2021 Memory Master Class

PREPARED REMARKS | MAY 5, 2021

MICHAEL SULLIVAN | Corporate Vice President, Investor Relations

Hello and thank you very much for joining us. I'm Mike Sullivan, head of Investor Relations.

[Page 4]

At our Investor Meeting on April 6, we announced plans to hold Master Classes to give you insights into how the roadmaps are changing, driving new technology inflections that are going to reshape our markets over the next several years. Today's event is all about memory, and our next Master Class will be on June 16th and cover logic technology.

In the second half, we plan to host Master Classes on specialty semiconductors – or what we call our ICAPS markets. Also heterogeneous design and advanced packaging. And inspection and process control. We have a lot of new products and technologies in the pipeline, and we'll introduce a number of them to you at these events.

[Page 5]

Here is today's agenda.

I'll lead us off with a quick overview of Applied's data and memory thesis. Then we'll have a brief fireside chat with an old colleague and friend of mine named Ed Doller who'll give us the memory landscape. Next, Kevin Moraes will host the technology section of the meeting. Then, Raman Achutharaman will summarize the key business opportunities for Applied in memory. Finally, Raman and Kevin will help address your questions about today's material.

[Page 6]

Now for the overview.

As we said at the Investor Meeting last month, 2021 is shaping up to be another record year of equipment investment by our customers. In fact, four of the past five years will be records.

[Page 7]

To help explain why this is happening, I wanted to revisit a presentation that we shared at the previous Master Class at the Maydan Technology Center in Santa Clara back in 2017 which was the breakout year for WFE. 2017 was when we finally surpassed the $35 billion dollar run rate that was first reached in 2000, when 300-millimeter wafers were introduced.

[Page 8 & 9]

We shared our expectation that a new era of computing would drive an explosion of data creation and sustainably higher demand for logic and memory. And we highlighted the strong correlation between data generation -- and the growth of both DRAM and NAND. This data thesis continues to guide our thinking today.

[Page 10]

As Gary showed at the Investor Meeting last month, we're still in the very early stages of the transition from an application-centric world -- where people generate most of the data -- to a data-centric world -- where almost all of the data growth comes from machines. The Internet of Things and smarter vehicles may drive a 10X increase in data generation over the next 4 to 5 years.

[Page 11]

As everything around us gets smarter, chip content is growing – from the smartphone in your pocket, to the house you live in, the car you drive, and the data centers that help keep everything orchestrated.

[Page 12]

Over the past 10 years and the past 20 years on average, foundry-logic has been over 55% of equipment spending. We believe this ratio is likely to continue into the future.

[Page 13]

And today, we'll focus on memory.

Here's a simplified chart of the memory hierarchy, showing how different technologies are used to deliver the optimum combination of speed and cost.

Today, we'll focus on DRAM and NAND technology and also give you a quick update on MRAM and PCRAM.

And to help us with the macro trends for these technologies, I'd now like to introduce Ed Doller.

Ed has been in the industry for 37 years, and he began his career as a memory technologist at IBM. He later joined Intel, and that's where he and I had quite a few investor meetings together when he was the CTO of the flash business. Ed participated in the spin-off of Numonyx which had flash and phase-change technology, and that brought him to Micron where he was chief strategist for the NAND business.

Ed, thank you for joining us.

ED DOLLER | Doller Consulting Group

My pleasure Mike – it's good to be working with you again.

Michael Sullivan: ***Ed, you've always had a great perspective on the memory markets. Which device types are going to lead the industry over the next 3-5 years?***

It will continue to be DRAM and NAND. Nothing compares to DRAM in random access performance and endurance. And certainly nothing compares to NAND in cost per bit. New memories are always lurking in the background, but DRAM and NAND revenue combined in 2020 was over $124 billion dollars, and if you look at all of the emerging memories combined, it was less than half of a billion

dollars. It's really tough to keep up with DRAM and NAND in cost per bit which are certainly fast-moving targets.

Michael Sullivan: ***What is the role of MRAM right now in that case?***

MRAM right now is certainly a technology that's being heavily invested in. Unfortunately, right now, it's cost is problematic in terms of its adoption. Certainly something that down the road, could be a promising technology given the fact that it has low latency compared to DRAM and high endurance but certainly nothing exactly compatible to DRAM. It might be good as a NOR flash replacement or an SRAM replacement for on-chip cache or etcetera, but certainly nothing that's going to challenge DRAM or NAND in the short term.

Michael Sullivan: ***And what about phase change memory – like 3D Crosspoint?***

Clearly Micron decided to move away from 3D Crosspoint, and that leaves Intel. Intel is looking at ways to promote the adoption and one of the thing that is being looked at is moving it from the DDRT bus, the DRAM bus to CXL. But it's clearly been slow to develop.

Michael Sullivan: ***OK, so DRAM and NAND are going to be the high-volume runners for the foreseeable future. Let's talk about bit scaling and cost. What are you seeing?***

The first thing to recognize is that Moore's Law is no longer working that well in memory. DRAM is clearly not on Moore's Law anymore. And NAND is close to Moore's Law, but once again when you factor in things like string stacking, putting multiple tiers of these 3D NAND stacks together, it starts to fall apart.

Michael Sullivan: ***Could you dive into that in a little more detail?***

Moore's Law is obviously about doubling the bits every two years. If you look at 3D NAND, we've had aggressive scaling and as we the move from 32 to 64 layers to 128 layers and beyond. But going from 64 to 96 layers is not Moore's Law. It's about doubling not adding 50 percent more. It's going to be hard to continue doing that especially as we add more and more layers. Think about going from 256 layers to 512 layers. You have to add 256 layers and that's something that certainly is going to be hard and we're going to be start hitting the laws of diminishing returns.

Michael Sullivan: ***What will that mean for the industry?***

It means we're going to need more wafer starts to keep up with any particular level of bit demand growth. If you want to maintain a Moore's Law bit output and you can't do it through scaling, you need to do it with building more factories and adding more capacity.

Michael Sullivan: ***Now let's go into the technology. What should people know about floating gate and charge trap NAND?***

Two very different storage technologies and they both have pros and cons. You have a choice: you either take the manufacturing pain that occurs with floating gate or the multilevel cell pain that occurs with charge trap. Floating gate NAND cells are made from polysilicon surrounded by oxide. The electrons can move around in the poly and give you a much more stable and better data retention provided you know how to manufacture the cell. There is a fundamental belief that floating gate is more

stable and better suited for data retention and stability which is good for things like enterprise solid state drives. Which of course is why a company like Intel, who is laser-focused on that market is the one manufacturer left that's on that particular technology.

Michael Sullivan: ***What about charge trap?***

In charge trap, you trap charge at the nitride oxide interface. Charge trap can be easy to store very discrete charges, which is good when you want to logically represent 3, 4 or even 5 bits per cell. However, the number of charge levels is 2 to the power of the number of bits that you want to store in each cell.

So to represent 3 bits per cell, you need to discern 8 voltage levels. 4 bits needs 16 levels, and 5 bits needs 32 levels. Certainly very hard to do, but that's why charge trap can be better suited to higher-capacity, lower-cost SSDs.

Michael Sullivan: ***Great. And with that, you've described logical scaling. How far can we take it, and what are the other ways to drive NAND bit growth?***

Right. Logical scaling is already getting more challenging to do with marginal returns. Scaling is being done by adding more bits, more multilevels. Four bits per cell is common, 5 bits is being discussed. 6 bits would need 64 individual voltage levels. I can tell you that in my mind, that's something that is, I don't want to say it's impossible, but it's close. But also when you start adding all these multiple voltage levels, it's tough on write performance, read performance and data retention. Again, it changes the fundamental dynamics of what it is you that you ultimately have.

Michael Sullivan: ***So what else can we do to increase bit density?***

The other two levers are vertical scaling and lateral scaling. As I said earlier, vertical scaling is already off of Moore's Law and generating diminishing returns. We're at 96-plus layers using single tiers, and companies are targeting as many as 500 layers using multi-tier architectures. But the progression is linear and not geometric like Moore's Law. One of the big challenges to vertical scaling is dielectric etch times which are at the practical limit of the technology and certainly would require a very tight relationship between wafer equipment manufacturers and the NAND manufacturers. You can't etch taller stacks in a single pass without dramatically increasing the etch times, which increases cost and lowers your wafer output in terms of gigabytes. So I think the most promising area from here is lateral scaling.

Michael Sullivan: ***OK, so what can we do to drive lateral scaling?***

Fortunately, there are a couple of drivers.

One, you can move the memory holes closer together with better patterning. Two, you can condense the contacts by moving from structures that look like a pyramid to structures that look more like a stadium. And three, you can get a lot more efficient with all of the logic circuitry that accompanies the memory cells to perform writes, reads, block management. We call that the periphery. To the extent that you can shrink the periphery or move it, is something that will ultimately lower the die size which is proportional to the cost.

There's CMOS under the array which is being done by some manufacturers but not all and there's other manufacturers that are putting the CMOS circuitry on other chips and then bonding them. There are certainly things that can be done to drive lateral scaling.

Michael Sullivan: ***Does NAND really need to have more advanced transistors or are we ok with polysilicon?***

To the extent that you can have advanced transistors, there obviously is a benefit. One of the companies that I talked about that has moved the logic to another chip , it has allowed them to use more advanced logic transistors which has allowed them to get better performance. That is certainly going to be something the industry is going to wrestle with, as they balance performance and power.

Michael Sullivan: ***That's very helpful. Now, what's the status of DRAM scaling? How far can we take it?***

In DRAM, 2D capacitor scaling is also getting very difficult, and we're well into the diminishing and marginal returns, but that's what memory companies do. They continue to shrink. The industry talks about one alpha, one beta. We don't even talk about 45 nanometer, 42 nanometer. This nomenclature has nothing to do with the metal pitch anymore. Clearly, we're seeing the diminishing returns occurring in DRAM. I think we can get to the 1D node before we have to consider need a radical new architecture. So 3D DRAM is something that DRAM manufacturers are looking at. It will get a lot more attention over the next several years.

Michael Sullivan: ***If the capacitor scaling is slowing down, is there anything else we can do before we get to 3D DRAM?***

Like the conversation around scaling NAND, scaling the logic that surrounds the memory cells, is something that is absolutely being focused on. In Micron's latest architecture, the majority of the die size reduction was in the periphery logic. High-k metal gate is smaller -- and it's faster and lower power. FinFET could further improve power, performance and area-cost. So I think the periphery is going to be very important to be engineered as we look at reducing the overall cost of DRAM technology in the future.

Michael Sullivan: ***Will 3D DRAM be like 3D NAND?***

In some ways yes, in more ways no. The vertical scaling won't be nearly as aggressive. And it will be completely different from a materials perspective. Here's the thing: 3D NAND memory cells get used once in a while, so you can use dielectric materials and sacrifice performance. Also, there are big architectural differences between the suppliers, and you can hide them using software. But in DRAM, you bang on the memory cells all the time. So you need to be fast and low power. You also need to have exacting interfaces. You can't afford to waste CPU cycles with long RAS to CAS delays. So instead of dielectrics, you'll deposit metals and shape and remove them with conductor etch. It's a very different kind of inflection compared to 3D NAND.

Michael Sullivan: ***And finally, what about advanced packaging?***

I think packaging is going to be super important. Stacking and interposers have already been helpful in smartphones. But now we need advanced packaging in high-performance computing, especially in AI.

What you really want is to get the memory as close to the logic as possible. With hybrid bonding, you can directly connect heterogeneous chips using copper. That will give us shorter and tighter interconnects, lower latencies, higher performance and lower power. And that will help us make up for the slowing of Moore's Law.

Michael Sullivan: ***OK Ed, it sounds like there are plenty of inflections to keep us busy over the next few years. Thank you for joining us, and let's keep in touch!***

Thanks Mike. Best of luck to you guys!

DR. KEVIN MORAES | Vice President, Products & Marketing, Semiconductor Products Group

Thank you, Mike, and hello everyone on the line. We have had a lot of fun putting the technical section of the Memory Master Class together and I hope you will enjoy it. Today you will meet Dr. Sony Varghese, and he will talk about the roadmap for DRAM memory.

Sony is Director of Strategic Marketing for Memory in the Silicon Products Group, where he identifies key challenges associated with inflections in the Memory Industry. He spent more than a decade in research and development at Micron Technology developing various memory technologies, before joining Applied Materials in 2017. And Sony will be followed by Dr. Sean Kang, senior Director in the Semiconductor Products Group.

Sean looks at technology inflections and market trends in the Memory Space. Sean has spent over 20 years in developing semiconductor technologies, including stints at Samsung and Cisco before joining Applied in 2016. Sean will talk about the 3D NAND roadmap.

With that introduction, Sony, over to you…

DR. SONY VARGHESE | Director of Strategic Marketing for Memory, Semiconductor Products Group

Thank you, Kevin.

The DRAM industry is going through significant developments to meet the demands of higher density, speed and lower power. Today we will review some of the challenges that the industry is facing and how Applied Materials is helping solve some of them.

[Page 15]

DRAM is a major part of the computer ecosystem and it is often the first point of contact for the processor to get information to perform the billions of calculations that it makes every second. There are various aspects of the DRAM that determines how much data it can store as well as how quickly it can transfer it to the processor.

In order to make it easy to understand these complexities, it's important to understand the various components in the DRAM chip. A typical DRAM chip has three major areas, the cell array, where the individual bits are stored in tiny capacitors. The cell array usually occupies about 60% of the chip area.

The logic or core area where devices such as sense amplifiers and word line decoders help determine how data is accessed from the cell array. The third part is the periphery which forms the communication links out of the DRAM chip. All these areas must scale to meet the industry's DRAM performance requirements. DRAM makers may choose scale one or more of these areas from one generation to the other.

[Page 16]

Let's start by trying to understand how a typical DRAM works. The DRAM cell is an elegant design, that uses a single transistor and capacitor to store a charge. The DRAM cell is very small. Current generation DRAM cell occupies an area that is smaller than 50x40nm. In comparison, it is almost 6 times smaller than an SRAM cell used as cache in a microprocessor.

The transistor is controlled by a word line or gate, that turns the transistor on or off to move electrons in and out of the capacitor. The capacitor is very small as well and holds a charge of about 7 to 10 femto farads at the current technology nodes. The charge is moved out of the DRAM cell using thin metal lines call the bit lines. The logic device called the sense amplifier is used to measure this minute charge.

Since the charge from the capacitor is already very small it is important to make sure that the transistors themselves are not a source of losses due to variability. This variability comes due to a mismatch between Vts of the transistors. As you can imagine, when you continue to shrink the DRAM cell from one generation to the other, the amount of charge stored in the capacitor continues to decrease. If the performance of the periphery transistor does not improve, the actual detected signal from the capacitor will continue to decrease as well.

In fact, DRAM makers have been slowing the rate at which periphery transistors have been scaling due to performance degradation. We must improve all aspects off this signal ecosystem by maximizing the charge that can be stored in the capacitor, reducing Sense amplifier variability and any loses due to the wiring in the circuitry.

[Page 17]

We have started to see the adoption of several new technologies by all the DRAM makers based on several announcements seen recently. With the recent ratification of the DDR5 specs by JEDEC, we are seeing more DDR5 products in the market. We saw the announcement of the first 1z DRAM that uses EUV to pattern one level. And most recently we saw the introduction of the first 1-alpha DRAM product. We have also started to see increasing adoption of HKMG for scaling the peripheral transistors. We may even see the introduction of new DRAM architectures in the future for continued scaling

[Page 18]

At a high level, the paths for increasing the density of DRAM can be viewed as four pronged. Number one is the reduction of cell area by reducing the dimensions of cell. This could result in deterioration of the silicon transistor performance as well as the charge stored in the capacitor. One other option to reduce the die area is by shrinking the peripheral circuitry. This circuitry comprises of logic transistors as well as the wiring that connects the various parts of the DRAM chip. Some of the challenges that are

being faced with scaling the periphery are being addressed by introducing new materials such as low k dielectrics for copper wiring as well as new transistor technology such as hi K metal gate.

Lately, stacking multiple dies as a way of increasing the density of DRAM has started to gain some additional momentum. This has been primarily driven by the introduction of through Silicon vias to connect chips. Unfortunately, we might be getting to a point where the physics of scaling planar DRAM might be hitting its limits.

This could mean that new design architectures that help scale the DRAM cells into the third dimension may be needed soon.

[Page 19]

Now, let's get into a little more detail about the challenges with scaling DRAM as well as solutions that the industry has been working on, by partnering with Applied Materials. As described earlier, the DRAM cell array where the data is stored is the largest part of the die. We can increase the density of the die by increasing the number of cells in a unit area. For this to happen we must reduce the dimensions of the cell silicon transistor. This, in addition to the reduction in gate and bitline pitches, helps scale the DRAM cell. The reduction in the DRAM cell area also means that the capacitor will get smaller.

[Page 20]

Before getting into the details of how a capacitor is formed, I would like to discuss how a DRAM capacitor works. A capacitor works by storing a charge in a high k dielectric that is sandwiched between two metal electrodes. The amount of charge stored in the capacitor is proportional to the surface area offered by the metal electrodes. Higher the surface area – Higher is the charge in the capacitor. The DRAM capacitor is formed by etching a deep hole in an oxide mold by using a material called as the Hardmask which acts as a stencil.

The high k dielectric material is then deposited between two thin metal electrodes to form a parallel plate capacitor. As the cell area decreases, the diameter of the capacitor decreases as well. This reduces the overall surface area of the capacitor.-The aspect ratio which is the ratio of the height of the capacitor to its diameter has been decreasing node over node. As the hole’s diameter gets narrower it gets more challenging to etch the deep holes and effectively remove the byproducts. Etch ion energies have to be increasing to maintain a steady etch rate.

This change in the process also results in the Hardmask material getting etched away before the full hole is fully formed. Increasing the thickness of the Hardmask only makes the aspect ratio worse and gets to a point of diminishing returns. Defects and variations in the dimensions of the capacitor also get worse because of these challenges.

[Page 21]

What we need is a material that won't etch as fast under the aggressive etch conditions. Draco is a new Hardmask material that we have developed that has greater etch selectivity to the capacitor oxide mold material. By using Draco, our customers have been able to reduce the thickness of the Hardmask down by 30%. The quality of the holes made in the Hardmask has to have the best uniformity and shape as possible since any distortions in it would transfer to the next mold etch process.

The Hardmask etched that is used in the industry currently is our Sym3 etch technology. A hot wafer chuck in addition to the high conductance chamber helps deliver holes that have excellent dimensional uniformity. The quality of the hole dimensions is so critical that we have developed advanced metrology tools so as to get a large amount of data, fast so that we can check if the process is performing as desired. Our unique metrology aids in making measurements inside the small holes without having to cut the wafer for electron microscopy imaging.

[Page 22]

To summarize, through the unique combination of our advanced DRACO Hardmask and Sym3 Hardmask etch tool we have been able help our customers achieve a 30% reduction in the thickness of the Hardmask used for capacitor etch. In addition to this, we have been able to deliver a 50% reduction in variability of the diameter of the capacitor holes. This has helped reduce the bridging defects by over 100X.

[Page 23]

The data from the DRAM cell is managed by periphery logic devices. The periphery circuitry primarily consists of highspeed logic transistors and the wiring that connects the various parts of the DRAM. Losses associated with this circuitry can impact the integrity of the information stored in the DRAM cell. It is important to clearly be able determine if the data stored in the capacitor is a binary 1 or 0.

[Page 24]

The technology that is used in the current periphery transistors is based on a polysilicon gate. The polysilicon is separated from the silicon by an oxide layer called as the gate dielectric. In order to be able reduce the dimensions of the transistor, the gate length has to decreased as well. A transistor with a smaller gate length also requires a thin gate dielectric to effectively control the transistor.

Unfortunately, as the thickness of the gate dielectric decreases, a film with a lower k value is prone to be become more leaky. This phenomenon to a large extent has prevented the polySiON based transistor from scaling. This is similar to a problem that the logic processor industry faced over a couple of decade ago. They addressed this by changing the dielectric to a higher k material and use a metal as the gate. This helped reduce the leakage in the transistor by several orders of magnitude.

Lately, DRAM has been adopting this technology to improve the transistor performance. When several different films are deposited on top of each other any defects in the films as well interfaces plays a big role in how the electron transport works. There are over 6 different materials used in the HKMG stack. Several different treatments are used in engineering this.

We are seeing rapid adoption of HKMG in high performance DRAM and expect greater adoption in other DRAM product segments in the near future.

[Page 25]

The integrity of the signal transmitted is only as good as its communication channel. A part of this communication channel is the tiny copper wiring that is used in DRAM. Over the years, as the distance between these tiny wires has gotten smaller and smaller, we are starting to see signal interference between adjacent wires due to capacitive coupling. The copper wiring is separated from each other by

a dielectric material such as silicon oxide. RC delay in circuitry is directly proportional to dielectric constant of the film. Lower the k value of the film, lower is this delay.

At reduced wiring pitches signal delay and power loses increase. Another problem we see due to shrinking wire geometries is there is an increase in current densities in the at these small dimensions. This results in increased electromigration that can be detrimental to reliability of the DRAM device.

[Page 26]

We have seen increased adoption of our industry leading low k dielectric product called Black Diamond in the DRAM backend copper wiring. Black Diamond offers a 25% reduction in dielectric constant compared to the current Silicon oxide films. The reliability of the backend copper metallization has been improved over 10x by the adoption of our Endura CuBS product along with CVD Cobalt. The use of Cobalt capping helps reduce copper electromigration at higher current densities.

[Page 27]

I'd like to switch gears and talk about how the DRAM industry can continue to scale but this time in the vertical direction. We'll discuss about die stacking and related technologies in a future master class.

Today we'll focus on increasing cell density in the vertical direction.

DRAM manufactures are finding that increasing costs of patterning as well as possibly hitting limits of physics, scaling in the two dimensions is getting more challenging. New device architectures, materials and process technology may be needed to extend DRAM scaling.

[Page 28]

DRAM chip bit densities had been increasing by about 25% node over node until a few years ago. Lately, due to the challenges with scaling the cell area, the node over node increase in bit density has trended down to about 20% now. The cell area is defined by the wordline gate and the bitline gate pitches. This rectangular area has to shrink at about 15% with every generation to achieve a density increase of about 20%.

We expect that self-aligned Quadruple patterning will reach the limits of its scaling beyond the n+4 node. DRAM makers have to find new ways to continue increasing DRAM bit density to meet consumer demand.

[Page 29]

NAND faced a similar scaling limit a few years ago. The semiconductor industry came together and addressed this scaling limit by finding new ways to store bits in the vertical direction. So, can DRAM pull this off as well? DRAM is a very different device compared to NAND. It is a high-speed device that is almost a 1000X faster than NAND. DRAM achieves this by using the high mobility silicon substrate as the starting material to form the channel.

Silicon on the other hand uses poly silicon as the channel material. Polysilicon is much more inferior compared to Silicon as a channel material. The high speed in DRAM also comes from the fact that a charge can be moved in and out of the capacitor quickly. DRAM is therefore what is called as a volatile memory since it will lose any information stored in it when the power is turned off. NAND on the other

hand is designed to store data for a long time and it achieves this by using a charge trap layer. The process of storing and erasing a bit is therefore very slow.

If DRAM has to scale in the vertical direction, new material innovations are required to enable high mobility, ultra-low defect channels in addition to other material and process innovations.

[Page 30]

Applied Materials has been a leader in new material innovation and it has had unit process leadership in several key technologies that are used to deposit high mobility materials through its offerings in areas such as in epitaxial growth, PVD and ALD. Once deposited, these films will have to be etched to access the bottom layers. We have leadership in the high aspect ratio conductor etch segment to enable the formation of these deep features in these high mobility materials.

3DDRAM is also expected to require advanced gapfill, selective removal and advanced doping capabilities. These capabilities must be integrated seamlessly to enable a successful solution. Our leadership in developing integrated Materials Solutions to co-optimize these various processes will also be a key enabler to offer holistic solutions to integration challenges that 3DDRAM is bound to face.

[Page 31]

The DRAM industry is expected to continue to innovate to meet the demands of increased bit density and speed. DRAM products that offer lower power and lower bit cost will require innovative process and integration solutions to help address this need.

Thank you. Over to you now, Sean.

DR. SEAN KANG | Senior Director, Semiconductor Products Group

Thank you, Sony!

Hello everyone and thanks for joining the 3D NAND memory master class. In this session, I'm going to explain 3D NAND memory challenges and roadmaps, along with Applied Materials' solutions.

[Page 33]

First let's start with – what does 3D NAND look like?

Today's 3D NAND die is around 12mm x 6mm, depending on its density. In this die, there are not only memory cells but also peripheral transistors which operate the memory cell. The top image shows a simplified 3D NAND cross section. A Conventional 3D NAND architecture has a cell array with staircase and peripheral transistors next to it.

The image in the bottom right corner shows a 3D NAND cell in more detail. It consists of, from the outside in:

- W/TiN for the gate electrode,
- AlO and blocking oxide, which controls the bandgap and blocks electron movement,

- The silicon nitrate trapping layer which can store electrons,
- and then the tunneling oxide
- At the very center is the polysilicon channel that makes the electrical connection across the string of memory cells

[Page 34]

Now that we understand what the 3D NAND structure looks like, let's talk in a little more detail about how 3D NAND works. There are 3 steps in NAND operation, Write, Erase and Read. "Write" means that the electrons move from the silicon channel to the trap silicon nitrate layer by applying 0V to the source and drain and around 20V to the gate electrode. To erase data, electrons need to get back to the silicon channel from the trap layers. By applying 20V to the source and drain and 0V to gate electrode, the electrons can be moved back. The last operation is Read. This is the process to detect the amount of electrons in the trap layer.

During the read operation, source and drain have very low voltage - less than 1V and the gate electrode needs some voltage which is smaller than the write voltage. The graphic on the right side shows NAND flash's own scaling method. Unlike in DRAM, NAND flash can have more than 1 bit in the 1 physical cell or 1 transistor by controlling the # of electrons in the trap layer. SLC stands for single level cell and it has only 1 bit per cell. It is the same as an empty or full cup of water. MLC stands for multi bit level cell and it can have 2 bits per cell.

In order to have 2 bits, 3D NAND cells would need to have 4 different levels of electrons similar to 4 different amounts of water in a cup. The last image near the bottom right shows TLC, or a triple level cell, which has 8 different levels and TLC is the most common product in the 3D NAND memory industry today.

[Page 35]

Next, let's briefly discuss the 3D NAND roadmap. For 3D NAND scaling, the industry is not talking about critical dimension anymore, but instead is counting stacks or pairs. From the very first 3D NAND product, the number of pairs has increased with each new generation and it continues to do so. However, total stack height increase is not proportional to the number of pairs. The reason being that with better understanding of 3D NAND, the industry has put in a lot of effort in reducing the thickness of each pair.

As shown in the table, the 1st generation of 3D NAND had around 70nm per pair, but now each pair thickness is down to less than 50nm. For scaling, every generation has key inflections to the structure and associated fabrication challenges. It is expected that all 3D NAND memory manufacturers will have a 2-tier based product in 2021 with the peripheral location changed to either under or over the memory cells. Recently, we have seen announcements from different memory manufacturers, about the availability of 3D NAND products with over 150 pairs. Others have announced the adoption of 2 tiers and QLC technology. We may even see the introduction of cell design changes in the future for continued scaling.

[Page 36]

Then, what can be done? How does 3D NAND scale further? There are clearly 3 levers - Lateral scaling, vertical scaling and increasing the number of bits per cell. First is lateral scaling. The first few generations of 3D NAND were mostly about going vertical and this still leaves significant opportunity for lateral scaling.

In lateral scaling, the industry is considering reducing the staircase area, peripheral circuit area and slit areas. Second is continued vertical scaling. Vertical scaling can be done by increasing the number of pairs, by reducing the thickness of each pair, or by fabricating the memory array in tiers. Third, is increasing the number of bits per cell, which I explained earlier. This method has a huge benefit for NAND flash memory, because we can increase bit density without further stacking or shrinking.

[Page 37]

What are the High Value Problems and solutions for lateral scaling?

For staircase, a single flight staircase is not scalable since staircase area increases proportionally with the number of stacks. Given that, the industry is adopting new staircase formations like a zig-zag shape, which will require multi-pair etch as well as large area gap-fill technology for the staircase etched out area. 2nd HVP is the Peripheral area, which accounts for ~ 30% of the die. Memory manufacturers are trying to reduce the peripheral area by moving the peripheral transistors to under or over the cell array.

3rd HVP is in trying to reduce the slit area. Slit area, which is not being used in a working cell, takes up more than 10% of the cell array today. By adding more memory holes, manufacturers can reduce the slit area. However, due to increased pitch between slits, new silicon nitrate exhume and ALD fill processes will be needed.

[Page 38]

Let’s start with the staircase area savings. The role of a staircase is to act as a pad for the gate electrode contact. Here we compare the traditional staircase with zig-zag staircase architectures. From the image at the top, you can see that a 48 pair 3D NAND staircase takes ~ 60 micrometer and without any inflections, and staircase length can be as long as ~120 micrometer for 96pairs. But in reality, a 96 pair staircase length is ~ 10 micrometer, as shown in the second image, which is almost a ~90% area savings!

How is this possible?

Let’s assume you need to make 10 contacts. You can do this either as a 1 X 10 or 2 X 5 or 5 X 2. By placing more contacts in the width direction rather than the length direction, you can save on the overall staircase length. This zig-zag like staircase formation brings others challenges, such as the need for a deeper etch with profile control as well as maintaining CD and etch rate uniformity. Nowadays, Memory manufactures are paying more attention to these types of staircase innovations and Applied Materials’ Sym3 has already been adopted for this in production at 3 major customers, continuing our leadership position in staircase etch.

[Page 39]

Another opportunity for lateral scaling is in the Peripheral CMOS area. In first-generation 3D NAND, most of the memory manufacturers placed the CMOS and array side by side and this consumes a significant amount of silicon area, ~ 30% of the die. But now more memory manufacturers are moving the peripheral transistors to either under or over the memory array, expecting around a 10-15% die area savings. What are the benefits and challenges of CMOS under array versus CMOS over array?

First, let's look at CMOS under array, also known as CuA. The primary benefit of CuA compared to CMOS over Array, or CoA, is that it has a lower manufacturing cost and yield loss. However, the CMOS junction in CuA must be fabricated before the cell formation and this is subject to an additional thermal budget. Compared to CuA, the CMOS transistor in CoA must be fabricated on a separate silicon wafer. Subsequently, the two wafers must be bonded to connect the critical metal connection pads.

The CoA process has the benefit of not being constrained by the thermal budget limitations and also allows for faster copper interconnect technology to be used. However, this brings a different set of challenges, such as more copper layers and copper CMP for wafer to wafer bonding.

[Page 40]

What technologies does the industry need to enable CuA and CoA? The CuA architecture requires more thermally stable junction formation. For this, optimized ion implantation and annealing processes are most crucial. With low energy and co-implant technology, we believe the industry can minimize lateral thermal diffusion and make thermally robust transistors.

Both CuA and CoA will require several interconnect steps compared to the conventional architecture. For CuA, there will be more than 2 additional tungsten interconnect layers and for CoA there will be more than 4 additional copper steps for wafer to wafer bonding.

[Page 41]

The Next key technology for CoA is wafer to wafer bonding, specifically hybrid bonding, which involves copper to copper and Dielectric to dielectric bonding. For Hybrid bonding, the key challenge is copper dishing control during CMP. A key requirement for hybrid bonding is less than 5nm dishing over the pad. If there is severe dishing, the copper layers cannot be bonded to each other.

Applied Materials' Reflexion LK copper CMP uses 3 platens for Bulk copper removal, barrier removal and topography correction, respectively, to overcome the CMP challenges. We have confirmed that wafer dishing ranges can be improved by more than 60% using our platform.

[Page 42]

The second approach to 3D NAND scaling is in the vertical direction. This is mainly accomplished by adding more pairs, but this becomes less cost effective, as the height and the aspect ratio increase. Shrinking the thickness of each pair, allows more layers at the same stack height. But in shrinking the stack height, it gets progressively harder to exhume the silicon nitrate, and to fill the space with metal in the replacement gate process.

A third approach to vertical scaling is to split the stack into tiers, first fabricating a lower tier, and then an upper tier. The processes for making multiple tiers are more complex and create challenges with the alignment of the two tiers as well as staircase integration.

[Page 43]

Let's get into more of the details of vertical scaling by adding pairs. With added pairs, thicker Hardmask deposition and etch are needed for post-processing. Ironically, a thicker Hardmask increases the overall stack height, resulting in a higher aspect ratio. In order to resolve this issue, Applied Materials has developed more selective and low-stress Hardmask films compared to conventional Hardmask, which enables the Hardmask to be thinner.

These new Hardmask films need new etch patterning processes. By deploying Applied Materials' Integrated Materials Solution (IMS) approach, we have optimized the etch process in parallel with the new Hardmask.

[Page 44]

After the staircase formation, the etched-out area must be filled with dielectric films. With the new zig-zag staircase architecture I explained earlier, the typical feature size is around 2 micrometer in width, and 10 micrometer in depth. This CD, 2 micrometer, is too large for an ALD process, but too small for a conventional CVD dielectric process due to the high aspect ratio. Given that, we have developed a new Plasma Enhanced High Aspect Ratio Process or, PE-HARP dielectric gap fill process.

This film has a very high deposition rate with tunable stress and also exhibits the lowest shrinkage post-process compared to the alternatives. In addition, our conformal PE-HARP process simplifies the next planarization step. Earlier CVD technologies left a thick overburden in areas of the wafer requiring additional etch back steps before CMP. With these new films, there will be no need for the additional etch back steps prior to CMP, which provides additional integration cost savings to the memory manufacturers.

[Page 45]

The next key process for vertical scaling is metal gap fill for high aspect ratio contact. In 3D NAND, CVD tungsten is widely used to make many connections in the vertical direction. With an increase in the stack height, this metal fill process faces two challenges. First, the deep contacts have a barrel-shaped profile. When the contacts are filled with W, the contact can be pinched off at the top, trapping corrosive gas inside, which can damage large parts of the die afterwards.

Second, these metals typically grow with a high tensile stress, around 2 giga pascal, which can deform the wafer and crack adjacent delicate features. To overcome these issues, Applied Materials pioneered Seam Suppressed Tungsten technology. This process includes a nucleation and treatment step that suppresses the film growth on the top and allows more uniform, seamless bottom up filling.

This process can also be tuned to produce a film with a stress level that is one third of the stress versus conventional CVD tungsten. We are seeing strong traction for this process in 3D NAND manufacturing with the increase in stack heights.

[Page 46]

Let me summarize the 3D NAND portion of this Memory Masterclass. NAND scaling for higher density and higher I/O speed requirements will continue and more innovative architectures and process technologies will be needed to keep pace with the demand for higher density and speed.

We strongly believe that Applied Materials is well positioned with leading products which will be strengthened for both unit processes as well as for integrated material solutions. This is an exciting time for NAND with multiple new technologies being adopted and an inflection to a new cell design on the horizon.

I couldn't be more excited than to be here working on these challenges at this time! Thank you!

DR. RAMAN ACHUTHARAMAN | Group Vice President, Semiconductor Products Group

Thanks Sean.

And that concludes the technical portion of the meeting.

It's great to meet with you today all once again and show you how we're enabling the PPACt roadmap for our customers in these inflections with our leadership products and unique solutions.

[Page 49]

I'll start by recapping how we've been doing in the Memory equipment markets as we increased our investments in new products aimed at particular inflections.

Our share has steadily increased with the success of Sym3, our Conductor Etch product, along with our CVD systems. Our success has come from cooptimizing Sym 3 etch and new CVD hard mask films for use in patterning applications in both 3D NAND and DRAM.

Our share is now well balanced across all the device types making us agnostic to the spending mix in any given year.

[Page 50]

We shared our vision and strategy with you at the investor meeting last month. Accelerating PPACt – which stands for Power, performance, area, cost, and time to market – is the key challenge facing the industry. Our goal is to be the PPACt enablement company. And we feel we are uniquely positioned to enable the PPACT roadmap for our customers. Briefly, there are three elements to our strategy.

First, leveraging our unit process leadership and broad portfolio of products

Second, developing enabling solutions by uniquely combining our leadership technologies

And third, using unique, massive data and analytics to create solutions for accelerating PPACt in new ways.

[Page 51]

Here's a framework to think about the evolution beyond unit processes. The first stage is co-optimization of processes, where we consider how to optimize upstream and downstream steps. About 40% of our products today have an element of co-optimization.

The second is integrated materials solutions. As we continue to scale and use more exotic materials, any exposure of the wafer to the environment can have a negative impact on device performance and yield. Our unique ability to connect our broad leadership products in vacuum provides interface innovations that have huge PPACt benefits. About 30% of our products come designed with this capability.

The third stage involves generating actionable data at scale and using AI to help cooptimize the many process variables to speed time to recipe development and widen process windows which enables higher yields. The Applied AIx platform enables us to provide new actionable insights to our customers.

[Page 52]

We just introduced our AIx platform last month. It combines our process systems with sensors, metrology tools, analytics and machine learning capabilities. And it’s scalable for use in new contexts and applications. There are 3 key elements to accelerating actionable insights for customers.

First, we cannot fix what we cannot measure. So we are adding new sensors and metrology to our chambers and systems. These give us the ability to gather process data in real time.

Second, we cannot fix what we cannot see. Our eBeam technology leadership enables us to measure on-wafer results 100 times faster with 50% better resolution over conventional methods. Combining the process sensor and metrology data with a million on-wafer measurements enables us to create actionable data that correlates process parameters with on-chip and on-wafer results.

Third, AI can see what human engineers cannot. As you heard earlier from Sony and Sean, process complexity is increasing. Each tool is capable of 10,000 process combinations, and integrated flows for applications like interconnects provide a million potential variations.

AI is excellent at pattern recognition. The AIx analytics platform can help engineers find the optimum recipes with the widest process windows. We can cut R&D times in half and widen process windows by one third. And once the recipes are proven in R&D, the same technology can be used to match the ideal results in ramp and high volume. Our goal is to accelerate PPACt at every stage of the process lifecycle, from R&D, to ramp and HVM.

[Page 53]

Today, we issued a press release announcing three ways our DRAM customers are using our technology to continue scaling DRAMs. For capacitor scaling, we announced Draco, our new hard mask material that allows customers to shrink DRAM capacitor size while maintaining very high aspect ratios to maximize charge. We also announced how our DRAM customers are shrinking the periphery in two ways.

First, Applied has reengineered our Black Diamond dielectric CVD film for the DRAM market and made it available on the highly productive Producer GT platform. Black Diamond for DRAM enables smaller,

more compact DRAM interconnect wires to move signals through the chips at multi-gigahertz speeds without interference and at lower power consumption.

Second, the DRAM industry is borrowing another page from the logic playbook by transitioning the periphery logic transistors from polysilicon gates to high-k metal gates. This gives them higher density plus higher performance and lower power.

All of these technologies are now shipping in high volume and fanning out to more DRAM customers.

[Page 54]

Now I'll help you connect the technology inflections you learned about in the Master Class to the growth opportunities in our business over the next several years.

The first lever for DRAM scaling we discussed today was memory cell scaling. Within the cell, the capacitor module is the most critical, and capacitor scaling is the industry's number-one high value problem. The key challenge and breakthrough was creating a new capacitor patterning solution. We've introduced 3 innovations to solve this problem.

First, we developed a new hard mask material, Draco, with 30% higher selectivity. With this new material, we can lower mask thickness by 30% which decreases aspect ratio and reduces the burden on the Etch process.

Second, we delivered a new Sym3 etch technology that is cooptimized to take advantage of the unique properties of Draco.

Third, we are using PROVision eBeam inspection and metrology which provides a unique, non-destructive way of quickly measuring the bottom CD dimensions.

This co-optimized solution for DRAM capacitor scaling helped us accelerate our customer's time to market and gave us significant growth in DRAM CVD and etch revenue. Our DRAM capacitor hard mask solution provides us with a 1-billion-dollar cumulative revenue opportunity from 2020 through 2024. We generated significant revenue in 2020, and we expect to grow revenue by 4X in 2024.

[Page 55]

The Second lever of DRAM scaling we discussed today was the periphery area where both the transistors and interconnect wiring are being reengineered to improve power, performance, area and cost. Applied became the leader in logic transistor and interconnect scaling over more than two decades, and this leadership is very applicable in DRAM today.

HKMG transistors can help double DRAM performance while lowering power by 10%. We have the industry's most proven solutions for HKMG transistor formation. Our Black Diamond CVD film is now fanning out across the DRAM market as well. As this happens, we have a 2-billion-dollar cumulative revenue opportunity from 2020 through 2024. We established strong customer pull in 2020, and we expect to grow revenue by 3X in 2024.

[Page 56]

Advanced packaging is now an important PPACt enabling technology for chip and system performance in logic and memory. This trend creates a new growth market for advanced process equipment and know-how. In DRAM, 3D die stacking using bump and through silicon vias is already prevalent. In the future, CMOS over array architectures for NAND will require a new technology called hybrid bonding with direct, copper-to-copper bonding of heterogeneous die.

Applied is #1 in advanced packaging. We are well positioned because we have the broadest portfolio, unique capabilities, and a full flow packaging lab where our customers can innovate with us.

Additionally, we have strong ecosystem collaborations and partnerships with customers, industry peers and research institutes. Packaging is already a significant business for us, with around $500 million dollars in revenue in 2020. We plan to more than double revenue by 2024. And we are still in the early innings of growth. Hybrid bonding will be highly enabling, and we see its growth accelerating in the outer portion of our model horizon and well beyond it.

[Page 57]

Before we transition to Q&A, I'd like to summarize today's Master Class with these key messages.

First, The AI Era of computing inflection is driving a new wave of industry growth and new opportunities for semiconductor innovation.

Second, It's clear that the future is not going to be like the past. AI computing workloads require semiconductor advances at a time when traditional Moore's Law scaling is slowing down.

The new PPACt playbook that will shape the industry's future will be enabled by advances in materials engineering and materials to systems co-optimization. We have aligned our strategy and investments to position Applied as the PPACt enablement company.

Third, we are uniquely positioned to accelerate the PPACt roadmap. We have the broadest and most enabling portfolio of unit process technology spanning materials creation, modification, removal and analysis.

This allows us to combine these technologies in unique and highly enabling ways that no one else can. With the AIx platform, we have developed an exclusive suite of solutions to provide actionable insight acceleration and significantly improve the time to market of new PPACt innovations.

And fourth, as we outlined today, multiple big technology inflections are occurring in memory, and our leadership in logic scaling is giving is a new wave of growth in memory.

We have significantly grown our memory share through 2020, and we have very strong momentum as we look ahead to 2024 and beyond. I've been working in product development for over 25 years, and I feel like there has never been a better outlook for growth and customer enablement with technologies that are right in our sweet spot.

So -- thank you kindly for your attention, and now Mike and Kevin, let's begin the Q&A.