YUFEI ZHAO: All right, today we're going to start a new topic an additive combinatorics. And this is a fairly central topic having to do with the structure of set addition. So the main players that we're going to be seeing in this chapter have to do with, if you start with a subset of some obedient group under addition-- not necessarily finite. So the obedient group that I'm going to keep in mind, the ones that will come up generally are integers, z mod n or the finite field model.

We're going to be looking at objects such as a sum set, so a plus b, meaning the set of elements that can be written as a sum, where you take one element from a and another from b. Likewise, you can also have a minus b defined similarly, now taking a minus b. We can iterate this operation. So kA, so 2A, 3A, 4A, for instance, means I add A to itself k times, not to be confused with a dilation, which we'll denote by k dot A. So this is notation for multiplying every element of A by the number k.

So given a subset of integers I can do these operations to the set. And I want to ask, how does the size of the set change when I do these operations? For example, what is the largest or the smallest? So how large or small can A plus A be for a given set size, A?

So if I allow you to use 10 elements, how can you make A plus A as big as possible? And how can you make it as small as possible? So this is not a hard question. How can you make it as big as possible? So what's the maximum size A plus A can be as a function of A?

Well, I'm looking at pairwise sums, so if there are no collisions between different pairwise sums, this is as large as possible. And then it's not hard to see that the maximum possible is the size of A plus 1, choose 2. So since at most, this many pairs and space possible if all sums are distinct. So for example, in integers, you can take 1, 2, 2 squared, and so on. So that will give you the span.

The minimum possible is also not too hard. We're allowed to work in a general obedient group. So in that case, the minimum could be just the size of A. The size is always at least the size of A. And this is tight if A is a subgroup. If you have a subgroup, then it's closed under addition. So the set does not expand under addition.

In the integers, you don't have any finite subgroups. So if I give you k integers, what's the smallest, the sum set can be?

AUDIENCE: 2k minus 1.

YUFEI ZHAO: 2k minus 1, right? So the example is when you have an arithmetic progression. So in integers, the minimum is 2k minus 1. And it's achieved for an arithmetic progression.

So let me just give you the one-line proof why you always have at least this many elements, is if A has elements sorted like this, then the following elements are distinct in the sum set. So you start with A plus A. And then you move A1 plus A2, A1 plus A3, and so on, to A1 plus Ak. And then you move the first element forward. OK, so here you already see 2k minus 1 distinct elements in A plus A.

OK, so these are fairly simple examples, fairly simple questions. So now let's get to some more interesting questions, which is, what can you say about a set if you know that it has small doubling? If it doesn't expand by very much, what can you tell me about the set? And for that, let me define the notion of a doubling constant.

So the doubling constant of A is defined to be the number which we often denote by k, the number obtained by dividing the size of A plus A by the size of A. And we would like to understand-- and this is the main question that's addressed in the upcoming lectures is, what is the structure of a set with bounded doubling constant?

So for instance, think of k as fixed. Let's say k is 100. If you know a set has doubling constant, at most, 100, what can you tell me about the structure of the set? So that's the main question. Let me show you in a second a few examples of sets the have bounded doubling constant.

So that's easy to check that those examples indeed have bounded doubling constant. And what this question amounts to is what is often known as an inverse question. So it's an inverse problem that asks you to describe in reverse-- so it's easy to check in the upcoming examples that all of those examples have bounded doubling constant. And what we want to say is, in reverse, that if a set has bounded doubling constant, then it must in some sense look like one of our examples. It's the harder inverse question.

OK, so let me give you some examples of sets with small doubling constant. One example we already saw earlier is that if you have an arithmetic progression. If you have an arithmetic progression, then the size of A plus A is always 2 times the size of A minus 1.

So the doubling constantly is always, at most, 2. That's pretty small. That's as small as you can get in arithmetic progressions is in the integers.

But if you start with an arithmetic progression and now I take just a subset of the elements of this progression, so if I take AP, and if I cross out a few elements, just a small number of elements from this progression, or even cross out most, but keeping a constant fraction of elements still remaining, I claim that's still a pretty good set. So if A can be embedded inside an AP such that the AP has size no more a constant factor and more than that of A, then the size of A plus A is, at most-- so we bound it by the size of P plus P, which is, at most, 2P.

So the doubling constant of A is, at most, 2C. So if you have a set which is at least 1/10 fraction of an AP, then you are doubling constant at most, 20, bounded. So this is another class of examples. So it's kind of a modification, some alteration of the arithmetic progression.

Another more substantial generalization of APs is that of a two-dimensional arithmetic progression. So you think of an arithmetic progression as equally spaced points on a line. But you can extend this in multiple dimensions, so like a grid.

So this is a two-dimensional arithmetic progression, but I still want to work inside the integers. So what we are going to do is project this picture onto the integers. So that's a two-dimensional arithmetic progression. And specifically, we have a set of the form, so x0 is the starting point, plus l1 of x1-- l1 times x1, and l2 2 times x2, where the little l's are integers, non-negative integers up to big L.

So that's a two-dimensional arithmetic progression. So the picture that you can have in mind is, on the number line, we can get, write down first an AP and then a few more points like that so that you can have a two-dimensional arithmetic progression. We say that this set, this two-dimensional arithmetic progression is proper if all terms are distinct.

And if that's the case, then I can write A plus A in a very similar format. So A plus A contains elements still of the same form, but now the indices go up to 2L minus 1. So you see that A plus A has size, at most, 4 times the original set, A. Also easy to see from this blue picture up there-- you expand that grid. It goes to, at most, 4 times the size. Yes?

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: So the question is, should it be?

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: 2x0?

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: What do you mean?

AUDIENCE: 2x0 plus l1 x0 plus 1?

YUFEI ZHAO: Ah, thank you, so 2x0, thank you. Yeah, 2x0, great. OK, so that's the size. And of course, you can generalize this example of a fairly straightforward way to d dimensional arithmetic progressions. And we call those things generalized arithmetic progressions.

So a Generalized Arithmetic Progression, which we will abbreviate by the letters GAP, is a set of numbers of the form as above, except now you have d different directions and indices, are also straightforward generalizations of what was earlier. So this is the notion of a generalized arithmetic progression. So think about projection of a d dimensional grid onto the integers.

And for GAPs, we say that it's proper if all the terms are distinct. We call d the dimension of the GAP. And for a GAP, whether it's proper or not, we call the size to be the product of the lengths.

And this is potentially larger. So this is larger than the number of distinct elements if it's not proper. So when I refer to the size of a GAP-- so I view the GAP more than just as a set, but also with the data of the initial point and the directions. If I talk about the size, I'm always referring to this quantity over here. Great.

So you see, if you take a GAP or a fraction of a GAP, then, as with earlier examples, you have small doubling. So if P is a proper GAP, of dimension d, then P plus P is, at most, 2 raised to power d times the size of P. And furthermore, if A is an arbitrary subset of P and such that A has size-- such that the GAP has size, at most, a constant fraction bigger than that of A, then A has small doubling as well.

So all of these are examples of constructions of sets where, for some fixed constant, the doubling constant, we can find a family of sets with doubling constant bounded by that number. And the natural question though is, are these all the examples? So have we missed some important family of constructions not covered by any of these examples?

And so that's the kind of inverse question I was referring to earlier. So all of these examples,

easy to check that they indeed have small doubling constant. Can you go in reverse? So can you ask the inverse question, if a set has small doubling constant, must it look like, in some sense, one of these sets?

It turns out this is not such an easy problem. And there is a central result in additive combinatorics known as Freiman's theorem which gives a positive answer to that question. So Freiman's theorem is now considered a central result in additive combinatorics. And it completely describes, in some sense, the sets that have small doubling.

And let me write down the statement. So if A is a subset of Z and has bounded doubling, then A is contained in a GAP of bounded dimension and size bounded by some constant times the size of the set. This is a really important result in additive combinatorics. The title of this chapter, "Structure of Set Addition," Freiman's theorem tells us something about the structure of a set with small doubling.

The next few lecturers are going to be occupied with proving this theorem. So this theorem will have-- its proof is involved and probably the most involved proof that we have in this course. And the proof will take the next several lectures. And we'll see a lot of different ingredients, a lot of really nice tools. Fourier analysis will come up at some point, but also other tools like the geometry of numbers and also some more classical additive combinatorics ideas.

But before starting on a proof, I want to offer a few remarks and historical remarks to just give you some more context about Freiman's theorem, but first, a few mathematical comments. In this conclusions of Freiman's theorem, I didn't mention properness. And that's mostly a matter of convenience. So you can, in fact, make the conclusion proper as well at the cost of increasing the number somewhat, but still constants depending only on k-- can guarantee properness as well. So there is an extra step involved which we'll not cover, because it's not entirely trivial, but it's also not too hard.

Freiman's original proof, so it's named after Freiman. So he proved that in the '60s. But at that time, the proof was considered rather obscure. It actually did not get the attention and the recognition that it deserved until much later. So this was kind of a forgotten result, a forgotten proof for a very long time until quite a bit later when Ruzsa-- Ruzsa's name will come up many times in this chapter.

Ruzsa came and gave a different proof of Freiman's theorem, and significantly cleaned up the proof, and offered many new ideas. So much of what we'll see today are results that we now

attribute to Ruzsa. And theorem sometimes is also called the Freiman-Ruzsa theorem.

But this result was really brought into-- brought as a highlight of additive combinatorics in the work of Gowers when he proved, that gave his new proof of Szemerédi's theorem, giving much better bounds. So he had to use quite a bit of serious additive combinatorics. And many of the ideas that went into Gowers' proof of Szemerédi's theorem came from this line of work, Freiman and Ruzsa. So and their work was, again, brought into prominence as a result of Gowers' Fields-Medal-winning work on Szemerédi's theorem.

So this is some of the history. And now Freiman's theorem is considered a central result in the area. You can see, it's a beautiful result. And it's also quite a deep result.

Let me mention a few things about bounds. So what do we know about this d of k and f of k? But first, an example-- so if the set A is dissociated in the sense of having no arithmetic structure, no coincidental sums colliding, so for example, if a of this form, then you see that-- also and we saw the size of A plus A, so A plus 1 choose 2. So in this case, the doubling constant is the size of A plus 1 divided by 2, so roughly on the same order as the size of A.

But what do you need to take in Freiman's theorem for d and for f? So how can I embed this A in generalized arithmetic progression? See, there is not a great way to do it. So I want to keep the size small. There is not a great way to do it.

So one way to do it is to use one direction for each of these elements. So contained in GAP-- now of course, there is always a trade off between dimension and size. But usually, not a great-- I mean, it's not such an important trade off. So but certainly it's contained in the GAP of dimension size of A minus 1 and size 2 to the size of A minus 1, by thinking about A as a cube.

And so you convince yourself that you basically cannot do much better. So the best possible bound that we can hope to prove is of the form d being, at most, linear in k, and f being, at most, exponential in k. So you see already, the bounds, that you have to lose some things. Yes?

AUDIENCE: Why can't we just make the dimension 1 and just let our arithmetic progression be 1 through 2 to the sine of A minus 1?

YUFEI ZHAO: OK, great, so that's a great question. So why can't we just make dimension 1 and have the entire thing be as part of a single linear arithmetic progression? So you can do that, but then I

can cook up other examples where I blow up this cube. So I ask you to think about how to do that. So you can try to blow up this cube so that you really do need the dimension to not be constant, so exercise.

So the best result is not quite this claim. So this is still open. So the best result so far is due to Tom Sanders, whose name came up earlier, as he has basically the best bound on Roth's theorem. And you know, many of these results are all related to each other.

So Sanders has-- so he showed that Freiman's theorem is true with d being, so basically k, but you lose a poly log factor. I think the big O is maybe 3, or 4, something like that, so not substantial. And then f of k is also basically exponential, but you lose a poly log factor in the exponent.

Just a minor note about how to read this notation-- so I mean, it's written slightly sloppily as log k raised to big O of 1. You should think k as constant, but somewhat big, because if k were 2, this notation actually doesn't make sense. So just think of chaos, as at least 3 when you read that notation.

All right, so we will prove Freiman's theorem. So this bound will show a worse bound. It actually will be basically exponentially worse, but it will be a constant. So it will be just a function of k. And that will take us the next several lectures.

So we'll begin by developing some tools that are, I think, of interest individually. And they can all be used for other things. So we'll develop some tools that will help us to show, eventually lead us to Freiman's theorem.

And I'll try to structure this proof in such a way that there are several goal posts that are also interesting. So in particular, just as what we did with Roth's theorem, we'll begin by proving a finite field analog of Freiman's theorem. So what would that mean, a finite field analog?

So what would a problem like this mean in F2 to the n? So in F2 to the n, so this is a finite field analog. If A plus A is small-- so I'm not trying to ask an inverse question. But what are examples of sets in F2 to the n that have small doubling?

AUDIENCE: 2 to the n.

YUFEI ZHAO: So 2 to the n, so you can take the entire space. Any other examples that have small doubling?

AUDIENCE: You can take a subspace.

YUFEI ZHAO: Exactly, I can take a subspace. So a subspace, well, it doesn't grow. So A plus A is the same as A. All right, so and also, as before, you can take a subset of a subspace. So then the analog of Freiman's theorem will say that A is contained in a subspace of size, at most, a constant times the size of A.

So this is the analog of Freiman's theorem in F2. And so we'll see, so this will be much easier than the general result about Freiman's theorem, but it will involve a subset of F2. And we'll see this theorem first. So we'll prove that next lecture.

Of course, this is much easier in many ways, because here, unlike before, I don't even have to think about what subspace to take. I can just take the subspace generated by the elements of A. All right, Any questions so far? Yes?

AUDIENCE: Is the f of k here still exponential in k?

YUFEI ZHAO: OK, so the question, is the f of k here still exponential in k? So the answer is, yes. And the construction is if you take A to be a basis.

OK, so let's start with some techniques and some proofs. So in this chapter, many things are named after Ruzsa. And at some point, it becomes slightly confusing which ones are not named after Ruzsa. But the first thing will be named after Ruzsa. So it's a Ruzsa Triangle Inequality.

All right, the Ruzsa Triangle Inequality tells us that, if A, B, and C-- so unless otherwise I tell you so, and I'll try to remind you each time, but basically, we're always going to be looking finite sets in an arbitrary obedient group always with an under addition-- then one has the inequality on their sizes of different sets. The size of A times the size of B minus C is upper bounded by the size of A minus B times the size of A minus C. So that's the Ruzsa Triangle Inequality.

Let me show you the proof. We will construct an injection from A cross B minus C to A minus B cross A minus C. Of course, if you can exhibit such an injection, then you prove the desired inequality. To obtain this injection, we start with an element a, d.

And for this a, d, so for each d, let me pick-- so if d is an element of B minus C, let us pick arbitrarily but stick with those choices a b of d in the set B and a c of d in the set C such that d

equals to b of d minus c of d. So because d is the set B minus C, it can be represented as a difference from one element from each set. So it may be represented in many ways.

But from the start, you pick a way to represent it. And you stick with that choice. And you label that function b of d and c of d.

Now I map a, d to the element a minus b of d and a minus c of d. So this is a map. I want to show that it is injective. Why is it injective? Well, to show something is injective, I just need to show that I can recover where I came from if I tell you the image. So I can recover a and d from these two numbers. So if-- sorry, new board.

OK, so well you basically can think about how you can recover a and d from the image elements. So if the image-- so I label that map phi. So that's phi up there. So if the image is given, then I can recover d.

So how can we recover the element d? So you subtract these two numbers. So d is minus x. And once you recover d, you can also then take a look at the first element. And you can recover a. So now you know d. I can now recover a.

OK, so then this is-- you can check this is an injection. And that proves the Ruzsa Triangle Inequality. OK, so it's short, but it's tricky. It's tricky.

OK, so why is this called Ruzsa's Triangle Inequality? Where is the triangle in this? The reason that it's given that name is that you can write the inequality as follows. Suppose we use rho A, B to denote this quantity obtained by taking the log of the size of A minus B divided by the square root of the product of their individual sizes, then the inequality says that the rho of B, C is, at most, rho of A, B plus rho of A, C, which looks like a triangle inequality.

So that's why it's called Ruzsa's Triangle Inequality, because this is-- don't take it too seriously, because this is not a distance. So rho of A, A is not equal to 0. But it certainly has the form of a triangle inequality, hence the name.

How should you think of Ruzsa's triangle inequality? So in this chapter, there's going to be a lot of symbol pushing around. And it's easy to get lost and buried in all of these symbols. And I want to tell you about how you might think about what's the point of Ruzsa's Triangle Inequality. How would you use it?

And the idea is that if you have a set with small doubling, we want to use Ruzsa's triangle

inequality and other tools to control its further doublings. So in particular, if-- so I'll say, applications.

So suppose you knew that 2A minus 2A is size, at most, k times A. So this is a stronger hypothesis than just A has small doublings. Even if you iterate it several times, you still have size, at most, constant times A.

I would like to start from this hypothesis and control further iterations, further subsets of A. And Ruzsa's Triangle Inequality allows us to do it, because by the Ruzsa's Triangle Inequality, setting B and C to be 2A minus A, we find that 3A minus 3A is, at most, 2A minus 2A squared over A, the size of A. So plug it in. This is what you get.

So if the size of 2A plus 2A is, at most, k times the size of A, then the size of 3A times 3A is-- blows up by a factor, at most, k squared. So it controls further doublings. And of course, we can iterate.

If we know set B and C to be 3A minus 2A, then what we get is 5A minus 5A is, at most, a size of 3A minus 3A square divided by the size of A. And so now you have a bound which is k to the 4 times A. And you can continue. You can continue.

OK, so this is all a consequence of Ruzsa's triangle. So starting with this hypothesis, now I get to control all the further doublings, the further subset iterations. I call them doublings, but they're no longer doubles, but further subsets. But this is a stronger hypothesis than the one that we start with in Freiman's theorem, because if you have that, then this 2A minus 2A is at least as large as the size of 2A. So can we start with just doubling constant and then obtain bounds on the iterations?

| it turns out you can. It will require another theorem. So this theorem is called Plunnecke inequality. But actually, these days, in literature, it's often referred to as Plunnecke-Ruzsa inequality. So Plunnecke initially proved it. But nobody understood his proof. And Ruzsa gave a better proof. And actually, recently, there was an even better proof. And that's the one I will show you.

So Plunnecke-Ruzsa inequality tells us that if A is subset of some obedient group, and has doubling constant, at most, k, then for all non-negative integers m and n, the size of mA minus nA is, at most, k to the m plus n times the size of A. So if you have bounded doubling, then the further iterations, the further subset iterations are also controlling size. I want you to think of

polynomial transformations in k as negligible.

So don't worry about that we're raising things here. k is constant. You should think of m and n as constant. So I'm changing k to some other constant. And in fact, I'm only changing it by a polynomial. So this is, like, almost no change at all. So this is tricky. So we'll do it after a short break.

All right, let's prove Plunnecke's inequality, Plunnecke-Ruzsa inequality. So the history of Plunnecke's inequality has some similarities with Freiman's theorem. So Plunnecke initially proved it, but his proof was hard to understand and was sort of left not understood for a long time until others like Plun and Ruzsa came in and really simplified the proof.

But even then, the proof was not so easy. And if I were teaching this course about 10 years ago, I would have just skipped this proof, maybe sketched some ideas, but I would have skipped the proof. And the proof, actually, it's a beautiful proof, but it uses some serious graph theory. It uses Menger's theorem about flows. You construct some graph. And then you try to understand its flows. It's very pretty stuff. And I do encourage you to look it up.

And then about eight years ago, Petridis found a proof, so a proof by Petridis, who was a PhD student that Tim Gowers at the time. And that was surprisingly short, and beautiful, and kind of surprised everyone that such a short proof exists, given that this theorem sat in that state for such a long time. And it's a pretty central step in the proof of Freiman's theorem.

We'll prove Plunnecke-Ruzsa via a slightly more general statement. So you see, it generalizes the earlier statement. Instead of having one set, it will be convenient to have two different sets. So let A and B be subsets of some obedient group, as usual.

If size of A plus B is, at most, k times the size of A, then mB minus nB has size, at most, k to the m plus n times the size of A for all non-negative integers m and n. So instead of having one set, so I have two sets, A and B. Of course, then you derive the earlier statement setting A and B to equal. So we'll prove this more general statement.

The proof uses a key lemma. And the key lemma says that if a subset x of A is non-empty and-- so if x is a non-empty subset of A that minimizes the ratio x plus B divided by size of x, and let k prime be this ratio, this minimum ratio, then so the conclusion says that x plus B plus C has size, at most, k prime times the size of x plus C for all sets C.

So that's the statement. I'll explain how you should think about the statement. These ratios

which you see in both hypotheses, how you should think about them is that there is this graph.

Let's say it's the group bipartite graph with the group elements on both sides. And the graph has edges, the bipartite graph, where the edges are from each vertex a drawn edge for each element of B. So I expand by B.

So if you have this graph and you start with some A on the left, then its neighbors on the right will be A plus B. And those ratios up there are the expansion ratios. so quantities like this, they are expansion ratios. You start with some set on the left and see by a what fraction does it expand if you look at the neighborhood.

So let's read the statement of the key lemma. It says, if you have a set x-- I look, so I have a set A. And I'm choosing a subset of A that minimizes the expansion ratio, so choose a non-empty subset that minimizes the expansion ratio. And if this minimum expense ratio is k prime, then, so x minimizes expansion ratio and expense ratios k prime, then x plus C also has expansion ratio, at most, k prime as well.

So that's the statement. I mentioned earlier that the previous proofs of this theorem went through some graph theory and Menger's theorem, that type of graph theory. You can kind of see where it might come in.

We're not going to do that. We're going to stick with additive combinatorics. We're going to stick with playing with sums, playing with additive combinatorics. So let's see how we can prove the statement up there, so using the key lemma.

So assuming key lemma, so let's prove the statement, the theorem up there. So take a non-empty subset x of A-- sorry, so x subset of A that minimizes the ratio x plus B divided by x. And let k prime be this minimum ratio.

Note that k prime is, at most, k, because if you plug in x equals the k, you get-- if you plug in x equals to A, you get k. But I'm choosing x to be possibly even lower. So k prime is, at most, k.

Now, applying the lemma, so applying the key lemma with C equals to B, we find that x plus 2B, so C, plug in B, x plus 2B has size, at most, k times size of x plus B. But the size of x plus B is, at most, k times the size of A. So we get k squared, so k times the size of x, at most, k squared x. So we're already in good shape.

If you iterate expansion twice-- so I imagine there is several chains of these bipartite graphs. If

you iterate this expansion twice, you still do not blow up by too much. So we can iterate further, so apply the lemma with C being now 2B, and then later 3B, and so on. So you find that x plus nB has size, at most, k raised to power n times the size of x for all non-negative integers, n.

What do we want to control? So we want to prove a bound on the size of mB minus nB. Take a look at the statement of Ruzsa Triangle Inequality.

Applying Ruzsa Triangle Inequality, we find that if we want to control mB minus nB, we can upper bound it by x plus mB x plus nB divided by the size of x. Because each of these two factors in the numerator are small expansions of x, now we can upper bound the whole expression by k to the m plus n times the size of x. And because x is a subset of A, we can do one more upper bound and obtain the bound that we are looking for. OK, so that proves the key lemma. It's OK?

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: Sorry, that proves the theorem, assuming the key lemma. Thank you, that's what I meant to say. Yeah, so that proves the theorem, assuming the key lemma. So now we do prove the key lemma.

Great, we need to prove the key lemma. And so Petridis' proof of the key lemma, it's quite surprising, in that it uses induction. And basically, we have not used induction in this course ever since the first or maybe the second lecture, and for good reason.

So everything in this course is fairly analytic. You know, you have these Roth bounds. And putting one extra vertex often doesn't really help. OK, so here, we're going to use induction on the size of C. OK, I just want to emphasize again that the use of induction here was surprising.

So if the base case-- always check the base case-- when C is 1, then plus C is a translation. So this shifts the set over. And so you can see that if you do plus C and minus 1, you raise the plus C. And the conclusion follows basically from the hypothesis. So in this case, x plus B plus C is equal to x plus B, which is, at most, k prime times the size x, by definition of-- so this actually is equal to the size of k. The base case is easy.

Now we do the induction step. So let's assume that the size of C is bigger than 1 and C is C prime plus an additional element, which we'll call gamma. So let's see this expression, x plus B

plus C, by separating it according to if its contribution came from C prime or not.

The contributions that came from C prime, I can write it like that. And then there are other contributions, namely those that came from this extra element. But I may have some redundancies in doing this. So I may have some redundancies coming from the fact that some of the elements in this set might have already appeared earlier.

So let me take out those elements by taking out elements where it already appeared earlier. So this means I'm looking at the set z being elements of x such that x plus B plus gamma is already a subset of x plus B plus C prime. So the stuff in yellow, I can safely discard, because it already appeared earlier.

So because of the definition of z, we see that z plus B plus gamma appears in x plus B plus C prime. So that union is valid. Now, z is a subset of x. So the expansion ratio for z is at least k prime, because we chose x to minimize this expansion ratio.

We would like to understand how big x plus B plus C. So let's evaluate the cardinality of that expression up there. The cardinality I can upper bound by the union of these sum of the sizes of the components.

So up there, so I just do a union bound on that expression up there. And now you see z is a subset of x. So I can split this expression up even further.

All right, now let's use the induction hypothesis. So we have some expression involving x plus B plus C prime. So now we apply induction hypothesis over here to this expression that has plus C prime.

And we obtain an upper bound which is k prime x plus C prime. And the two expressions on the right, well, one of them here is, by definition, coming from the expansion ratio of x. And then the other, we gave a bound just now.

OK, so we're almost there. So we are trying to upper bound the size of this quantity. So we decomposed it into pieces according to its contribution coming from this extra element. And we analyzed these pieces individually.

But now I want to understand the right-hand side, so x plus C. So let's try to understand the right-hand side. See, the x plus C, I can likewise write it as earlier by decomposing it into contributions from C prime and those from the extra element.

And as earlier, we can take out contributions that were already appearing earlier, which we now recall W plus gamma, where W is the set of elements in x, such that x plus gamma is already contained in x plus C prime. So this part was already included earlier. We don't need to include it any more.

A couple of observations that were different from earlier-- now this union I claim and say disjoined union. So this union is a disjoined union. So there is actually no more overlaps. And furthermore, W is contained in the set z from earlier. Any questions?

All right, therefore, the size of x plus C is equal to, because this is a disjoined union, x plus C prime plus the size of x minus the size of W, and which is-- so W, because W is contained in z, is x plus C prime plus the size of x minus the size of z. Now you compare these two expressions. And that proves the key lemma. OK? That's it. Yeah?

AUDIENCE: Can you explain one more time why it's a disjoined union?

YUFEI ZHAO: OK, great, so why is this a disjoined union? Now, I have the set here. So I'm looking at this x plus gamma. So think about, let's say, gamma equals to 0. So we translate, think about if gamma equals to 0.

So I include x, but if some element of x was already here, I take it out. So here is x plus C prime. And let's say this set is x. This W then would there be their intersection.

So now x minus W is just this set. So it's a disjoined union. So the points are, here, you're adding single elements, where there, you're adding some sets. So you cannot necessarily take a whole partition, necessarily. But here it's OK.

It's tricky. Yeah, it's tricky. And you know, this took a long time for people to find. It was found about eight years ago. And yeah, it was surprising when this proof was discovered. People did not expect that this proof existed.

And it's also tricky to get right. So the details-- I do it slowly. But the execution, like, the order that you take the minimalities is important. It's easy to mess up this proof. OK, any questions?

Let me show you, just as an aside, an application of this key lemma. So earlier we saw Ruzsa's Triangle Inequality. And you may wonder, what if you replace the minus signs in the theorem by plus signs?

I mean, if you replace the right-hand side, the two pluses by minuses, the same proof works. But if you replace all the minus signs by plus signs, you see, the proof doesn't work anymore. Just give yourself a moment to convince yourself that. If you just replace all the minus signs by plus signs, it doesn't work anymore, but it's still true.

So this is more of an aside. We will not use it. But it's nice. It's fun. So we have the inequality A B plus C bounded by A plus B A plus C.

So hopefully you've convince yourself that if you follow our notes with the previous proof, you are you're not going to get it. You're not going to prove this this way. It's still true. So how can we prove it? So we are going to use the key lemma. So first, the statement is trivial if A is empty. So let's assume that's not the case.

Let x be a subset of A that minimizes the expression or the expansion ratio x plus B divided by x as in the key lemma. So let k denote the quantity A plus B over A, so the expansion ratio for A, and k prime be the expansion ratio for x. So the quantities came up earlier. k prime is, at most, value of k, because of our choice of x.

So the key lemma gives x B plus C-- OK, and this, it's really amazing what's happening. It seems like we're just going to throw in some extra stuff. So I'm going to upper bound it by x plus B plus C. I'm just going to throw in some extra stuff. And then by the lemma, I can upper bound this expression by k prime times the size of x plus C. So that's what the lemma gives you.

And because x is a subset of A, we can upper bound it by the size of A plus C. And now k prime is, at most, the size of k. k prime is, at most, k. So you have that.

But now look at what the definition of k is. And that's it. So that's how you can prove this harder version of Ruzsa's Triangle Inequality, Yes, question?

AUDIENCE: Are there equality cases for this?

YUFEI ZHAO: All right, question, are there equality cases for this? Yes, so I mean, if you're in a subgroup, then all things are equal, although, if A, B, and C are all the same subgroup of some finite obedient group.

AUDIENCE: What if you're working in the integers?

YUFEI ZHAO: Great, yeah, so the question is, what if you're working in integers? That's a good question. I mean, you can suddenly get expansion ratio of two if you have-- no, OK.

Right, yeah, so that's a good question. Can you get equality cases? If you set A, B, and C to be sets of very different sizes, AP is a very different set. Yes?

AUDIENCE: If you set B being true for the single element and B and C to just be sets that have full extension so that B plus C is not [? B. ?]

YUFEI ZHAO: Great, yeah, so you take A to be a single-element set, then it could be that B plus C is the same as the size of B times the size of C if B and C have no additive interactions. Yeah?

AUDIENCE: Are there other known proofs of this that are less involved?

YUFEI ZHAO: OK, are there other known proofs of this? I don't know. I'm not aware of other proofs. It would be nice to find a different proof. More questions? Yeah?

AUDIENCE: How did come up with this?

YUFEI ZHAO: How did he come up with this? You know, Petridis did a very long PhD. He spent, I think, seven or eight years in his PhD. And he eventually came up with this proof. So he must have thought a lot about this problem.

But the already proofs are still nice. The earlier proofs, I think they are worth looking at. They are looking at expansion ratios in graphs. So you take a sequence of graphs, multi-partite graphs. And you think about expansion. And you think about flows.

It's, again, not easy at all, but maybe more motivated if you're used to think about expansions and flows in graphs. And this one really distills the core ideas of that proof, but looks something you can teach in half a lecture, whereas before this proof came about, I could have taught the proof, but most likely, I would have just skipped it.

To just give you a sense of what's coming up ahead, so going forward, the first thing we'll do in the next lecture is we'll show-- we'll see the proof of the Freiman's theorem in the finite field setting, so in F2 to the n. There is one more thing, one more very quick lemma called the covering lemma, Ruzsa are Covering Lemma, that I will tell you. And then once we have that, then we can prove Freiman's theorem in the finite field setting.

But then moving on to the integers, we'll need to understand how to think about the integers.

Well, if you start with a subset of integers, they could, even if you have a small number of elements, they could be spread out, really, all over the place. But because you only care about the additive structure within the integers, you can try to model that very spread-out set of integers to something that is very compact. So there is something called the modeling lemma, Ruzsa's Modeling Lemma, that we'll see next time. And that will play a pretty important role.

Before finishing off, I also want to mention that Freiman in his work, so he had this result. And he also wrote a book I think called *The Structural Theory of Set Addition,* or something like that, that emphasized this connection. He tried to draw this analogy sort of comparing additive combinatorics to geometry in the sense of cline, where in order to understand sets, you don't think about sets.

You think about maps between sets, which was kind of an obscure idea at the time. But we'll see next lecture that this actually is a very powerful, it's a very influential idea to really think about a sets of integers under transformations that only preserve their additive structure. So we'll see this next time.