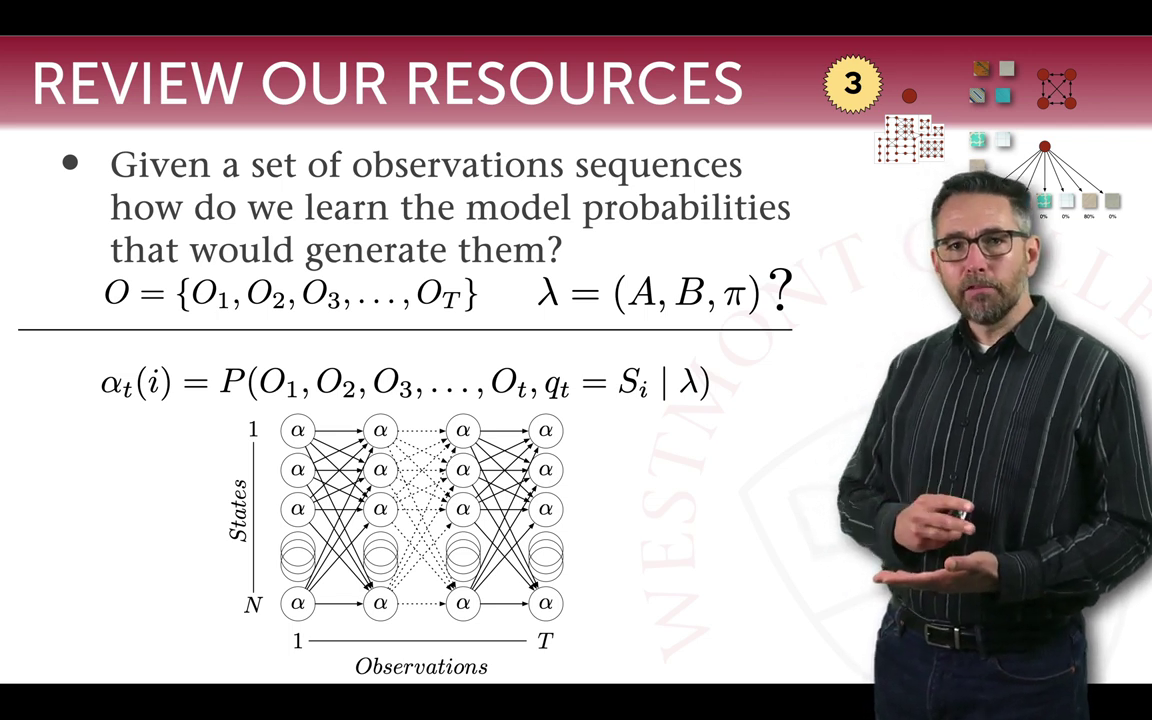
**Baum Welch Method (Training a Hidden Markov Model)**

Q: Given a set of observation sequences, how do we learn the model probabilities that can generate the sequence? In other words:

Given O={O1, O2, …, OT} what are the values of π, A and B for λ(π, A, B)?

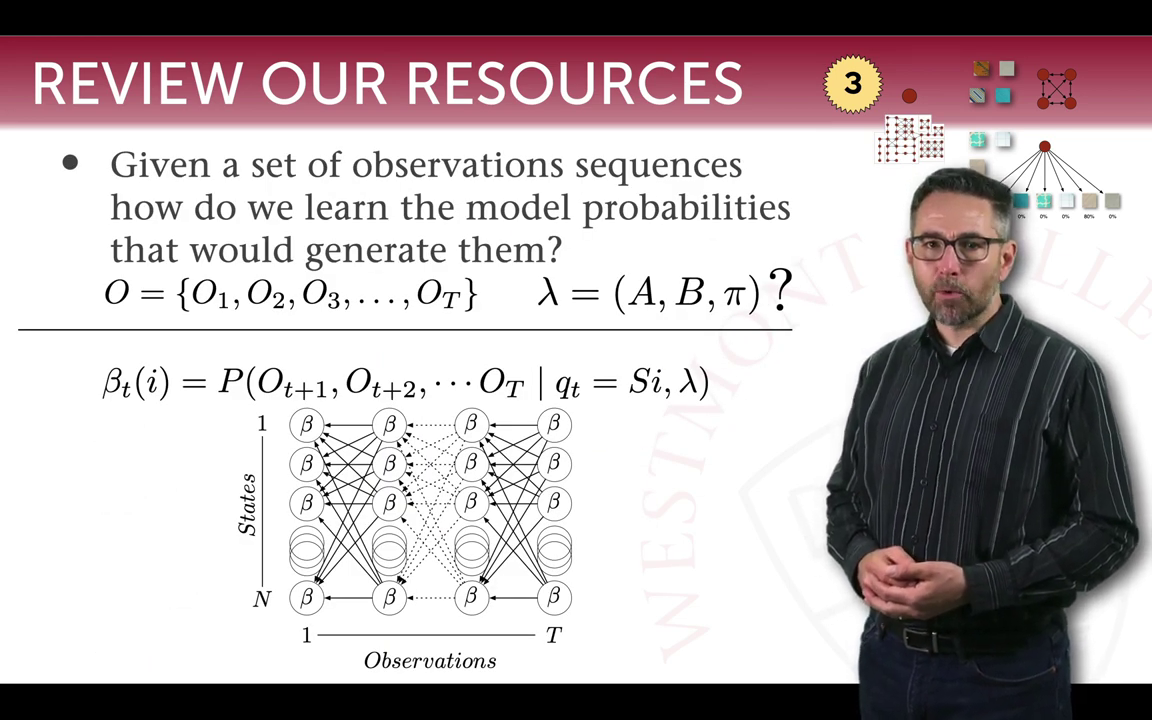
Let us consider the forward probability

αt(i) = P(qt=Si | {O1,…, Ot}, λ)



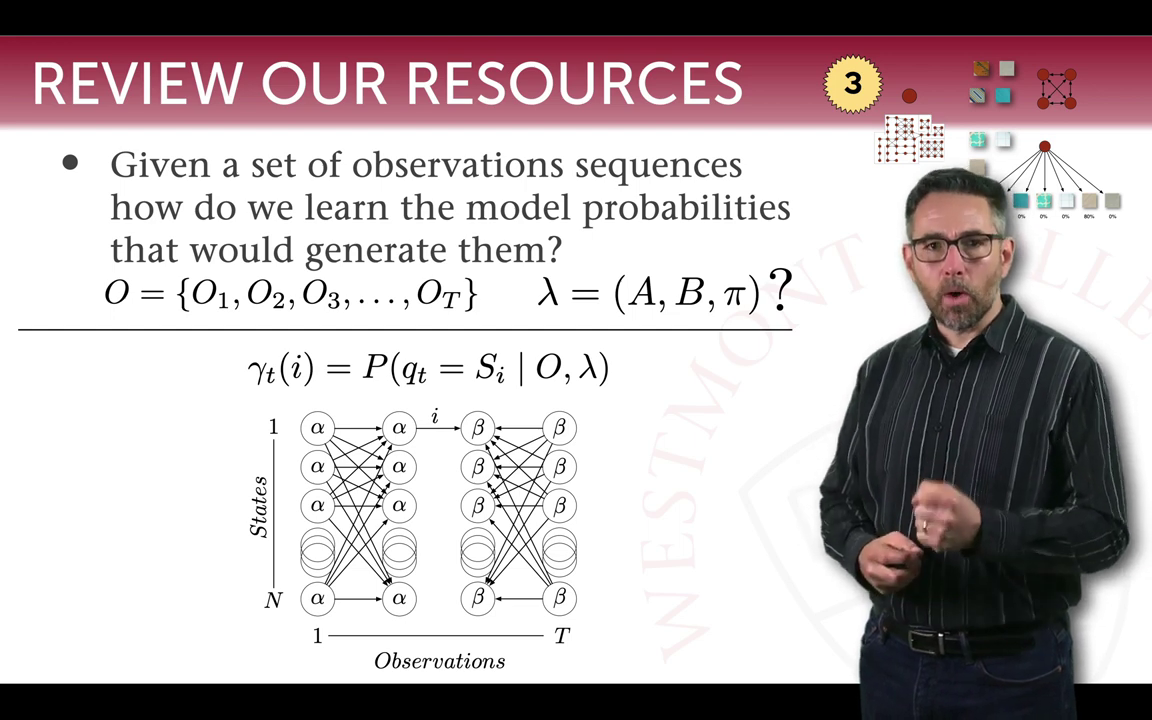
Similarly, the backward probability can be given as:

β t(i) = P({Ot+1,…,OT}| qt=Si , λ)



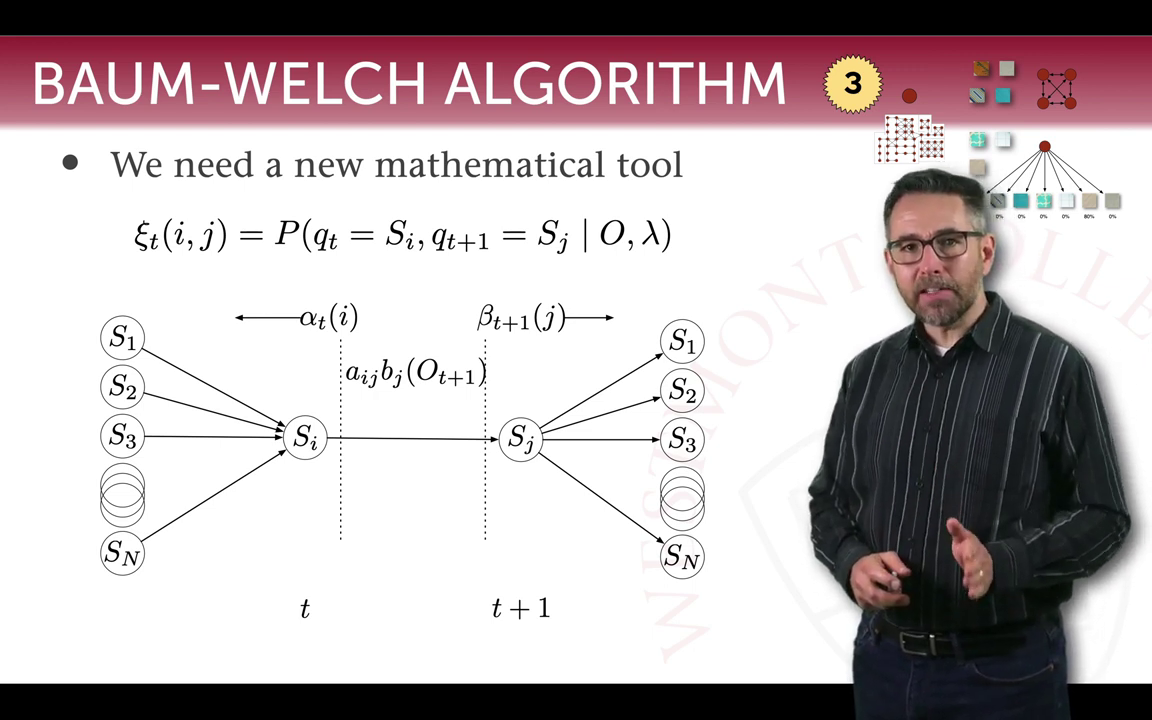
Now, considering the forward and backward parameters, we can derive

γ t(i) = P(qt=Si | O, λ)



We also add another parameter

ξ t(i,j) = P(qt=Si , qt+1=Sj | O, λ)



ξ t(i,j) =

And ξ t(i,j) is related to γ t(i) as

γ t(i) =

Since γ t(i) is the probability of being in state Si at time t, if we sum over all t, we get the expected number of times Si is ever visited.

And since ξ t(i,j) is the probability of transitioning from Si to Sj at time t, if we sum over all t, we get the expected number of times Si ever transitions to Sj.

So how do we get our desired λ from all this?

Well, we can write that = γ 1(i).

And

We iteratively update λ(π, A, B) until convergence criteria is met and terminate.

Predicting: Find the most probable sequence of hidden states using *Viterbi* algorithm (Decoding), then find the next most likely state transition and the most likely observation given the state transitioned to.