Foliable Operational Structures for General Probabilistic Theories

Lucien Hardy
Perimeter Institute,
31 Caroline Street North,
Waterloo, Ontario N2L 2Y5, Canada

December 23, 2009

Abstract

In this chapter a general mathematical framework for probabilistic theories of operationally understood circuits is laid out. Circuits are comprised of operations and wires. An operation is one use of an apparatus and a wire is a diagrammatic device for showing how apertures on the apparatuses are placed next to each other. Mathematical objects are defined in terms of the circuit understood graphically. In particular, we do not think of the circuit as sitting in a background time. Circuits can be foliated by hypersurfaces comprised of sets of wires. Systems are defined to be associated with wires. A closable set of operations is defined to be one for which the probability associated with any circuit built from this set is independent both of choices on other circuits and of extra circuitry that may be added to outputs from this circuit. States can be associated with circuit fragments corresponding to preparations. These states evolve on passing through circuit fragments corresponding to transformations. The composition of transformations is treated. A number of theorems are proven including one which rules out quaternionic quantum theory. The case of locally tomographic theories (where local measurements on a systems components suffice to determine the global state) is considered. For such theories the probability can be calculated for a circuit from matrices pertaining the operations that comprise that circuit. Classical probability theory and quantum theory are exhibited as examples in this framework.

1 Introduction

Prior to Einstein’s 1905 paper [1] laying the foundations of special relativity it was known that Maxwell’s equations are invariant under the Lorentz transformations. Mathematically the Lorentz transformations are rather complicated and it must have been unclear why nature would choose these transformations over the rather more natural looking Galilean transformations. Further, there
was a understanding of the physical reasons for Galilean transformations in terms of boosts and the additive nature of velocities. We find ourselves in a similar situation today with respect to quantum theory. Regarded as a probabilistic theory, it is much more complicated from a mathematical point of view than the rather natural equations of classical probability theory. And further, we can motivate classical probability by ordinary reasoning by imagining that the probabilities pertain to some underlying mutually exclusive set of possibilities. The situation with respect to the Lorentz transformations was resolved by Einstein when he showed that they follow from two very reasonable postulates: that the laws of physics are the same in all reference frames and that the speed of light in vacuum is independent of the motion of the source. Once we see Einstein’s reconstruction of the Lorentz transformations we have a sense that we understand why, at a fundamental level, nature prefers these over the mathematically simpler Galilean transformations. We need something similar for quantum theory [2].

The subject of reconstructing quantum theory has seen something of a revival in the last decade [3]. Generally, to reconstruct quantum theory we write down a set of basic axioms or postulates which are supposed to be well motivated. They should not appear unduely mathematical. Then we apply these in the context of some framework for physical theories and show that we obtain quantum theory. This framework itself should be well motivated and may even follow from one or more of the given postulates.

The purpose of this chapter is to set up one such framework. This will be a framework for general probabilistic operational theories. There is a large literature on this (see Section 2). To construct the mathematics of such a framework we must first specify what we mean by our operational structure. Only then can we add probability. The mathematics associated with the part of this where we add probability has become fairly sophisticated. However, a fairly simple minded point of view is usually taken with respect to setting up the operational structure upon which the whole endeavor is founded. The picture normally adopted is of a system passing sequentially through various boxes representing operations or, more generally, of many systems where, at any given time, each system passes through a box with, possibly, the same box acting on more than one system at once (see Fig. 1). This simple picture is problematic for various reasons. First, there is no reason why the types and number of systems going into a box be equal to the types and number emerging. Second, the notion of system itself is not fully operational. Third (and most significantly), this circuit is understood as being embedded in a background Newtonian time and this constitutes structure in addition to a purely graphical interpretation of the diagram (it matters how high on the page the box is placed since this corresponds to the time at which the operation happens).

To deal with these three points we set up a more general framework where (1) we allow the number and types of systems going into a box to be different from the the number and types of systems going out (2) give an operational definition of the notion of system (3) define our temporal concepts entirely in terms of the graphical information in the diagram. This third point gives rise to
a natural notion of spacelike hypersurface in such a way that multiple foliations are possible. Hence, we call this a foliable operational structure.

It is worth being careful to formulate the operational structure well since such structures form a foundation for general probabilistic theories. Different operational structures can lead to different probabilistic frameworks. Once we have an operational structure, we can introduce probabilities. We then proceed along a fairly clear path introducing the notions of preparation, transformation, measurement, and associating mathematical objects with these that allow probabilities to be calculated. This gives an example of how an operational framework can be a foundation for a general probabilistic theory. The foliable framework presented here is sufficient for the formulation of classical probability theory, quantum theory and potentially many theories beyond.

However, the foliable operational structure still, necessarily, has a notion of definite causal structure - when a system passes between two boxes that corresponds to a timelike separation (or null in the case of photons). We anticipate that a theory of quantum gravity will be a probabilistic theory with indefinite causal structure. If this is true then we need a more general framework than the one presented here for quantum gravity. Preliminary ideas along this line have been presented in [4]. In future work it will be shown how the approach taken in this chapter can be generalized to theories without definite causal structure - that is non-foliable theories. First we must specify a sufficiently general non-foliable operational structure and then add probabilities (see [5] for an outline of these ideas).

2 Related work

The work presented here is a continuation of work initiated by the author in [6] in which a general probabilistic framework, sometimes called the r-p framework
(because these vectors represent effects and states), was obtained for the purpose of reconstructing quantum theory from some simple postulates. In \[4, 7\] the author adapted the \(r_p\) for the purpose of describing a situation with indefinite causal structure to obtain a general probabilistic framework that might be suitable for a theory of quantum gravity. The idea that states should be represented by joint rather than conditional probabilities used in these papers is also adopted here. Preliminary versions of the work presented here can be seen in \[5, 8\].

In this work we consider arbitrary foliations of circuits. Thus we take a more space-time based approach. There are many other space-time based approach in the context of a discrete setting in the literature (particularly work on quantum gravity). Sorkin builds a discrete model of space-time based on causal sets \[9\]. Work in the consistent (or decoherent) histories tradition \[10, 11, 12, 13\] takes whole histories as the basic objects of study. A particularly important and influential piece of work is the quantum causal histories approach developed by Markopoulou \[14\]. In this, completely positive maps are associated with the edges of a graph with Hilbert spaces living on the vertices. Blute, Ivanov, and Panangaden \[15\] (see also \[16\]), motivated by Markopoulou’s work, took the dual point of view with systems living on the edges (wires) and completely positive maps on the vertices. The work of Blute et al., though restricted to quantum theory rather than general probabilistic theories, bares much similarity with the present work. In particular, similar notions of foliating circuits are to be found in that paper. Leifer \[17\] has also done interesting work concerning the evolution of quantum systems on a causal circuit.

Abramsky and Coecke showed how to formulate quantum theory in a category theoretic framework \[18\] (see also \[19\]). This leads to a very rich and beautiful diagramatic theory in which many essential aspects of quantum theory can be understood in terms of simple manipulations of diagrams. The diagrams can be understood operationally. Ideas from that work are infused into the present approach. Indeed, in category theoretic terms, the diagrams in this work can be understood as symmetric monoidal categories.

The \(r_p\) framework in \[6\] is actually simple example of a framework for general probability theories going back originally to Mackey \[20\] and has been worked on (and often rediscovered) by many others including Ludwig \[21\], Davies and Lewis \[22\], Araki \[23\], Gudder \[24\], Foulis and Randall \[25\].

Barrett elaborated on \(r_p\) framework in \[26\]. He makes two assumptions - that local operations commute and that local tomography is possible (whereby the state of a composite system can be determined by local measurements). In this work we do not make either assumption. The first assumption, in any case, would have no content since we are interested in the graphical information in a circuit diagram and interchanging the relative height of operations does not change the graph. Under these assumptions, Barrett showed showed that composite systems can be associated with a tensor product structure. We recover this here for the special case when we have local tomography but the more general case is also studied. In his paper Barrett shows that some properties which are thought to be specific to quantum theory are actually properties of
any non-classical probability theory.

More examples of this nature are discussed in various papers by Barrett, Barnum, Leifer, and Wilce in [27, 28, 29] and in [30, 31] the general probability framework is further developed.

The assumption of local tomography is equivalent to the assumption that $K_{ab} = K_a K_b$ where $K_{ab}$ is the number of probabilities needed to specify the state of the composite system $ab$ and $K_a$ ($K_b$) is the number needed to describe system $a$ ($b$) alone (this is the content of Theorem 5 below). Theories having this property were investigated in a paper by Wootters [32] (see also [33]) in 1990 who showed that they are consistent with the relation $K_a = N_a^r$ where $N_a$ is the number of states that can be distinguished in a single shot measurement (this was used in [6] as part of the axiomatic structure).

In 1994 Popescu and Rohrlich [34] exhibited correlations that maximally violate Bell’s inequality but do not permit signalling. These correlations are more nonlocal than quantum correlations. Barrett asked what principles would be required to prescribe such no-signalling correlations to the quantum limit [26]. Pawlowski et al. [35] have shown that the Popescu Rohrlich correlations (and, indeed, any correlations more nonlocal than quantum theory) violate a very natural principle they call the information causality principle. And Gross et al. [36] have shown, as speculated by Barrett [26], that the dynamics in any theory allowing Popescu Rohrlich correlations are trivial.

Another line of work in this type of framework has been initiated by D’Ariano in [37] who has a set of axioms from which he obtains quantum theory. In a very recent paper by Chiribella, D’Ariano, and Perinotti [38] set up a general probabilistic framework also having the local tomography property. Like Abramsky, Coecke and co-workers, Chiribella et al. develop a diagrammatic notation with which calculations can be performed. They show that theories having the property that every mixed state has a purification have many properties in common with quantum theory.

There have been many attempts at reconstructing quantum theory, not all of them in the probabilistic framework of the sort considered in the above works. A recent conference on the general problem of reconstructing quantum theory can be seen at [3].

3 Essential concepts

3.1 Operations and circuits

The basic building block will be an operation. An operation is one use of an apparatus. An operation has inputs and outputs, and it also has settings and outcomes (see Fig. 2). The inputs and outputs are apertures which we imagine a system can pass through. Each input or output can be open or closed. For example, we may close an output by blocking the aperture (we will explain the significance of this later). The settings may be adjusted by knobs. The outcomes may be read off a meter or digital display or correspond to a detector clicking
Figure 2: An operation having knob settings, measurement outcomes, and inputs (at the bottom of box) and outputs (at the top).

or lights flashing for example. It is possible that there is no outcome readout on the apparatus in which case we can simply say that the set of possible outcomes has only one member. The same apparatus may be used multiple times in a given experiment. Each separate use constitutes an operation.

Each input or output is of a given type. We can think of the type as being associated with the type of system that we imagine passes through. The type associated with an electron is different than that associated with a photon. However, from an operational point of view, talk of electrons or photons is a linguistic shortcut for certain operational procedures. We might better say that the type corresponds to the nature and intended use of the aperture. Operations can be connected by wires between outputs and inputs of the same type. These wires do not represent actual wires but rather are a diagrammatic device to show how the apertures on the operations are placed next to one another - this is something an experimentalist would be aware of and so constitutes part of the operational structure. If we actually have a wire (an optical fibre say) this wire should be thought of as an operation itself and be represented by a box rather than a wire. Likewise, passage through vacuum also should be thought of as an operation. The wires show how the experiment is assembled. Often a piece of self-assembly furniture (from Ikea for example) comes with a diagram showing an exploded view with lines connecting the places on the different parts of the furniture that should be connected. The wires in our diagrams are similar in some respects to the lines in these diagrams (though an experiment is something that changes in time and so the wires represent connections that may be transient whereas the connections in a piece of furniture are static).

There is nothing to stop us trying to match an electron output with a photon input or even a small rock output with an atom input (this would amount to firing small rocks at an aperture intended for individual atoms). However this would fall outside the intended use of the apparatus and so we would not expect our theory to be applicable (and the apparatuses may even get damaged).

We will often refer to tracing forward through a circuit. By tracing forward we mean following a path through the circuit from the output of one operation, along the wire attached, to the input of another operation and then from an output of that operation, along the wire attached, to the input of another operation and so on. Such paths are analogous to future directed time-like trajectories in
We require that there can be no closed loops as we trace forward (i.e. that we cannot get back to the same operation by tracing forward). This is a natural requirement given that an operation corresponds to a single use of an apparatus (so long as there are no backward in time influences). It is this requirement of no closed loops that will enable us to foliate.

In the case that we have a bunch of operations wired together with no open inputs or outputs left over then we will say we have a circuit (we allow circuits to consist of disconnected parts). An example of a circuit is shown in Fig. 3.

As mentioned above, we assume that any input or output can be closed. This means that if we have a circuit fragment with open inputs and outputs left over we can simply close them to create a circuit. This is useful since the mathematical machinery we will set up starts with circuits. Closing an output can be thought of as simply blocking it off. The usefulness of the notion of closing an output relates to the possibility of having no influences from the future. This will be discussed in Sec. 3.3. Closing an input can be thought of as sending in a system corresponding to the type of input in some fiducial state. We will not make particular use of the notion of closing a input (beyond that it allows us to get circuits from circuit fragments) and so we need not be more specific than this. We could set up the mathematical machinery in this chapter without assuming that we can close inputs and outputs without much

Figure 3: A bunch of operations wired together form a circuit if there are no open inputs or outputs left over. We require that there be no closed loops as we trace forward. We have not drawn in the settings or the outcomes (these will usually be taken to be implicit in these circuit diagrams). There are some closed inputs and outputs.

spacetime physics.
more effort but the present approach has certain pedagogical advantages.

3.2 Time in a circuit

We do not think of time as something in the background but rather define it in terms of the circuit itself. We take the attitude that two circuits having the same circuit diagram (in the graphical sense) are equivalent. Hence there is no physical meaning to sliding operations along wires with respect to some background time. This is a natural attitude given the interpretation of wires as showing how apertures are placed next to each other rather than as actual wires.

We define a synchronous set of wires to be any set of wires for which there does not exist a path from one wire to another in the set if we trace forward along wires from output to input. See Fig. 4 for examples.

We call a synchronous set of wires a hypersurface, \( H \), if it partitions the circuit into two parts, \( \gamma_{H}^{-} \) and \( \gamma_{H}^{+} \) that are not connected other than by wires in the hypersurface. Each of the wires in the hypersurface has an end connected to an output (the “past”) and an end connected to an input (the “future”). \( \gamma_{H}^{-} \) is the part of the circuit to the “past” and \( \gamma_{H}^{+} \) is the part of the circuit to the “future” of the hypersurface. A hypersurface, as defined here, is the circuit analogue of a spacelike hypersurface in spacetime physics.

We say two hypersurfaces are distinct if at least some of the wires are different. We say that hypersurface \( H_{2} \) is after hypersurface \( H_{1} \) if the intersection of the past of \( H_{1} \) (this is \( \gamma_{H_{1}}^{-} \)) and the future of \( H_{2} \) (this is \( \gamma_{H_{2}}^{+} \)) has no operations in common. If we can get from every wire in \( H_{2} \) by tracing forward from a wire in \( H_{1} \) then \( H_{2} \) is after \( H_{1} \) (there are, however, examples of \( H_{2} \) after \( H_{1} \) that are not like this).

A foliation is a ordered set of hypersurfaces \( \{ H_{t} \} \) such that \( H_{t+1} \) is after \( H_{t} \). A complete foliation is a foliation that includes every wire in the circuit. It is easy to prove that complete foliations exist for every circuit. Define an
**initial wire** to be one connected to an output of an operation having no open inputs. Take the set of all initial wires (this cannot be the null set as long as we have at least one connecting wire in this circuit). These wires form a hypersurface \( H_1 \). Consider the set of operations for which these wires form inputs. Since, according to the wiring constraints, there can be no closed loops, there must exist at least one operation in this set which has no inputs from wires connected to outputs of other operations in this set. Substitute the wires coming out of one such operation for the wires going into the operation in \( H_1 \) to form a new spacelike hypersurface \( H_2 \) (this is after \( H_1 \)). This can be repeated until all wires have been included forming a complete foliation. This proves that complete foliations always exist. There can, of course, exist other complete foliations that are not obtainable by this technique.

Although we do not use a notion of a background time to time-order our operations, it is the case that these foliable structures are consistent with a Newtonian notion of an absolute background time. Simply choose one complete foliation and take that as corresponding to our Newtonian time. They are, however, more naturally consistent with relativistic ideas since, for a general circuit, there exist multiple foliations.

### 3.3 Probability

Now we are in a position to introduce probability into the picture. Probability is a deeply problematic notion from a philosophical point of view [39]. There are various competing interpretations. All these interpretations attempt to account for the empirical fact that, in the long run, relative frequencies are stable - that if you toss a coin a million times and get 40% heads, then if you toss the same coin a million times again, you will get 40% heads again (more or less). It is not the purpose of this chapter to solve the interpretational problems of probability and so we will adopt the point of view that probability is a limiting relative frequency. This gives us the basic mathematical properties of probability:

1. Probabilities are non-negative
2. Probabilities sum to 1 over a complete set of mutually exclusive events.
3. Bayes rule, \( \text{Prob}(A \& B) = \text{Prob}(A | B) \text{Prob}(B) \), applies.

We could equally adopt any other interpretation of probability that gives us these mathematical properties and set up the same theoretical framework.

Typically an experimentalist will have available to him some set of operations, \( O \), he can use to build circuits. On each operation in the circuit are various possible settings (among which the experimentalist can choose) and various outcomes one of which will happen. We say the circuit is setting specified if each operation is given. We say the circuit is setting-outcome specified if the setting and outcome on each operation is specified. A setting-outcome specified circuit corresponds to what happens in single run of the experiment. We define:
Figure 5: A set of operations is closable if any circuit built from it has a probability associated with it depending only on that circuit and if that probability is independent of any extra circuitry.

**Closable sets of operations.** A set of operations, $O$, is said to be closable if, for every setting-outcome specified circuit that can be built from $O$,

(i) there is a probability depending only on the particulars of this circuit that is independent of choices made elsewhere, and

(ii) if we open closed outputs on this circuit and add on extra circuitry then the probability associated with the original bit of setting-outcome specified circuitry (ignoring outcomes associated with the extra bit) is unchanged for any such extra circuitry we can add (see Fig. 5).

Part (i) of this definition concerns choices made elsewhere. These could be choices of settings on operations in other circuits (disjoint from this one), choices of what circuits to build elsewhere, or choices not even having to do with circuits built from the given set of operations. We might also have said that the probability associated with a setting-outcome specified circuit is independent of the outcomes seen in other circuits. This is a very natural assumption since otherwise the probability attached to a particular circuit could be different if we restrict our attention to the case where we had seen some particular outcomes on other circuits. In the case where (i) and (ii) hold and also the probability is independent of the outcomes in other circuits we will say the set of operations is *fully closable*. It turns out we can go a very long way without assuming this. Further, we will prove that in Section 6.6 that if a very natural condition holds then closable sets are, in any case, fully closable.

Part (ii) imposes a kind of closure from the future - choices on operations only connected to a part of a circuit by outgoing wires (or even choices of what circuitry to place after outgoing wires) do not effect the probabilities for this circuit part. This could almost be regarded as a definition of what we mean by wires going from output to input. We do not regard it as a definition though
Figure 6: If there is no-signalling between A and B then the correct circuit is that shown in (a). However, if there is signalling from B to A then (a) could not be the correct circuit. Instead, the correct circuit would have to be something like that shown in (b). The framework is perfectly capable of incorporating signalling. Hence the assumption of no-signalling is not an assumption of the framework but rather corresponds to asserting that the correct circuit for the given no-signalling situation is the one in (a).

It is interesting to consider examples of sets of operations which are not closable. Imagine for example that, among his operations, an experimentalist has an apparatus which he specifies as implementing an operation on two qubits but actually it implements an operation on three qubits - there is an extra input aperture he is unaware of. If he builds a circuit using this gate then an adversary can send a qubit into the extra input which will effect the probability for the circuit. Thus, the probability would depend on a choice made elsewhere. In this case we could fix the situation by properly specifying the operation to include the extra input. The notion of closability is important since it ensures that the experimentalist has full control of his apparatuses. It is possible, at least in principle, that a set of operations cannot be closed by discovering extra apertures. By restricting ourselves to physical theories that admit closability we are considering a subset of all possible theories (though a rather important one).

3.4 Can no-signalling be an axiom?

Many authors have promoted no-signalling as an axiom for Quantum Theory. It may appear that part (ii) of the definition of closability sneaks in a no-signalling assumption here. In fact this is not the case. Indeed, a no-signalling axiom would actually have limited content in this circuit framework. Consider the circuit shown in Fig. 6(a). A no-signalling axiom would assert that a choice at one end, B say, cannot effect the probability of outcomes at the other end, A. However, this is actually implied by part (ii) since B is connected to the AC part of the circuit by an out-going wire. Hence, it looks like we are assuming no-signalling. However, this is not an assumption for the framework but only a
consequence of asserting that the correct circuit is the one in Fig. 6(a). Imagine that there actually was signalling such that probabilities for the $AC$ part of the circuit depended on a choice at $B$ then, under the assumption that the operations are drawn from a closable set, it is clear that the situation cannot be described by Fig. 6(a). Rather, the situation would have to be that shown in Fig. 6(b) where there is an extra wire going from $B$ to $A$ (or something like this but with more structure). This framework is perfectly capable of accommodating both signalling and no-signalling situations by appropriate circuits and, consequently, we are not sneaking in a no-signalling assumption.

This reasoning leads us to question whether a no-signalling axiom could do any work at all. Indeed it is often unappreciated in such axiomatic discussions that the usual framework of Quantum Theory does allow signalling. One can write down nonlocal Hamiltonians which will, for example, entangle product states. Of course, in Quantum Field Theory one incorporates a no-signalling property by demanding that field operators for space-like separated regions commute so that such nonlocal Hamiltonians are ruled out. However, this is an example where we have a given background. In general, a no-signalling axiom with respect to some particular given background would restrict the type of circuits we allow. For example the circuit in Fig. 6(b) would not be allowed unless $A$ was in the future light cone of $B$. In quantum field theory we have an example where a no-signalling axiom with respect to a Minkowski background restricts the types of unitary evolution and measurement that are allowed. However, it is often claimed that the abstract Hilbert space framework itself (which makes no mention of Minkowski spacetime) can be derived using no-signalling as one of the axioms. It is this more ambitious claim we question. In fact we will see that we can define this abstract framework of quantum theory for any circuit (as long as there are no closed loops) including no-signalling and signalling situations (as in Fig. 6(a) and (b)). Hence a no-signalling axiom clearly cannot be regarded as a constraint on this abstract framework.

This criticism of the usefulness of no-signalling as an axiom does not apply to a recently proposed generalisation of this principle in an extraordinary paper by Pawlowski et al. [35]. They introduced the information causality principle. It was shown that this very compelling principle limits violations of the Bell inequality to the quantum limit. Imagine that Alice receives $n$ classical bits of information. She communicates $m$ classical bits to Bob. Bob is expected to reveal the value of one of the $n$ classical bits though neither Alice or Bob know which one this will be in advance. The information causality principle is that Alice and Bob can only be successful when $n$ is less than or equal to $m$. For $m = 0$ this is the no-signalling assumption we have criticized. The information causality principle can be read as implying that if the task cannot be achieved for $m = 0$ then it cannot be achieved for any other value of $m$. This principle would be useful in prescribing what is possible in the framework described in this chapter. Consider two fragments of a circuit that cannot be connected by tracing forward (these fragments are analogous to spacelike separated regions). The information causality principle implies there is no way of accomplishing the above task for any $m$ with $n > m$ between these two circuit fragments. That we
cannot do this for \( m = 0 \) is already implied (assuming that we have the correct circuit).

Although a simple no-signalling “across space” principle is of limited use for the above reasons, we do employ what might be regarded as a no-signalling backward in time principle since we do not allow closed loops in a circuit.

### 3.5 Systems and composite systems

We wish to give an operational definition of what we mean by the notion of a system. We may find that whenever we press a button on one box, a light goes on on another box. We can interpret this in terms of a system passing between the two boxes. We find this happens only when we place the boxes in a certain arrangement with respect to one another (which we think of as aligning apertures). Given this we clearly want to associate systems with wires. Hence, we adopt the following definition:

**A system** of type \( ab\ldots \) is, by definition, associated with any set of synchronous wires of type \( a, b, c, \ldots \) in any circuit formed from a closable set of operations.

We may refer to a system type corresponding to more than one wire by a single letter. Thus we may denote the system type \( abc \) by \( d \).

The usefulness of the notion of closable sets of operations is that it leads to wires being associated with the sort of correlation we expect for systems given our usual intuitions about what systems are. Nevertheless, our definition of system is entirely operational since wires are defined operationally.

It is common to speak of composite systems. We define a composite system as follows:

**A composite system**, \( AB \), is associated with any two systems (each associated with disjoint sets of wires) if the union of the sets of wires associated with system \( A \) and system \( B \) forms a synchronous set.

This definition generalizes in the obvious way for more than two systems. A system of type \( aabc \) can be regarded as a composite of systems of type \( aa \) and \( bc \) or a composite of systems of type \( aac \) and \( b \), or a composite of systems of type \( ac, a \) and \( b \), to list just a few possibilities. Systems associated with a single wire cannot be regarded as composite.

A hypersurface consists of synchronous wires and so can be associated with a system (or composite system). A complete foliation can therefore be associated with the evolution of a system through the circuit (though the system type can change after each step). This evolution can also be viewed as the evolution of a composite system.
Figure 7: A circuit fragment is a part of the circuit having inputs wires in a synchronous set and output wires also in a synchronous set. A circuit can be divided up many ways into circuit fragments corresponding to preparations (no open inputs), transformations, and effects (no open outputs).

4 Preparations, transformations, and effects

4.1 Circuit fragments

We can divide up a circuit into fragments corresponding to preparations, transformations, and effects as shown in Fig. 7. By the term circuit fragment we mean a part of a circuit (a subset of the operations in the circuit along with the wires connecting them) having inputs coming from a synchronous set of wires and outputs going into a synchronous set of wires. We allow lone wires in a circuit fragment (wires not connected to any operations in the fragment). An example of a lone wire is see in the circuit fragment in the rectangle on the left in Fig. 7. The lone wire corresponds to the identity transformation on that system and contributes an input and output to the circuit fragment. Generally, we take the term circuit fragment to imply that the settings and outcomes at each operation associated with these circuit fragments have been specified. A circuit fragment is, essentially, an operation at a course grained level. Preparations correspond to a circuit fragment having outputs but no open inputs. Transformations have inputs and outputs. Effects have inputs but no open outputs. Note that preparations and effects are special cases of transformations.
4.2 States

A preparation prepares a system. For a given type of system there will be many possible preparations. We will label them with $\alpha$. This label tells us what circuit fragment is being used to accomplish the preparation (including the specification of the knob settings and outcomes on each operation). Associated with each preparation for a system of type $a$ will be a state (labeled by $\alpha \in \text{Prep}_a$). We can build a circuit having this preparation by adding an effect for a system of type $a$. There are many possible effects labeled by $\beta \in \text{Eff}_a$. The label $\beta$ tells us what circuit fragment is used to accomplish the effect along with the knob settings and outcomes at each operation. Associated with the circuit is a probability $p^{\alpha\beta}$. We define the state associated with preparation $\alpha$ to be that thing represented by any mathematical object that can be used to calculate $p^{\alpha\beta}$ for all $\beta$. We could represent the state by the long list

$$P_\alpha^a = \left( \begin{array}{c} \vdots \\ p^{\alpha\beta} \\ \vdots \end{array} \right) \quad \beta \in \text{Eff}_a \quad (1)$$

However, in general, physical theories relate physical quantities. Hence, it is only necessary to list a subset of these quantities where the remaining quantities can be calculated by the equations of the physical theory. We call the forming of this subset of quantities physical compression. In the current case, we expect the probabilities in this list to be related. We consider the maximum amount of physical compression that is possible by linear means. Thus we write the state as

$$p_\alpha^a = \left( \begin{array}{c} \vdots \\ p^{\alpha\beta} \\ \vdots \end{array} \right) \quad \beta \in \Omega_a \subseteq \text{Eff}_a \quad (2)$$

where there exist vectors $r^\beta_a$ such that

$$p^{\alpha\beta} = r^\beta_a \cdot p^\alpha_a \quad \text{for all } \alpha \in \text{Prep}_a, \beta \in \text{Eff}_b \quad (3)$$

We call the set $\Omega_a$ the fiducial set of effects for a system of type $a$. The choice of fiducial set need not be unique - we simply make one choice and stick with it. Since we have applied as much linear physical compression as possible $|\Omega_a|$ is the minimum number of probabilities required to calculate all the other probabilities by linear equations. The vectors $r^\beta_a$ are associated with effects on a system of type $a$. For the fiducial effects they consist of a 1 in the $\beta$ position and 0’s elsewhere.

An important subtlety here is that we define states in terms of joint rather than conditional probabilities. This makes more sense for the circuit model since, generally, we want to calculate a probability for a circuit. If we want to calculate conditional probabilities we can use Bayes’ rule in the standard way.
4.3 Transformations

Now consider a preparation $\alpha$ which prepares a system of type $a$ in state $p^\alpha_a$ followed by a transformation $\beta$ which outputs a system of type $b$. We can regard the preparation and transformation, taken together, as a new preparation $\alpha\beta$ for a system of type $b$ with state $p^\alpha\beta_b$. Now follow this by a fiducial effect $\gamma \in \Omega_b$. The probability for the circuit $\alpha\beta\gamma$ can be written

$$p^{\alpha\beta\gamma} = r^\gamma_b \cdot p^\beta_b = r^\gamma_a \cdot p^\alpha_a$$

where $r^\beta_a$ is the effect vector associated with the measurement consisting of the transformation $\beta$ followed by the effect $\gamma$. Given the special form of the fiducial effect vectors it follows that the state transforms as

$$p^{\alpha\beta}_b = b^b Z^\beta_a p^\alpha_a$$

where $b^b Z^\beta_a$ is a $|\Omega_b| \times |\Omega_a|$ matrix such that its $\beta$ row is given by the components of the effect vectors $r^{\beta\gamma}$. We use a subscript, $a$, for the inputted system type and a pre-superscript, $b$, for the outputted system type. Hence, a general transformation is given by a matrix acting on the state. If the matrix transforms from one type of system to another type of system with a different number of fiducial effects then it will be rectangular.

The general equation for calculating probabilities is

$$p^{\alpha\beta\gamma\delta} = (r^\delta_c)^T b^c Z^{\gamma\beta}_b b^b Z^{\alpha}_a$$

where $T$ denotes transpose. If we have more than one transformation then, by a clear extrapolation of the above reasoning, we can write

$$p^{\alpha\beta\gamma\delta} = (r^\delta_c)^T b^c Z^{\gamma\beta}_b Z^{\alpha}_a$$

and so on. Now since the $Z$ matrices can be rectangular we can think of $p^\alpha_a$ and $(r^\gamma_a)^T$ as instances of a transformation matrix. The state $p^\alpha_a$ can be thought of as corresponding to the transformation which turns a null system (no system at all) into a system outputted by this preparation and we can change our notation to $a^a Z^\alpha$ instead (a column vector being a special case of a rectangular matrix). Likewise we can change our notation for the row vector $(r^\gamma_a)^T$ to $Z^\gamma_b$ (a row vector being a special case of a rectangular matrix). Then we can write

$$p^{\alpha\beta\gamma\delta} = Z^\gamma_c b^c Z^{\beta\alpha}_a$$

The agreement of output and input system types is clear (by matching pre-superscript with subscripts between the $Z$’s).

The label $\alpha$ labels the circuit fragment along with the knob settings and the outcomes. Sometimes it is useful to break these up into separate labels. Thus we write

$$\alpha \equiv (F, \varphi, l)$$

where $F$ denotes the circuit fragment before the settings and outcomes are specified, $\varphi$ denotes the settings on the operations, and $l$ denotes the outcomes.
In particular, this means that we can notate the effects associated with the different outcomes of a measurement as \( \alpha_i = (F, \phi, l_i) \).

Let \( L \) be the set of possible outcomes \( l_i \) for a given measurement (with fixed \( F \) and \( \phi \)). We can subdivide the set \( L \) into disjoint sets \( L_k \) where \( \bigcup_k L_k = L \). We could choose to be ignorant of the actual outcome \( l_i \) and rather only record which set \( L_k \) it belongs to. In this case the transformation effected can be denoted \( \alpha_k \). Since we have used linear compression, we must have

\[
\sum_{i \in I_k} b \gamma_i a = \sum_{i \in I_k} b \gamma_{\alpha_i} a
\]

where \( I_k \) is the set of \( i \)’s corresponding to the \( l_i \)’s in \( L_k \). Since we can always choose to be ignorant in this way, we must include such transformations in the set of allowed transformations.

The matrices corresponding to the set of allowed transformations must be such that, when closed expressions such as (8) are calculated, they always give probabilities between 0 and 1. This is an important constraint on this framework.

### 4.4 The identity transformation

One transformation we can consider is where we do nothing. The wires coming in are the same as the wires coming out and no operation has intervened. We will denote this transformation by \( 0 \). Then we have, for example,

\[
a \gamma^a_0 = \sum_{i \in I_k} b \gamma_{\alpha_i} a
\]

This is a \(|\Omega_a| \times |\Omega_a|\) matrix and must be equal to the identity since, as long as it is type matched, it can be inserted as many times as we like into any expression where non-trivial transformations act also.

### 4.5 The trace measurement

One effect we can perform on a preparation \( \alpha \) is to close all outputs. This forms a circuit and hence there is an associated probability, \( p^{\alpha-} \) (where \( - \) denotes that the outputs have been closed). This is an effect and hence we must have

\[
p^{\alpha-} = r_\alpha^a \cdot p_\alpha^a
\]

where the vector \( r_\alpha^a \) corresponds to this effect. We call this the trace measurement (terminology borrowed from quantum theory where this effect corresponds to taking the trace of the density matrix). It follows from part (ii) of the condition for a closable set of operations that this is the probability associated with the preparation part of the circuit even if the outputs are open and more circuitry is added.

In the case that \( r_\alpha^a \cdot p_\alpha^a = 1 \) we say that the state is of norm one. In general we do not expect states to be of norm one since they consist of joint rather than conditional probabilities and hence we require only that \( 0 \leq r_\alpha^a \cdot p_\alpha^a \leq 1 \).
We can normalize a state by dividing by $r_a \cdot p_a^\alpha$. We denote the normalized state by
\[ \bar{p}_a^\alpha = \frac{p_a^\alpha}{r_a \cdot p_a^\alpha} \] (13)
We cannot guarantee that $\bar{p}_a^\alpha$ belongs to the set of allowed states (i.e. that there exists a state for which $p_a^\alpha = \bar{p}_a^\alpha$) since preparations for all states parallel to $p_a^\alpha$ may be intrinsically probabilistic.

5 Mixtures

5.1 Forming Mixtures

Imagine that we have a box with a light on it that can flash and an aperture out of which a system of type $a$ can emerge. With probability $\lambda_i$ we place preparation $\alpha_i$ in the box such that the system (which we take to be of type $a$) will emerge out of the aperture and such that the light will flash if the outcomes corresponding to this preparation are seen. The state prepared for this one $i$ is $\lambda_i p_a^\alpha_i$. If $\lambda_i = 0$ then the state prepared is the null state $0_a$ which has all 0’s as entries. If we use this box for a set $\alpha_i$ with $i \in I$ such that $\sum_{i \in I} \lambda_i \leq 1$ then the state prepared is
\[ \sum_{i \in I} \lambda_i p_a^\alpha_i \] (14)
This is a linear sum of terms since we have linear compression. This process of using a box may be beyond the experimental capacities of a given experimentalist. It certainly takes us outside the circuit model as previously described. However, we can always consider taking mixtures like this at a mathematical level.

A technique that can be described in the circuit model is the following. Consider placing a single preparation circuit into the box described above where $\alpha_j = (F, \varphi, l_j)$ where $l_j$ labels the outcomes. We can arrange things so that the light flashes only if $j \in J$ (where $J$ is some subset of the $j$’s). The state is then given by
\[ \sum_{j \in J} p_a^{\alpha_j} \] (15)
This technique does not require having a coin to generate probabilities $\lambda_i$ and neither does it require the placing of different circuit fragments into a box depending on the outcome of the coin toss.

The most general thing we can do is a mixture of the two above techniques. With probability $\lambda_i$ we place a circuit $F_i$ with settings $\varphi_i$ and outcomes $l_{ij}$ in the box for $i \in I$ and $j \in J_i$. The state we obtain is
\[ \sum_{i \in I, j \in J_i} \lambda_i p_a^{\alpha_{ij}} \] (16)
We can absorb the $\lambda$'s into the $p$'s and relabel so we obtain that a general mixture is given by

$$ p^\alpha = \sum_i p_i^\alpha $$

(17)

We have the constraints that $r^- \cdot p_i^\alpha \geq 0$ and $\sum_i r^- \cdot p_i^\alpha \leq 1$.

Note that if we write $p_i^\alpha \equiv \mu_i \overline{p}_i^\alpha$ where $\overline{p}_i^\alpha$ is normalized, and include an extra state $p_0^\alpha \equiv 0$ (the null state) then we have

$$ p^\alpha = \sum_i \mu_i \overline{p}_i^\alpha $$

where $\mu_i \geq 0$, $\sum_i \mu_i = 1$ (18)

where the sum now includes the null state. Hence we can interpret a general mixture as a convex combination of normalized states and the null state.

5.2 Homogeneous, pure, mixed, and extremal states

If two states are parallel then they give rise to the same statistics up to an overall weighting and if we condition on the preparation then they have exactly the same statistics. We define a **homogeneous state** as one which can only be written as a sum of parallel states. Thus $p_i^\alpha$ is homogeneous if, for any sum,

$$ p_i^\alpha = \sum_{i' \in I} \eta_{ii'} p_i^\alpha $$

(19)

we have $p_i^\alpha = \eta_{ii'} p_i^{\alpha'}$ for all $i, i' \in I$. A state which is not homogeneous (i.e. which can be written as a sum of at least two non-parallel states) is called a **heterogeneous state**.

Given a particular homogeneous state, there will, in general, be many others which are parallel to it but of different lengths. We call the longest among these a **pure state**.

A **mixed state** is defined to be any state which can be simulated by a probabilistic mixture of distinct states in the form $\sum_j \lambda_j p_i^\alpha$ where $\lambda_j \geq 0$ and $\sum_j \lambda_j = 1$. A pure state is not a mixture (since the $p_i^\alpha$'s would have to parallel to the given pure state, and therefore, given the $\sum_j \lambda_j = 1$ condition, equal to the given pure state). Homogeneous states which are not pure are mixtures. Heterogeneous states are also mixtures. **Extremal states** are defined to be states which are not mixtures. Pure states are extremal. The null state is also extremal. If all pure states have norm equal to one (i.e. $r^- \cdot p_i^\alpha = 1$) then there are no more extremal states beyond the pure states and null state. However, if some pure states have $r^- \cdot p_i^\alpha < 1$, then there may be additional extremal states.

Usually treatments of convex sets of states do not make these distinctions. More care is necessitated here because states are based on joint rather than conditional probabilities.

Any state, extremal or mixed, can be written as the sum of homogeneous states. This means that there must exist at least one set of $|\Omega_a|$ linearly independent homogeneous states all of which can be pure. There cannot exist sets with more linearly independent states than this.
5.3 Optimality of linear compression

We state the following theorem

**Theorem 1** If we allow arbitrary mixtures of preparations then
(1) linear compression is optimal and (2) optimal compression is
necessarily linear.

The first point follows since there must exist $|\Omega_a|$ linearly independent states
(otherwise we could have implemented further linear compression). We can
take an arbitrary mixture with $\sum \lambda_i \leq 1$ then these $\lambda$'s are all independent
and hence we need $|\Omega_a|$ parameters and the compression is optimal. To prove
the second point consider representing the state by a list of $|\Omega_a|$ probabilities
$p_{\alpha\beta}'$ with $\beta' \in \Omega_a'$ where we do not demand that a general probability is given
by a linear function of these probabilities. Represent this list by the vector $q_{\alpha}'$.
Now
\[ p_{\alpha\beta}' = r_{\beta}' \cdot p_{\alpha} \quad \text{for all } \beta' \in \Omega_a' \] (20)

Hence $q_{\alpha}' = C p_{\alpha}'$ where $C$ is a square matrix with real entries. $C$ must be
invertible since otherwise we could specify $q_{\alpha}'$ with fewer than $|\Omega_a|$ probabilities.
Hence
\[ p_{\alpha\beta} = r_{\beta} \cdot C^{-1} p_{\alpha} \] (21)

for a general $\beta$. Hence the probability is linear in $q_{\alpha}'$ and so the compression is
linear.

6 Composition

6.1 Preliminaries and notation

As we discussed above, systems associated with more than one wire can be
thought of as composite. The $p$, $r$, $Z$ framework just discussed can be enriched
by adapting it to deal separately with the components of composite systems
(rather than regarding all the wires at each time step as constituting a single
system). The advantage of this is that we can break up the calculation into
smaller parts and thereby define a theory by associating matrices to smaller
transformations. Ultimately, we would like to have a matrix associated with
each operation in the set of allowed operations which can be used to calculate
the probability for any circuit. A transformation is now associated with a matrix
such as
\[ \begin{bmatrix} b & a & c & d \end{bmatrix} Z_{\alpha}^{\alpha} \]

This transformation inputs a system of type $acb$ and outputs a system of type $bacd$. The label $\alpha$ denotes the circuit fragment used to do this including the
knob settings and the outputs at each operation in the fragment. The ordering
of the symbols representing the system (such as $bacd$) is significant in that it
is preserved between transformations to indicate how the wires are connected.
Thus, the matrix for a transformation comprised of two successive transformations is written
\[ dZ^\alpha_{bacd} Z_{acbd} \]

(23)

The system types in between the two transformations must match (as in this example) since we employ the convention that the wires are in the same order from the output of one transformation to the input of the next. We can allow that the symbols for the systems (such as \( c \)) actually correspond to composite systems (so that they correspond to a cluster of wires). For example, it might be that \( c = aabb \). Some transformations consist of disjoint circuit fragments and it is useful to have notation for these. We write

\[ \langle b \rangle \langle ac \rangle \langle c \rangle \langle d \rangle \]

(24)

\[ Z_{(a)c(b)} \]

(25)

\[ \langle b \rangle \langle ac \rangle \langle c \rangle \langle d \rangle \]

(26)

We see that the wires match (for example wire 6 is of type \( c \) as an output from the first transformation and an input into the second transformation). The integers labeling the wires are of no significance and any expression is invariant under any reassignment of these labels (this is a kind of discrete general covariance).

6.2 Commutation

Consider the situation shown in Fig. 8 By inspection of this diagram we can write

\[ \langle b \rangle \langle ac \rangle \langle c \rangle \langle d \rangle \]

(28)

where the 0 denotes that we do nothing (the identity transformation). The first equation here is obtained by first transforming from hyperplane \( H_1 \) to \( H_2 \) (past operation \( A \)) and then from \( H_2 \) to \( H_4 \) (past operation \( B \)). To get the second equation we evolve from \( H_1 \) to \( H_3 \) first (past \( B \)) then from \( H_3 \) to \( H_4 \) (past \( A \)). There are a few points of interest here. First note that we can break
down a compound transformation into its parts. Second, we see that there is a kind of commutation - it does not matter whether we update at A or B first. This property is not assumed but derived from more basic assumptions and definitions. In the special case where wires c and d are of type a and b respectively (so the transformations do not change the system type) we have the commutation property

\[ [ab^{03}, ab^{00}] = ab^{03} ab^{00} - ab^{00} ab^{03} = 0 \]  

The usual commutation relation is, then, a special case of the more general relation in (28) where the local transformations may change the system type.

The fact that we can break up a compound transformation into smaller parts is potentially useful. But there is a stumbling block. The matrix \( cb^{0} \) transforms past the A operation. However, it is a \( |\Omega_{cb}| \times |\Omega_{ab}| \) matrix. That is, we still have to incorporate some baggage because we include wire b. It would be good if we could write

\[ cb^{0} = c^{0} \otimes b^{0} \]  

where \( b^{0} \) is just the identity matrix (as discussed above). By considering the sizes of the matrices it is clear that (30) implies \( |\Omega_{ab}| = |\Omega_{a}| |\Omega_{b}| \). In Sec. 7.1.1 we show that equation (30) holds in general if \( |\Omega_{ab}| = |\Omega_{a}| |\Omega_{b}| \) is true (for any system types a and b). We will also see that this condition corresponds to a very natural class of physical theories. If (30) holds true we can break any circuit down into its basic operations appending the identity transformation as necessary. Then we can calculate the probability associated with any circuit from the transformation matrices associated with the operations.

Figure 8: We can consider the evolution of this circuit with respect to different foliations.
Figure 9: A bipartite system 12 of type $ab$ is prepared by some preparation $\gamma$ and subjected to effects $\alpha_i$ and $\beta_j$ on subsystem 1 and subsystem 2 respectively.

### 6.3 Homogenous states and composite systems

Consider a composite system 12 consisting of systems of types $a$ and $b$ with preparation $C$ labeled $\gamma$. The state prepared by this is $p^{\gamma_{ab}}$. If we block the output 2 of the preparation then we have a preparation for a system of type $a$. Let the state so prepared be $p^{\gamma_{a}}$. Even if we do not block output 2 it follows from part (ii) of the condition for closable sets of operations that this state $p^{\gamma_{a}}$ gives us the correct probabilities for all measurements on system 1 alone (that do not involve system 2). We call $p^{\gamma_{a}}$ the reduced state for system 1. It is, effectively, the state of system 1 taken by itself.

We prove the following theorem.

**Theorem 2** If one component of a bipartite system is in a homogenous state (i.e. the reduced state for this system is homogeneous) all joint probabilities for separate effects measured on the two systems factorize.

Systems 1 and 2 can be subjected to measurements $A$ and $B$ respectively (see Fig. 9). We also consider the possibility of closing either or both outputs from $C$. Let measurement $A$ have outcomes $l_i$, $i \in I$ the effects for which are labeled $\alpha_i \equiv (F^A, \phi^A, l^A_i)$. Similarly, for $B$ we have outcomes $l^B_j$ for $j \in J$ and effects labeled by $\beta_j \equiv (F^B, \phi^B, l^B_j)$. By part (ii) of the condition for a closable set of operations we have

$$p^{\gamma_{--}} = \sum_{i \in I} p^{\gamma_{\alpha_i \beta_j}} = \sum_{j \in J} p^{\gamma_{\beta_j}}$$

(31)

and

$$p^{\gamma_{-\beta_j}} = \sum_{i \in I} p^{\gamma_{\alpha_i \beta_j}}$$

(32)

If the output 2 from $C$ is closed then we say that the state prepared (for the system of type $a$) is $p^{\gamma_{a}}$ (the reduced state). We say that the preparation due
to $C$ and $B$ with outcome $l_j^B$ (this is a preparation circuit fragment) is $p_{a}^{γβ_j}$. It follows from part (ii) of the condition for a closable set of operations that

$$p_{a}^{γ} = \sum_{j \in J} p_{a}^{γβ_j}$$

(33)

If $p_a^{γ}$ is homogeneous then all the $p_{a}^{γβ_j}$ states must be parallel to it. Hence we can write $p_{a}^{γβ_j} = η_j p_a^{γ}$. We must have

$$p^{γα,β} = r_a^{α_i} \cdot p_{a}^{γβ_j} = η_j r_a^{α_i} \cdot p_a^{γ} = η_j p^{γα}$$

(34)

Summing this over $i$ and using (31, 32) we obtain $p^{γ−β} = η_j p^{γ−}$. Hence

$$p^{γα,β}p^{γ−} = p^{γα}p^{γ−β}$$

(35)

Here $p^{γ−}$ is the probability of the preparation being successful. Dividing this through by $(p^{γ−})^2$ and using Bayes’ rule we obtain

$$\text{prob}(l_i^A l_j^B | \text{prep}) = \text{prob}(l_i^A | \text{prep})\text{prob}(l_j^B | \text{prep})$$

(36)

Hence we see that if one system is in a homogeneous state then joint probabilities factorize between the two ends (obviously the result also holds if both components are in homogeneous states). An obvious corollary is

**Corollary 1** If the state of a bipartite system $12$ of type $ab$ with preparation $γ$ is of norm one and the reduced state of either or both components is pure then

$$p^{γαβ} = p^{γα}p^{γβ}$$

(37)

where $α$ is any effect on $1$ alone and $β$ is any effect on $2$ alone.

This is true since pure states are necessarily homogeneous and because $p^{γ−} = 1$ since the state is of norm $1$.

Now we consider a related but slightly different situation. Imagine we have a preparation consisting of two disjoint parts one of which prepares a homogeneous state.

**Theorem 3** If a preparation, $γδ$, consists of two disjoint circuit fragments $γ$ and $δ$ which prepare systems of type $a$ and $b$ respectively, and one of these circuit fragments, $γ$, taken by itself prepares a homogeneous state, $p_a^{γ}$, then the state prepared by closing the outputs of the second circuit fragment of $γδ$ is parallel to $p_a^{γ}$ (and therefore also homogeneous).

The proof of this theorem is based on the same idea as the previous theorem. We can put $δ = (F, ϕ, l)$ where $F$ denotes the actual circuit fragment, $ϕ$ the settings, and $l$ the outcomes. Then we can put $δ = (F, ϕ, \tilde{l})$. This is the circuit fragment associated with not seeing $l$. Either $l$ or $\tilde{l}$ must happen, and hence

$$p_{a}^{γ} = p_{a}^{γδ} + p_{a}^{γ\tilde{δ}}$$

(38)

Since $p_a^{γ}$ is homogeneous both $p_{a}^{γδ}$ and $p_{a}^{γ\tilde{δ}}$ must be parallel to it and hence are also homogeneous.
6.4 Fiducial measurements for composite systems

Assume that system 1 is prepared by some preparation $\gamma$ in a homogeneous state $p_1^\gamma$ and, similarly, system 2 is prepared in homogeneous state $p_2^\delta$. Assume these are two separate preparations corresponding to two disjoint circuit fragments. Hence we can consider the joint preparation $\gamma\delta$ giving rise to the state $p_{12}^{\gamma\delta}$.

A natural question is, what is the relationship between the states $p_1^\gamma$, $p_2^\delta$, and $p_{12}^{\gamma\delta}$? By virtue of Theorem 3 we know that the reduced state at either end is homogeneous and hence, by virtue of Theorem 2 (equation (35) in particular) and the fact that the subsystems are in homogeneous states we can write

$$r_{ab}^{\alpha\beta} \cdot p_{ab}^{\gamma\delta} = p^{\gamma\delta\alpha\beta} = \mu_{\gamma\delta} \gamma^\alpha p^{\gamma\delta\beta} = \nu_{\gamma\delta} (r_1^\alpha \otimes r_2^\beta) \cdot (p_1^\gamma \otimes p_2^\delta)$$

where we obtain

$$\nu_{\gamma\delta} = \frac{p^{\gamma\delta\alpha\beta}}{p^{\gamma\delta\alpha\beta}}$$

by putting $\alpha = -$ and $\beta = -$.

We know that there must exist $K_a \equiv |\Omega_a|$ linearly independent homogeneous states for system 1 (they can all be chosen to be pure). Let $\gamma \in \Omega_a$ be the preparations associated with one such set of linearly independent homogeneous states for system 1 (where $|\Omega_a| = |\Omega_a|$). Likewise we have $K_b \equiv |\Omega_b|$ linearly independent homogeneous states for system 2 with preparations $\delta \in \Omega_b$ (where $|\Omega_b| = |\Omega_b|$). The $K_aK_b$ vectors

$$p_1^\gamma \otimes p_2^\delta$$

for $\gamma \in \Omega_a \times \Omega_b$ (41)

are linearly independent and, similarly, the $K_aK_b$ vectors

$$r_1^\alpha \otimes r_2^\beta$$

for $\alpha\beta \in \Omega_a \times \Omega_b$ (42)

are linearly independent. It follows from (39) and some simple linear algebra that the $K_aK_b$ vectors

$$p_{ab}^{\gamma\delta}$$

for $\gamma\delta \in \Omega_a \times \Omega_b$ (43)

are linearly independent as are the $K_aK_b$ vectors

$$r_{ab}^{\alpha\beta}$$

for $\alpha\beta \in \Omega_a \times \Omega_b$ (44)

From this we can prove the following

Theorem 4 For composite systems we can choose $\Omega_{ab}$ such that

$$\Omega_a \times \Omega_b \subseteq \Omega_{ab}$$

(45)

This theorem has an immediate corollary:

Corollary 2 For composite systems $K_{ab} \geq K_aK_b$
Here $K_a = |\Omega_a|$. This inequality follows from the fact that we have at least $K_aK_b$ linearly independent states in $\Omega_a \times \Omega_b$. The set relation (45) follows since the effects in $\Omega_a \times \Omega_b$ are linearly independent and hence that we can choose $K_aK_b$ of the $K_{ab}$ fiducial effects in $\Omega_{ab}$ to correspond to local effects. By a local effect we mean one comprised of disjoint circuit fragments, one on system $a$ and one on system $b$. It follows from (45) that we can write

$$\Omega_{ab} = \tilde{\Omega}_{ab} \cup \tilde{\Omega}_{ab}$$

where $\tilde{\Omega}_{ab} = \Omega_a \times \Omega_b$ (46)

The fiducial effects in $\tilde{\Omega}_{ab}$ are local. Hence, we can write a general bipartite state, with preparation $\varepsilon$, as

$$p_{\varepsilon ab}^a = \tilde{p}_{\varepsilon ab}^a \oplus \tilde{p}_{\varepsilon ab}^a$$

(47)

where the elements of $\tilde{p}_{\varepsilon ab}^a$ are the probabilities corresponding to effects in $\tilde{\Omega}_{ab}$ and the elements of $\tilde{p}_{\varepsilon ab}^a$ are the probabilities corresponding to the effects in $\tilde{\Omega}_{ab}$. Note that $\tilde{p}_{\varepsilon ab}^a$ lives in the tensor product space of the vector spaces for component systems because $\tilde{\Omega}_{ab} = \Omega_a \times \Omega_b$. If systems 1 and 2 are both in homogeneous states then it follows from equation (39) that

$$\tilde{p}_{\varepsilon \gamma \delta}^a = \nu_{\gamma \delta} p_{\varepsilon a}^\gamma \otimes p_{\varepsilon b}^\delta$$

(48)

A similar result holds even if only one system is in a homogeneous state (this follows from theorem 1). These results for bipartite systems generalize in the obvious way to composites having more than two component systems (for three systems we use $\Omega_{abc} = (\Omega_a \times \Omega_b \times \Omega_c) \cup \Omega_{abc}$).

It is easy to see that if each system is subject to its own local transformation (so the circuit fragments corresponding to the transformations are disjoint) then the state updates as

$$p_{\varepsilon \gamma \delta}^a = [(c^\gamma Z_a^\alpha \otimes d^\delta Z_b^\beta) \tilde{p}_{\varepsilon ab}^a] \oplus c^\delta d^\gamma Z_{ab} \tilde{p}_{\varepsilon ab}$$

(49)

where the form of the $c^\delta d^\gamma Z_{ab}$ matrix depends, in general, on the particular theory (it acts on $p_{\varepsilon ab}^a$ to give $\tilde{p}_{\varepsilon ab}^a$). This equation follows by considering the case where both systems are in homogeneous states. Then we have $K_aK_b$ linearly independent vectors $\tilde{p}_{\varepsilon \gamma \delta}^a = \mu_{\gamma \delta} p_{\varepsilon a}^\gamma \otimes p_{\varepsilon b}^\delta$ with $(\gamma \delta \in \Omega_a \times \Omega_b)$. The $\tilde{p}$ part of the state must remain as a tensor product like this after the local transformations to ensure consistency with equation (39) (notice that if it did not then there would exist a correlation-revealing measurement contradicting Theorem 2). But the vectors $p_{\varepsilon a}^\gamma \otimes p_{\varepsilon b}^\delta$ span the space of possible $\tilde{p}_a$ vectors and so (49) must be true generally.

If system $d$ is the null system (so the transformation on system 2 is an effect) then the $Z$ matrix has no elements and we can write

$$p_{\varepsilon \gamma \delta}^a = (c^\gamma Z_a^\alpha \otimes -Z_b^\beta) \tilde{p}_{\varepsilon ab}$$

(50)

(Note that if the $\tilde{Z}$ matrix did have elements then the two sides of this equation would be column vectors of different lengths.) In particular, the reduced state
of system 1 is given by
\[ p_\alpha^\varepsilon \equiv p_\alpha^{\varepsilon a} = [\varepsilon Z_\alpha^0 \otimes (r_\alpha^\varepsilon)^T] p_{ab} \]  
(51)
(recall that $\varepsilon Z_\alpha^0$ is the identity). Hence the reduced state at either end depends only on the elements in the $\tilde{p}_{ab}$ part of $p_{ab}$.

If both systems $c$ and $d$ are null then (50) becomes
\[ p_\epsilon^{\alpha\beta} = (r_\alpha^\alpha \otimes r_\beta^\beta) \cdot \tilde{p}_{ab} \]  
(52)
This also follows directly from (39) and the fact that the vectors $p_\alpha^\alpha \otimes p_\alpha^\beta$ span the space of possible $\tilde{p}_{ab}$ vectors. This equation tells us that all local effects are linear combinations of the fiducial effects corresponding to the $\tilde{\Omega}_{ab}$ part of $\Omega_{ab}$. Hence, all effects $r_{ab}^\gamma$ with $\gamma \in \tilde{\Omega}_{ab}$ are nonlocal: the corresponding circuit fragments cannot consist of disjoint parts acting separately on $a$ and $b$. Theories in which the state can be entirely determined by local measurements are called \textit{locally tomographic}. This gives us an important theorem

**Theorem 5** Theories having $K_{ab} = K_a K_b$ are locally tomographic and vice-versa.

This corresponds to the case where $\tilde{\Omega}_{ab}$ is the null set.

### 6.5 Homogeneity and uncorrelatability are equivalent notions

We define:

\textbf{An uncorrelatable state} is one having the property that a system in this state cannot be correlated with any other system (so that any joint probabilities factorize).

Let $p_{ab}^\epsilon$ be an uncorrelatable state. Let $p_{ab}^\gamma$ be a state for a bipartite system having the property that its reduced state is $p_{ab}^\gamma$. If the bipartite system is prepared in any such state then the joint probabilities will, by definition, factorize. We will prove

**Theorem 6** If we allow arbitrary mixtures then all homogeneous states are uncorrelatable and all uncorrelatable states are homogeneous.

That is homogeneity and uncorrelatability are equivalent notions. It follows immediately from Theorem 2 that homogeneous states are uncorrelatable. To prove that uncorrelatable states are homogeneous we assume the converse. Thus assume that the heterogeneous state $p_{ab}^\varepsilon = p_{ab}^{\gamma_1} + p_{ab}^{\gamma_2}$ (where the two terms are non-parallel) is uncorrelatable. Since the state is assumed to be uncorrelatable we must be able to write the state of the composite system as
\[ p_{ab}^\gamma = \mu_{\gamma} [p_{ab}^\alpha \otimes p_{ab}^\beta] \oplus \tilde{p}_{ab} \]  
(53)
since otherwise the probability does not factorize for all $\alpha\beta \in \Omega_a \times \Omega_b$. Hence

$$p_{\alpha\beta}^\gamma = \mu_{\gamma_1}[p_{\gamma_1 a} \otimes p_{\gamma_1 b}] \oplus \hat{p}_{\alpha\beta}^\gamma + \mu_{\gamma_2}[p_{\gamma_2 a} \otimes p_{\gamma_2 b}] \oplus \hat{p}_{\alpha\beta}^\gamma$$ (54)

But, using (51), we see that the reduced state, $p_{\alpha a}$, for system 1 of (54) is parallel to the reduced state, $p_{\alpha a}'$, of

$$p_{\alpha a}' = \mu_{\gamma_1\delta_1}[p_{\gamma_1 a} \otimes p_{\gamma_1 b}] \oplus \hat{p}_{\alpha\beta}^\gamma + \mu_{\gamma_2\delta_2}[p_{\gamma_2 a} \otimes p_{\gamma_2 b}] \oplus \hat{p}_{\alpha\beta}^\gamma$$ (55)

where we choose any two distinct $p_{\delta_1 a}$ and $p_{\delta_2 a}$ having normalization such that $\mu_{\gamma_1\delta_1} = \mu_{\gamma_2\delta_2}$. It is possible to choose two distinct states like this if there exist systems that are non-trivial in the sense that they require more than one fiducial effect. (If all systems are trivial then all states are homogeneous and so, by Theorem 2, all states are uncorrelatable in any case.) We allow arbitrary mixtures and so can take mixtures with the null state to make sure $p_{\delta_1 a}$ and $p_{\delta_2 a}$ have normalization so that $\mu_{\gamma_1\delta_1} = \mu_{\gamma_2\delta_2}$. The state (55) is preparable by taking a mixture of the preparations $\gamma_1\delta_1$ and $\gamma_2\delta_2$. This state is clearly correlated. By taking a mixture with the null state for the longer of $p_{\gamma_1 a}$ and $p_{\gamma_1 a}'$ we obtain two equal states one of which is uncorrelatable by assumption and one of which is correlatable by the above proof. Hence our assumption was false and it follows that uncorrelatable states are homogeneous.

### 6.6 Probabilities for disjoint circuit

If we have two disjoint setting-outcome specified circuits, $\alpha$ and $\beta$, then expect the joint probability to factorize

$$p^{\alpha\beta} = p^{\alpha} p^{\beta}$$ (56)

A simple application of Bayes’ rule shows that (56) is equivalent to demanding that the probability associated with a circuit is independent of the outcomes seen at other disjoint circuits. This is an extremely natural condition since otherwise we would have to take into account all the outcomes seen on all other disjoint circuits in the past which form a part of our memory before writing down a probability for the circuit. On the other hand, one can easily envisage a situation in which the probability is not independent of outcomes elsewhere. For example, the eventual outcome of a spinning coin might be correlated with the outcome of an apparently disjoint experiment which is, incidently, influenced by photons scattered from the coin while it spins. More generally, if there are hidden variables, then there may be correlations between outcomes even though the marginals are independent of what happens at the other side. It is not clear that disjointness of the circuits is enough to prevent such correlations. In view of this, it is interesting that the following theorem holds.

**Theorem 7** If there exists at least one type of system which can be prepared in a pure state of norm one then the probability associated with any circuit is independent of the settings on any other disjoint circuit.
Let the preparation associated with this pure state of norm one be \( \gamma \). Consider the circuit \((\gamma -)(\gamma -)(\alpha)(\beta)\) consisting of two instances of the circuit obtained by performing the trace effect on a preparation \( \gamma \), and two more disjoint circuits \( \alpha \) and \( \beta \). Thus we have four disjoint circuits in total. We can regard this circuit as consisting of the effect \( -\alpha \) on one of the \( \gamma \) preparations and the effect \( -\beta \) on the other \( \gamma \) preparation. Then we have

\[
p^{(\gamma -)(\gamma -)(\alpha)(\beta)} = p^{\gamma -\alpha}p^{\gamma -\beta}
\]

by Theorem 2 (equation (35) in particular). But \( p^{\gamma -} = 1 \) since the state is of norm one. Hence, using Bayes’ rule, (56) follows and the theorem is proved.

We say that a set of operations is fully closable if it is closable and if the probability for a circuit is independent of the outcomes seen at other disjoint circuits. It follows from Theorem 7 that closable sets of operations admitting at least one pure state of norm one are fully closable. In the case that we have a fully closable set of operations it is clear that we can write the \( \tilde{P}_{ab} \) part of the state associated with disjoint preparations as

\[
\tilde{P}_{ab}^{\gamma \delta} = P_{\alpha}^{\gamma} \otimes P_{\beta}^{\delta}
\]

It is interesting to note that (48) is an example of this with \( \nu_{\gamma \delta} = 1 \) which clearly follows from (40) when probabilities for disjoint circuits factorize.

6.7 Examples of the relationship between \( K_{ab} \), \( K_a \), and \( K_b \)
If \( N_a \) is the number of states that can be distinguished in a single shot measurement then it is reasonable to suppose \( N_{ab} = N_aN_b \). This is true in all the examples we will discuss. In classical probability theory \( K_a = N_a \). In quantum theory \( K_a = N_a^2 \). Hence \( K_{ab} = K_aK_b \) and so, by Theorem 5, we have local tomography in these theories. In real Hilbert space quantum theory, where the state is represented by a positive density matrix with real entries, we have \( K_a = N_a + N(N - 1)/2! \). This has \( K_{ab} > K_aK_b \) which is consistent with Corollary 2. However, quaternionic quantum theory has \( K_a = N_a + 4N(N - 1)/2! \) which has \( K_{ab} < K_aK_b \). This is inconsistent with Corollary 2 and hence quaternionic quantum theory cannot be formulated in this framework. Since we have made very minimal assumptions (only that we have closable sets of operations in an operational framework) it seems that quaternionic quantum theory is simply an inconsistent theory (at least for the finite \( K_a \) case considered here).

7 Theories for which \( K_{ab} = K_aK_b \)
7.1 Motivation for local tomography
Of the examples we just considered, the two corresponding to real physics are both locally tomographic having \( K_{ab} = K_aK_b \). This is a very natural property for a theory to have (it is one of the axioms in [4]). It says that, from a counting
point of view, no new properties come into existence when we put two systems together. It allows a certain very natural type of locality so that it is possible to characterize a system made from many parts by looking at the components. It implies that the full set of states for a composite system requires the same number of parameters for its specification as the separable states (formed by taking mixtures of states prepared by disjoint circuits). Given that this is such a natural constraint we will study it a little more closely. We will also give axioms for classical probability and quantum theory since they are examples of this sort.

7.1.1 Operation locality

An extremely useful property of locally tomographic theories is that they are local in the sense that the state is updated by the action of local matrices at each operation. We will call this property operation locality. We see from (47) that if $\tilde{\Omega}_{ab} = \emptyset$ then $\tilde{p}^\gamma_{ab} = 0$ and $p^\gamma_{ab} = \tilde{p}^\gamma_{ab}$ and hence according to (49) we see that under local transformations at each end (corresponding to disjoint circuit fragments) the state will update as

$$p^\gamma_{\alpha\beta} = (c^\alpha Z^\alpha_a \otimes d^\beta Z^\beta_b) p^\gamma_{ab}$$

(59)

Hence

$$cdZ^\alpha\beta_{ab} = c^\alpha Z^\alpha_a \otimes d^\beta Z^\beta_b$$

(60)

for transformations corresponding to disjoint circuit fragments. In particular this implies

$$cb^\alpha 0_{ab} = c^\alpha Z^\alpha_a \otimes b^\alpha 0_{b}$$

(61)

This is equation (30) we speculated about in Section 6.2. This means an operation has a trivial effect on systems that do not pass through it. If we have a fully closable set of transformations (as long as there exists a least one state of norm-one this follows from Theorem 7) then we can specialize this equation to the case of null input states (where $a = -$ and/or $b = -$) since the state prepared by disjoint preparations is a product state. We will assume this in what follows.

The great thing about (60) is that it can be used to calculate the probability for any circuit using a $Z$ matrix for each operation. To do this we choose a complete foliation and then use the tensor product to combine operations at each time step. One way of calculating the probability $p^{\alpha\beta\gamma\delta\epsilon\zeta}$ for the example shown in Fig. 10 is

$$Z^\zeta_{fg}(f^\epsilon Z^\epsilon_f \otimes g^\zeta Z_{cd})(f^\delta Z^\delta_f \otimes g^\epsilon Z_{bc}) (f^\gamma Z^\gamma_f \otimes g^\delta Z_{ab}) (f^\zeta Z^\zeta_f \otimes g^\gamma Z_{cd} \otimes d^\beta Z^\beta_b \otimes c^\beta 0_{c} \otimes d^\alpha Z^\alpha_a \otimes d^\alpha Z^\alpha_d)$$

(62)

While it is very satisfying that the calculation can be broken down like this, it is unfortunate that we have to pad out the calculation with lots of identity matrices like $c^\alpha Z^\alpha_a$. This means that there are more matrices than operations in this calculation. Relatedly, we have to be very careful what order we take the product of all these matrices (it has to correspond to some complete foliation).
Figure 10: We show how to calculate the probability associated with this simple example in the text using a $Z$ matrix for each operation.

In the causaloid approach [4, 7, 5] we will have neither of these problems. We simply take what is called the causaloid product of a vector associated with each operation without regard for the order and without having to pad out the calculation with identity matrices.

7.2 Classical probability theory

It is very easy to characterize classical probability theory in this framework. It is fully characterized by the following two axioms:

**Composition** $K_{ab} = K_a K_b$

**Transformations** Transformation matrices, $Z^a$, have the property that the entries are nonnegative and the sum of the entries in each column is less than or equal to 1.

To see this is equivalent to usual presentations of classical probability theory note the following. We can interpret $K_a = N_a$ as the maximum number of distinguishable states for this classical system (for a coin we have $N_a = 2$, for a die $N_b = 6$). The state is given by a $p^a = Z^a$ and is a column vector. The
sum of the entries in this vector must be less than or equal to 1. They can be interpreted as the probabilities associated with each of the distinguishable outcomes (for a die they are the probabilities associated with each face). The trace effect is given by \((r_\alpha^\top) = Z_\alpha\). It is a row vector. The value of each entry is 1 and this is consistent with the constraint that the sum of the columns cannot be greater than 1. Norm preserving transformations are stochastic matrices. Since \(K_{ab} = K_a K_b\), we have the operation locality property and so we can calculate the probability for an arbitrary circuit from matrices for the operations that comprise it.

7.3 Quantum theory

To give the rules for Quantum Theory we need a few definitions first. Let \(\mathcal{H}_{N_a}\) be a complex Hilbert space of dimension \(N_a\). Let \(\mathcal{V}_{N_a}\) be the space of Hermitian operators that act on this. All positive operators are Hermitian. Furthermore, it is possible to find a set of \(N_a^2\) linearly independent positive operators that span \(\mathcal{V}_{N_a}\). Let \(\hat{P}_a^\alpha\) for \(\alpha \in \Omega_a\) be one such set. Define

\[
\hat{P}_a = \begin{pmatrix}
\vdots \\
\hat{P}_a^\alpha \\
\vdots 
\end{pmatrix} \quad \alpha \in \Omega_a
\]  

(63)

A positive map \(\hat{\mathcal{S}}_a\) from \(\mathcal{V}_{N_a}\) to \(\mathcal{V}_{N_c}\) is one which acts on a positive operator \(\rho_a \in \mathcal{V}_{N_a}\) and returns a positive operator \(\rho'_c \in \mathcal{V}_{N_c}\) for any positive operator \(\rho_a\). The map \(\hat{\mathcal{S}}_a\) is completely positive if \(\hat{\mathcal{S}}_a \otimes I_b\) is a positive map from \(\mathcal{V}_{N_a N_b}\) to \(\mathcal{V}_{N_c N_b}\) for any \(b\) where \(I_b\) is the identity map on \(\mathcal{V}_{N_b}\). Further, we want our maps to have the property that they do not lead to probabilities greater than 1. We will demand that they must be completely trace non-increasing when they act on density matrices. This means that \(\hat{\mathcal{S}}_a \otimes I_b\) must be trace non-increasing for any \(b\). Quantum theory is fully characterized by the following two axioms.

Composition \(K_{ab} = K_a K_b\)

Transformations Transformation matrices are of the form

\[
\hat{Z}_a^\alpha = \text{Trace}(\hat{P}_c \hat{\mathcal{S}}_a^\alpha (\hat{P}_a^T)) [\text{Trace}(\hat{P}_a \hat{P}_a^T)]^{-1}
\]  

(64)

where \(\hat{\mathcal{S}}_a\) is completely positive and completely trace non-increasing and \(T\) denotes transpose.

This is a much more compact statement of the rules of quantum theory than is usually given. We will make a few remarks to decompress this. First note that \(\text{Trace}(\hat{P}_c \hat{\mathcal{S}}_a^\alpha (\hat{P}_a^T))\) is a \(K_c \times K_a\) matrix having \(\beta \gamma\) element \(\text{Trace}(\hat{P}_c \hat{\mathcal{S}}_a^\alpha (\hat{P}_a^T))\) with \(\beta \in \Omega_c\) and \(\gamma \in \Omega_a\). By defining

\[
p_a^\delta = \text{Trace}(\hat{P}_a \rho_a^\delta)
\]  

(65)
where $\rho_a^\delta$ is the usual quantum state, and using $\tilde{\rho}_a^\alpha = \tilde{P}_a \cdot s^\alpha$ (since $\tilde{\rho}_a$ must be
given by some sum of the linearly independent spanning set), it can be shown
after a few lines of algebra that

$$p^{\delta \alpha} = c Z^\alpha p_a^\delta \iff \tilde{\rho}_a^{\delta \alpha} = c s^\alpha (\tilde{\rho}_a^\delta) \quad (66)$$

Hence we get the correct transformations with (64). Note that, as in the classical case, we can write a state as $p_a^\alpha = a Z^\alpha$. This state is associated with a
completely positive map $\alpha Z^\alpha$ where the absence of an input label implies that
we have a null input system which corresponds to a one dimensional Hilbert
space. This must have trace less than or equal to one since otherwise $a Z^\alpha \otimes b I_b$
would be trace increasing (this is the reason we impose that the map should
be completely trace non-increasing rather than just trace non-increasing). Also
note that for $a Z^\alpha$ (the identity map) we get the identity for $a Z^\alpha$ in (64) as we
must. The composition rule $K_{ab} = K_a K_b$ implies that $N_{ab} = N_a N_b$ (since we
see from inspection of the rank of the matrices that $|\Omega_a| = N^2_a$) and hence the
tensor product structure for Hilbert spaces corresponds to the tensor product
structure discussed in Sec. 7.1.1. The fact that we have $K_{ab} = K_a K_b$
means that we have the operation locality property. We can calculate the probability
for a general circuit using these $Z$ matrices for each of the operations. Hence
our list of postulates for Quantum Theory is complete.

### 7.4 Reasonable postulates for quantum theory

The objective of this chapter has been to set up a general probabilistic frame-
work. It is worth mentioning that we can give the following very reasonable
postulates which enable us to reconstruct quantum theory within this framework.

**Information** Systems having, or constrained to have, a given information car-
rying capacity have the same properties.

**Composites** Information carrying capacity is additive and local tomography
is possible (i.e. $N_{ab} = N_a N_b$ and $K_{ab} = K_a K_b$).

**Continuity** There exists a continuous reversible transformation between any
pair of pure states.

**Simplicity** Systems are described by the smallest number of probabilities con-
sistent with the other postulates.

We can show from the first two postulates that $K = N^r$ where $r = 1, 2, \ldots$.
The continuity postulate rules out the classical probability case where $K = N$.
The simplicity postulate then implies that we have $K = N^2$. We construct the
Bloch sphere for the $N = 2$ case using, in particular, the continuity postulate.
Then the information postulate and composites postulate are used to obtain
quantum theory for general $N$. We refer the reader to [6, 8] for details.
8 Conclusions

We have exhibited a very natural framework for general probabilistic theories in an operational setting. We represent experiments by circuits and have been particularly careful to give operational interpretations to the elements of these circuits (operations are single uses of an apparatus and wires represent apertures being placed next to one another). By considering closable sets of operations we are able to introduce probabilities and then to set up the full theory wherein $Z$ matrices are associated with circuit fragments. The special case of locally tomographic theories has the operation locality property so that we can combine $Z$ matrices corresponding to circuit fragments that are in parallel using the tensor product. This enables us to break down a calculation into smaller parts. The framework here is still lacking. Most crucially, it is only able to take into account one particular way in which operations can be connected (corresponding to placing apertures next to each other) but there are many other ways. A more general theory is under development to allow more for other types of connections (see [5]). The framework is discrete and hence is not readily adaptable to quantum field theory. It would be very interesting either to develop a continuous version or to show how quantum field theory can be fully understood in such a discrete framework. Algebraic quantum field theory can be understood in operational terms (see for example Haag [40]). However, putting the issue of discreteness aside, it is a rather less general operational theory than that presented in this chapter and so there may be advantages to studying quantum field theory in the framework presented here. It is worth saying that there is a tension between operationalism and use of the continuum in physics. From an operational point of view, the continuum is best understood as a mathematical tool enabling us to talk about a series of ever more precise experiments. It is possible that such a series may, eventually, be better described with other mathematical tools.

There are two types of motivation for considering general probabilistic theories. First, we may be able to better formulate and understand our present theories within these frameworks. It may be possible to write down a set of postulates or axioms which can be used to reconstruct these theories within such a framework. For the case of quantum theory there has been considerable work of this nature already. It would be interesting to see something similar for general relativity. In particular, there ought to be a simple and elegant formulation of general relativity for the case where there is probabilistic ignorance of the value of quantities that might be measured in general relativity, (let us call this probabilistic general relativity). Such a theory might be best understood in a general probabilistic framework (though probably more general than the one presented in this chapter) [5]. The second reason to consider general probabilistic theories is to try to go beyond our present theories. The most obvious application would be to work towards a theory of quantum gravity (see [4, 7]). The program of constructing general probabilistic theories and then constraining them using some principles or postulates may free us from the hidden mathematical obstacles to formulating quantum gravity that stand in the way of the more
standard approaches such as string theory and loop quantum gravity.

Acknowledgements

Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI. I am grateful to Vanessa Hardy for help with the manuscript.

References

[1] A. Einstein. Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, **17** (1905) 891-921.

[2] C.A. Fuchs. Quantum mechanics as quantum information (and only a little more). *arXiv:quant-ph/0205039* (2002). [http://arxiv.org/abs/quant-ph/0205039](http://arxiv.org/abs/quant-ph/0205039).

[3] Conference: Reconstructing quantum theory, organised by Philip Goyal and Lucien Hardy. PIRSA:C09016 (2009). [http://pirsa.org/c09016](http://pirsa.org/c09016).

[4] L. Hardy. Probability theories with dynamic causal structure: A new framework for quantum gravity. *arXiv:gr-qc/0509120* (2005). [http://arxiv.org/abs/gr-qc/0509120](http://arxiv.org/abs/gr-qc/0509120).

[5] L. Hardy. Operational structures as a foundation for probabilistic theories. PIRSA:09060015 (2009). [http://pirsa.org/09060015](http://pirsa.org/09060015).

[6] L. Hardy. Quantum theory from five reasonable axioms. *arXiv:quant-ph/0101012* (2001). [http://arxiv.org/abs/quant-ph/0101012](http://arxiv.org/abs/quant-ph/0101012).

[7] L. Hardy. Towards quantum gravity: A framework for probabilistic theories with non-fixed causal structure. *Journal of Physics, A40* (2007) 3081-3099.

[8] L. Hardy. Operational structures and natural postulates for quantum theory. PIRSA:09080011 (2009). [http://pirsa.org/09080011](http://pirsa.org/09080011).

[9] R. Sorkin. Spacetime and causal sets. In J. DOlivo et. al., editor, Relativity and Gravitation: Classical and Quantum. World Scientific, (1991).

[10] R.B. Griffiths. Consistent histories and the interpretation of quantum mechanics. *Journal of Statistical Physics*, **36** (1984) 219-272.

[11] M. Gell-Mann and J.B. Hartle. Classical equations for quantum systems. *Physical Review D*, **47** (1993) 3345-3382.

[12] R. Omnès. The Interpretation of quantum mechanics. Princeton Univ. Press, 1994.
[13] J.B. Hartle. Spacetime quantum mechanics and the quantum mechanics of spacetime. *Gravitation and Quantizations: Proceedings of the 1992 Les Houches Summer School*, ed. by B. Julia and J. Zinn-Justin, North Holland, Amsterdam, (1995).

[14] F. Markopoulou. Quantum causal histories. *Classical and Quantum Gravity*, **17** (2000) 20592077.

[15] R.F. Blute, I.T. Ivanov, P. Panangaden. Discrete quantum causal dynamics. *International Journal of Theoretical Physics* **42** (2003) 2025-2041.

[16] P. Panangadan. Discrete quantum causal dynamics. PIRSA:09060029 (2009). [http://pirsa.org/09060029/](http://pirsa.org/09060029/)

[17] M. Leifer. Quantum causal networks. PIRSA:06060063 (2006). [http://pirsa.org/06060063](http://pirsa.org/06060063)

[18] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science (LICS 04)*, (2004) 415425.

[19] B. Coecke. Where quantum meets logic, . . . in a world of pictures! PIRSA:09040001 [http://pirsa.org/09040001](http://pirsa.org/09040001)

[20] G. Mackey. *Mathematical Foundations of Quantum Mechanics*. Benjamin (1963).

[21] G. Ludwig, *An Axiomatic Basis of Quantum Mechanics* volumes 1 and 2. Springer-Verlag (1985, 1987).

[22] E. B. Davies and J. T. Lewis, An operational approach to quantum probability, *Communications of Mathematical Physics* **17** (1970) 239-260.

[23] H. Araki. On a characterization of the state space of quantum mechanics. *Communications of Mathematical Physics* **75** (1980) 124.

[24] S. Gudder, S. Pulmannová, S. Bugajski, and E. Beltrametti. Convex and linear effect algebras. *Reports on Mathematical Physics* **44** (1999) 359-379.

[25] D. J. Foulis and C. H. Randall, Empirical logic and tensor products, in H. Neumann, (ed.), *Interpretations and Foundations of Quantum Theory*, Bibliographisches Institut, Wissenschaftsverlag, Mannheim (1981).

[26] J. Barrett. Information processing in generalized probabilistic theories. *Physical Review A* **75** (2007) 032304.

[27] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Cloning and broadcasting in generic probabilistic models. arXiv.org [quant-ph/0611295](http://arxiv.org/abs/0611295) (2006).

36
[28] H. Barnum, J. Barrett, M. Leifer and A. Wilce. A general no-cloning theorem, Phys. Rev. Lett. 99 240501 (2007).

[29] H. Barnum, J. Barrett, M. Leifer, and A. Wilce. Teleportation in general probabilistic theories. arXiv.org:0805.3553, (2008). [http://arxiv.org/abs/0805.3553](http://arxiv.org/abs/0805.3553)

[30] H. Barnum and A. Wilce. Information processing in convex operational theories. Proceedings of QPL-DCMIV, Electronic Notes in Theoretical Computer Science (2008).

[31] H. Barnum and A. Wilce. Ordered linear spaces and categories as frameworks for information-processing characterizations of quantum and classical theory. arXiv:0908.2354 (2009). [http://arxiv.org/abs/0908.2354](http://arxiv.org/abs/0908.2354)

[32] W.K. Wootters. Local accessibility of quantum states, in Complexity, entropy and the physics of information edited by W. H. Zurek (Addison-Wesley, 1990)

[33] W.K. Wootters, Quantum mechanics without probability amplitudes. Foundations of Physics 16 (1986) 391.

[34] S. Popescu and D. Rohrlich. Nonlocality as an axiom. Foundations of Physics 24 (1994) 379.

[35] M. Pawlowski, T. Paterek, D. Kazlikowski, V. Scarani, A. Winter and M. Źukowski, et al. A new physical principle: Information causality. Nature 461 (2009) 1101.

[36] D. Gross, M. Mueller, R. Colbeck, O.C.O. Dahlsten. All reversible dynamics in maximally non-local theories are trivial. arXiv:0910.1840v1 (2009). [http://arxiv.org/abs/0910.1840](http://arxiv.org/abs/0910.1840)

[37] G.M. D’Ariano. How to Derive the Hilbert-Space formulation of quantum mechanics from purely operational axioms. arXiv:quant-ph/0603011 (2006). [http://arxiv.org/abs/quant-ph/0603011](http://arxiv.org/abs/quant-ph/0603011)

[38] G. Chiribella, G.M. D’Ariano, P. Perinotti. Probabilistic theories with purification. arXiv:0908.1583 (2009). [http://arxiv.org/abs/0908.1583](http://arxiv.org/abs/0908.1583)

[39] D. Gillies, Philosophical theories of probability, Routledge (2000).

[40] R. Haag, Local quantum physics: Fields, particles, algebras. Springer (1992).