NPTEL

NPTEL ONLINE CERTIFICATION COURSE

REINFORCEMENT LEARNING

POMDP Introduction

with

Prof.Balaraman Ravindran

Department of Computer Science and Engineering

Indian Initiation of technology, Madras

So far right so we have been making an assumption that the, the agent is operating in some kind of an MDP right and then we are learning policies for trying to solve it optimally and so on so forth so one thing is that we assume there at every point of time the agent has access to the full state information. The agent knows exactly where it is in the state space and we had assumed that the state space is Marco right.

So and as long as a chain knows where it is in the state currently it has sufficient information to make decisions so did be violated at any point of time. The assumption that agent knows exactly where it is in the state, paramedics when we are talking about value functions and stuff like that right but if you think about it we still assume that the agent knows where it is in the state space.

Except that when it looks up the value function it looks it up in a parameter ice form right but still the age it was aware of where it was in there in the state space and then it just looked it up in a parameterized wherever is not like the agent did not have information about where it was in the state space at all times we assume that the agent knew where it was right so,

So that is the crucial assumption we have made also now right but it turns out that such an assumption is often not valid in practice right because of a variety of reasons so there was one instance where I did in the discussion so if we did violate this whole knowing where you are for sure thing. I mean remember we are talking about state space here right so when you talked

about hierarchies, so there was an instance where we actually were willing to give up information about where we are exactly.

So that we can learn more compact policies right I was telling you if you have to so when you are solving the navigate problem you do not really care whether the passenger is in the car or which one of the RGB why locations the passenger isn't right you just want to navigate to reach there so I do not have to worry about the passenger location not do I have to worry about the passenger destination in fact I gave up information right so that I could have a more compact representation for the navigate policy.

It is easier to learn and so on so forth so in some sense even though I gave up on knowing exactly where I am in the state space right it was not a great loss because with regard to the policy of learning how to navigate to our or how to navigate to why the features that I ignored were actually irrelevant right so even though I could not make a distinction between whether the passenger is in the car or is that are why solving navigate but that is only within the context of that navigate option or navigate subtask and it is fine because a navigate scepter doesn't need it but still in the global context.

We assume that we know exactly where the passenger we just chose to ignore it here right so all of this is fine so now let us go back and talk about a robot navigation problem so if you think about in any standard robot that is actually trying to navigate what are the kind of features the robot will be getting so typically there is some kind of sonar is all around it and there are some bump sensors and things like that so it is possibly going to get some sonar reading so it like okay the nearest obstacle in that direction is one meter away in this direction is two point three meters away in that direction is some point eight meters away.

And maybe there is an obstacle at the back or the wall at the back some three meters away or something like this is all the information is going to get right and that's incredibly confusing to it right so it won't be able to make distinguish between say standing here that there's a wall behind me that's a wall in front of me and all of that right and then standing here at possibly am get still getting the same kind of signals.

You know I might send us on our this face on our that way right and then say that okay there are openings on both sides something is close to in front of me something is close behind me right so

that is basically all the information I am getting right so it is not that I'm choosing to ignore information here right it is just that the information is not even being measured it is too expensive to measure.

I need to put in a lot I mean much, much more you know sophisticated sensing equipment so that is going to be more expensive right so, so I am not able to even since enough information to tell me where I am think of another case where I could look around and let us say know everything right but suppose I am in one of those you know modern-day office buildings we have a lot of cubicles in them right a turn on the robot it looks around it can see they are cubicle so it knows exactly where it is in the cubicle.

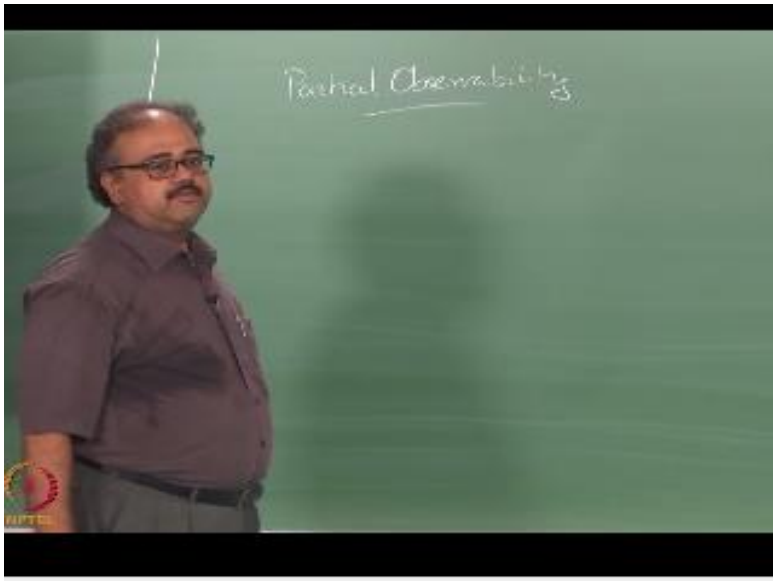
Right but there is no way of knowing where the cubicle is with reference to the rest of the building because all the cubicles look the same unless somebody has managed to put up some poster or something like that inside the cubicle all the cubicles are going to look the same so you look around I know that I am in a cubicle I know where I am in the cubicle but I do not know which cubicle I am in until I start moving around right in or if I have a GPS coordinates and I know the exact GPS locations of all the cubicles but that is another, another kind of sensor that you need to have right.

They so think of humans and we'll be talking about robots so far but think of animals right we suffer perceptual lacing all the time this is called perceptual lacing so I get the same perception but from multiple different places right so it's like the, the perception is earliest just like in signals right and humans suffer from perception of lacing all the time so we continuously keep refining where we think we are you know especially if you put you put you in this one of those confusing office buildings right so you keep walking around and then see how okay now here I saw a mostly most, most things most efficient in tank and things to useful stuff like toilets and so on so forth right okay.

So okay, okay now I saw that that's where the men's toilet is again now I know where I am so kind of thing in fact whenever I go to one of this big IT Park like campuses I invariably get lost so unless there is somebody guiding your own it is very hard to find your way around this the links right. They are built entirely to be characterless right so that all the buildings says it makes it easy for you to shift companies I suppose all your office's look the same anyway so this kind of so what is happening here is some kind of partial information is available to you about the state.

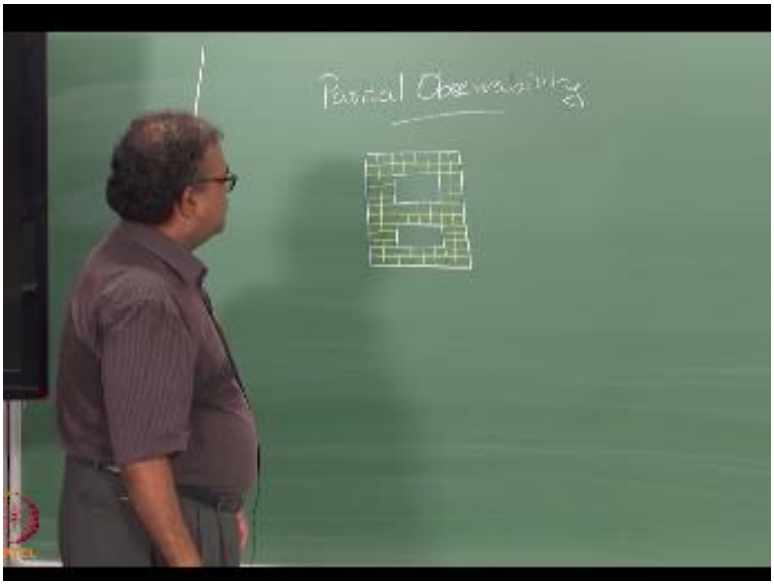
Right so I do not have the full information about the state they only have partial information about the state so we call this

(Refer Slide Time: 7.54)



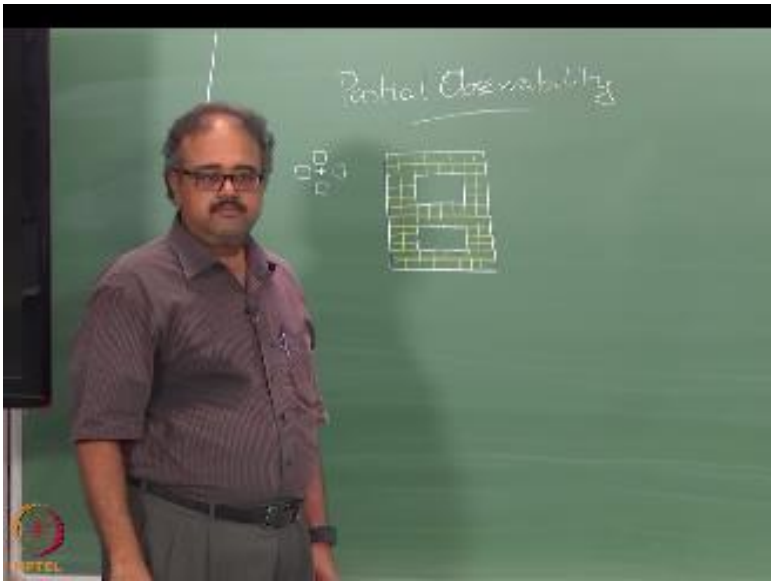
Partial observe ability right so if you try typing in absorbility in any of your editors which does spell check it will tell you it is not a word right but yeah it's a technical term is also so partial observe ability is essentially not having full information about the state space. So let us let us take a simple example

(Refer Slide Time: 8.31)



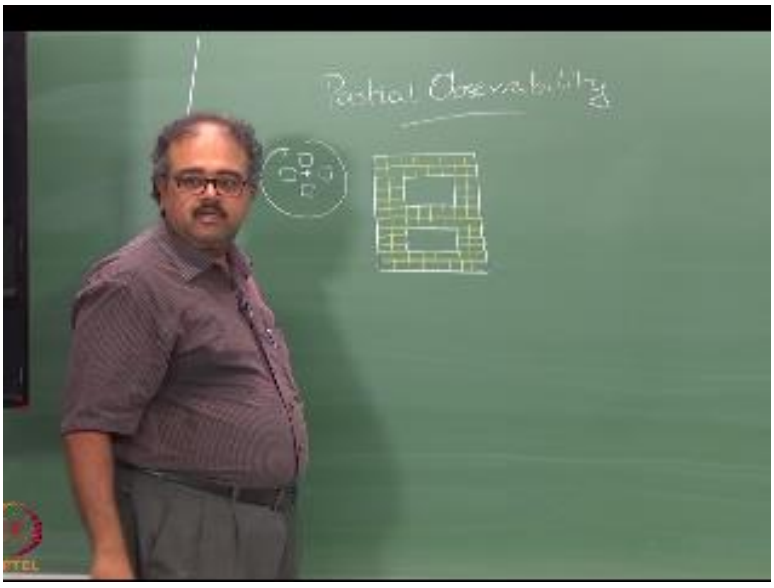
So the time being assume all the cells are of the same size okay so I do not want to erase and redraw all the cells are of the same size. So an agent has a partial information it can sense the following okay it can sense if the suppose

(Refer Slide Time: 9.57)



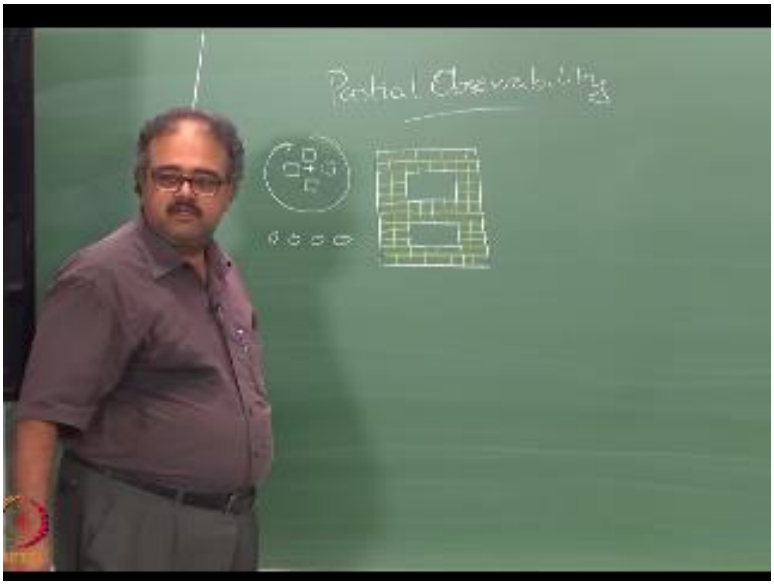
An agent is here right it can sense this, this square this square this coil so these are the four things it can sense okay.

(Refer Slide Time: 10.06)



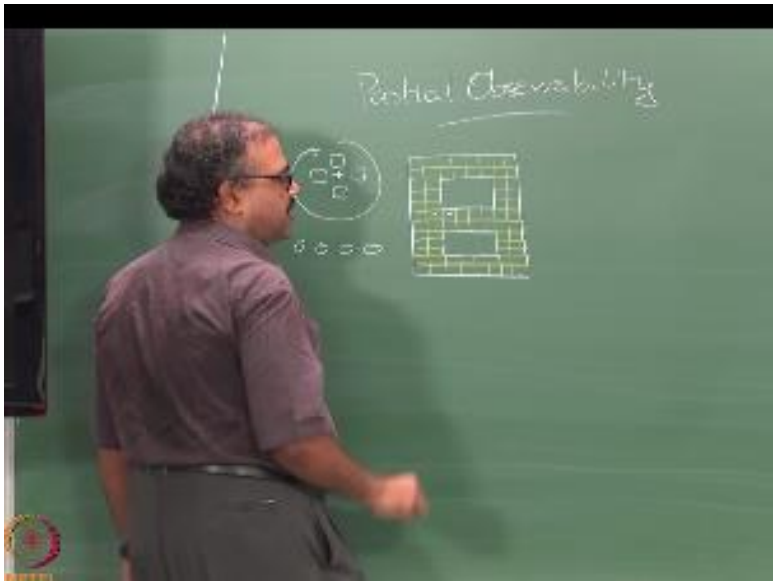
Let us say it goes in this order, so if all of them mark clear all the four places are clear I am going to say so

(Refer Slide Time: 10.21)



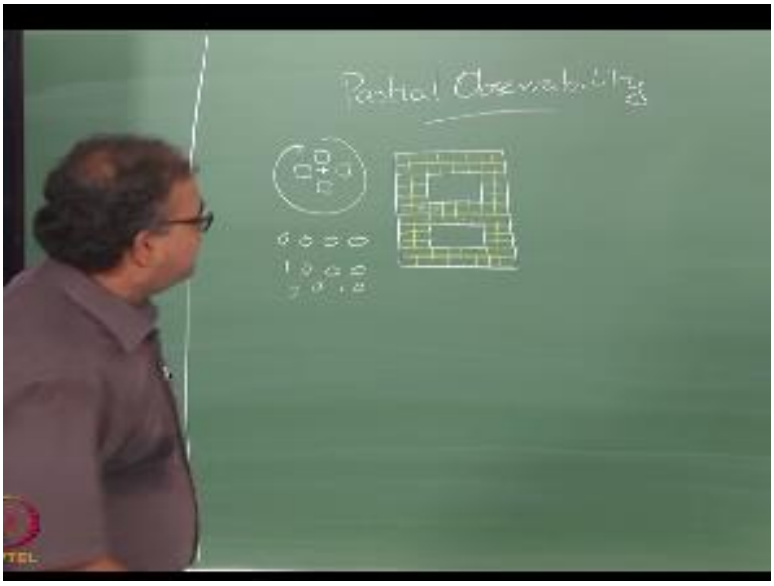
0, 0, 0, 0 that means there are no obstacles anywhere around me in my immediate vicinity right so 0 means I can move in the direction without hitting into anything

(Refer Slide Time: 10.34)



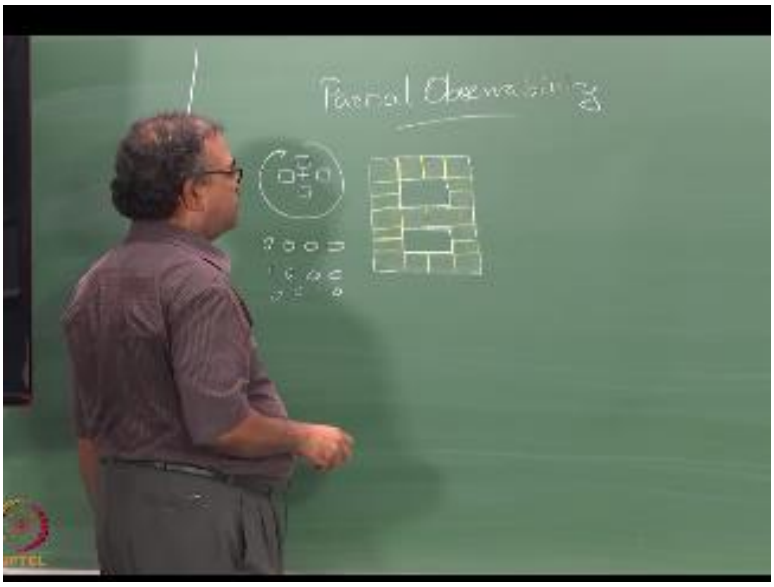
Right so for example here it would be 0000, 0000 okay.

(Refer Slide Time: 10.48)



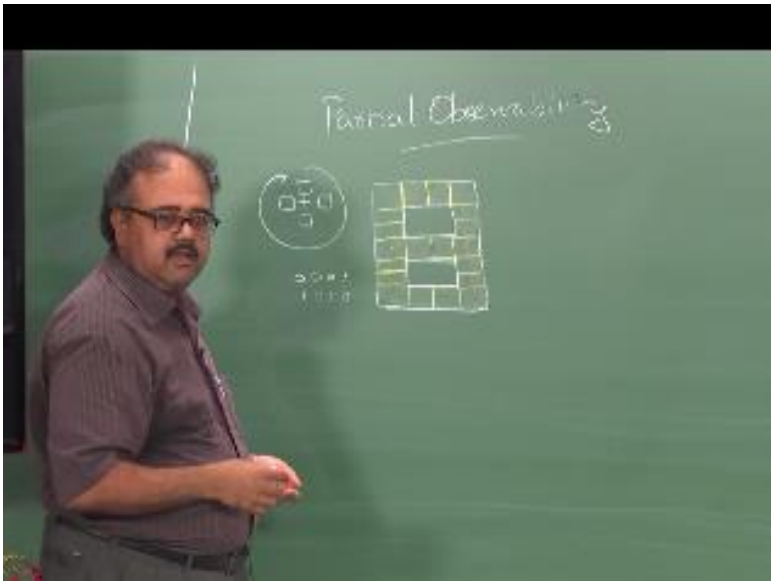
So then turn into problems with this way looks like

(Refer Slide Time: 11.13)



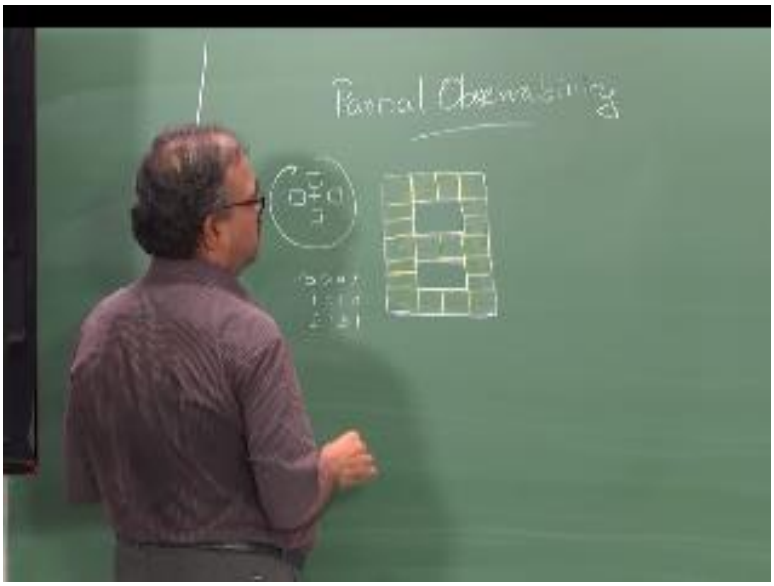
I have to erase something with you now we are going to have some fun so is there any place at is 0000 here. So what about here,

(Refer Slide Time: 12.24)



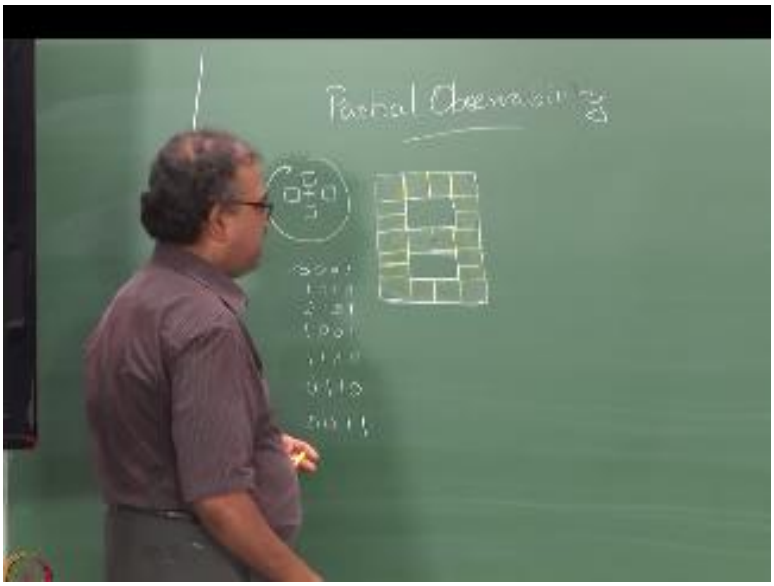
Here would be so on so forth right. So there are how many states do I have that I said anything else I have

(Refer Slide Time: 12.42)



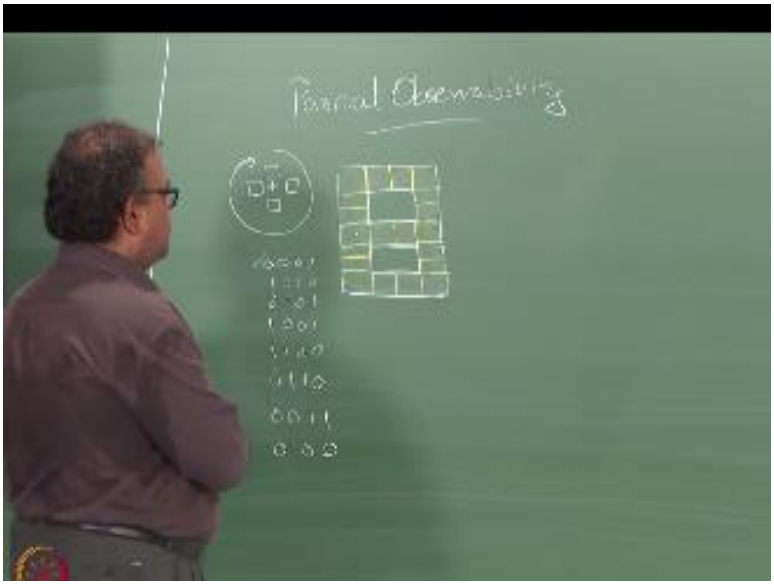
I have a 0 10 1 yes 1001 anything else

(Refer Slide Time: 13.03)



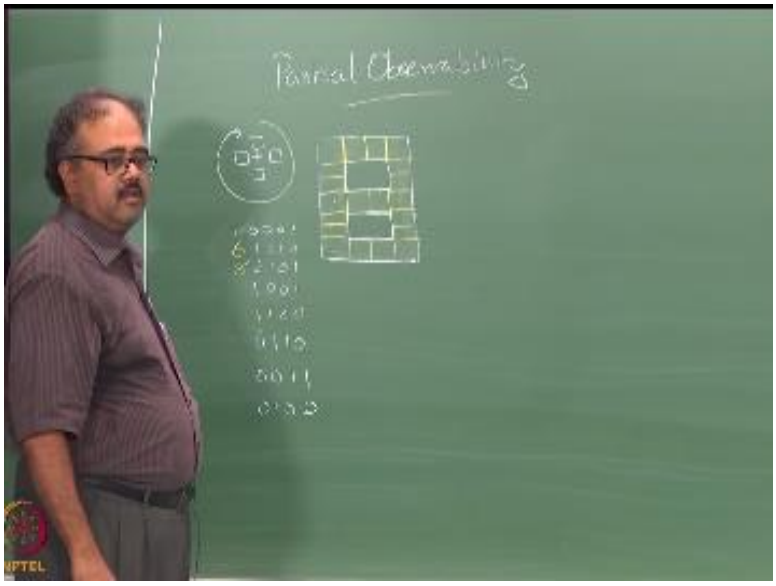
That's about it right yeah so good now can you give me a count of how many of each of these are there 001 is where that anywhere else one more missing one is it any other 001 okay

(Refer Slide Time: 13.35)



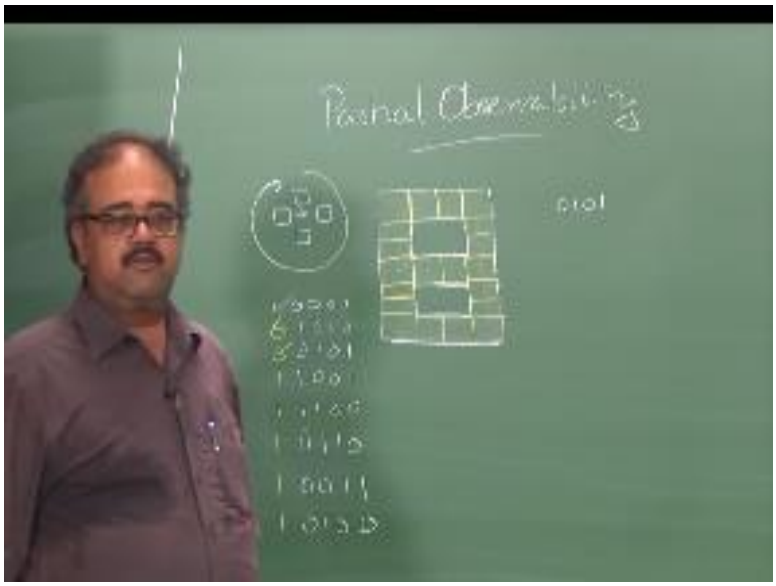
One then 1010 top and bottom 0101 left and right

(Refer Slide Time: 14.03)



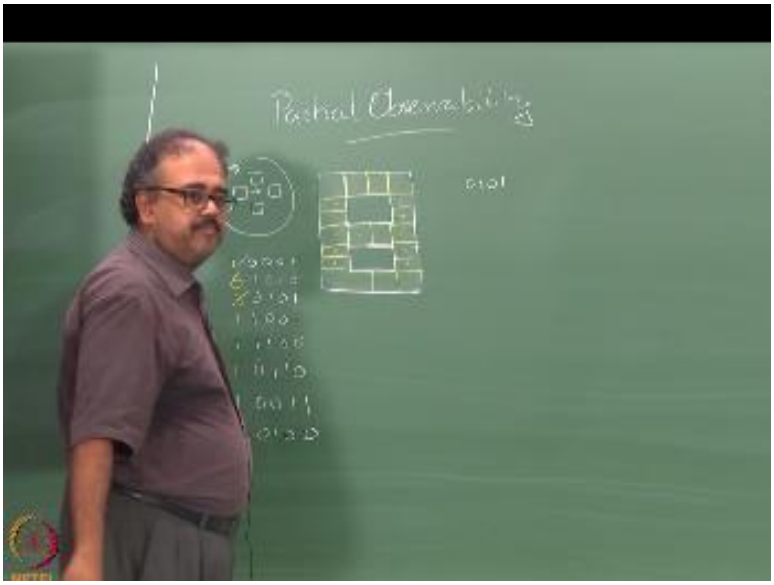
1001 to top and left 110 so this is all the input that you'll get I will get these four digits so I will get this for four numbers right I will get 0011 or something like that that will be the input I get suppose I tell you that you are currently in state.

(Refer Slide Time: 14.36)



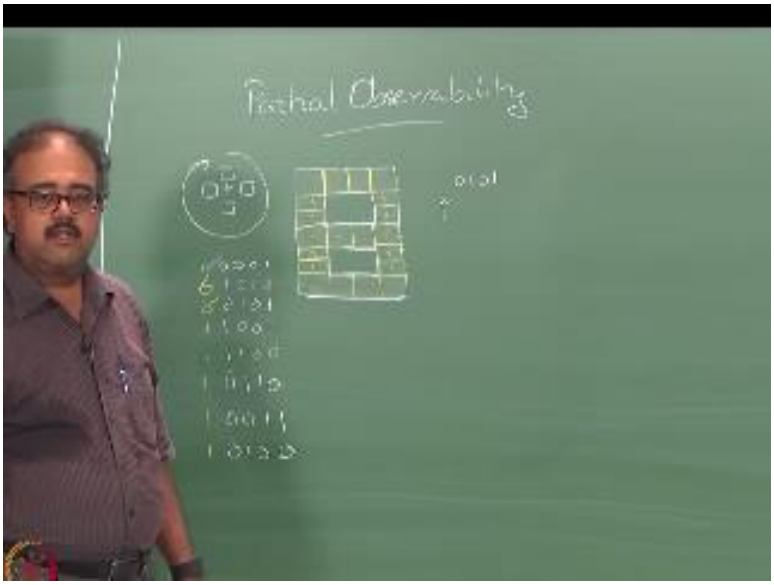
That is observation that you have right can you tell me where you are I can be in any one of four places or eight places eight places right so I could be

(Refer Slide Time: 14.53)



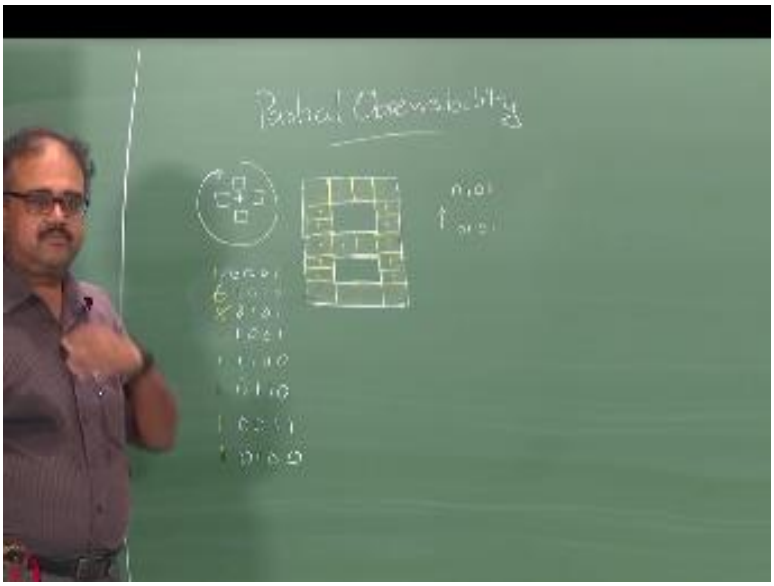
Here, here, here, here, here, here, here and here right now.

(Refer Slide Time: 15.04)



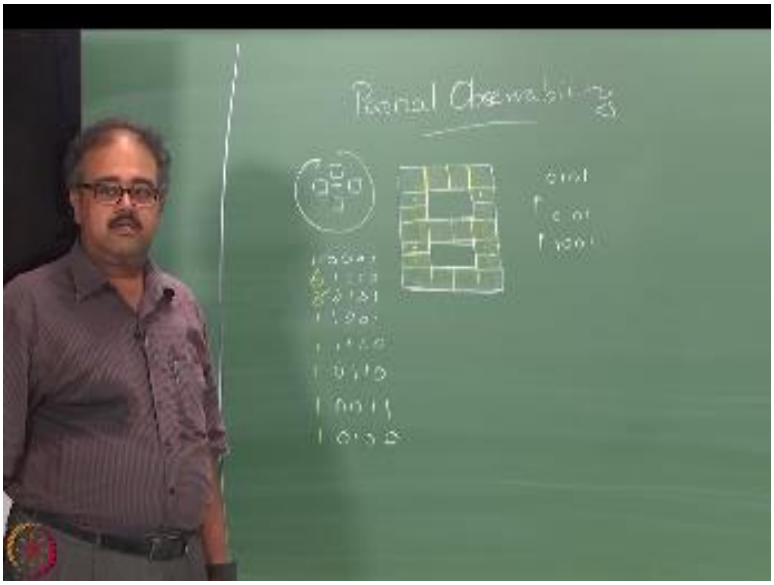
I will tell you something I telling I am going north right

(Refer Slide Time: 15.16)



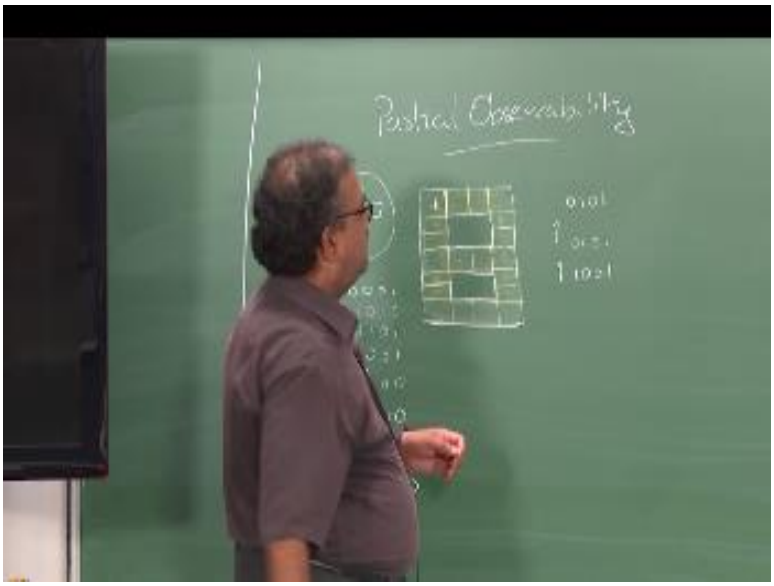
I am Telling I am going north and what do I see next I am seeing so where am I now one of those four places right and because I saw 0101 then I went up then I see 0101 and again so I made one of those four places now

(Refer Slide Time: 15.42)



I go up again right and I see 1001 so where was I earlier so this is where I am now but in fact

(Refer Slide Time: 16.16)



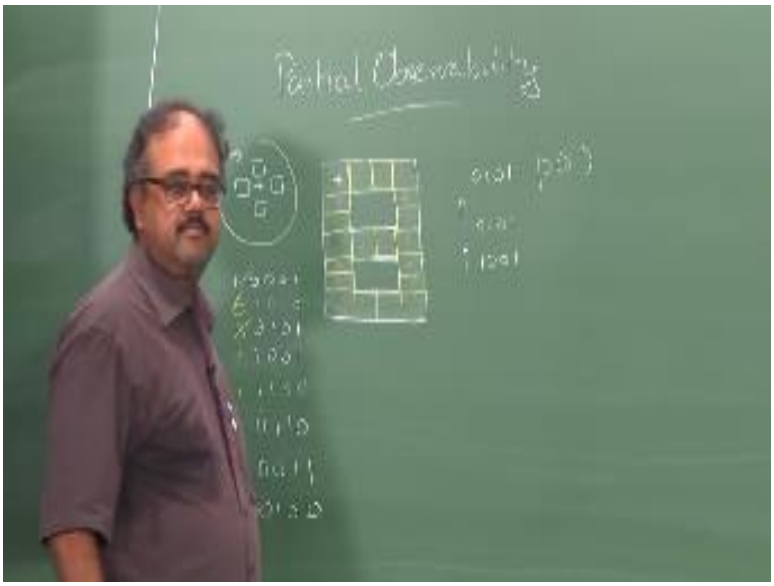
The only problem with this example is you do not need the history that even without the history you can find out that you are here that is why I tried drawing a more complicated grid but then it became little too complicated so I need to have something simpler but the point is you could see that as we keep more history around when our uncertainty about where we are it comes down significantly right.

So in this case it turns out that this is a uniquely identifying observation it could very well be that the sequence is uniquely identifying even though the individual observations by themselves may not be any clarifying right but in this case It turns out that the individual observation is uniquely identifying so what is it take away lesson here yeah so the point here is that even though if you

have partial observability so one way of tackling the partial observability is to use some amount of history.

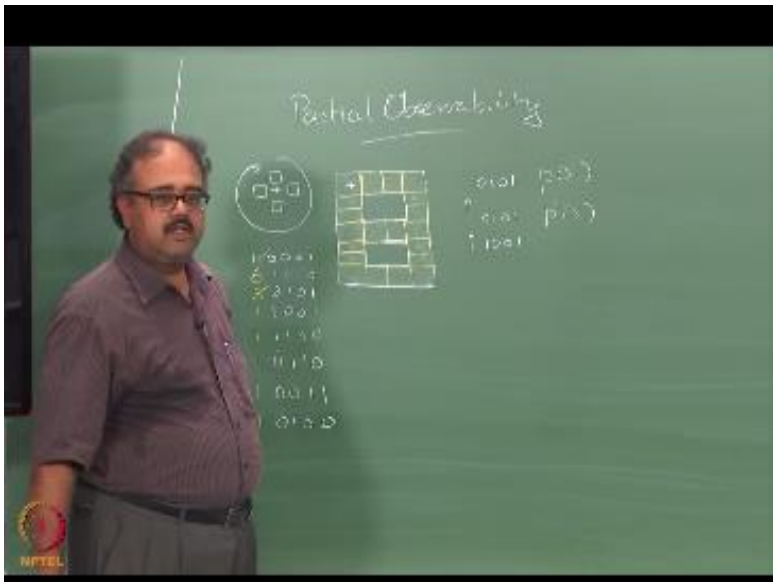
I saw one way of tackling this partial observability so you some amount of history and the history allows you to figure out where you are right so what really happened if you think about it when you started out here right so I had some kind of a probability

(Refer Slide Time: 17.49)



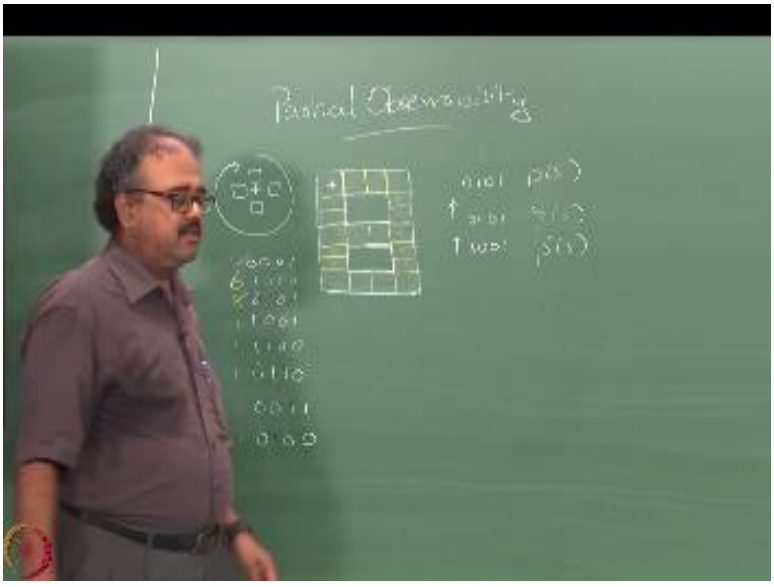
Over where I am in the state space it I had some probability over where I am in the state space and what was the probability distribution it was 181818181818818 0 everywhere else then I moved right and then I made a updated distribution

(Refer Slide Time: 18.05)



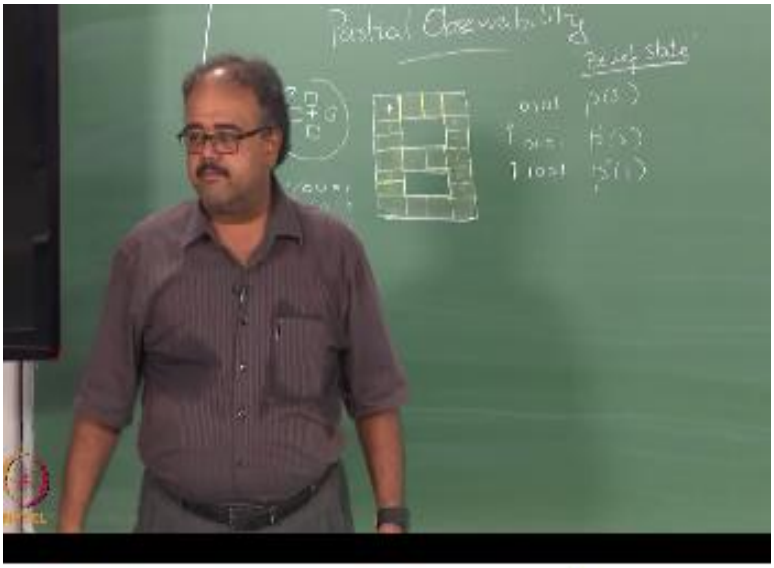
That had one fourth one fourth one fourth one fourth okay then I moved

(Refer Slide Time: 18.18)



Again I made another updated distribution with basically as 1 here and 0 of develops so what happens is as I keep getting more and more observations my belief about where I am in the state space keeps getting more and more focused. Right so this kind of this probability distribution that we maintain is called

(Refer Slide Time: 18.50)



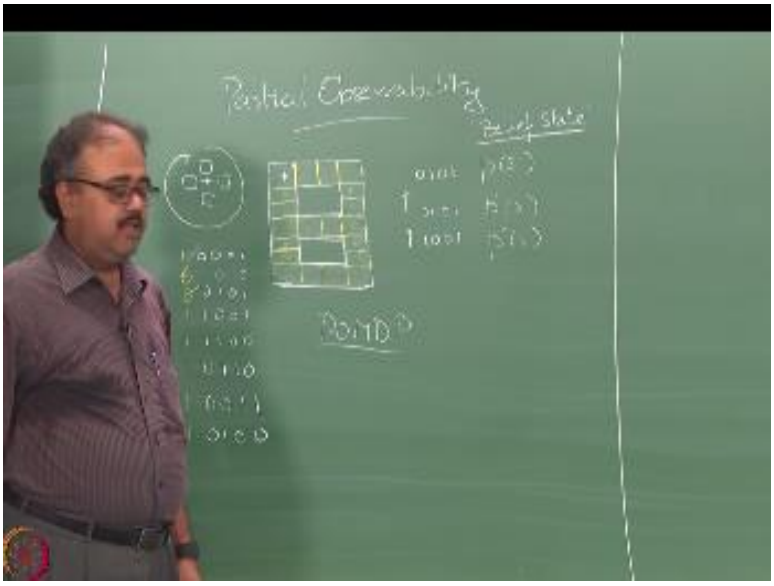
The brief stay right so I have an MDP assumption we are making here is I have an MDP right so there is some state space below underneath it right so because we are now giving probability distributions to certain locations in state right so that means there is an underlying state space so I am assuming there is an MDP with, with the underlying states and everything right but I am not

able to observe the states of the MDP I can only make observations which are some function of the state right.

And based on that I am going to form some kind of a belief about where I am in the state-space right and I can try to do what, yeah but what do you want to do with it ,what do you do with having a belief state, can use it to solve and use it to get a policy, it's one of the things that you can do right so that's the interesting question to ask.

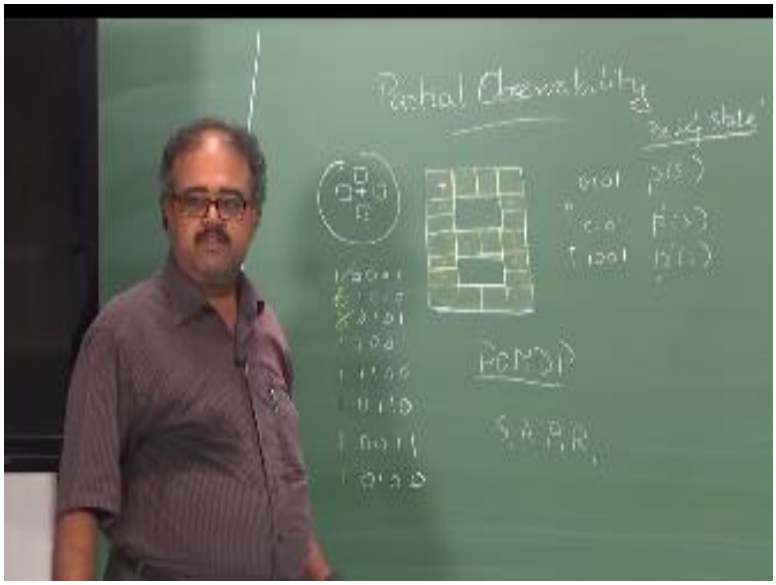
So what I am going to talk about this for the next this class in the next class is different ways of solving partially observable problems okay. Differentiation solving partially observable problems so a standard model that people use for this kind of partially observable problems

(Refer Slide Time: 20.32)



Is called a palm DP so you do not really need a palm DP okay for some of the methods that we use you really do not need a palm DP but the palm DP is a good model to know right so what is a palm DP it is an MDP right plus some additional components for and handling the partial observability so what are the things we are going to have you're going to have

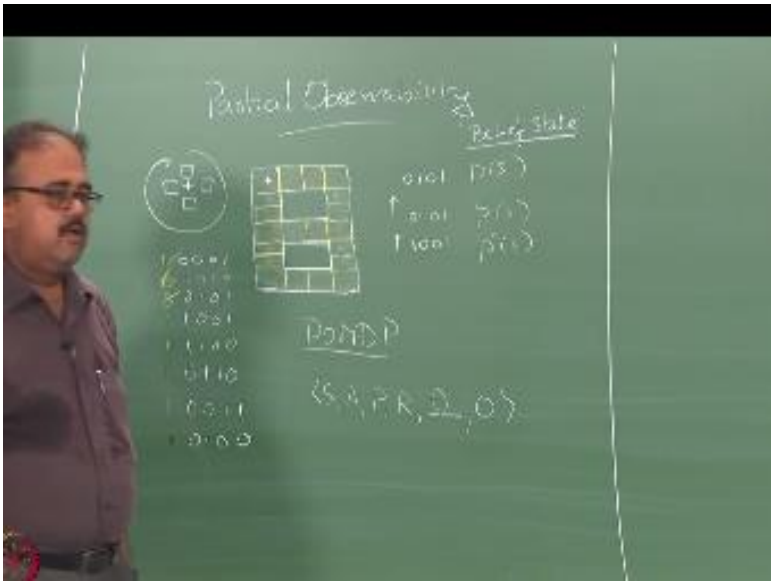
(Refer Slide Time: 20.56)



S we are going to have A we are going to have P we're going to have R so all of the standard MDP settings and then you are going to have Ω which is a space of symbols space of observations right so for example in this case Ω would be it will be 0 1 power 4 or you could explicitly enumerate these things and say this is all my Ω is because I at least for this given world I can't see anything else right so it's a smaller subset half of it basically it since of all possible sixteen combinations.

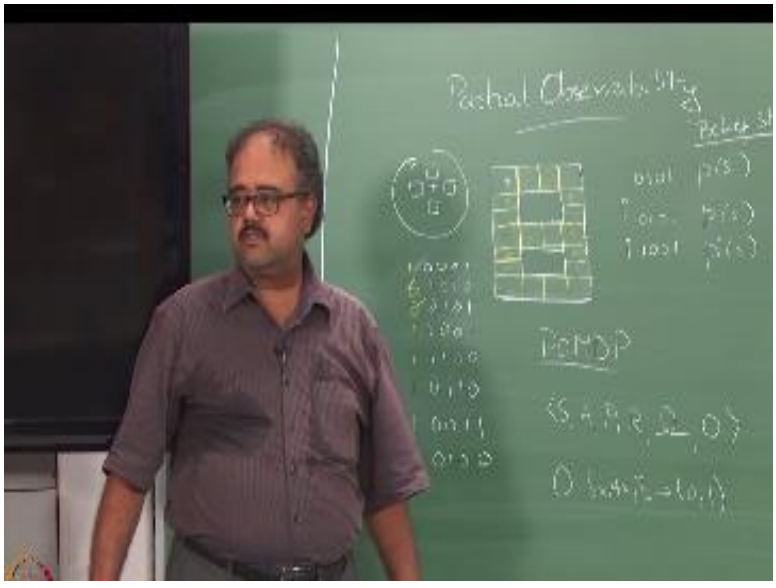
I have only eight combinations that can occur in this in this world you can now obviously imagine a world where you could have obstacles in their right so that you get more, more of these combinations occurring and soon so forth right so that is my Ω and then

(Refer Slide Time: 22.08)



I have my observation function that tells me something about how the states and the Ω related so essentially observation function.

(Refer Slide Time: 22.18)



Is a mapping from States action Ω to 01 so it could be probabilistic. And so it could be a function of the state on the action right so one way of thinking about it is you can see if I am here and I move up what will be the observation I will see that kind of marginalizes over my next state distribution at or you could also think of defining an observation function that says okay I'm here what is observation I am going to see what is the probability of 1001 given that my state is this.

Which is 1 what is the probability of1001 given that my state is this 0 that in this particular case we have assumed that the probability is 1 or 0 right but in general I could have some noisy sensors so for example if each one of my senses as a small noise factor right they still report 0 or 1 but they can report 0 or 1 with a with the noise of say 0.1 right.

So essentially so 10 percent of the times if it is 0 it will report one if it is one it will report now immediately the observation space has gone up to 60 it's too powerful because any, any anything can come on at any time right and then if I am here what happens so the probability of 0101 is only 0.9^4 right so there is a probability that I will actually see 0001 rate which is0.9 cubed in 2.1 so three of the things are reported correctly one is incorrect.

So like that so I can now start having probability distributions over my observations rate observations did not necessarily be exactly the same observations now can we go back and do he thought exercise that we did the simple exercise not thought to make simple exercise that we did right so I had 0 1 0 1what would be my belief state.

That it only still have a support of eight listen it's a lot more complicated right those eight states will still have the bulk of the probability because they will have 0.9^4 is it right. There are many states are some to can't be 0.9 44 right. No, no what can be 0.9 before 2 all this series is 0 1 0 and we not have 0.95 for weight assuming that all the situation is very wealth yah into an way yeah sum of them will be $0.9^{4.}$

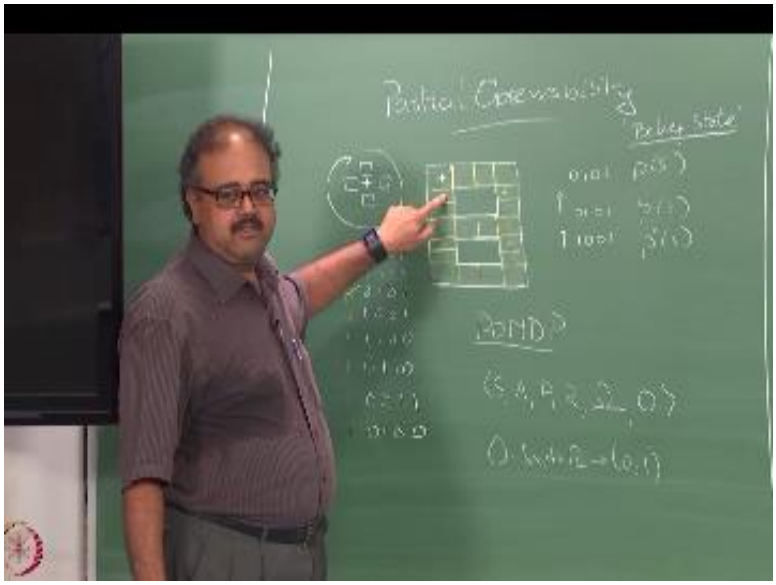
That I can get point in before permit I can get 0 1 0 1 I fall the sensors are correct right with 0.9^4 where will the 0.9^4. Where will all the census be correct is one of these things right but that is

not the only places I can get 0 1 0 1 I can get 0 1 0 and anywhere else also in fact I could get 0 1 0 1 anywhere in the world.

Let's just set more, more and more sensors have to misfire right so what about here everything has to misfire 0.1^4 right so, so this, this these places are the ones which I'm least likely to be if I get 1010 right suppose if I suppose I go here 00 01 so what happens one sensor is correct. The 0 sensor this correct is 0.1^3 in 2.9 so the probability of me being there is pretty low and.

Now I moved up and I got 0 1 0 1 now what do you think and becomes even more weird because I will have to worry about where I was and I am assuming now at least my motion is deterministic and what if my motion is non-deterministic that's another confusion that will have to worry about and see we motion is deterministic right so if I had been here and would have

(Refer Slide Time: 27.48)



Right I will be here that's for sure but I need to have been here so what is the probability that I was here that is what we just computed $0.9^4/8$ so with that probability I could be here it's at all the new observations also could be noisy right so what is the probability that I am here and I actually make the 0101 observation that's another 0.9^4 so I have to multiply by that 0.9 right so,

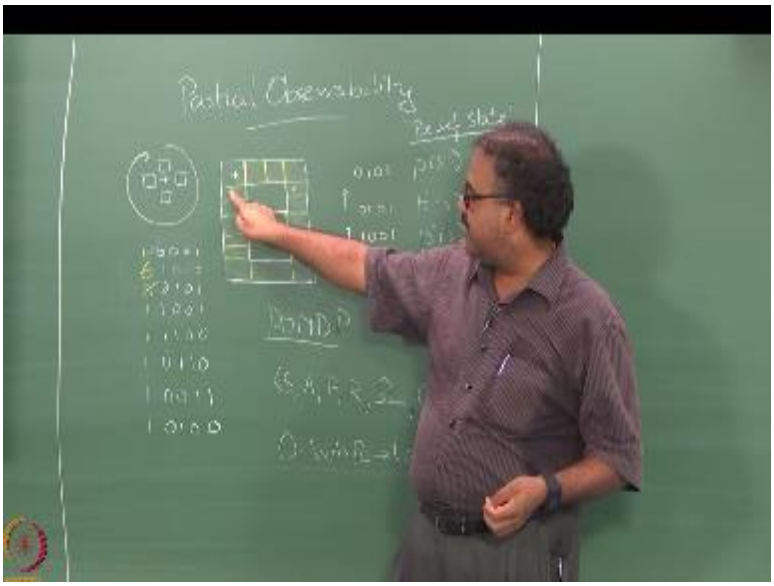
so it becomes more and more complicated the computations become more and more complicated right.

So that is essentially one way we have to think about it how will you do this belief update in a compact fashion how will you do it what will be low probability yeah okay. See one of the things I should we should we should remember this me moving from here to here would be the other one which is a high probability 0101 observation right but that will not happen because this has a low probability of actually having occurred initial.

So what will happen is the earlier what happened from 181818 it became ¼ , ¼ right that will not happen now but still this will be the four states that will have the highest probability become small over and lower so in fact they'll have much higher probability than the what you started off with distributing to the eight states so this will be closer to ¼, ¼ but will be discounted a little bit because you still have to give away a lot of things.

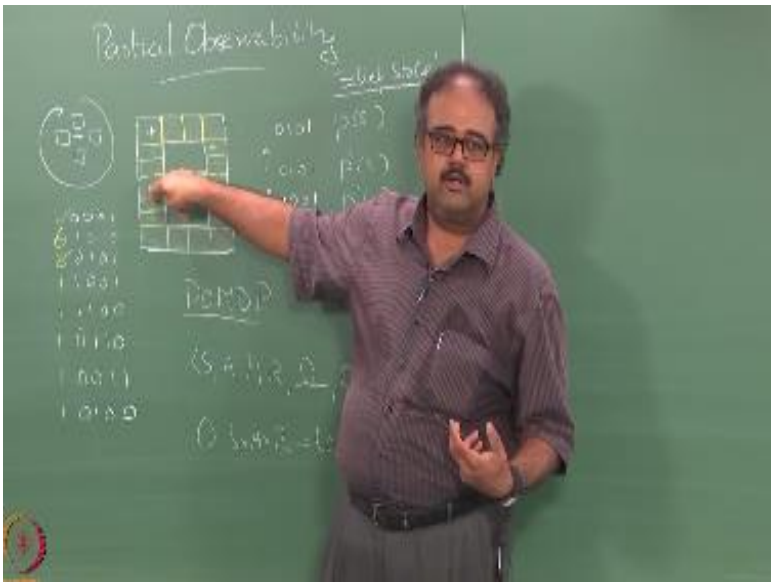
So which will be the which will have high probabilities here right so the probability of you being here will be high this will be the highest I do not know no question about it these four will be the highest next up it will be this guy because some post probability that you could have been here and he moved up here and this is the right observation to make here right

(Refer Slide Time: 29.52)



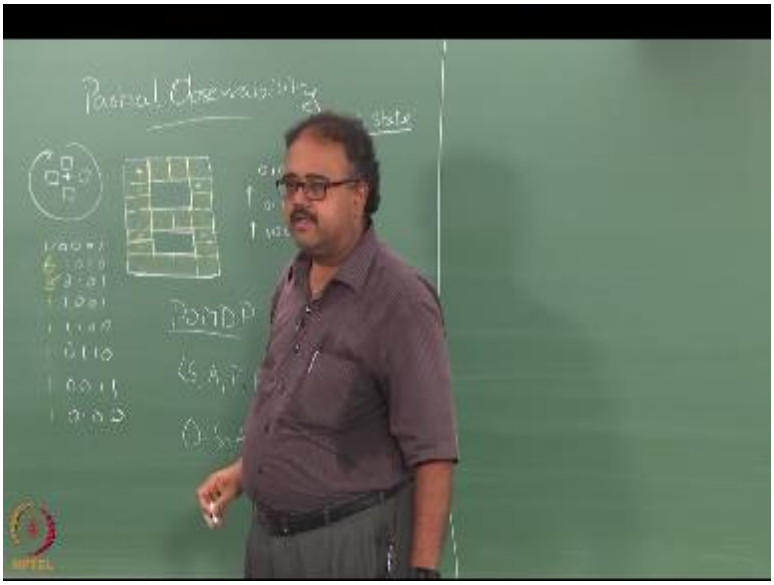
And earlier that you could have been here and we moved up okay the probability of this is very low so those are the things that you'll have to look at so this in fact there will be an equal probability of you being here and here

(Refer Slide Time: 30.15)



Because earlier the probability of you being here was low so from a low probability place you come to a high probability place in this case from a high probability place you go to a low probability place so both cases the probability will be equally bad right but they will probably with the second highest probabilities among these states right so.

(Refer Slide Time: 30.36)



So this will be highest right and then the second highest would be this and everything else will have much lower probability and of course the once in the cross corridors what are almost vanished to 0 so we did 0.1 0.1^4 into another 0.1^4 it is gone there right that were gone almost to 0 now what about the third one I go up again and I get 1001.

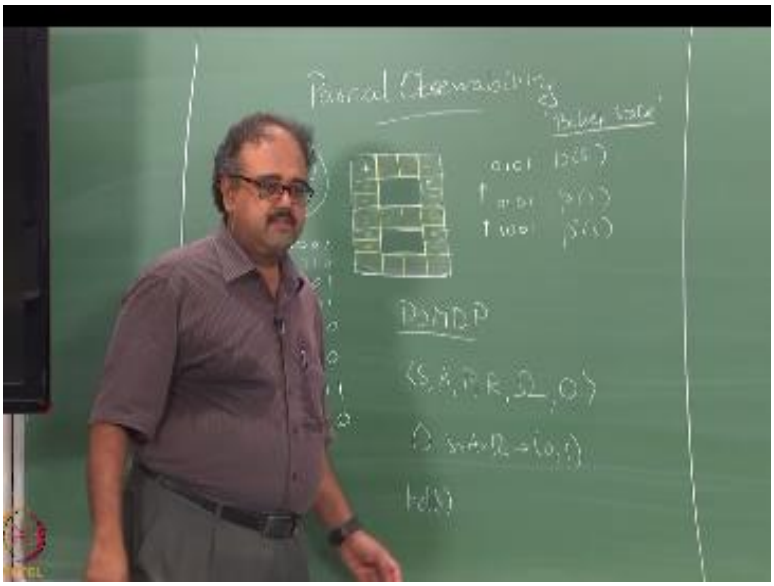
Which one which one will vanish to 0 this why will that vanish to 0 but is where I am now no, no at this point where I'm a was the question for the previous observation well if you want to go back and update the probability of very we're in the second step yes it will go to 0 but the probability of where I am right now so that will be the highest probability right in fact all other places the problem it will start going down.

But still they wouldn't have gone down to 0 it's only the cross Corridors would have gone down to 0 the vertical corridors you'll still have a non 0 probability of you being somewhere else right so this should be familiar to people who are in the robotics class how many of you are still

around 4 okay. Fine so it should be familiar to you all this belief state stuff right so we looked at it because it's used a lot in robot navigation because that's the reality of things right so that's how things are with robots so this kind of belief state computation is too long okay so,

So this is just to get you an appreciation of how hard it is so we really need a compact way of updating the doubly so believe stuff sometimes written as

(Refer Slide Time: 33.10)



That tells you that it's a belief that you are in the state yes right now it's basically the probability that you are occupying state S.

IIT Madras Production

Funded by

Department of Higher Education

Ministry of Human Resource Development