ML4Q&A with Markus Müller
Episode #18
Published on July 31, 2024

TRANSCRIPT

00:00:00 Intro

[Mira] Hi, my name is Mira,

[Chris] and I am Chris

[Mira] we are associated members of ML4Q

[Chris] and you're listening to ML4Q&A – a show where members from the Matter and Light for Quantum Computing Cluster talk about their careers, their research and the future of quantum.

[Mira] Today we talk with Markus Müller, Professor at RWTH Aachen and expert in the topic of Quantum Error Correction.

[Chris] I'm happy we finally have an episode dedicated to this. Quantum error correction will be essential for building a large-scale quantum computer. Markus has done the theory support for many of the pioneering logical qubit experiments in different platforms, for example ion traps and superconducting qubits.

[Mira] So he can give us an overview of the field on the topic but we will also ask him about his latest research. I would also like to learn about his recent work. As we know, QEC is based on mid-circuit measurements but this work introduces a way to avoid that completely! This is good news for some architectures and chip designs.

00:01:06 Welcome Markus

[Chris] It is my pleasure to welcome Markus Müller on the ML4Q podcast. Nice to have you here.

[Markus] Hi, nice to be here. Thanks for inviting me.

[Chris] Yes, of course. And we want to talk with you about your early career history. And then we want to ideally have a brief history of quantum error correction and then sort of explain to people the building blocks of a fault tolerant quantum computer. No big deal, right?

[Markus] Yeah, sounds like a good plan. Let's get into it.

00:01:35 Reflections on the early-career experience

[Chris] Exactly. Let's start with your early career. So now you're an expert on quantum error correction, but I guess you didn't start out that way. Shall we talk about where you started and how you got into this?

[Markus] Yeah, sounds good. Yeah, I actually did my diploma thesis back then in solid state physics. And then I had the feeling I wanted to change. And I had a look around and then I stumbled over this idea of actually quantum simulation back then, not related to quantum computation. The idea to take a controllable quantum system to simulate complex systems that are hard to simulate or calculate on classical computers.

And then I looked essentially at groups, and in the end got in touch with Peter Zoller's group at the University of Innsbruck, who invited me to visit the group. And that's kind of how I got into this field at that time, not having an idea at all about many things, and certainly not about error correction.

[Chris] You didn't go into this with a big plan but you found an interest you thought quantum simulation is an interesting problem and then.

[Markus] Exactly. So in a PhD I was mostly focusing actually on atomic physics and quantum optics and how we can use these tools to design quantum gate operations or the building blocks of quantum computers. And also of quantum simulators. And then in collaborations, also with visitors, I came across the surface code, which is one of the quantum error correcting codes. At that time, I didn't understand a lot.

And then another visitor, actually, my later postdoc advisor in Madrid, Miguel Ángel Martín Delgado, he was visiting another colleague and he gave some lectures where I also didn't understand a lot, but he had nice pictures and these colorful visualizations of quantum correcting codes and that kind of sparked my interest to learn more in this direction.

[Mira] And did it in some way shape your experimental knowledge? I mean now I think you understand the hardware very well because you have one side. You have this theory of error correction and then you also know how to go about implementing it.

[Markus] Yes, so, well, let's say during my PhD, I was working on theory proposals for implementations in part. But some of the systems back then, for instance, neutral atoms, which we thought about, Rydberg atoms, were not yet at the level where we could really implement or colleagues could implement this in the lab.

And then it was, again, some kind of coincidence, if you like, namely that a postdoc back then in Rainer Blatt's group, an experimental ion trap group at the University of Innsbruck, got hands on some notes I had written on how one could do some open system dynamics with a few qubits and he thought, well, I could do this actually in my lab. And he started to do this after his normal project in the evening. And that's kind of how I got into collaboration with trapped-ion experimentalists.

Then when you find the common language and you understand how experimentalists think and find ways of how to translate your concepts to what it really means in the lab, what kind of gates or laser pulse have to be applied, then

you find some common ground and learn more and more about the platforms, what they can do. What their limitations are. And this can be also inspiring to think again from a theory point of view, OK, they have this interesting limitation. Can I perhaps design a way around this, or are there specific approaches where we can work with this?

[Chris] So what kind of quantum simulations did you think about early on? You said open system already. What were the models that you were looking at?

[Markus] Yeah, I started to work on trapped ions, excited to Rydberg states, so highly excited electronic states. And these are then nice quantum systems because you have well localized quantum spins or qubits that interact strongly.

And then the typical models that you can engineer with this and look at their dynamics would be spin models. So chains or other structures of interacting spins where one could then look at how excitations move through such many body systems. These were kind of the models we were interested in in these early days and then based on the work on designing protocols for quantum gate operations on individual atoms or groups of atoms.

Then we started to think about how we can take these quantum gate operations for a gate-based digital quantum simulator back then. And the open system aspect, the open system means that we are thinking not only about perfectly well-isolated quantum systems, but we also want to mimic the dynamics of quantum systems that are coupled to an environment, and that is what we call open system dynamics.

And we thought about strategies of how we can simulate such dynamics in a controlled way in physical platforms such as Rydberg atoms.

[Chris] So in a funny way, this is already not unrelated to error correction, right? Because if you have, in many ways, the way that, I mean, let's get into the history of error correction in a bit.

But let's say we already said surface code at some point. So the surface code you can also think of as a lattice of spins. And you can think of the excitations in this lattice and so on. So in some ways, there's a mapping between spin models and quantum error correction.

[Markus] Absolutely, that's kind of the prime example of a 2D quantum spin model with interesting ground states and then interesting excitations, as you said.

And this is kind of how I stumbled into this field, because we were thinking about how we can, for instance, prepare ground states of this model by engineering dynamics that cools effectively these many-body systems so that we can prepare or potentially even stabilize such interesting states.

And as you say, we come to this more technically in detail. Essentially what we have to do in error correction is also to detect errors, excitations, remove them by some specifically designed dynamics that removes noise or entropy from your system.

[Chris] And the contact with experiment was really something that sort of just happened organically. So Innsbruck apparently has the right sort of environment for theorists and experiments to randomly bump into each other.

[Markus] Yes, exactly. And it's the right environment. And it's always, I mean, not only Innsbruck, but generally it's always also a question of the right people, you know, if they are kind of willing to speak to a theorist and then kind of, for instance, on the experimental side and then to kind of see how one can get together, find a common language.

And yeah, this was then the way in which many of these things then took shape and also for myself and to really learn what some mathematical concepts like general quantum maps or so, what this now really means when you want to do this on a group of qubits experimentally.

[Mira] So after your PhD you mentioned you moved to Spain. So did you already start working on quantum error correction there or was it much later?

[Markus] Yes, I started to work there on quantum error correction because my postdoc supervisor there, he was one of the authors of the papers that proposed an alternative class to the surface codes that we mentioned already earlier, so-called color codes. And I wanted to learn about this.

And the idea was then to connect this for the first time, these purely theoretical codes at that time and develop ideas of how I could realize them. That was part of my postdoc work there, which was not only limited to quantum error correction, but that's when I indeed started to dive more deeply into error correction.

[Chris] And funny enough, you kept the connections with experimental groups, right? Because these color codes very soon became a thing in ion trap quantum computers, for the most part.

[Markus] Exactly, yes, yeah. This was then when we, together with, again, the experimental ion trap group, of Rainer Blatt in Innsbruck, we developed and then implemented for the first time encodings in the smallest color codes back then with seven qubits.

I still remember this time. It was exciting. I was there for three weeks in Innsbruck, and this was the time when my experimental colleagues would calibrate these machines on Monday, Tuesday, and on Wednesday, I would say, hey, Marcos, I know it seems to work. Come over, and then we would be sitting there really literally in the evening, in the night taking data, and then it's a fascinating moment when you see, okay, for the first time, experimental signatures by some fluorescence data of, okay, now we've prepared here for the first time this complete logical qubit.

This was really the first complete error correcting code in a scalable system, as small as the seven qubit color code back then. With a lot of shortcomings still, which we had to improve on, over the subsequent years but this is very nice and I still enjoy this a lot when you see really a theory idea coming to existence in the lab.

[Chris] I think it's amazing that you got into this pioneering place of working on this at the moment where it goes from idea to reality.

So I guess this was also like for you, this must have been just very lucky because essentially you became a theorist who can work with experimentalists.

How is it then later? Is it now that people just call you or is it that you call them like because now you have not only work now with ion trap qubits, but also with superconducting qubits and with Rydberg atoms.

So essentially the qubit architectures that have realized error correction you have talked with them or published with them.

[Markus] Yes, exactly. Yeah, how does it work? Do people call me? Yes, now sometimes people call me.

And also, again, sometimes you start to work with some people, for instance, in a project where we were aiming at realizing logical qubits based on color codes with trapped ions, kind of in a competing consortium led by Andreas Wallraff at the ETH Zurich who focuses on superconducting qubits.

We were kind of just discussing there and then, you know, I helped them with some kind of characterization tools for their surface code experiments with superconducting qubits. And this is, you know, how collaborations sometimes start. And, you know, now, on the next steps, we will also be working together there on kind of taking the next steps of superconducting qubits.

The nice thing is that you can often transfer many concepts to different platforms. If you understand both a bit the theory and then you can understand what the native or the available operations are in different platforms.

[Chris] Funny enough, I was kind of an innocent bystander because my PhD supervisor around the time that I started and up close to now, I guess, was also involved in this IARPA project on pushing towards logical qubits. And I think indeed the players in the IARPA project were, if I remember right, IBM, Wallraff, Rainer Blatt, and Leo DiCarlo at Delft.

So I got to see some of the meetings and some of the excitement, although I never personally worked on the error correction route. But apparently, yeah, I mean, this IARPA push was really productive, right?

[Markus] Exactly. It's a very special type of project, so in some sense there's quite some pressure to achieve what you've promised, but it's also, I think, very nice because the funding agency is also really interested in understanding how you do things. And if things don't work out why they don't work out and so on. So this is a very special aspect that I appreciate a lot about the US funded projects.

[Chris] Yeah, I was thinking that it's amazing that this US funding which was a significant amount of money both in Delft and in Zurich even though these places have money anyway; like this project was really an accelerator also in Innsbruck I guess. And it's amazing that they fund these projects in Europe.

I mean, really a big part of these projects were European based with American money. But at the same time, the people who are governing this from America, I guess there's Robin Blume-Kohout and these people, they really have a deep understanding that the governance of these projects is extremely strong - was my impression.

[Markus] Typically these program managers are former postdocs from leading groups in the US and they know exactly what it is about. They are really experts also in the field and in the area of the program.

And then the aspect that as you correctly said that you don't have to be based in the US to receive this funding is very nice. It doesn't come with strings attached in the sense that we can still publish all results and so on; it's this mentality; in this case of the US, they want to see when new ideas come up, when new concepts come up. And that's one way of doing this, to see what else is going on worldwide. And of course, be among the first to see if some new promising avenues are being discovered.

[Mira] And since we already spoke about you having worked with different platforms, do you have a favorite?

[Markus] Oh, that's a dangerous question. No, I mean, as I said, I come from an AMO, atomic molecular optical physics background. So I'm, let's say, in what regards modeling really of the physical devices, I'm much more deeply acquainted with atoms and ions. And that's where I've worked a lot on.

And now I have started to learn also about, you know, more the physics of superconducting qubit-based platforms. But I think that the nice thing is that, you know, different platforms have different capabilities, and perhaps we come to this in more detail later, but I think it's kind of an open question if any of these platforms we've mentioned so far will in the end kind of be the one based on which we will build quantum computers.

This is like a race where sometimes some platform is leading and then others are catching up also because new technical developments take place or new theoretical approaches come up such as new codes or so that are particularly suited for one architecture or another. So I enjoy working a lot with the different platforms.

[Chris] Maybe to wrap up the sort of early career or formative part of the podcast. Can we just ask like one question about like, I mean, you worked with amazing mentors. How was the importance of this for your career and what would your advice be to young students on how to find their way into a good position in the quantum ecosystem?

[Markus] Okay, that's a difficult question, but clearly for me it was very important to have mentors and to get advice. Of course, by my PhD advisor Peter Zoller, but also Rainer Blatt on the experimental side, I think I can really call them like a very long-term mentor for myself, but also some, let's say, younger colleagues that were, for instance, just a few years ahead of myself in the career.

One example here is Sebastian Diehl, you know, a professor here at ML4Q in Cologne who actually did a postdoc in Innsbruck with Peter when I started there as a PhD student. The advice that you get there is a little bit that they make you, these mentors make you look a little bit outside of your specific one project that you're working on at the moment and they make you think, okay, wouldn't it make sense? For instance, I got this advice. Yeah, wouldn't it make sense to look into positions in the UK? You know, it's good if you start to build up your own group, you're visible on your own. So take this next step after a postdoc to become more independent. And this was certainly good advice, sometimes kicking your ass a little bit to get out of the comfort zone and think of how to take the next step.

Nowadays, I think there are fascinating opportunities and what is really nice at the moment is that, you know, even if you are not sure whether you want to pursue an academic career, it in many cases makes absolute sense to, you know, take, for instance, do a PhD in the wider field of quantum information or quantum knowledge, because there's simply so many opportunities now also to not only work in academia, but also in industry and continue to do their research on quantum computing, on quantum technology.

I think it's always important not to over-optimize kind of career-wise, but in the end choose something that you're really interested in and have motivation to work on. That was for me the case when I, for instance, went to Madrid. Yeah, I really wanted to learn these codes. And I also wanted to live in Spain again. I admit that. And which was a great time. And I think you have to have this intrinsic interest. And then the rest, then, yeah, also in part comes by interaction with other people.

And I think it's very hard to kind of write on a roadmap for your own career. I almost know no people where this has worked out. Or afterwards one can always invent nice stories that tell you the path that you've taken but in reality it's often much more driven by random interactions with people and then being open and starting new things.

00:21:44 A deep dive into quantum error correction

[Mira] So after Spain, you were for a short period of time in the UK and then Germany was calling you back.

[Markus] Yeah. Indeed. I mean, as you know, this was then shortly after I arrived in UK, then UK decided, okay, I want to leave Europe, all this with Brexit, which would have quite severely affected my research, because I had gained at that time an ERC grant, but I was also involved in European projects back then already in the flagship project on ion trap quantum computing and these ties, you know, would have either been cut as it happened with other collaborators in the UK.

And also the climate changed in a way, and I'm not speaking about the rain in the UK. But with this Brexit debate, where I felt there are perhaps other places where I prefer to be, wouldn't have necessarily needed to be going back to Germany, but this was of course a very nice opportunity. When this position opened up here and I got offered a professor position in Aachen and Jülich.

[Markus] Yes, yeah. OK, why is it important? That's perhaps the easier question.

In the end what we want is like for classical computers, we want to rely on the machine now we want to that we can program, we can run any algorithm ideally on these machines, and we want that the machine is working reliably so that we can trust the output and trust the result and errors during the computation will be picked up and corrected in runtime so that we don't have to be lucky to kind of get a good run with the correct result.

Now the problem is if we now take individual physical qubits, the different systems we've already mentioned, ions or atoms, there could be different electronic states encoding a qubit. If we take those bare physical qubits as the fundamental units for our computation and directly try to run quantum computations by doing gates on single atoms, on two atoms, or ions, we have the problem that as soon as one error happens on such a physical qubit, it will spoil the result of the computation.

And these errors are ultimately unavoidable. They will always occur no matter how hard the experimentalists work in isolating the system. An example of such an error would be if you have an excited electronic state, that there can be spontaneous emission where the electron simply falls from an excited state into a low-lying state, emitting a photon, and this happens randomly. We don't know when this happens. And then the information encoded into superpositions of these electronic states, for instance, would be lost.

And there the idea of quantum error correction comes in, which is essentially generalizing concepts that we also use in classical error correction, namely redundancy to protect information. And then the basic idea is instead of storing the state of one qubit in a single physical qubit, we can take groups of qubits and store information now in collective states of these groups of qubits. And the types of states are what we then call different encodings or different quantum error correcting codes.

And the key idea is on a qualitative level, if now something happens to one or only a few of my constituent particles, say, atoms or ions for instance, then the information is still sufficiently spread out so that from this redundant coding we can still recover this information and get back to the originally encoded state.

[Chris] Yeah, shall we just, I mean, we can just briefly discuss like the bit flip example, right? You encode one qubit in three.

[Markus] Yes, exactly, yeah. And this is like when we make a copy of classical information, let's say, instead of storing information in zero and one, we could store information in three zeros and three ones, right? And if now one of these bits suffers an error, we would end up in a classical repetition code with a string of, for instance, zero, zero, one.

And now by looking at these, what we can now see is, okay, well, they are not aligned anymore. Now somehow, okay, first and second qubit, they are still aligned, both in zero, but now the second is in zero, the third is in one, they are misaligned.

And now from this kind of alignment information, we can make a guess of what the most likely error was that has happened on the system. And in this case, you know, we would be guessing if there was only one error on the system, if the error rate is low - that's a reasonable assumption - then this must very likely happened on my bit number three. And then we apply, as a corrective operation, simply a flip to this bit three.

And this idea can now also be leveraged to the quantum state where we would not have classical bits anymore but qubits. And there the key challenge is that we now have to be careful when we interrogate our system about this alignment, which we get from measuring so-called stabilizers, multi-qubit operators. So that we do this in a way that does not collapse by doing these measurements the quantum states that encode our logical information. We only want to get the minimal information required for us to be able to make this guess of which error has happened, but not reveal the kind of the information that is encoded at the logical level in this group of qubits.

[Chris] Would you say, like I guess, in the 90s when people started talking or even early 80s with Feynman when people talked about quantum computing; in the beginning I guess before this was discovered you could have said, well, what you're describing is an analog computer. Qubits can have analog errors meaning you can have even a very small angle error. Which means and you know if you look around we don't see many analog computers because you cannot run complicated calculations on analog computers. They're inherently not correctable. So this must be complete bullshit.

But somehow then Peter Shor, who also did the algorithm with the factoring, which sparked a lot of interest, he also realized that actually quantum is analog, but in a way that is digital enough to correct it.

[Markus] Yes. You're right. Quantum is analog. So the typical, you know, time evolution as we call it, you know, the quantum system is very continuous, you know. And this might be here an example that you alluded to, you know, perhaps there's a background, a weak field or so in your lab that is on or perhaps a bit of light still shining on your qubit. So this starts to rotate now slowly. And these are continuous errors because we, you know, we can rotate by more, by less, we don't know.

And, and now what we need to do is we have to have a way to correct. And for that we have to now force our quantum system to take a decision, essentially, has there been an error? And then it should be, for instance, this complete bit flip that you alluded to earlier, or has there been no error?

And the key insight, and also in the construction of the first codes by Peter Shor, as he said, is that the measurement is actually the part of the physical process that discretizes the errors. When we have our quantum system, an error might continuously accumulate, we couple it to an auxiliary system, and in the moment when we measure this auxiliary system, we make a partial projection of our quantum

state where we either collapse on the result that corresponds to that there are no errors that happened, or a full bit flip has happened.

And then we know what to do in both cases. In the case without error, we leave the system untouched, and in the other case, we know exactly: okay, we have to kind of flip it completely back from one state to the other.

It's this discretization which really lies at the heart of any quantum error correction protocol and which renders this into a digital quantum computer despite the fact that the noise can be continuous.

[Chris] How hard was this realization of Peter Shor, do you think? I mean of course now to all of us it is kind of trivial because it's almost like ancient history in quantum computing.

But like, I mean, he must have at some point looked, I can have the system of nine qubits. And I can create a situation where if I rotate any of these qubits by a little bit, I can correct one error on any of these qubits and it will be perfect.

How hard do you think was this step?

[Markus] I think it's a big conceptual step. I don't know how it was for Peter Shor. I don't know about this history.

But when I explain error correction in a lecture or in a talk, particularly to people who are not experts in quantum error correction, this question always comes up. Namely, yeah, but you said you now place this bit flip there - what about a continuous error?

So this is something that I think everybody who kind of starts to learn quantum error correction at some point stumbles over and then kind of gets this insight when you explain. Okay. Yeah, it's really the measurement that discretizes the error. I think it's conceptually a very important step.

[Chris] So we have Shor and I mean, essentially what he drew a set of measurements that, so you say you have my nine qubits, but then I have some extra qubits and I map parities of my qubits on my extra qubits and I only measure the extra qubits, right? That was the architecture he came up with.

And these ideas were very swiftly, sort of from 95, they were very quickly generalized to establish the formalism.

[Markus] Exactly. So, first of all, these correlations or parities, as you call them, this was exactly this kind of alignment of qubits that he realized that we would need this not only for one type of bit flips, but also that the phases in quantum states, the relative phases also play a role.

So what he essentially did is he took this idea of the three qubit code or 3-bit code that we discussed and essentially constructed a version of 9 where he would have this protection against both types of errors.

And then indeed by now coupling this to auxiliary qubits he would read out this parity information - are these groups of qubits still aligned or misaligned in order to make a

guess of the error that has occurred? And now I actually have forgotten what your question was.

[Chris] I wanted to ask, like, I mean, Shor gave essentially this initial idea of quantum error correction. He just presented an idea, nine qubits, eight ancillas, I suppose, and you have a qubit. So you need to do eight measurements, but the measurements need to be multi-qubit measurements. But this was very quickly generalized, right, into the stabilizer formalism. So now we are, essentially, this is now driving the field in some sense.

[Markus] Exactly. So with regard to the stabilizer formalism, this is a nice mathematical framework where we don't have to write down the quantum states anymore. This becomes essentially very, very lengthy. You know, we would have to write these states on many, many pages of paper. But we just have to look at these parity operators and keep track of, you know, whether parity is even or odd. So just a bunch of numbers, essentially.

And these numbers, like what you know from the hydrogen atom as quantum numbers, n, l, m for energy, angular momentum, and so on. These are enough to fully specify the quantum state. So, these are kind of the fingerprints, if you like, of the state of the quantum error correcting code.

Then you write, based on these basic codes by Peter Shor and then also by Steane, the seven qubit code, then more complicated constructions, people came up with more complicated constructions, for instance, based on concatenation.

And concatenation means that I take each of these qubits of my code and again encode each of the qubits in another code. This is like a layer structure like shells you know of an onion and the more shells you put there the more protection you get against errors and it's then culminated into the development of what is now called or known as the threshold theorem, stating that if one does this construction properly -- and properly means in a way that errors do not spread out in an uncontrolled way -- this adding more layers and more layers gives you more protection. The argument holds, and what one can then show theoretically - and this was a big achievement - is that one can run quantum computations on noisy quantum computers in principle for arbitrary long times by using these codes, despite the fact that the errors in your hardware have a non-zero level.

[Mira] We've been talking about errors, but I think we only mentioned the bit flip error, which has a classical analog, but there's also a phase flip error, which does not. And what kind of errors are we then talking about that can be corrected in a quantum computer?

[Markus] Yes, you're right. We have to at least be able to correct bit flips and phase flips. And if we can do both of them, we can also do combinations of them. These kind of span all the computational errors. Anything that can go wrong within your subspace where your qubits live can be described by these.

However, there are also other errors in reality. For instance, if you think about atoms, you can also really lose the carriers of quantum information. An atom can be lost from an optical lattice or a tweezer where you store it, and then your entire qubit is

away. That's another type of error that we have to deal with. This is called quantum erasure or qubit loss.

And then another kind of error that one also has to deal with is with leakage. In this case not leaking water from pipes but leaking your qubit out of this two level system into other levels, because in reality there's no physical system which is only built from two levels. There are always other levels. Perhaps they have a different energy splitting or so, so you can kind of in an approximate way treat your physical system as a two level system, but when you operate on this, then often you can have the problem that some part of your qubit gets stuck in other levels and this is what we call leakage.

And what then has been realized or also studied is that one can also correct for leakage and qubit loss errors and the combination with computational errors. And then it comes really to the question, what are dominating error sources and then for different platforms where different errors are more or less important, different types of codes can be more or less suitable to protect your information.

[Chris] So you already mentioned that what people realized in this concatenation early work was that there's a certain amount of errors that you can still fix. And if you have more errors than that, namely for example, if you do majority voting with three bits and you have two errors, you will not be able to vote your way out of it. So people realized that there's a threshold.

[Markus] Exactly, yeah. So when we talk about the three qubit code, it's exactly as you say. Then if there's more than half of the qubits, an error, then we can't correct. So the idea is, okay, let's take more qubits, yeah.

But then we have now an extensive, large, many-qubit system. And the notion of the threshold means that now I operate on this. And kind of what it is as a promise to the experimentalist, is that there is a critical error strength in the operations, in the gates, in the measurements, in the initialization of qubits, and as long as if all these building blocks, these steps, can be done with a sufficiently low error rate, then, and low error rate means below a critical value, that's the threshold, then we can scale up the systems and by scaling up the systems to more and more qubits, actually they will work better.

If we are above the threshold, if my experimental operations are still too noisy, building larger and larger quantum devices will not help in suppressing the errors.

[Chris] Exactly. The threshold basically tells you, if you are below threshold, it means you can think about building a larger quantum computer. If you're above threshold, you should actually build a smaller quantum computer, maybe.

[Markus] Yes, or work on your... First on your operations, indeed.

[Chris] And so the initial thresholds, like just in our little history lesson, the initial thresholds when people came to the experimentalists, people must have been very disappointed because the initial error rates were one error in a million operations, roughly.

[Markus] Exactly, yes. One in a million, ten to the minus six error rates. A million gates have to work before one fails. This is very daunting experimentally.

[Chris] Exactly. So the late 90s experimentalists must have been laughing and saying, well, this will never happen.

[Markus] Yes. And that's where then the topological codes come into play. These are lattice systems where qubits are, for instance, arranged on a square lattice and information is spread out and protected in global properties of this many-body system.

And besides being very beautiful from a theoretical point of view, they are very relevant from a practical point of view; because what people have shown in theory works first is that the thresholds there were much less daunting, so on the order of a percent or 0.1 percent so that means in practice: okay, one in hundred gates or one in a thousand gates can fail and that's of course then still very challenging experimentally but less than requiring to run a million gates before the first failure.

[Chris] So we went into essentially… we went into the new millennium roughly, this I think is around 2000, and people were saying: well if you want to build a quantum computer if you can have one error in a thousand operations you can make it happen. So, essentially, at that point it became clear that we could do this.

But your initial work on our color codes, which is a family of stabilizer codes, happened in the 2010s, right?

[Markus] Exactly.

[Chris] So there was another 10 years between saying, ah, this should be doable to the first people doing it.

[Markus] Exactly. If you think about, let's say, the first proposals to build an actual quantum computer were in 1995. The first proposal was actually by Cirac and Zoller on how to build a trapped iron, ion-based quantum computer, and then other platforms followed.

And in the early 2000s, as you said, by Kitaev and others, these topological codes were introduced.

But you have to think in only 2003, for instance, the first two-ion entangling gates were realized. So, a mini quantum computer with just two qubits, not even big enough to have a three-qubit code back then.

And then it took another about 10 years or so until the systems were large enough so that one could control on the order of 10 physical qubits. And that's then the range where we can start to think about encoding one qubit in a seven-qubit code and nine-qubit code together with some auxiliary qubits to measure the errors.

But also back then, I should say, when we did this first encoding, this was not fault tolerant. It was a proof of principle. The error rates back then were still too large for this error correction to be actually useful. So, we are kind of above threshold, if you like, with experimental capabilities at that point.

00:44:02 Getting to the promised land of fault-tolerant quantum computing

[Mira] And coming to fault tolerance, what exactly is fault tolerance and how is an architecture, which is fault-tolerant different, from a non-fault tolerant quantum error correction?

[Markus] Yes. So, fault-tolerant is essentially a design principle of how you make qubits communicate to each other.

I mean, since the pandemic, we know what fault-tolerant communication is. Imagine, you know, that one person carrying a virus talks to all the other people, you know, in the end, on the next day, all other people will be affected so that the illness has spread.

So what you do instead, when you don't think about people, but about qubits, you want to have qubits interact in such a way that if on some qubit an error sits, it can perhaps spread to a second qubit.

That's unavoidable when we actually do two qubit gates. But, we want to prevent that it kind of spreads too far and affects our entire qubit register so that the single error directly infects all other qubits.

And for this we need certain quantum circuit constructions where we realize the task, for instance, of error correction in such a way that we prevent this uncontrolled spread of errors. And this, one can now formulate mathematically what it means, but that's essentially the design principle of fault-tolerance.

So you can think of this as a machinery, not yet at the physical level, that you have to develop for a quantum error correcting code. A code by itself gives you a promise. I can correct these many errors.

But now you have to run the code in the right way by using these circuits that are built for respecting fault-tolerance principles. And then only as a next step, then you can think about, OK, how I would run this on specific hardware. It's really this: how to run quantum correction codes in the right way, what is under this term fault-tolerance.

[Mira] So it's like suppressing error propagation in some way.

[Markus] Exactly. It's suppressing error propagation by suitable design where and in which order you place and between which qubits you place gates.

This comes with an overhead, because typically non-fault-tolerant constructions require less gates, less qubits. But there you have the problem that they will not allow you to scale the system up eventually.

In order to get into this promised land of operating below threshold, you have to equip your quantum error correcting codes or the quantum computers with fault-tolerant circuit designs.

[Chris] And so this is a mathematical building block. And one of the key terms is transversality in gates, right? And a transversal gate is a gate that doesn't spread errors. Can you say it like that?

[Markus] That's right. Yes. There are some codes, for instance, the seven qubit color cod -- the smallest code we did back then in Innsbruck -- where if you want to do a gate operation, that's now a bit technical, a Clifford gate operation, this is a single qubit gate, in this encoded logic qubit, what it means is that this logical gate operation can be done by doing qubit-wise the respective physical gate operations. So, that means the experimental colleagues would apply a laser pulse one by one on each of the seven physical qubits.

And now what is kind of clear is: if in one of these laser pulses an error occurs, it will only affect one of the qubits. And one error on one of the qubits of my code can still be corrected.

And this principle of transversal gates can also be extended to entangling gate operations. You can think of two codes of, say, two seven qubit codes. And when you want to do an entangling operation between both codes, what you do is pairwise physical entangling operations between the first qubit of the first code and first qubit of the second code. Second qubit of the first code, second qubit of the second code, and so on, like in a ladder, if you think.

Again, if in one of these physical qubit gate operations, an error happens or spreads from one code to the other, you easily realize, yes, okay, in the worst case, you end up with one error on one code and one error on the other code, and again, this is an error that the code can still correct.

[Chris] But of course, we cannot really have nice things always. The bad news is that we cannot realize a quantum error correcting code that can make a universal quantum computer with all fault-tolerant and protected operations.

[Markus] Yeah, we can do this with fault-tolerant and protected operations, but not using only transversal strategy, exactly.

There's unfortunately a theory, a no-go theorem that tells us, yeah, as nice as this transversal approach is, there will be always at least one gate that we need to do in a different way. And we need this gate to have a universal logical gate set.

A universal logical gate set means it's a discrete set of operations from which we can program any algorithm. So, otherwise our quantum computer will not be useful and we will not even have, for instance, quantum speed up if we are not able to do so-called non-Clifford gates.

And then is there the question of how one can get around this no-go theorem and there are different approaches. Two leading approaches are called magic state preparation and injection, and another approach is based on code switching.

00:50:20 About magic state preparation and injection

[Chris] But the bad news is that actually to run your fault-tolerant universal quantum computer both of these things are rather costly, right?

So this magic state distillation essentially is a way to prepare a certain state that is not a Clifford, which allows you to do this non-Clifford gate. And it is a probabilistic protocol, meaning you have to repeat until success. And it costs a lot of qubits.

[Markus] Yes, exactly. So the idea of the magic state distillation, as you correctly said, is that our magic state is a resource state.

You can think of this like a fuel for a Non-Clifford quantum gate. And if we had very good states, then we can use quantum teleportation to inject these states and thereby do these types of quantum gates that we cannot get in a transversal way.

Now, in order to get these magic states, it's like getting a good Schnaps. You start by distilling from kind of rather poor ingredients. And in this case, you're on to states that you prepare, and they will have a lot of errors. And then there are protocols of how you can now, for many of these copies, with a good chance get still one copy, wasting many of these, but one good copy get it out with a lower error rate.

And then if you do this distillation step two or three times, then you can in theory at least suppress the errors a lot and get a very, very clean copy there. But, the overhead, the extra cost involved there is pretty large. In early proposals, actually, what one realizes is that actually most of the qubits of a quantum error corrected quantum computer would be actually used for the preparation of these resource states.

[Mira] And from a recent work, we've seen that it was quite easy. I'm talking about the configurable Rydberg Atom array.

We've seen that it was quite easy to apply these transversal gates because the logical states could be moved around. But that is not the case in some other architectures.

So how dependent are these fault-tolerant architectures or like the quantum error correcting codes themselves on the architecture of the platform and are you usually able to first come up with a theory which I mean usually theorists they do not they come up with a theory and then they look after implementing it so have you ever really had issues in implementing a certain theory with a certain platform?

[Markus] Yes, let's say the quantum correcting codes are really related to different architectures. So one example of one leading architecture are superconducting qubits or spin qubits as solid state systems that sit on a 2D planar layer of your hardware where qubits are at fixed positions and can talk to their neighbors.

And the topological codes that we have been speaking about earlier, they are ideally suited for this type of embedding. But when it comes, for instance, to the preparation of magic states by some newer techniques, this would have been very hard or is actually not possible at the moment to do this on these architectures where you have this only local qubit connectivity between neighbors.

Other architectures such as the neutral atoms that you've mentioned or trapped ions where you can also do gates between any pair of ions in the same ion crystal or you can also move ions around in larger trap structures. They allow you to have a dynamical qubit configurability so that you can move a qubit somewhere else or even groups of qubits as shown in the recent neutral atom work. And thereby, this gives you more flexibility in implementing, for instance, quantum error correcting codes where you do not have this nearest neighbor structure built into your quantum error correcting code itself.

So yes, these things play together and one can always think about how you can get around it with some additional overhead, but in principle, yes, some architectures are really more suitable for certain encodings, for instance, or ways to do quantum computing.

00:55:23 About code switching

[Chris] You said the alternative to magic state distillation is this code switching, and for that you do need to, in some ways, reconfigure your lattice.

[Markus] Yes, exactly. So the idea of code switching is, as we said earlier, there is no code where I can do everything in a fault-tolerant way by transversal gates. But perhaps I have one code where I can do part of these gates, and I have another code where I can do a complementary part of the gates. So that the two codes together or the gate sets that are supported together form a universal gate set.

And then the idea is essentially: if I want to run a quantum computation in the setting, if I need one gate, I have to be in one encoding, do my gate there, and if I now need the other, say, non-Clifford gate, then I now have to switch during runtime from one encoding into another encoding and do in these other codes the other gate operations before I, for instance, switch back again to the other code for another gate.

And now in this code switching, you said correctly, I have to kind of reconfigure my code, my lattice, if you like. And for instance, one example would be that you switch from a planar two-dimensional color code to a three-dimensional color code, which would be where you would have to find an embedding in 3D, which can, for instance, be done in principle with neutral atoms. And this would, again, be something that is hard in a solid state - purely planar architecture.

[Chris] Because embedding a 3D connectivity into a 2D system is kind of inefficient.

[Markus] Exactly. It's inefficient. You can do it on a very small scale, but it doesn't scale. So if you try to press something that is really three-dimensional into 2D, things that are local in 3D will become very, very distant in your 2D embedding.

00:57:23 What about decoding?

[Chris] Maybe we can just quickly again, just for clarity, come back to the structure.

So we discussed the codes, which is essentially an arrangement of these stabilizers and the qubits that you have. Then you have the implementation of the code, which is mapping this to some operations that you do, as we said, that has to be fault tolerant.

But the part that we haven't talked about yet is the decoder. So you have these errors, and whenever you need to do this code switching or sort of these non-transversal gates, you have to make sure that your error level going into these operations is low. So it's not enough to just keep track of your errors. At some points in your quantum computer, you need to actually fix them.

[Markus] Yes, exactly. So what we do constantly is to keep up with the errors, we constantly measure these so-called stabilizer operators with some helper qubits, auxiliary qubits, and this gives us classical information.

And now what we do is, and this is called decoding, as you said, is we have to process this classical information to make a guess of what the error was that has happened on our quantum hardware. And then, also before actually doing non-Clifford quantum gate operations on our logical qubits, we have to apply the respective operations.

And then what you see immediately is: this has to be reliable and fast enough. We want this classical processing and feedback on our quantum hardware to keep up with the speed of our quantum computer and happen fast enough so that this outweighs again new errors that pop up while our quantum hardware is waiting.

And there have been different decoding strategies for different codes that are efficient from a mathematical point of view so that, let's say a computer scientist would now say, yeah, my work is done. It scales only polynomially with the size of my quantum computer or my logical qubit.

But that is not good enough in practice, because it has to be, for instance, for superconducting qubit processors, one has to have all these classical processing on the order of a few microseconds being done. And this requires then, from the theory and computer science point of view, development of new decoders.

But you also really have to think on a hardware level now, or even electronics engineers have to think about: What are the requirements for my classical control hardware of my quantum computer to be able to keep up and do this decoding process and then trigger the respective correction operations on my hardware fast enough?

[Mira] And the decoders: are they also hardware specific?

[Markus] They can be and some interesting approaches, for instance, to use machine learning based decoders, neural networks to train them, to feed the information we get about errors from measuring these stabilizers into a classical neural network, train this to kind of network to learn to recognize and make suggestions of what kind of corrections to apply.

And this is a very nice approach that works typically on small systems and it's able to discover correlations in the noise processes of my specific hardware that we might not actually know.

But these machine-learning based approaches are then able to discover and use such correlations to come up with optimized corrections that take into account these correlations.

01:01:15 Discussing recent work on fault-tolerant resource-optimized schemes

[Mira] And you mentioned ancilla qubits, these auxiliary qubits which help you understand where or what kind of an error occurred. But this is usually done by measuring the ancilla qubit. But then you have a new idea where you don't really need to measure the ancilla qubit. So can you explain a little how that works?

[Markus] Yeah, the idea is, as you said, we usually map the information about errors onto these helper qubits and measure those. And we do this in this indirect way in order not to collapse our quantum states encoding our precious logical information.

But in many architectures, measurements are much slower than gate operations. To give you a number, for instance for neutral atoms, they can be done in perhaps a few hundred microseconds or a millisecond. Whereas gate operations, they can be sub-microsecond or even only a few nanoseconds, so you have stronger working directions. So the gates are fast and then the quantum computer somehow has to wait for these measurements that affect a thousand times slower or so. That's not very nice.

So the idea that we had, and it was kind of known theoretically that it should be possible in principle, is to then map the information onto this register of helper or ancillary qubits but then use quantum gates that are controlled by the state of the auxiliary qubit register as quantum feedback operations directly onto our logical qubits.

And this is the advantage. We don't have to measure. We can afterwards just throw away these helper qubits and bring in fresh qubits or reset those qubits. And resetting qubits is much, much easier and much faster, or holding a large enough reservoir of fresh qubits, than actually measuring.

And this is a way that can be potentially much faster, allow much faster error correction, and thereby allow faster processing speeds of our logical quantum computer.

And what we managed in this work now is to design fault-tolerant resource-optimized schemes where this quantum feedback is done in a fault-tolerant way. If one does this naively, then the problem is that by doing this quantum feedback one actually introduces more errors than one corrects and then again the quantum error correction is essentially useless. And the other difficulty was to find circuits that respect this fault tolerance property and do this quantum feedback in a suitable way.

[Chris] It's really beautiful work. It's rare that in the podcast we discussed something that just came out, in this case the paper came out in PRX quantum.

[Mira] Two weeks ago.

[Chris] This is probably the most recent work that we have ever discussed. We can slowly start wrapping up, but maybe we should discuss one of the things for you.

2022 was a really big year because there were back-to-back Nature papers where you were a co-author, right? Both again on two different systems. Should we talk a little bit about this? What happened there and how were these results?

[Markus] Yeah, so yeah, this was a great privilege to be involved in two collaborations. One was with the Zurich team by Andreas Wallraff that managed for the first time to run the smallest version of topological surface code and for the first time demonstrated this ability to quickly measure error information via these auxiliary qubits in a pipeline approach, showing really repeated rounds of error correction on a full quantum error correcting code - close to the, not quite yet, but very close to the threshold of becoming useful.

And the other work I was involved in then was with trapped ions. And there we realized that new schemes which had been proposed on how to reduce the overhead for this magic state preparation that we talked about earlier, could allow us to realize for the first time a magic state preparation and injection. And thereby a demonstration for the first time, in any architecture actually, of a universal logical gate set from which in principle one could now build arbitrary quantum algorithms on larger scale quantum computers.

[Chris] So again, in some ways the ideas are from, let's say, 2000-ish, so between the ideas of sort of a fault-tolerant quantum computer and the first proof of principle demonstrations of all the building blocks we had about 20 years.

[Markus] Yes, this is true for, you know, for the work on the surface code. The theory idea was really like, you know, realized 20 years later. It's not quite true for the magic state injection because there are these newer protocols based on so-called flag qubits. These are other type of helper qubits. These are pretty new, only going back like five years or so where people only very recently came up with this alternative way of preparing magic states.

And this is really what allowed us to do this in an experiment with 16 ions rather than, you know, with more than 100, which would have been by doing this with a traditional approach of magic state distillation.

[Mira] And would you say that now the progress is happening very fast because recently we had this work by the Lukin group and I would think they have really set the bar high for all other architectures?

[Markus] Yes, the progress at the moment is happening extremely fast. I'm very excited because now quantum error correction has become a reality and kind of the experimental progress proves people wrong that say: Well, error correction is very nice, but it will start in 20 years. No, it is starting actually now.

And as you correctly said, you know, by the recent experiments by Mikhail Lukin's group and collaborators, what they managed to show for the first time is that they could run quantum algorithms on small, error detected logical qubits, up to 48, I would call this mini logical qubits, where they did not yet fully show universal quantum computation but what they showed already is that this encoding information in a few physical qubits opens a completely new ability: Namely that at the end of your quantum computation you can actually detect whether errors have happened.

This is something you can't do if you run the algorithm directly on physical qubits. And when you detect errors you can throw away some runs where you think too many errors have happened. And with that you kind of keep only the good runs of which you think they are good, but that gives you a way, a new handle of how you can boost the accuracy of your encoded quantum computation.

And what they were able to show for the first time is that by running this on these encoded small logical qubits, they were actually doing better than the best way they could run it on the unencoded physical qubits.

So this is what I would call the new era of NISC type small scale quantum algorithms on not yet fully error corrected qubits, in this case only error detected, but where really one should think about now how to start running quantum computations not anymore on physical qubits but on small logical qubits. Because in this intermediate era now, where we can start with playing with error detection and so on, this really also is from a NISC point of view, noisy intermediate scale quantum devices, in my opinion, the way to go. To squeeze the most out of the machines that we have now before we in the end, you know - this will take still some time - have quantum computers with millions of physical qubits.

[Chris] I mean, yeah, what you said that the encoded qubits do better, let the logical qubits do better, one of the things that shocked me in Delft as a beginning student learning about error correction was how hard, like, there's the threshold.

And what we usually mean by the threshold in error correction is just that more becomes better but there's another threshold which you could define as you beat the best physical qubit. And that threshold is actually harder and in many systems it is not reached at all.

[Markus] That's right and it's also very hard to reach it in small codes. That's particularly hard. It becomes even easier in some sense if you work with larger codes of course and you have to control more physical qubits.

Of course, the caveat now in the experiments by the Lukin team, for instance, is that the gates on encoded logical qubits, not for all encodings, but they are generally not better necessarily than the physical gates. But working with the logical qubits, this allows you to do this error detection and throw away some of the runs.

This is really a new handle before you actually want to run fully error-corrected operations on larger logical qubits beating the physical qubit operations. This has been shown already in other experiments. For instance, logical entangling gates can already work better than physical and technical gates on encoded logical qubits. But it has not been shown for a universal logical gate set yet.

01:12:02 Outlook on future projects

[Chris] Yeah. So you have like, I don't know, we can, we can probably come to an end. You have a very exciting time ahead. Just to finish up what, what are, what are your current, um, what are the current things that you are most excited about?

And what will be your outlook for, for the next, let's say two, two, three years?

[Markus] Yeah, I mean, the things I'm quite excited about at the moment is measurement-free error correction that you alluded to. We want to really see how far we can take this. And in other aspects that I'm also very interested in are to think about other paradigms of how we can do run robust quantum information processing.

For instance, we are working on quantum machine learning based approaches such as quantum mechanical versions of cellular automata that you might know from the Game of Life, you know, where some cells evolve kind of locally according to some rules.

And what we are trying to investigate, and have found first schemes that work, is whether we can use this quantum cellular automata to also learn some dynamics that is useful and that clean up, for instance, quantum states in an autonomous, again, measurement-free way.

[Chris] Doing away with the decoder, so to speak.

[Markus] There you would not need any decoder anymore. Exactly. You don't measure information. The error correcting dynamics emerges from microscopic or quasi-local quantum dynamics on groups of qubits.

And this is again something where we're trying to see how far we can take this. I find this also fundamentally very fascinating to ask the question: Can you have local quantum dynamics, dissipative quantum dynamics that can stabilize global many-body quantum states encoding logical quantum information?

[Chris] Thanks a lot for giving us this tour of quantum error correction. It was a huge pleasure.

[Markus] The pleasure was mine. Thanks a lot for having me and for the exciting discussion.