YUFEI ZHAO: Today we want to look at the sum product problem. So for the past few lectures, we've been discussing the structure of sets under the addition operation. Today we're going to throw in one extra operation, so multiplication, and understand how sets behave under both addition and multiplication.

And the basic problem here is, can it be the case that A plus A, A times A, which is, analogously, the set of all pairwise products of elements from A-- can these two sets be simultaneously small, that is, the same for some single A? Can we have it so that A plus A and A times A are simultaneously small? For example, it's easy to make one of them small.

We've seen examples where if you take A to be an arithmetic progression, then A plus A is more or less as small as it gets. But for such an example, you see A times A is pretty large. It's actually not so clear how to prove how large it

Is. And there are some very nice proofs. And this problem has actually been more or less pinned down. But the short version is that A times A has size close to its maximum possible.

So it turns out the size of A times A is almost quadratic. So this number is actually now known fairly precisely. So this problem of determining the size of A times A for the interval 1 through N is known as the Erdos multiplication table problem.

So if you take an N by N multiplication table, how many numbers do you see in the table? So that turns out to be sub-quadratic, but not too sub-quadratic. So this problem has been more or less solved by Kevin Ford. And we now know a fairly precise expression, but I don't want to focus on that. That's not the topic of today's lecture. This is just an example.

Alternatively, you can take A times A to be quite small by taking A to be a geometric progression. Then it's not too hard to convince yourself that A plus A must be fairly large in that case. And the geometric progression doesn't have so much additive structure, so A plus A will be large.

So can you make A plus A and A times A simultaneously small? So there's this conjecture that the answer is no. And this is a famous conjecture in this area, known as the Erdos similarity conjecture on the sum product problem, which states that for all finite sets of real numbers, either A plus A or A times A has to be close to quadratic size.

So that's the conjecture. It's still very much open. Today I want to show you some progress towards this conjecture via some partial results. And it will use a nice combination of tools from graph theory and incidence geometry, so it nicely ties in together many of the things that we've seen in this course so far.

So Erdos and Szemeredi proved some bound, which is like 1 plus c for some constant c. Today we'll show some bounds for somewhat better c's. So you'll see.

The first tool that I want to introduce is a result from graph theory known as the "crossing number inequality." So you know that planar graphs are graphs where you can draw on the planes so that the edges do not cross. And there are some famous examples of non-planar graphs, like K5 and K 3, 3.

But you can ask a more quantitative question. If I give you a graph, how many crossings must you have in every drawing of this graph? And the crossing number inequality provides some estimate for such a quantity.

So given the graph G, denoted by cr, so crossing of G, to be the minimum number of crossings in a planar drawing of G. There is a bit of subtlety here, where by a planar drawing, do I mean using line segments or do I mean using curves? It's actually not clear how it affects this quantity here. That's a very subtle issue.

So for planar graphs, there's a famous result that more or less says if a planar graph can be drawn using continuous curves, then it can be drawn using straight lines. But the minimum number of crossings, the two different ways of drawings, they might end up with different crossing numbers. But for the purpose of today's lecture, we'll use a more general notion, although it doesn't actually matter for today which one we'll use-- so planar drawing using curves.

Draw the graph where edges are continuous curves. How many crossings do you get? The crossing is a pair of edges that cross.

You can ask-- it's just a cross over point that can-- it doesn't matter. So there are many different subtle ways of defining these things. They won't really come up for today's lecture.

The crossing number inequality is a result from the '80s, which give you a lower-bound estimate on the number of crossings. If G is a graph with enough edges-- the number of

edges is, let's say, at least four times the number of vertices-- then the number of crossings of every drawing of G is at least the number of edges cubed divided by the number of vertices squared. And there's an extra constant factor, which is some constant. So the constant does not depend on the graph.

In particular, if it has a lot of edges, then every drawing of G must have a lot of crossings. So the crossing number inequality was proved by two separate independent works, one by Ajtai, Chvatal, Newborn, Szemeredi and the other by Tom Leighton, our very own Tom Leighton. So let me first give you some consequences of this theorem, just for illustration.

So if you have an n-vertex graph with a quadratic number of edges, then how many crossings must you have? You plug in these parameters into the theorem. See that it has necessarily n to the 4th crossings.

But if you just draw the graph in some arbitrary way, you have at most n to the 4 crossings, because a crossing involves four points. So when you have a quadratic number of edges, you must get basically the maximum number of crossings. The leading constant term factor is an interesting problem, which we're not going to get into.

Let's prove the crossing number inequality. First, the base case of the crossing number inequalities is when you can draw a graph with no crossings. And those are planar graphs.

So for every connected planar graph, if it has at least one cycle-- and you'll see why in a second, why I say this-- if with at least one cycle, so that's not a tree, we must have that 3 times the number of faces is at most 2 times the number of edges. So here, we're going to use the key tool being Euler's formula, which we all know as the number of vertices minus the number of edges plus the number of faces equals to 2. We're here for face, because I draw a planar graph, and so I count the faces. Here there are two faces, outer face, inner face, count edges and vertices, so you have Euler's formula up there.

And plug in Euler's formula for a planar graph with at least one cycle, so we can obtain this consequence over here, because every face is adjacent to at least three edges. If you go around the face, you see these three edges, and every edge is counted exactly twice, is adjacent to exactly two faces. So you do the double counting, you get that inequality up there.

So plugging these two into Euler gets you that inequality up there. Plugging these two into Euler, we get that the number of edges is almost 3 times the number of vertices minus 6. So

for this leaves that inequality, but plug it into Euler, plug in this into Euler, you get this. So we have that the number of edges is at most 3 times the number of vertices for every graph G.

So here, we require that the graph is planar and has at least one cycle, but even if we drop the condition that it has at least one cycle but just require that it's planar, every planar graph G satisfies this inequality over here. So in other words, you might have heard before, in a planar graph, the average degree of a vertex is almost 6. So in particular, the crossing number of a graph G is positive if the number of edges exceeds 3 times the number of vertices. It's not planar, so it has at least one crossing every drawing.

And by deleting an edge from each crossing, we get a planar graph. You draw the graph. You have some crossings.

You get rid of an edge associated with each drawing. Then you get a planar graph. If you look at this inequality and you account for the number of edges that you deleted, we obtain then the inequality that the number of edges minus the number of crossings is at least 3 times the number of vertices. So we obtain the inequality that the lower bounds in number of crossings as the number of edges minus 3 times the number of vertices, this one.

So that's some lower bound on the crossing number. It's not quite the bound that we have over there. And in fact, if you take a graph with a quadratic number of edges, this bound here only gives you quadratic lower bound on the crossing number, some lower bound. But it's not a great lower bound.

And we would like to do better. So here's a trick that is a very nice trick, where we're going to use this inequality to upgrade it to a much better inequality, bootstrap it to a much tighter inequality. So this involves the use of the probabilistic method.

Let me denote by p some number between 0 and 1, to be decided later. And starting with a graph G, let's let G prime, with vertices and edges being V prime and E prime, be obtained from G by randomly deleting some of the vertices, or rather randomly keeping each vertex with probability p, independently for each of these vertices. So you have some graph G.

I keep each vertex with probability p. And I delete the remaining vertices. And I get a smaller graph. I get some induced subgraph.

And I would like to know what can we say about the crossing number of the smaller graph in comparison to the crossing number of the original graph? For the smaller graph, because it's

still a planar graph so G prime-- so it's still a graph. It's not a planar graph, but it's still a graph, so G prime still satisfies this inequality up here.

So G prime still satisfies that the number of crossings in every drawing of G prime is at least the number of edges of G prime minus 3 times the number of vertices of G prime. But note that G prime is a random graph. G was fixed, given. G prime is a random graph.

So let's evaluate the expectation of both quantities, left-hand side and right-hand side. If this inequality is true for every G prime, the same inequality must be true in expectation. Now what do we know about all the expectations of each of these quantities?

The number of vertices in expectation-- that's pretty easy. So this one here is p times the original number of vertices. The number of edges is also pretty easy. Each edge is kept if both endpoints are kept. So this expectation on the number of edges remaining is also pretty easy to determine.

The crossing number of the new graph-- that I have to be a little bit more careful of, because when you look at the smaller graph, maybe there's a different way to draw it that's not just deleting the sum of the vertices from the original graph. So even though the original graph might have a lot of crossings, when you go to a subgraph, maybe there's a better way to draw it. But we just need an inequality in the right direction. So we are still OK.

And I claim that the crossing number of G prime is in expectation at most p to be 4th times the crossing number of G. Because if you keep the same drawing, then the expected number of crossings that are kept-- each crossing is kept if all four of its end points are kept. So each crossing is kept with probability p to the 4th.

So you can draw it in expectation with this many crossings. Maybe it's much less. Maybe there's a better way to draw it, but you have an inequality going in the right direction.

Looking at that inequality up there in yellow, we find that the crossing number of G is at least p to the minus 2 E minus 3p to the minus 3. And this is true for every value of p between 0 and 1. So now you pick a value of p that works most in your favor.

And it turns out you should do this by setting these two equalities to be roughly equal to each other. So setting p between 0 and 1 so that 4 times the-- basically, set these two terms to be roughly equal to each other. And then we get that this quantity here is at least the claimed

quantity, which is E cubed over V squared up to some constant factor, which I don't really care about.

In order to set p, I have to be a little bit careful that p is between 0 and 1. If you set p to be 1.2, this whole argument doesn't make any sense. So this is OK.

So we know p is at most one as long as E is at most 4p. I mean, the 4 here is not optimal, but if 4 were 2, then it's not true. So if E is 2V, you can have a planar graph, so you shouldn't have a lower bound on the crossing number.

So this is the proof of the crossing number inequality. As I said, if you have lots of edges, then you must have lots of crossings. Any questions?

So let's use the crossing number inequality to prove a fundamental result in incidence geometry. Incidence geometry is this area of discrete math that concerns fairly basic-sounding questions about incidences between, let's say, points and lines. And here's an example.

So what's the maximum number of incidences between endpoints and end lines, where by "incidence" I mean if p-- so curly p-- is a set of points, and curly l is a set of lines, then I write I of p and l to be the number of pairs, one point, one line, such that the point lies on the line. So I'm counting incidences between points and lines. You can view this in many ways. You can view it as a bipartite graph between points and lines, and we're counting the number of edges in this bipartite graph.

So I give you end points, end lines. What's the maximum number of incidences? It's not such an obvious question. So let's see how we can approach this question.

But first, let me give you some easy bounds. So here's a trivial bound-- so here, I want to know if I give you some number of points, some number of lines, what's the maximum number of incidences. So a trivial bound is that the number of incidences is at most the product between the number of points and the number of lines.

One point, one line, at most one incidence. So that's pretty trivial. We can do better.

So we can do better because, well, you see, let's use this following fact, that every line-- so every pair of points determine at most one line. I have two points. There's at most one line that contains those two points.

Using this fact, we see that the number of-- so let's count the number of triples involving two points and one line such that both points lie on the line. So how big can this set be? So let's try to count it in two different ways.

On one hand, this quantity is at most the number of points squared, because if I give you two points, then they determine this line-- so at most the number of points squared. But on the other hand, we see that if I give you a line, I just need to count now the number of-- let me also require that these two points are distinct. So if I give you a line, I now need to count the number of pairs of points on this line.

So I can enumerate over lines and count line by line how many pairs of points are on that line. So I get this quantity over here. On each line, I have that contribution.

And now, using Cauchy-Schwartz inequality, we find that this squared term is at least the number of incidences divided by the number of lines. And the remaining minus 1 term contributes just to the number of incidences. So the first is by Cauchy-Schwartz.

So putting these two inequalities together, we get some upper bound on the number of incidences. If you have to invert this inequality, you will get that the number of incidences between points and lines is upper bounded by the number of points times the number of lines raised to power 1/2 plus the number of lines. So that's what you get from this inequality over here.

By considering point-line duality-- so whenever you have this kind of setup involving points and lines, you can take the projected duality and transform the configuration into-- lines into points and points into lines, and the incidences are preserved. So I also have an inequality. By duality-- I also have an inequality where I switch the roles of points and lines.

So I is already the numbers. I don't need to put an extra absolute value sign. So the number of points and lines is upper bounded by the number of lines times the square root of a number of points plus an extra term, just in case there are very few lines. So these are the bounds that you have so far.

And the only thing that we have used so far is the fact that every two points determine at most one line, and every two lines meet at at most one point. So these are the bounds that we get. And in particular, for end points and end lines, we get the number of incidences is-- they go off n to the 3/2.

This should remind you of something we've done before. So in the first part of this course, when we were looking at extremal numbers, where did 3/2 come up?

AUDIENCE: [INAUDIBLE] like C4?

YUFEI ZHAO: C4, yeah. So if you compare this quantity to the extremal number of C4, it's also n to the 3/2. And in fact, the proof is exactly the same. All we're using here is that the incidence graph is C4-free

So in fact, this is an argument about C4-free graphs. So this fact here, every two points determine at most one line, is saying that if you look at the incidence graph, there's no C4. That's all we're using for now. Any questions?

So is this the truth? Now, back when we were discussing the extremal number for C4-free graphs, we saw that, in fact, this is the correct order. And what was the construction there?

So the construction also came from incidences, but incidences of taking all lines and points in the finite field plain, Fq squared. If you look at all the lines and all the points in a finite field plain, then you get the correct lower bound for C4. But now we are actually working in the real plane, so it turns out that the answer is different when you're not working the finite field.

We're going to be using the topology of the real plane. And we're going to come up with a different answer. So it turns out that the truth for the number of maximum number of incidences in the plane, for points and lines in the real plane, is not exponent 3/2, but turns out to be 4/3.

And this is a consequence of an important result in incidence geometry, a fundamental result, known as the Szemeredi-Trotter theorem. So the Szemeredi-Trotter theorem says that the number of incidences between points and lines is upper bounded by this function where you look at the number of points times the number of lines, and each raised to power 2/3 and plus some additional terms, just in case there are many more lines compared to points or way more points compared to lines.

So that's the Szemeredi-Trotter theorem. And as a corollary, you see that n points, n lines give you at most n to the 4/3 incidences, in contrast to the setting of the finite field plain, where you can get n to the 3/2 incidences. So somehow, we have to use the topology of the real plane for this one.

And I want to show you a proof-- turns out not the original proof, but it's a proof that uses the crossing number inequality to prove Szemeredi-Trotter theorem. You see, in crossing number inequality, we are using the topology of the real plane. Where?

AUDIENCE: Euler's formula.

YUFEI ZHAO: Euler's formula, right. So the very beginning, Euler's formula has to do with the topology of the real plane. Now, this bound turns out to be tight. So let me give you an example showing that the 4/3 exponent is tight.

And the example is, if you take p to be this rectangular grid of points, and L to be a set of lines-- so I'm going to write the lines by their equation, where the slope is an integer from 1 through k and the y-intercept is an integer from 1 through k squared. And you see here that every line in L contains exactly k points from P. So we got in total k to the 4th incidences, which is on the order of n to the 4/3. So n to the 4/3 third is the right answer.

Now let me show you how to prove Szemeredi-Trotter theorem from the crossing number inequality. It turns out to be a very neat application that's almost a direct consequence once you set up the right graph. And the idea is that we are going to draw a graph based on our incidence configuration.

So first, just to clean things up a little bit, let's get rid of lines in L with 1 or 0 points in P. So this operation doesn't affect the bounds. So you can check. These lines don't contribute much to the incidence bound, and only contributes to this plus L. So you can get rid of such lines.

So let's assume that every line in L contains at least two points from P. And let's draw a graph based on this incidence structure. So if I have-- so suppose these are my points and lines.

I'll just draw a graph where I keep the points as the vertices, and I put in an edge. It's a finite edge that connects two adjacent points on the same line. So I get some graph.

Let me make this graph a bit more interesting. So I get some graph. And how many crossings, at most, does this graph have?

So the number of crossings of G is at most the number of lines squared, because a crossing comes from two lines. So here, you have a crossing. A crossing comes from two lines. Number of crossings is at most number of lines squared.

On the other hand, we can give a lower bound to the number of crossings from the crossing number inequality. And to do that, I want to estimate the number of edges. And this is the reason why I assume every line contains at least two points from P, because a line with now k incidences gives k minus 1 edges.

And if k is at least 2, then k minus 1 is at least k over 2, let's say. I don't care about constant factors. So by crossing number inequality, the number of crossings of G is at least the number of edges cubed over the number of vertices squared, which is at least the number of incidences of this configuration cubed over the number of points squared. Actually, number of vertices is the number of points. And number of edges, by this argument here, is on the same order as the number of incidences.

Putting these two facts together, we see-- there was one extra hypothesis in crossing number inequality. Provided that this hypothesis holds, which is that the number of incidences is at least 8 times the number of points, so that the original hypothesis holds. So putting everything together, and rearranging all of these terms, and using upper and lower bounds on the crossing number, we find that the number of incidences is upper bounded by-- the main term you see is just coming from these two, but there are a few other terms that we should put in, just in case this hypothesis is violated, and also to take care of this assumption over here, so adding a couple of linear terms corresponding to the number of points and the number of lines.

If this hypothesis is violated, then the inequality is still true. So this proves the crossing numbers inequality. Any questions?

So we've done these two very neat results. The question is, what do they have to do with the sum product problem? So I want to show you how you can give some lower bound on the sum product problem using Szemeredi-Trotter theorem.

So it turns out that the sum product problem is intimately related to incidence geometry. And the reason-- you'll see in a second precisely why they're related, but roughly speaking, when you have addition and multiplication, they're are kind of like taking slope and y-intercept of an equation of a line. So there are two operations that are involved. So turns out, many incidence geometry problems can be set up and a way-- so many sum product problems can be set up in a way that involves incidence geometry.

And a very short and clever lower bound to the sum product problem was proved by Elekes in

the late '90s. So he showed the bound that if you have a subset of finite, subset of reals, then the sum set size times the product set size is at least A to the 5/2. As a corollary, one of these two must be fairly large. The max of the sum set size and the product set size is at least a to the 5/4.

Let me show you the proof. I'm going to construct a set of points and a set of lines based on the set A. And the set of points in R2 is going to be pairs x comma y, where the horizontal coordinate lies in the sum set, A plus A, and the vertical coordinate lies in the product set, A times A. And a set of lines is going to be these lines-- y equals to a times x minus a prime, where a and a prime lie in A.

So these are some points and some lines. And I want to show you that they must have many incidences. So what are the incidences?

So note that the line y equals to a times x minus a prime-- it contains the points a prime plus b and ab, which lies in P for all b in A. You plug it in. If you plug in a prime plus b into here, you get ab. And this point lies in P, because the first coordinate is the sum set.

The second coordinate lies in the product set. So each line in L contains many incidences. So each line in L contains a incidents. So this line, each line in L contains a incidences.

Also, we can easily compute the number of lines and the number of points. The number of points is A plus A size times the size of A times A. And the number of lines is just the size of A squared.

So by Szemeredi-Trotter, we find that the number of incidences is lower bounded by noting this fact here. We have many incidences. So the number of lines, each line contributes a incidences.

But we also have an upper bound coming from the Szemeredi-Trotter theorem. So plugging in the upper bound, we find that you have-- so now I'm just directly plugging in the statement of Szemeredi-Trotter. The main term is the first term. You should still check the latter two terms, but the main term is the first term.

So plugging in the values for P and L, we find this is the case, plus some additional terms, which you can check are dominated by the first term. So let me just do a big O over there. Now you put left and right together, and we could obtain some lower bound on the product of

the sizes of the sum set and the product set, thereby yielding allocations.

So this is some lower bound on the sum product problem. And you see, we went through the crossing number inequality to prove Szemeredi-Trotter, a basic result in incidence geometry. And viewing sum product as an incidence geometry problem, one can obtain this lower bound over here. Any questions?

I want to show you a different proof that was found later, that gives an improvement. And there's a question, can you do better than 5/4? So it turns out that there was a very nice result of Solymosi sometime later that gives you an improvement.

Solymosi proved in 2009 that if A is a subset of positive reals, then the size of A times A multiplied by the size of A plus A squared is at least size of A to the 4th divided by 4 ceiling log of the size of A, where the log is base 2. So don't worry about the specific constants.

A being in the positive reals is no big deal, because you can always separate A as positive and negative and analyze each part separately. So as a corollary to Solymosi's theorem, we obtain that for A, a subset of the reals, the sum set and the product set, at least one of them must have size at least A raised to 4/3 divided by 2 times log base 2 size of A raised to 1/3 third. So basically, A to the 4/3 minus little one in the exponent, so better than before. And this is a new bound.

I want to note that in this formulation, where we are looking at lower bounding this quantity over here, this is tied up to logarithmic factors, by considering A to be just the interval from 1 to n. If A is the interval from 1 to n, then the left-hand side, A plus A, is around size n. So you have n squared. And A times A is also, I mentioned, around size n squared. So this inequality here is tight. The consequence is not tight, but the first inequality is tight.

So in the remainder of today's lecture, I want to show you how to prove Solymosi's lower bound. And it has some similarities to the one that we've seen, because it also looks at some geometric aspects of the sum product problem. But it doesn't use the exact tools that we've seen earlier.

It does use some tools that were related to the lecture from Monday. So last time, we discussed this thing called the "additive energy." You can come up with a similar notion for the multiplication operation, so the "multiplicative energy," which we'll denote by E sub, with the multiplication symbol, A. So the multiplicative energy is like the additive energy, except that

instead of doing addition, we're going to do a multiplication instead.

So one way to define it is the number of quadruples such that there exists some real lambda such that a, comma, b equals to lambda c, comma, d. So basically the same as additive energy, except that we're using multiplications instead. By the Cauchy-Schwartz inequality-- and this is a calculation we saw last time, as well-- we see that if you have a set with small product, then it must have high multiplicative energy.

So last time, we saw small sum set implies high additive energy. Likewise, small product set implies high multiplicative energy. In particular, the multiplicative energy of A, you can rewrite it as sum over all elements x in the product set of the quantity, which tells you the number of ways to write x as a product, this number squared and then summed over all x.

By Cauchy-Schwartz, we find that this quantity here is lower bounded by the size of A to the 4th divided by the size of A times A. So to prove Solymosi's theorem, we are going to actually prove a bound on the energy, instead of proving it on the set. We're going to prove it on the energy.

So it suffices to show that the multiplicative energy is at most 4 times the sum set size times-- so let me divide the energy by log of A. So when you plug this into this inequality, it would imply that.

So it remains to show this inequality over here upper bounding the multiplicative energy. There's an important idea that we're going to use here, which is also pretty common in analysis, is that instead of considering that energy sum here, we're going to consider a similar sum, except we're going to chop up the sum into pieces according to how big the terms are, so that we're only looking at contributions of comparable size. And so this is called a "dyadic decomposition."

The idea is that we can write the multiplicative energy similar to above, but instead of summing over x in the product set, let me sum over s in the quotient set. So you can interpret what this quotient A is. This is the set of all A divided by B, where A and B are in A. A is a set of positive reals, so I don't need to worry about division by 0.

So what remains, then, is the intersection of s times A and A squared. Remember, s times A is scaling each element of A by s. So we have this quantity over here.

So I want to break up the sum into a bunch of smaller sums, where I want to break up the sum

according to how big the terms are, so that inside each group, all the terms are roughly of the same size. And easiest way to do this is to chop them up into groups where everything inside the same collection differs by at most a factor of 2. So that's why it's called a dyadic decomposition, going from 0 to-- the maximum possible here is basically A.

So let's look at i going from 0 to log base 2 of A. So this is the number of bins. And partition the sum into sub-sums where I'm looking at the i-th sub-sum consisting of contributions involving terms with size between 2 to the i and 2 to the i plus 1. Break up the sum according to the sizes of the summands.

By pigeonhole principle, one of these summands must be somewhat large. So by pigeonhole, there exists a k such that setting D to be the s such that that corresponds to the k-th term in the sum. So one has that this sum coming from just contributions from D is at least-- so it's at least the multiplicative energy divided by the number of bins.

All of that many bins-- by pigeonhole, I can find one bin that's a pretty large contribution to the sum. And the right-hand side, we can upper bound each term over here by 2 to the 2k plus 2, and the number of terms as the size of D. Let me call the elements of D S1 through Sm, where S1 through Sm are sorted in increasing order.

Now let me draw you a picture of what's going on. Let's consider for each element of D, so for each i and m, let's consider the line given by the equation y equals to s sub i times x. Let me draw this picture where I'm looking at the positive quadrant, so I have a bunch of points in the positive quadrant.

And specifically, I'm interested in these points whose coordinates, both coordinates are elements of A. And I want to consider lines through points of A, but I want to consider lines where it intersects this A cross A in the desired number of points. And we find those set, and then let's draw these lines over here, where this line here, L1 has slope exactly S1, and L2, L3, and so on.

I want to draw one more line, which is somewhat auxiliary, but just to make our life a bit easier. Finally, let's let L of m plus 1 be the vertical line, or rather be the vertical ray, which goes to the minimum element of A above Lm. So it's this line over here. That's Lm plus 1.

So in A cross A, I draw a bunch of lines. So now all the lines-- so all these lines involve some point of A and the origin, but I don't draw all of them. I draw a select set of them. And what we

said earlier says that the number of lines, the number of points on each of these strong lines, is roughly the same for each of these lines.

Let's let capital L sub j denote the set of points in A cross A that lie on the j-th line. So that's L1, L2, and so on. I claim that if you look at two consecutive lines and look at the sum set of the points in A cross A that intersect, you're looking at two lines, and you're adding up points on those two lines. So you form a grid.

So you end up forming this grid. And the number of points on this grid is precisely the product of these two point sets. Moreover, the sets Lj plus L sub j plus 1 are disjoint for different j.

And this is where we're using the geometry of the plane here. Because the sum of L1 and L2 lies in the span, the sum of L2 and L3 in a different span, so they cannot intersect. So they lie in-- so since they span disjoint regions, L1 plus L2 lies here, L2 plus L3 lies there, and so on. But they're all disjoint.

Now let's put everything that we know together. Remember, the goal is to upper bound the multiplicative energy as a function of the sum set. So in other words, we want to lower bound the sum set. So I want to show you that this A plus A has a lot of elements. There's a lot of sums.

And I have a bunch of disjoint contributions to these sums. So let's add up those disjoint contributions to the sums. You see that the size of A plus A squared is the same as the size of the product set A plus A.

So this is Cartesian product. Here is-- this is a Cartesian product, in other words, the grid that is strong up there. I add this product to itself. So I should get the same set here.

But how big is this sum set? That grid, that lattice grid added to itself, how big should it be? I want to lower bound the number of sums.

And the key observation is up there. We can look at contributions coming from distinct spans. In particular, this sum here, so this sum set here, size is lower bounded by these distinct Lj plus L j plus 1's. I threw away a lot. I only keep the lines on the L's, and I only consider sums between consecutive L's. That should be a lower bound to the sum set of the grid with itself.

But you see, and here, we're using these different-- for different j's, these contributions are destroyed. But by what we said up there, Lj plus L j plus 1 is a grid. So it has size Lj times L j

plus 1. And the size of each Lj is at least 2 to the k. So the sum here is at least m times 2 to the 2k.

But we saw over here that the energy lower bounds this 2 to the 2k. So we have a lower bound that is the multiplicative energy of A divided by 4 times the log base 2 of the size of A. So don't worry so much about the constant factors. That's just the order of magnitude that is important.

And that's it. Yep.

AUDIENCE: How do you know that the size of big L sub m plus 1?

YUFEI ZHAO: Great. The question is, what do we know about the size of big L sub m plus 1? So that's a good point. The easiest answer is, if I don't care about these constant factors, I don't need to worry about it.

You can think about what is the number of points on this line above that. It's essentially the number of elements of A above the biggest element of s m, above s m. It's a good question. I think we don't need to worry about it. I'm being slightly sloppy here. Yeah.

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: I think the question is, how do we know for j equals to m that you have this bound over here?

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: Great. So yes.

AUDIENCE: [INAUDIBLE]

YUFEI ZHAO: So there are some ways to do it. You can notice that the vertical line has at least as many points as the first slanted line. So details that you can work on. So this proves Solymosi's theorem, which gives you a lower bound on the sum set and the product set sizes and the maximum of those two.

It's based on-- it's very short. It's very clever. It took a long time to find. And it gave a bound on the sum product problem of 4/3 that actually remained stuck for a very long time, until just fairly recently there was an improvement that gives-- so by Konyagin and Shkredov where they improved the Solymosi bound from 4/3 to 4/3 plus some really small constant c. So it's

some explicit constant. I think right now-- so that's being proved over time, but right now, I think c is around 1 over 1,000 or a few thousand. So it's some small but explicit constant.

It remains a major open problem to improve this bound and prove Erdos' similarity conjecture, that if you have n elements, then one of the sums or products must be nearly quadratic in size. And people generally believe that that's the case. Any questions?

So this concludes all the topics I want to cover in this course. So we went a long way. And so the beginning of this course, we started with extremal graph theory, looking at the basic problem of if you have a graph that doesn't contain some subgraph, triangle, C4, what's the maximum number of edges. In fact, that showed up even today.

And then we went down to other tools, like Szemeredi's regularity lemma that allows us to deduce important arithmetic consequences, such as Roth's theorem. It's also an extremal problem if you have a set without a three-term arithmetic progression, how many elements can it have? And so the important tool of Szemeredi's regularity lemma then later showed up in many different ways in this course, especially the message of Szemeredi's regularity lemma, that when you look at an object, it's important to decompose it into its structural component and its pseudo-random component.

So this dichotomy, this interplay between structure and pseudo randomness, is a key theme throughout this course. And it showed up in some of the later topics as well, when we discussed spectral graph theory, quasi-randomness, graph limits, and also in the later Fourier analytic proof of Roth's theorem. All of these proofs, all of these techniques, involve some kind of interplay between structure and pseudo-randomness.

In the past month or, so we've been looking at Freiman's theorem, this key result in additive combinatorics concerning the structure of sets under addition. And there, we also saw many different tools that came up, and also connections I mentioned a few lectures ago, connections to really important results in geometry to group theory. And it really extends all around.

And a few takeaways from this course-- one of them is that graph theory, additive combinatorics, they are not isolated subjects. They're connected to a lot within mathematics. And that's one of the goals I want to show you in this course, is to show these connections throughout mathematics and some to analysis, to geometry, to topology.

And even simple questions can lead to really deep mathematics. And some of them I try to

show you, try to hint at you, or at least I mentioned throughout this course. And what we've seen so far is just the tip of the iceberg.

And there is a lot of still extremely exciting work that's to be done. And I've also tried to emphasize many important open problems that have yet to be better understood. And I expect in some future iteration of this course, some of these problems will be resolved, and I can show the next generation of students in your seats some new techniques, new methods, and new theorems.

And I expect that will be the case. This is a very exciting area. And it's an area that is very close to my heart. It's something that I've been thinking about since my PhD. The bulk of my research work revolves around better understanding connections between graph theory, on one hand, and additive combinatorics on the other hand. It's been really fun teaching this course, and happy to have all of you here. Thank you.

[APPLAUSE]