively as a single system whose classical state set is $\Sigma_{n-1} \times ... \times \Sigma_0$. If a quantum state of this system is represented by $|\psi\rangle$, and all of the systems are measured, then each possible outcome appears with probability $|\langle a_{n-1}...a_0|\psi\rangle|^2$.

For example, if systems X and Y are jointly in the state $3/5|0\rangle|a\rangle - (4i)/5|1\rangle|b\rangle$, then measuring both systems with standard basis gives probability 9/25 for $(0, a)$ and 16/25 for the other.

Partial Measurements. A qsv of (X, Y) takes the form $|\psi\rangle = \sum_{(a,b)\in\Sigma\times\Gamma} \alpha_{ab}|ab\rangle$, where $\{\alpha_{ab}|(a, b) \in \Sigma \times \Gamma\}$ is a collection of complex numbers such that $\sum |\alpha_{ab}|^2 = 1$ (so $|\psi\rangle$ is a unit vector. If we only measure the first system X, the probability for each outcome $a \in \Sigma$ to appear is $\sum_{b\in\Gamma} |\langle ab|\psi\rangle|^2 = \sum_{b\in\Gamma} |\alpha_{ab}|^2$. Once we measure X to be a, its quantum state becomes $|a\rangle$, but what happens to the quantum state of Y?
We can express $|\psi\rangle = \sum_{a\in\Sigma} |a\rangle \otimes |\phi_a\rangle$, where $|\phi_a\rangle = \sum_{b\in\Gamma} \alpha_{ab}|b\rangle$. Then, the probability for each outcome a when X is measured is $\sum_{b\in\Gamma} |\alpha_{ab}|^2 = |||\phi_a\rangle||^2$. Then, once X is measured to be a, the quantum state of (X, Y) becomes $|a\rangle \otimes \frac{|\phi_a\rangle}{|||\phi_a\rangle}$. $|a\rangle \otimes |\phi_a\rangle$ represents the component of $|\psi\rangle$ which is consistent with X being a.
As an example, consider $1/\sqrt{2}|00\rangle - 1/\sqrt{6}|01\rangle + i/\sqrt{6}|10\rangle 1/\sqrt{6}|11\rangle$. When X is measured, we write $|\psi\rangle = |0\rangle \otimes (1/\sqrt{2}|0\rangle - 1/\sqrt{6}|1\rangle) + |1\rangle \otimes (i/\sqrt{6}|0\rangle + 1/\sqrt{6}|1\rangle$. Based on the description above, the probability for the outcome 0 for X is $||1/\sqrt{2}|0\rangle - 1/\sqrt{6}|1\rangle||^2 = 2/3$, and then the state of (X, Y) becomes $|0\rangle \otimes (\sqrt{3}/2|0\rangle - 1/2|1\rangle)$. Similarly, the probability for X to be 1 is 1/3, and the resulting state is $|1\rangle \otimes (i/\sqrt{3}|0\rangle + 1/\sqrt{2}|1\rangle$.
The two-system solution can be extended to three or more systems by dividing the systems into two collections, those that are measured and those that are not, and we can then treat this almost identically to the two-state case. There is a worked example but it looks quite annoying to write down, check the website for review.

Unitary operations. Unitary matrices over the Cartesian products of the state sets are quantum operations on compound systems. If we let $\Sigma = \{1, 2, 3\}$ and $\Gamma = \{0, 1\}$, then one unitary matrix operating on (X, Y) is $\begin{bmatrix} 1/2 & 1/2 & 1/2 & 0 & 0 & 1/2 \\ 1/2 & i/2 & -1/2 & 0 & 0 & -i/2 \\ 1/2 & -1/2 & 1/2 & 0 & 0 & -1/2 \\ 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/2 & -i/2 & -1/2 & 0 & 0 & i/2 \\ 0 & 0 & 0 & -1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix}$ (it is 6 by 6 since the Cartesian product has 6 elements).
Unitary operations on three or more systems work similarly, with rows and columsn corresponding to the product of the classical state sets. The three-qubit operation $\sum_{k=0}^{7} |(k+1 \mod 8\rangle\langle k|$ is one such operation. This operation is unitary and also invertible (reversible), with conjugate transpose $\sum_{k=0}^{7} |(k-1) \mod 8\rangle\langle k|$.
Unitary operations can also be performed independently on a collection of individual systems. The combined action is represented by the tensor product of the unitary matrices (any tensor product of unitary matrices is unitary, linear algebra proof). We can also use the identity matrix to "do nothing" to some systems while altering others. For example, performing a Hadmard operation on X and ignoring Y is equivalent to doing $H \otimes I_\gamma = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \otimes Id_2 = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{bmatrix}$ on (X, Y). Not every unitary operation is the tensor product of unitary operations, however.

The swap operation. The swap operation on a pair (X, Y) exchanges the contents of the two systems but otherwise leaves them alone. So $\mathrm{SWAP}|a\rangle|b\rangle = |b\rangle|a\rangle$. The matrix associated with this operation is $\mathrm{SWAP} = \sum_{c,d\in\Sigma} |c\rangle\langle d| \otimes |d\rangle\langle c|$. As a simple example, if X and Y are qubits the matrix resulting from this Dirac notation is the identity with the second and third rows switched.
Another example is the controlled-unitary operation.