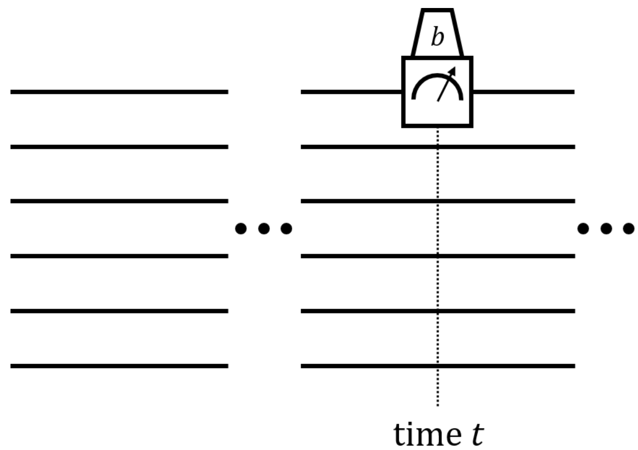
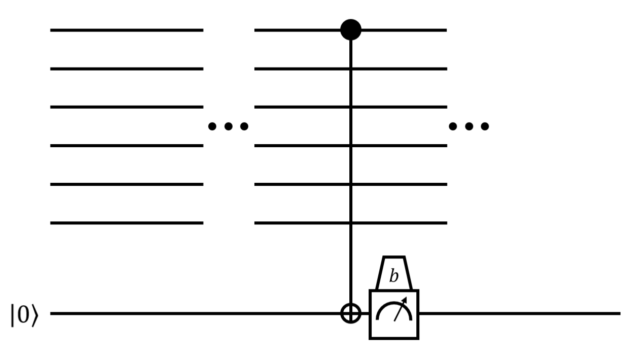
2. **[Principle of Deferred Measurement.]** The point of this problem is to show that if one has a quantum circuit with (partial) measurement gates in the middle, one can (without much loss in efficiency) replace it with an equivalent quantum in which all the measurement gates are at the end. This is nice, because a very useful simplifying assumption in quantum computation is that measurement gates only occur at the end of the computation.

So suppose we have some n-qubit quantum circuit, and we look at the first intermediate measurement gate that is applied; say (without loss of generality) it is applied to the 1st qubit, at time step t. Let $|\psi\rangle$ denote the quantum state just prior to time t. Now when the measurement gate is applied, two things happen: First, one classical bit of information — call it b — appears on the measurement gate's readout. Second, the state collapses according to the usual rules.



(a) [$**$] Suppose we do the following: First, we introduce a new $(n+1)$st qubit, initialized in the state $|0\rangle$. Second, we replace the measurement gate on qubit #1 at time t with a CNOT gate whose control qubit is #1 and whose target qubit is #$(n+1)$. Finally, we immediately apply a measurement gate to the $(n+1)$st qubit, and treat its readout as "b". Assume we then henceforth ignore the (collapsed) $(n+1)$st qubit. Show that this gives an exact simulation of the original circuit's operation. (Hint: you may want to somehow write $|\psi\rangle = \alpha\,|0\rangle \otimes |\psi_0\rangle + \beta\,|1\rangle \otimes |\psi_1\rangle$.)

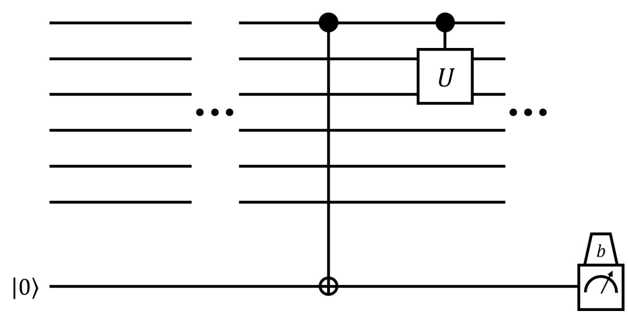


Remark: As we saw in class, operations that are applied to disjoint sets of qubits commute (this was the $(U\otimes I)\cdot(I\otimes V) = (I\otimes V)\cdot(U\otimes I) = U\otimes V$ stuff in the case of applying unitary gates, and similarly for the commuting of partial measurements). Thus, if we're never going to do anything with that $(n+1)$st qubit again, we can imagine that instead of measuring it immediately (just after time t), we instead delay its measurement to the very end of the computation. In this way, we've effectively deferred the first intermediate measurement of the quantum circuit to the end. By repeating this for all intermediate measurement gates, we can always move all measurement gates to the end (at the cost of adding one extra qubit and CNOT per deferral).

(b) [$**$] Let's look back at the original circuit and see "what was done" with the first qubit after it was measured. In some cases, that measurement gate was there because we genuinely wanted to know the 1 bit of classical information, b. In other cases, we don't care about b's value per se; rather, we just want to do different subsequent quantum operations to the other qubits, depending on whether $b = 0$ or $b = 1$. (The Quantum Teleportation scenario is a bit like this.) In other words, the rest of the quantum circuit might include something like

$$\textit{do } U \in \mathbb{C}^{4\times 4} \textit{ to qubits } 2 \textit{ and } 3 \textit{ if } b = 1, \textit{ else do nothing.}$$

Given that the post-measurement qubit's state is precisely $|b\rangle$, one can instead think of the above conditional-instruction as a "controlled-U gate applied to qubits 1 (control), 2 and 3 (targets)", rather than as some interactive intervention wherein U is applied or not applied, depending on the readout b.



So suppose we're in this case, where we don't really care to know b, we're simply doing some "controlled-U" quantum gates based off the outcome. And suppose we apply the Deferred Measurement trick from part (a). Since we don't actually care to know the classical bit b, do we really have to do the measurement of the $(n+1)$st qubit at the end? (I.e., will the circuit work just as well if we ignore that qubit?) If yes, give an example illustrating the necessity. If not, answer this: do we really have to do the CNOT, either?