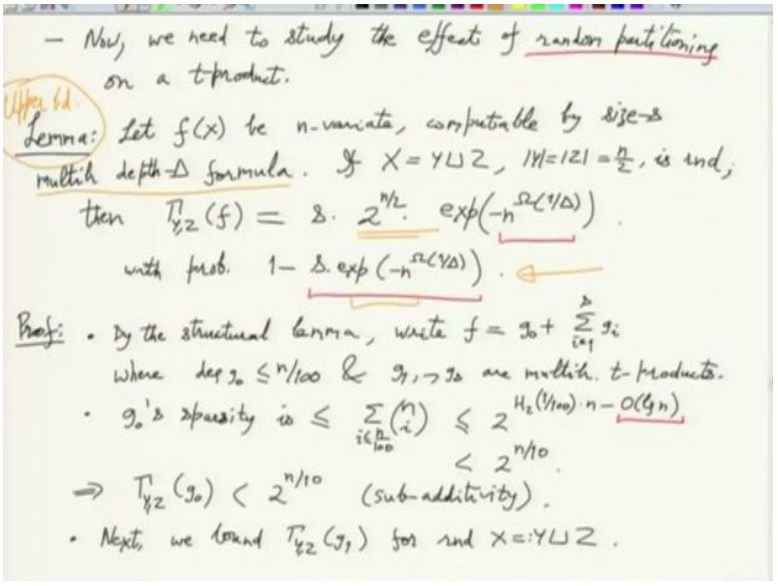
Arithmetic Circuit Complexity
Prof. Nitin Saxena
Department of Computer Science and Engineering
Indian Institute of Technology – Kanpur

Lecture – 18
Arithmetic Circuit Complexity

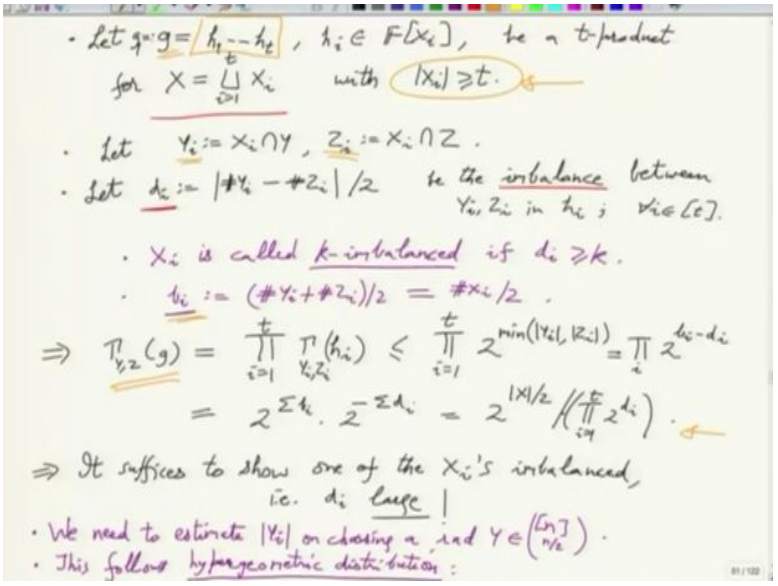
(Refer Slide Time: 00:17)



So last time we were in the middle of this lemma, upper bound lemma. So this is to do with the multi linear constant depth circuits or formulas. So what it is saying is, if you randomly partition your variable set into equal parts Y and Z. Then the measure is with high probability here the probability is this. So with probability, so if the circuit size s is smaller than this exponential and then essentially with constant probability, very close to 1, the measure will be smaller than $2^{n/2}$.

$2^{n/2}$ is the max possible measure and there will be a reduction by this factor of $2^{n/2}$ something. So, that is the upper bound and which would mean that if you take a polynomial f whose measure is large then you will get a lower bound on s. So the way we would do this is multi linear t product decomposition. So we will write f as essentially a sum of multi linear t products, plus there is a g 0 term which is the kind of term to ignore these are the low degree monomials which does not contribute much to the measure. So, then the question reduces to studying $\Gamma(g_i)$.

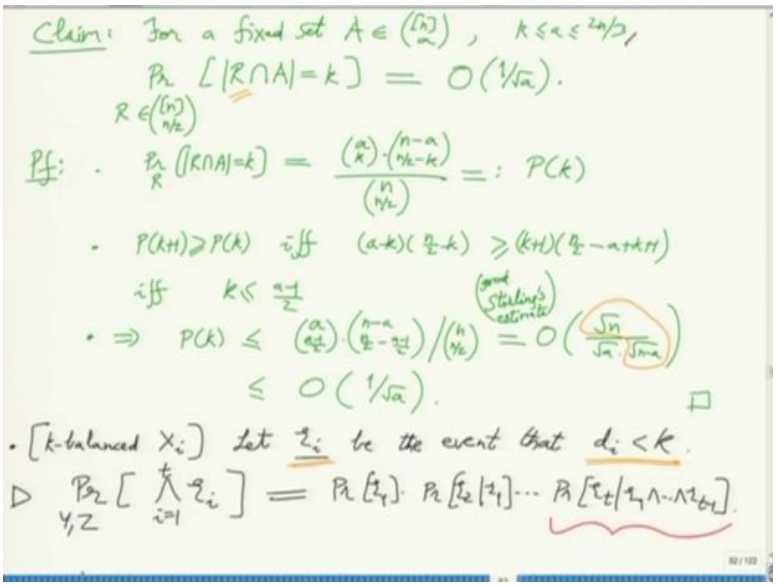
(Refer Slide Time: 02:21)



So, that is what we were doing, so we are looking at this g that is a product of h_1 to h_t each h_i depends on at least t. So, the variable set of h_i is x_i and that we are assuming is at least t and we pick Y which is a subset of $n/2$ variables of X randomly and then we will study what goes in Y and what goes in Z for the X_i variable set. So, this imbalance will be called d_i and we will say that X_i is k imbalanced, if this imbalance or discrepancy is more than K is at least k.

b_i we have defined to be essentially the size of X_i, there is a factor of half. So, this calculation showed this simple upper bound that the measure on g is the max measure divided by product of 2 raise to discrepancies. So, basically some of the discrepancy is or some of the imbalance is what you get as negative in the exponent with $n/2$. So, we will now show that one of the discrepancies is large, which will mean that $\Gamma(g_i)$ small which will mean that Γ of multi linear depth delta circuit is small.

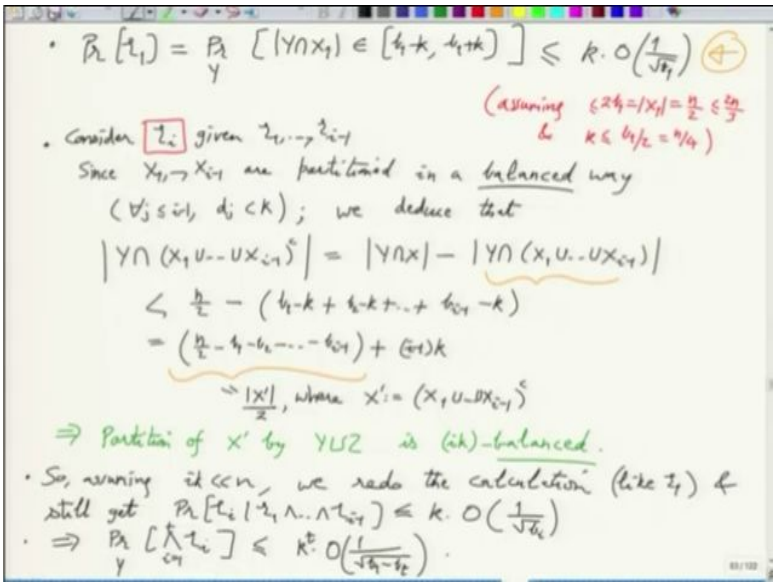
(Refer Slide Time: 04:16)



For that the analysis is through hyper-geometric distribution, so what happens when you pick a random subset R, so how many elements of a fixed set A will be picked? So, the k of them will be picked with probability $1/\sqrt{a}$, a is the size of A. We did this simple calculation and so now after that root over, we will analyse the probability of all these events happening simultaneously.

Which is all the discrepancies being smaller than k? That is the event. So, the first probability is that d_1 is smaller than k then second is assuming d_1 is smaller than k, d_2 is also smaller than k and so on. We will now study this expression, this conditional probability. So probability of ε_1, what is that? That is the base case.

(Refer Slide Time: 05:31)



So the probability of ε_1, is it is basically the probability over this random choice of Y $n/2$ subset of 1 to n, and discrepancy less than k means that X_1 elements that you pick in Y are between the b_1 - k and b_1 + k. So, these are the essentially 2k values that you are allowing to Y intersection X_1 and so, from hyper-geometric distribution you will get this to be less than equal to k times order of 1 over square root of, what is the size of the fixed subset? X_1 sizes is $2b_1$.

That lemma had some assumptions you just have to check that those assumptions hold. So, there was this assumption that when you are looking at the cardinality to be k the subset should be bigger than k and subset itself should not be much bigger than 2n by will should be at most 2n / 3. So those assumptions hold in our case. So, we are assuming, subset which is $2b_1$ is less than equal to? So, $2b_1$ is the size of X_1 which is less than equal; which is equal to $n/2$.

And so that is clearly less than equal 2n/ 3 and that is at least. So, what is b_1- k what k in this case? Give that, so this is the discrepancy bound, we will pick it to be very small. So we will make sure that it is we will take it to be less than equal to b_1 / 2, b_1 / 2 means n / 4. We will pick k to be quite small, so this the moderate assumptions will all be satisfied. So for every value between b_1 - k and b_1 + k, you can apply the hyper-geometric probability calculation. This is not very important; those assumptions are anyways quite weak.

So assuming all that you have this bound for probability of ε_1 that is b_1 less than k. So let us now move to the general case which is consider ε_i given the previous events 1 to i -1. So, since X_1 to X_i - 1 or partitioned in a balanced way because when you assume that ε_1 to ε_i - ε_1 events have happened, then basically what you get is for all j less than equal to i - 1, d_j less than k. So a discrepancy is less so we read it as they being partitioned in a balanced way. So we deduce that; so we will calculate this now.

$$|Y \cap (X_1 \cup \cdots \cup X_{i-1})^c|$$

So, basically things that remain $X_1 \cup \cdots \cup X_{i-1}$, well they are disjoint because we are looking at this t product. So these variable sets are subsets are disjoint and so what remains, we want to see its intersection with Y. So this will be

$$|Y \cap (X_1 \cup \cdots \cup X_{i-1})^c| = |Y \cap X| - |Y \cap (X_1 \cup \cdots \cup X_{i-1})$$

So what is $|Y \cap X|$? That is n / 2, you would be able to get any quality. But what can you say about the other part?

First of all this expression, Y intersection, this union compliment we want it to be large or small.Remember what you wanted to study ε_i, you want this to be small, you want an upper bound here and if you look at this Y intersection X_1 and X_2 and X_{i-1}. You have a lower bound for this, because for these the partition happened in a balanced way. So, you actually have both lower and upper bounds.

So, what is the lower bound, so, this will be less than put each of these lower bounds. So, Y intersection X_1 is at least b_1 - k and X with X_2 at least b_2 - k and b_i - 1 - k. Which is $(n/2 - b_1 - b_2 - \cdots - b_{i-1}) + (i-1)k$. What is this term? It should be a plus. So this is essentially the size of $X`$. Which I am defining as, write this properly, compliment the prior variable subsets where union is what remains, that is $X`$.

So what this inequality is saying is that why will not pick it is not $X`$ this should be $b_1$. When we had defined itself as half it is $|X`|/2$ and that that is where there was some semantic issue. So now it is fine, so this is $|X`|/2$ so what this is saying is that Y will use from $X`$ around $|X`|/2$ many elements so the partition with respect to $X`$ will be (ik)-balanced. All these things will make sense only when k is small.

Otherwise you do not get anything. So the way we will read this is that, partition of $X`$ by Y Z parts is (ik)-balanced. So, even when you are looking at this intermediate probability or probability of the intermediate event the partition is nearly balanced. So, the situation is not very far from ε_1 even at i^{th} stage the probability calculations will be very similar to the probability calculation for ε_1 event.

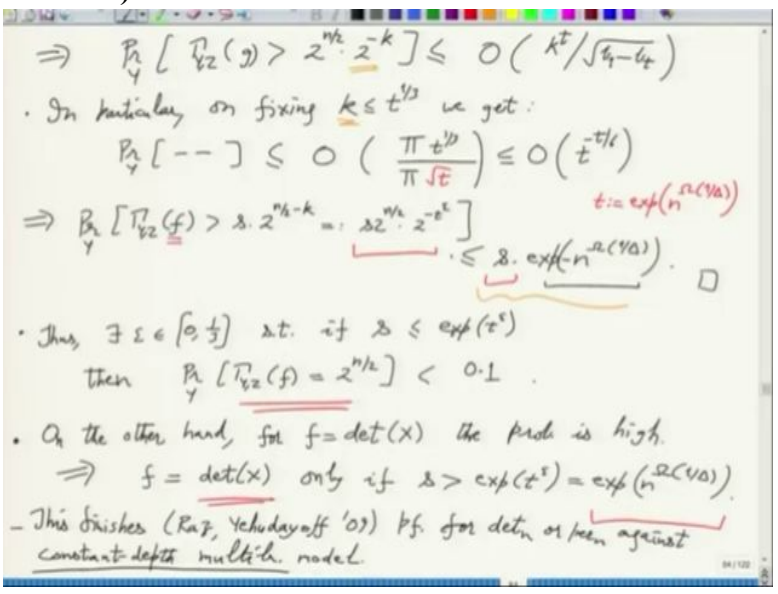
So, this will practically simplify our calculations, so we will just behave like we will just do what we did with ε_1 we will get the same expression. So, assuming this (ik) to be much

smaller than n, we redo the calculation like ε_1 and still get that the probability of ε_i, assuming that prior events have happened, so this is the same as we got here. So which will be $k \cdot O(1/\sqrt{b_i})$.

So, I mean this can only be understood so quickly by the extreme case the extreme cases, the partition is perfectly balanced for all ε_1 to ε_{i-1}. So, if the partition perfectly balances them then at ε_i also for the remaining variables it is perfectly balanced. So, you get the same probability estimate, as you got for ε_1 and what was i? Yeah, so we could put here ik, but, you could put here ik.

But I think i is also very small in the end so it will not matter, because i is upper bounded by t and we will pick t to be very small it will not matter in the end. So you will soon see, so once we have this we can now get the probability of intersection. So the probability of the intersection of all these events is then just take the product. So that is $k^t \cdot O(1/\sqrt{b_1 - b_t})$, so all these X_i being balanced so for that this is the probability which means that which implies that the estimate that we had gotten for Γ, that is large.

(Refer Slide Time: 19:46)



So let us write that in those terms so this means that probability over Y,

$$Pr_Y \left[\Gamma_{Y,Z}(g) > 2^{n/2} \cdot 2^{-k} \right] \leq O\left(\frac{k^t}{\sqrt{b_1 - b_t}} \right).$$

So this is because this 2^{-k}, this is coming from the sum of the discrepancies you had in the denominator or in the exponent which was in the denominator. So we are assuming all of them to be quite small but that has to happen for this measure to be large. So, this is certainly upper bounded by the probability we just calculated.

We now have the probability estimate for the measure being large. So, all that remains is you fix the parameters and show that this is a very low probability, as long as the size is small. So in particular on fixing this k to be $t^{1/3}$, we get the probability. So, what will you get? In the numerator you will get, you basically get a product of $t^{1/3}$ the numerator and in the denominator, what are you getting? What is $\sqrt{b_i}$?

So, how is that related to t, at some point we had some assumption, we had this. It is at least t / 2 and t also remember we had something in the no, but did we fix t in the proof, we have not fixed it, but it will be fixed to this lemma statement basically whatever expression you are seeing in the lemma statement, so let us come back to this. So this is $\sqrt{t}$ and i going from 1 to t. So you get this. So what is this?

This is 1 / 3 - 1 / 2 so that is 1 / 6. So $t^{-t/6}$, $t^{-1/6}$ and the whole thing raised to t since it is a product of t things. So now we will, we must have fixed it before. But you mean the first line, because no so this is the probability here interested in that the measure is in the back case measure could be large. So, if the measure is larger than $2^{n/2-k}$, what would it mean? It would mean from our older formula there was this formula.

So, from this formula, it would mean that all these b i are small. They are in fact smaller than k / t, the sum is smaller than k because if the sum was larger than k then measure will be smaller. So the sum of the discrepancy being smaller than k that you have estimated here, this place so we just reinterpret that error probability. It is basically the measure being large that was our original goal. Anyways and then we fix k to $t^{1/3}$ we get this probability to be t^{-t}.

And obviously we should fix t to be exponential in $n^{1/\Delta}$. So, think of t as $(2^n)^{1/\Delta}$. For that larger t your $-t$ correspondingly small probability will be small. Why did you fix k. No So,

the thing is, it is already determined so square root t is what you have as b_i. So, k has to be smaller than square root t, if it is $\sqrt{t}$ or exceeds is does not be error term is 1 useless I mean the error probability is 1 then.

So, you want it to be smaller than $\sqrt{t}$. So, you can pick anything which you like smaller than half and I think I prefer 1 / 3. So, one technical point is that these constants which we are using, for example also in Ω, they are absolute constants. They do not depend on anything, they do not depend on Δ. They do not depend on k or t or whatever. These are absolute constants, there is no hidden constants like in some older proofs case.

So just rewriting this probability over Y that Γ is greater than $s \cdot 2^{n/2-k}$. So, now we are moving to f which is low degree monomials plus sum of s many t products like g. So, now you will have $s \cdot 2^{n/2-k} = s \cdot 2^{n/2} \cdot 2^{-t^{\varepsilon}}$ and you are reducing it by some exponential like epsilon 1 / 3. So, probability that the measure of final polynomial f. The multi linear depth Δ circuit that measure is larger than this, this is $s \cdot exp(-n^{\Omega(1/\Delta)})$.

So that finishes the proof of the lemma, so in this last statement there are two heavy things hidden in the notation. One is that the measure is large by that much amount, so closer to $s2^{n/2}$ this even the probability is also dependent on s. So, the probability is upper bounded by s times this ε^{-t} kind of a thing. So, if f is larger than the series to minus t then this probability estimate is giving you nothing.

The size is actually being used in both the places. The measure being large there s appears and also the in the probability error probability s appears. Which is fine because we any ways assume s to be small I mean you we will say that if there is a multi linear depth Δ circuit that is small then this error term will also be error probability will also be small. So in other words usual for usual partitions for random partitions, the measure will be small.

So it is in the correct direction but all this is really dependent on s. Otherwise, it does not make sense. For g it was, correct. Yes. In that way, correct. So there might be many g is so if you take way too many g is, then you do not get any error probability. So that is a technical

thing in this estimate that s is appearing on both sides. But that is consistent with our final goal. So if this part was worse if it was too large then we could would not have been able to use this for any lower bound purposes.

The good thing is that in both the sides these f 3 parts are comparable and that is what will also give us the lower bound on this. So thus there exists this absolute constant ε such that. If s is smaller than exponential in t^{ε} then the probability over Y of the measures being $2^{n/2}$ is less than 10%. This we can deduce from the above so $\varepsilon = 1/3$ may not work.

But you may suitably reduce ε below 1 / 3 to absolute constant such that this probability here falls to 1 / 10 and here we are saying that the measure is maximized. So, measure being maximized happens with very low probability. So, what should your f t now? For the lower bound? Do you remember this probability for f equal to determinant? When f was determinant then I think this probability was quite high. So, on the other hand, which implies that f can compute determinant, how much was it?

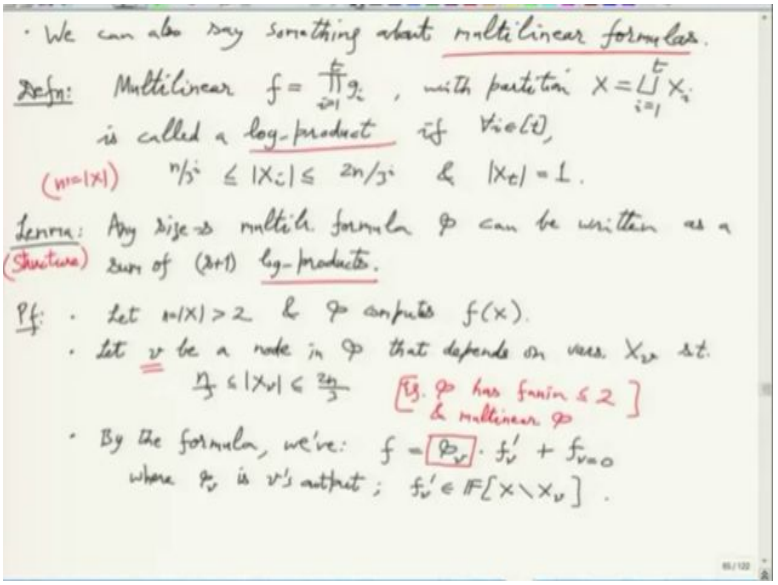
That cannot be f the determinant only if $s \geq t^{\varepsilon}$ exponential in t^{ε} which is exponential in $n^{\Omega(1/\Delta)}$. So f for determinant we get this size lower bound for multi linear depth Δ circuits. Note that you are getting $2^{n^{1/\Delta}}$ so for Δ constant this is an okay bound. It is not optimal but still you can call it exponential. You may even take delta to be just below log n even for depth Δ , $o(\log n)$ you will get some lower bound but that will not be exponential.

So depending on how close to log n you get the depth. So basically at Δ go to log in this will completely break down and which is a depth that is realistic. To compute determinant $n \times n$ determinant you would have to go to log n depth and that depth this tool breaks down. It does not give you anything but below that depth this gives you something, so, that is the limitation of this technique. So this finishes Raz and Yahudayoff proof for determinant, $n \times n$ determinant and the same thing for permanent against constant depth multi linear models.

Any question so we have done it first for multi linear depth 3 and then we generalize that idea to other constant depth, higher constant depth. So, do you want to generalize it further? So,

now, we can try to generalize it to formulas, multi linear formulas if it is so, there the thing is that we are not saying anything about the depth, depth can be arbitrary but then it will not be a circuit will be a formula. So, it is a generalization in a slightly different direction.

(Refer Slide Time: 37:48)



So, we can also say something about multi linear formulas, but the bound will be worse. It will not be that kind of exponential lower bound as we got for constant depth formulas, multi linear constant formulas. The t product idea will not work because that calculation we have done that we cannot make it work for Δ close to log n and in a multi linear formula that could be the depth could be log n.

So, a different product is considered called log product, so again it is a product of t multi linear and disjoint support are disjoint variable set g_i's with partition. Since it is multi linear the variable, underlying variable subsets are disjoint. So that induces a partition into X_1 to X_t. so this will be called a log product. So what do you think is the definition of a log product? So, remember that whatever you define here multi linear formula should reduce.

There should be a structural lemma, the structural lemma last class was not clear so it seems hard to come up with a definition that such a structural lemma would exist. So, previously we had said that each of these exercises at least t. So, we certainly have to relax that and now, what we will say is I mean X_1, X_2 ... X_t the sizes may be gradually reducing will allow the

size to gradually fall down and that will basically be by 1 / 3. So, you start with something and then you start with n / 3.

And then you allow reduction by 1 / 3 factor so this logarithmic fall. So, multi linear formula you can reduce by induction to this so again we are inductive proof, it is kind of consistent with the log depth. So they are the variables if you start from the bottom the variables are growing by constant multiple at each step. If for all i the size of X_i is at least $n/3^i$ and at most $2n/3^i$ and eventually X_t is 1. Only one variable in X_t, n is the size of X as always. That is a log product.

So, now we have to do two things. We have to show that formulas have this structure and they can be written as sum of log products. So, these are on log of n because t is log. Yes, so as a t is around $\log n$. As a t product this is you can only say t = 1 because X_t is 1.You cannot anything more; so it is not that the X_i are all large. They start large and then they gradually fall. It is actually pretty precise because you are saying that X_i is between n by something and 2n by something.

So, you have a good idea of how big and X_i used this can be achieved that is first and second, we will be studying the properties of Γ on this product which will again be by discrepancy. So any size s multi linear formula φ can be written as a sum of s + 1 log products. What is the inductive proof? X_1 size you want n / 3 to 2n / 3. In this case we do not need anything except this formula structure.

Yes, so you identify. So, again from the bottom when you see the variable subset is going to slowly increase because you start with variables subset single n and then it you reach n. So as you are going up it has to keep increasing so at some point you get a node where the variable subset is between n / 3 and 2n / 3 it could be too large maybe. So, do it this way like this sum, the sum that we have that in just be the inductive from the top that sum of s+1 things can be addition things from to the top.

And then, so, then you take each multiplication gate that have a number of inputs. You have unbounded number of inputs, but you know some of them together will span between 1 / 3 and 2 / 3 in the total variable and then the remaining which stands in the remaining variable. So use that, identity you get for formulas, f = gv whatever. You need that identity that is why you get a sum because of that identity inductively.

So, just for sanity let n be slightly, moderately large n should be at least 3.Otherwise division by 3 will not make sense and φ computes a polynomial f on the variable subset X variable set X, this case of n, 2 or less you can here is a base case it is a trivial base case. Otherwise when so, and when is n is 3 or large then we will induct on the size I would say on s. So, let v be a node in φ that depends on variable subset X_v such that $n/3 \leq |X_v| \leq 2n/3$. So, why will this v exist?

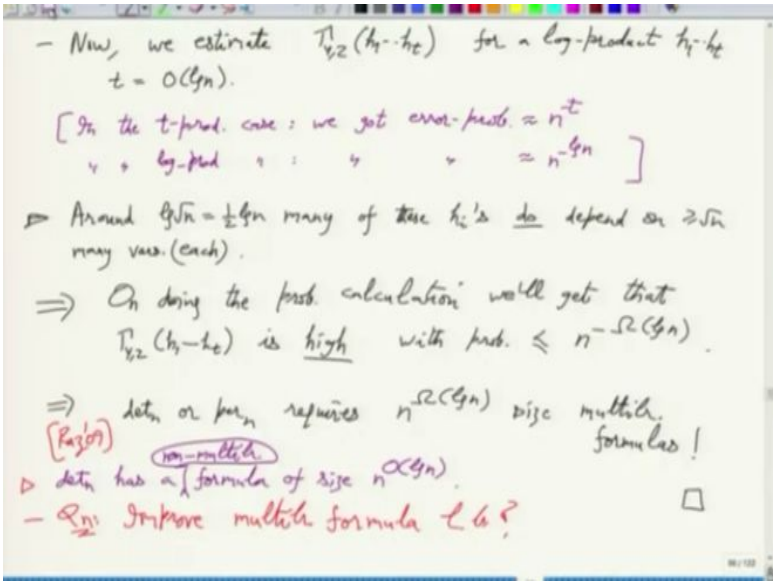
This X_v greater equal to n / 3 is clear because you are going from variables 1 variable to n variable. So at some point you have to cross n / 3. But why this upper bound to 2n / 3. For that, I think you have to assume some, maybe as you fanin to be 2 and 3. So fanin less than equal 2 will be now. So, you can assume the canonical form. So, φ has fanin at least I think fanin 2 should be required for this and it should also be enough.

Would make this deduction get node something. No, but we are not talking about degree or a number of ways to cannot jump to something below under the about correct? That I'm not; and multi linear it syntactically multi linear every node you have multi linearity. So at some point at least one of the variable subsets will cross n / 3 and the first time that happens each of them is n / 3. The resulting variable subset cannot exceed 2n / 3 that is the full argument.

So that identity which you always use. So by the formula, we have this usual identity f is whatever is computed at v, that is $f = \varphi_v \cdot f_v` + f_{v=0}$ . So where $\varphi_v$ is v is the output $f_v`$ is has variable subset Xv complement? $X \setminus X_v$. That is it; you do not know anything else so with respect to v, you can decompose this formula f formula for f. So φ_v becomes your starting point of log product. This is basically what you have done in the induction step.

That φ_v is your let us say g_1 and g_2 to g_t you will just get inductively from f_v`. It is now believed that measures do not exist, for VP different from VNP. That is the structural lemma, this definition of log product is good enough to be applied to multi linear formulas. But now the next question is how does Γ behave on log products. So, Γ will actually not give you a stunning lower bound. It will just give you something and I will also skip those calculations. We just have to repeat this hyper-geometric stuff.

(Refer Slide Time: 52:40)



So, now, we estimate $\Gamma_{Y,Z}$ on a log product $h_1, \cdots, h_t$, t is naturally at most log and support the variable subset of these h_i gradually falling from n / 3 to 1. So actually can you guess from this big calculation we did last time. What this measure will be is expected to be, you want to show that is small? No. So, here the problem is that, maybe I should write that down, that is an important thing.

So no all that is wrong. In the t product case we got the error probability something like 2^{-t}, very vaguely speaking. You got 2^{-t}, t was something like $n^{1/\Delta}$ that is a mistake I think this is not X. But t should not be X_t should be $n^{1/\Delta}$. This should be it, that was a mistake. So, in the t product case we took t to be $n^{1/\Delta}$ and that is what you then get in the exponent of error probability.

So, the higher delta is the worse I mean, the weaker result you are getting. So, just from that, actually, you can see that now, you will get something. Maybe I should further change this. It

was something like n^{-t} that you got now in the law product case, t is not that high. I mean the things that you are multiplying are not so many. So you will get error probability just like $n^{-\log n}$.

So error probability determines now everything, if your error probability is too high, or your result is worse in terms of probability. So then you can only say that s is at least $n^{\log n}$. You do not get an exponential result, you just get a quasi poly lower bound. So previously, by picking t to be, let us say $\sqrt{n}$ and you could have gotten $n^{-\sqrt{n}}$ lower bound. Now it is not possible.

So now you will be stuck into $n^{\log n}$. But what is the estimate? That remains the same but does not help because you are getting that the error probability is s times this. So if as soon as s exceeds into the log n, you are saying that your event happens with probability at most one. Which is not saying much so, that is the technical thing in the previous proof and it kills you here.

So, the way I mean some more details here. So around half of log n this is coming from $\log(\sqrt{n})$, so around $\log(\sqrt{n})$ many of these h_i do depend on greater than equal to root in many variables. Which of these sizes is the first ones if you look at the first log root n many h_i is obviously they all depend on at least $\sqrt{n}$ many variable each.

So, on this now you apply the hypergeometric calculation so, these X_i `s are big enough and when you will do a random partitioning. When you will randomly pick Y the discrepancies in one of these would be large. So, that part of the calculation is good but then the probability you get is bad and you have to pick the worst thing. So, you will only get in to the log n.

So on doing probability calculation we will get that $\Gamma(h_1 \cdots h_t)$ is high with probability smaller than $n^{-\Omega(\log n)}$. Which multiplied by s, that becomes your error probability so as soon as s exceeds into the log n you are in trouble. This is then a bad error probability so, what you get is determinant $n \times n$ or permanent $n \times n$ requires $n^{\log n}$ size multi linear formulas. So this is also a result by Raz.

In fact this he did first, this super polynomial or quasi polynomial lower bound for multi linear formulas. Then it was generalized to constant depth multi linear with Yehudayoff later. So as far as I remember this is still an open question improving this lower bound for multi linear formulas. To something like $2^{n^{\varepsilon}}$ for a constant ε something almost exponential instead of quasi polynomial.

What is the best formula you know for determinant, what is the size? What is the formula complexity of determinant? So, just because there is a poly size circuit for determinant circuit has this nice property that you can bring it down to log n depth with the multiplicative or product fanin 5 and then you can blow it up and make it a formula. So when you do this blow up you get an end to the log n size formula.

So determinant has a formula of size n to the log n but it is non multi linear of course because this even the circuit was non multi linear. This is for VP. Any circuit of size s because the results are general you do depth reduction bring it to this nice form and then make it a formula, I mean making a formula means that wherever you see fanout 2 you make copies, wherever you see a fanout of 2.

So, in the canonical form fanout is at most 2 just like fanin is at most 2. So, wherever you see fanin 2 you make a copy of both these things. So, every fanout you are blowing up by s the size and the depths are the levels are only log n. So you blow up to $s^{\log n}$, that is a general reduction. So hence determinant has a formula of this size. So, this $n^{\log n}$ is not so arbitrary.

There is a philosophical reason why it is coming, because actually this kind of non multi liner formulas do exist and this technique is getting stuck. There even in the case of multi linear assumption, but you would expect that when you have multi linear multi linearity exemption then nothing should exist it should be just to raise to n size, but that is not known. So, improve multi linear formula lower bound or upper bound, I mean both the things are open.

So, I think next topic I should start tomorrow. So from tomorrow we will move to a different model depth 4. Till now you have seen 2 measures, I mean the basic measures are only 2 and each of them were in their most beautiful form at the level of depth 3. So we started with this partial derivative, this multi depth 3 lower bound over finite fields, depth 3 or finite fields that was that gave one measure and this was multi linear depth 3 that gives another measure.

(Refer Slide Time: 01:05:56)

Now, we will go to depth 4 and define and work with the measure there and that model that we will study and prove lower bounds it's highly successful model in the last 10 years. May be even less last 6, 7 years. So this is degree restricted depth 4 circuit. So this will be $\Sigma\Pi^{a}\Sigma\Pi^{b}$, where we will restrict this a and b to slightly smaller than what it can be. So for, in the depth reduction you saw that this a and b you can bring down to $\sqrt{d}$ for a degree d polynomial computable by a sizes circuit. So, we can bring it down to depth 4 where the size will blow up to $s^{\sqrt{d}}$ and a and b will both be $\sqrt{d}$. So, actually we will take a b $\sqrt{d}$ and prove lower bounds. So that is the degree restricted depth 4 model mates, it seems to be a weaker model than depth 4 because degree could be as high as the size.

There is no reason for it to be square root of the final degree. So at the face of it, it is a restriction and for this we will see a measure and then use it to prove strong lower bounds; quite strong lower bonds.