SEMICON West 2024 Technology Breakfast

KEY POINTS | July 9, 2024

Intro:

- Beginning at our 2016 Analyst Day, we sensed that AI and IoT would drive a new era of growth for the industry.
- At the time, wafer fab equipment was about $30 billion, and it was around that same level for a decade and a half.
- But since then, the market has tripled.

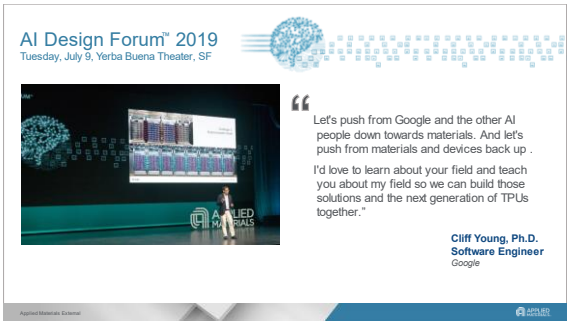
- At our 2017 investor meeting, we had our first AI panel for investors and debated topics like whether cloud computing or edge computing would drive more growth for the industry.
- And whether AI would drive logic or memory growth more.
- We really didn't know at the time.

SEMICON West 2024 Technology Breakfast

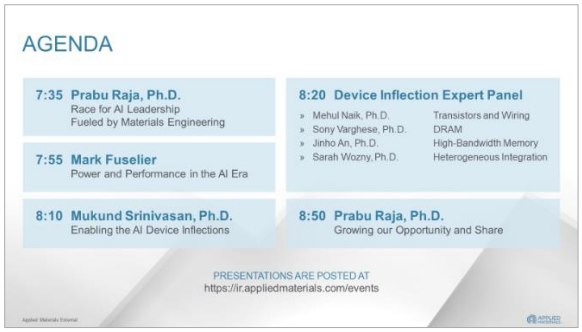
KEY POINTS | July 9, 2024

- Then around 6 years ago today, right here at SEMICON West, we hosted an event called “AI Design Forum: The Future of Computing, from Materials to Systems.”
- Bill Dally from NVIDIA gave a keynote, explaining how AI came to outperform humans in image recognition.
- Bill said deep learning would revolutionize most industries – and that its progress would come from more energy-efficient hardware and algorithms.

- At the same event, Applied Materials’ Gary Dickerson talked about the need for a 1,000-fold increase in performance per watt.
- And because classic Moore’s Law had begun to slow, he introduced a “new playbook” for performance, power and cost.
- The new playbook was based on new architectures like GPUs and TPUs, new materials, new 3D structures like Gate-All-Around transistors – along with advanced packaging.
- And that new playbook is absolutely what’s fueling AI computing today.



- Then, exactly 5 years ago today at SEMICON West, we had the second AI Design Forum where Cliff Young of Google explained why Google was designing TPUs – which is because running the new AI workloads on standard microprocessors was going to swamp all of Google's data center capacity.
- Cliff also looked to the future of AI and said the entire computing industry needed what he called co-design, from systems to materials and from materials to systems.
- He said, he'd love to learn about our field and teach us about his field so we can build those solutions and the next generation of TPUs together.
- That spirit of co-innovation – from materials to systems – is exactly what we've been driving ever since.

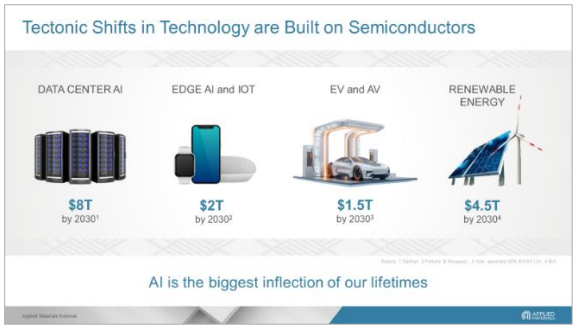


- Here is today's agenda.
- Prabu Raja will talk about the opportunities and challenges of the AI era.
- He'll also talk about our capabilities as a company, and how we're connecting our capabilities in new ways to co-innovate with our customers and industry partners to enable even better AI.
- Then, Mark Fuselier from AMD will lay out his company's AI computing technology vision and tell us what he needs more of from our industry.
- Mukund Srinivasan from Applied Materials is going to talk about how we co-innovate with our customers on these items, and he'll also invite a panel of our technology experts to give you the roadmaps for better AI across logic, DRAM, high-bandwidth memory and advanced packaging.
- And Finally, Prabu will tell us what these roadmap and R&D innovations will mean for our business.

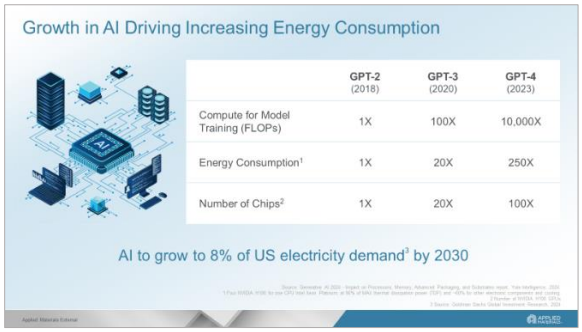
SEMICON West 2024 Technology Breakfast

KEY POINTS | July 9, 2024

- It's a very exciting time to be in the semiconductor industry.
- Today's event is all about artificial intelligence.
- We are in the middle of an AI race.
- Today, I will talk about what the AI race means to the wafer fab equipment industry and to Applied Materials.

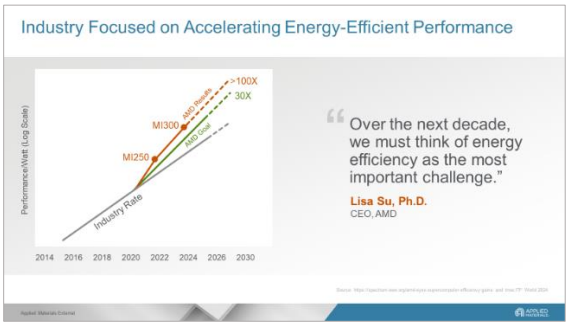


- Tectonic shifts in technology driven by AI and energy transformation are fueling the economy.
- Data center AI is forecast to become a $8 trillion market by 2030.
- Edge AI and IoT is forecast to grow to a $2 trillion market in the same timeframe.
- Every one of these areas is a race – a multibillion-dollar race.



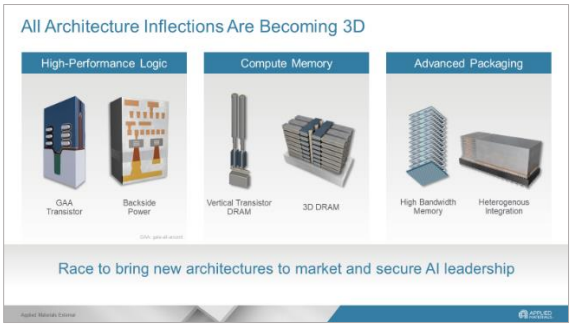
- There has been exponential growth in new large language models.

- In 5 years, compute for training models has grown 10,000x, energy consumption 250x and the number of chips has grown 100x.
- Some forecasts predict total AI energy consumption will grow to 8% of US electricity demand by 2030.
- This level of growth in energy consumption is not sustainable.
- Therefore, the race for more energy-efficient computing is crucial for AI.

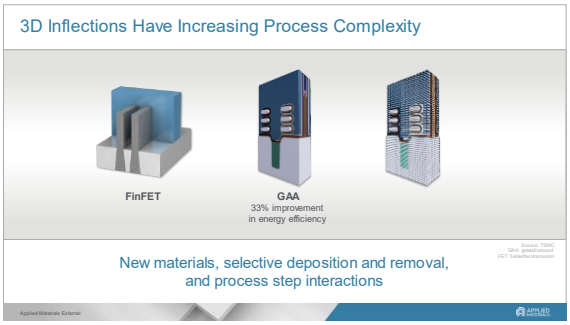


- Semiconductor companies are racing to accelerate energy-efficient performance.
- AMD has stated a goal of a 30x improvement in energy efficiency by 2020 to 2025, and is on track to achieve 100x by 2027.
- Mark Fuselier will talk about how AMD is meeting this goal.

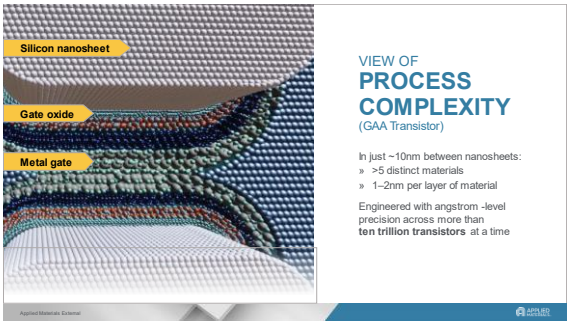
- Four major technologies are enabling the AI race: leading-edge logic, high-performance DRAM, high-bandwidth memory and advanced packaging.
- Companies are racing to implement the best of these technologies.



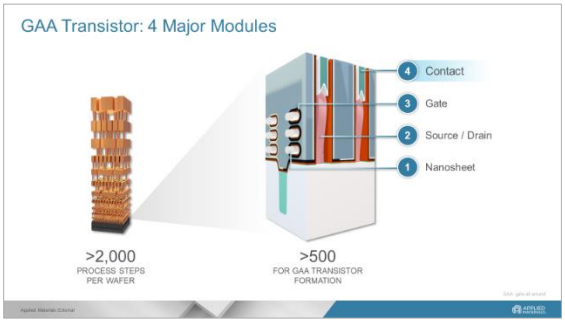
- The industry is adopting architecture inflections that enable energy efficiency.
- Logic is moving to a new transistor architecture: Gate-All-Around.
- In wiring, the new architecture is backside power.
- DRAM is moving to a vertical transistor architecture and then to 3D.
- Advanced packaging is being used for high-bandwidth memory and heterogeneous integration.
- All these inflections are going 3D, which means more materials engineering.



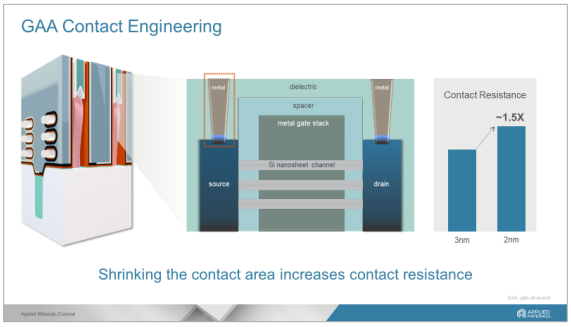
- 3D devices like the Gate-All-Around transistor architecture provide tremendous improvements, as much as 33% lower power than FinFET; however the manufacturing process is far more complex.
- 2D devices can be made with lithography, deposition and etch.
- 3D devices need many new materials, new processes and new selective deposition and removal technologies.
- And these new materials and adjacent process steps can have significant interactions that require co-optimization and integration.
- This is where Applied comes in.



- Let's take a closer look at one example of the complexity of the Gate-All-Around transistor.
- The space between the nanosheet channels is only 10 nanometers.
- More than 5 different materials are used to create the gate in that tiny area.
- Replicating this flawlessly across 10 trillion transistors on a wafer is an incredible feat of materials engineering.

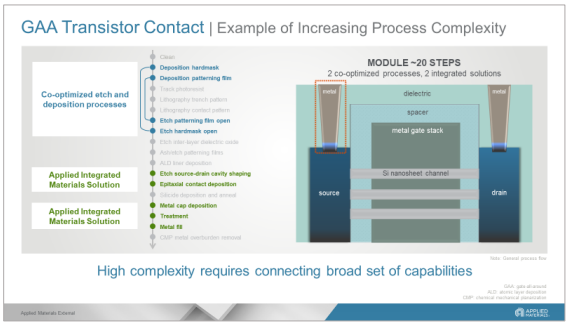


- Building a Gate-All-Around transistor involves 4 major modules: the nanosheets, source/drain, gate and contact.
- Let's drill down on the contact, which connects transistors to the wiring and is the narrowest connection on the entire chip.

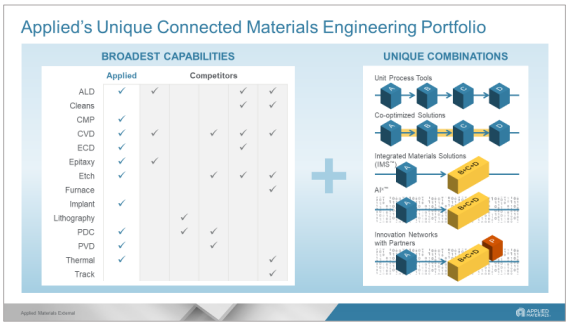


- This is a different view of the Gate-All-Around transistor that shows how we form the contact.
- As the contact area shrinks and becomes narrower node over node, the contact resistance increases.

- For example, from 3nm to 2nm contact resistance increases 1.5x.
- More resistance degrades performance.

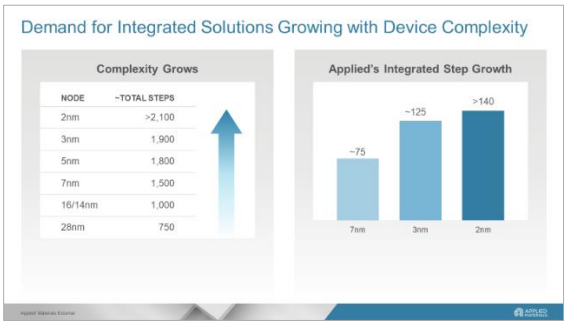


- Building the Gate-All-Around transistor contact involves around 20 steps, and Applied Materials participates in about half of them.
- There is a strong interaction between steps, both upstream and downstream.
- These steps need to be connected and co-optimized.
- The steps highlighted in green indicate the more critical interactions.
- For these steps Applied offers Integrated Materials Solutions (IMS) where multiple different steps can be performed in one system under vacuum with no air exposure.
- Applied has the unique ability to integrate different technologies and create IMS products.

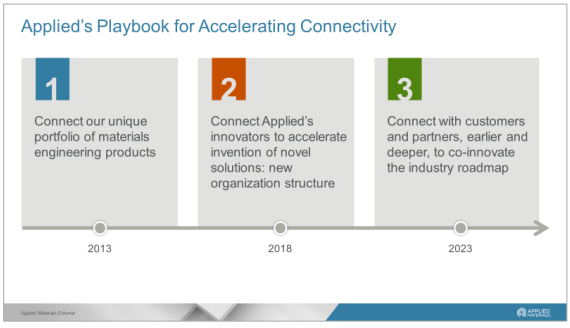


- Applied has the broadest portfolio of materials engineering technologies and can connect these technologies to create unique combinations of co-optimized solutions and IMS.

- Here is one example from the Gate-All-Around contact module showing our IMS which integrates etch to shape the source-drain cavity and selective epitaxy for contact deposition, with both steps done within the same platform.
- Our IMS lowers contact resistance by 50% compared to a non-integrated system, and lower contact resistance means better performance.
- This is just one example of integration in a transistor; there are many such examples in logic, DRAM and advanced packaging.



- Device complexity is increasing the number of process steps at every node.
- The number of Applied's integrated steps is also increasing as complexity drives the need for more connected process steps.

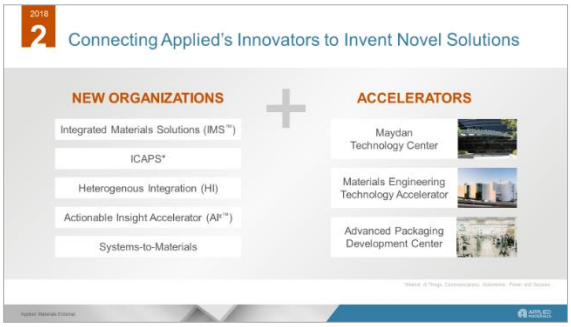


- Applied Materials saw this trend early, and in 2013 we started building and shaping our connected portfolio.
- By 2018, we formed a new organization structure to enable a more connected portfolio.
- We connected innovators across the company and brought device integration talent into the company.
- Then in 2023, we announced the EPIC Platform where we will be connecting with customers, partners and universities to accelerate critical solutions that enable the semiconductor roadmap.

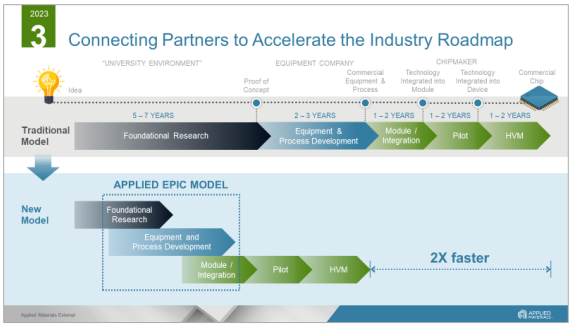
SEMICON West 2024 Technology Breakfast

KEY POINTS | July 9, 2024

- Step 1 was building and shaping our connected portfolio.
- We expanded from unit processes – one technology in one tool – to co-optimized processes where two or more technologies are co-optimized in adjacent systems, and then to integrated processes where two or more co-optimized technologies are in one system.
- Today, Integrated Materials Solutions represent about 30% of our Semiconductor Systems revenue which is up from about 20% in 2019.



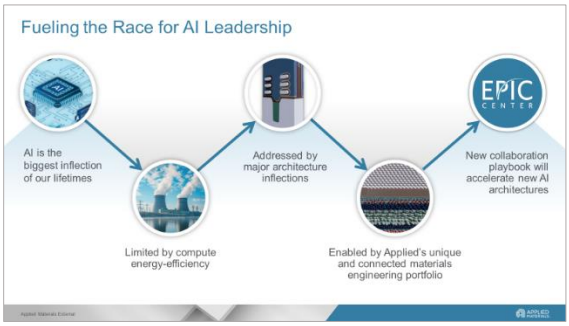
- In 2018, we formed new organizations to connect our innovators across the company.
- To accelerate breakthroughs we need to validate innovations in devices and created integrated labs.



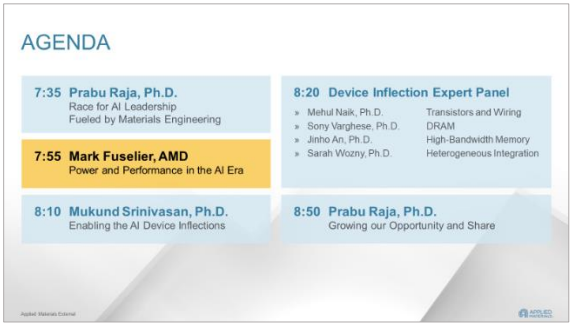
- In 2023, we announced the EPIC Platform which is designed to strengthen our connection with partners to accelerate the industry roadmap.
- EPIC represents a new model where we move from serial innovation to parallel innovation and commercialization of new technologies.

- It's designed to accelerate co-innovation with our customers, partners and universities.
- Co-innovation is a powerful concept to accelerate energy-efficient computing.

- At Applied, we collaborate with our major customers on their semiconductor roadmaps.
- Here is a quote from Jaihyuk Song, Corporate President and Device Solutions CTO at Samsung Electronics, commenting on Samsung's collaboration with Applied Materials.

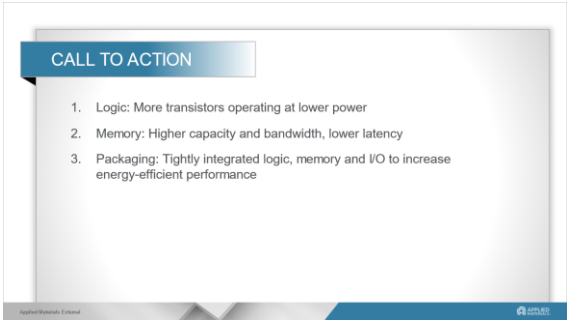


- In summary, the race for AI leadership is limited by energy-efficient compute.
- Energy-efficient compute will be addressed by major chip architecture inflections that are enabled by Applied's unique, connected materials engineering portfolio.
- Our EPIC Platform represents a new collaboration playbook designed to accelerate new AI architectures.



- Next, we'll hear from Mark Fuselier of AMD.

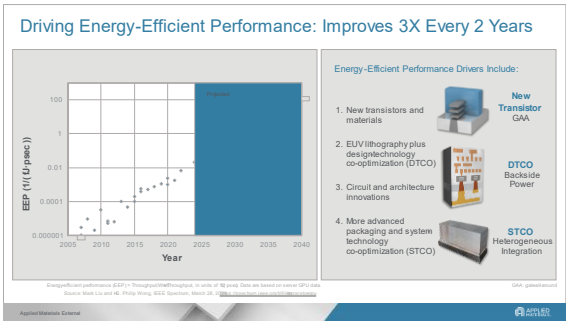
- Now, Dr. Mukund Srinivasan will explain how we are working with our customers to develop more energy-efficient technologies.
- Mukund joined Applied Materials 11 years ago, and he has been in the industry over 25 years.
- Starting about 6 years ago, he became the GM for our IMS group.
- His teams work with leading logic and memory customers to develop the next generations of technology, through co-optimized and integrated solutions.



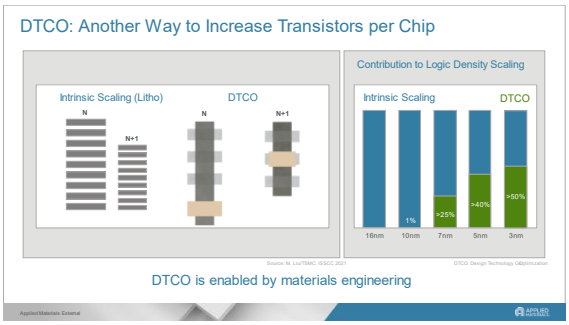
- Thank you Prabu.
- What stood out for me in Mark's presentation are 3 call-to-actions for energy-efficient performance.
- First, accelerate the transistor and wiring roadmap.
- Second, improve DRAM bit density, bandwidth and latency.
- Third, accelerate advanced packaging to improve energy-efficient performance.

- Here is the agenda.

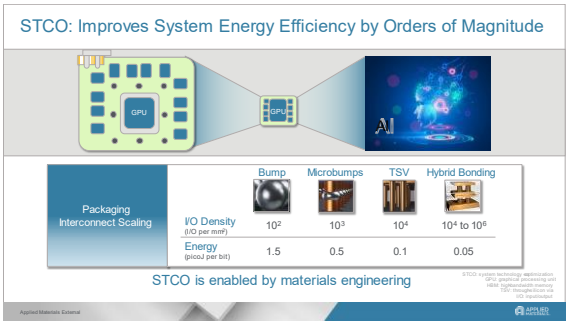
- First, I will talk about our key R&D goals and ways we work with our customers to deliver the semiconductor roadmap.
- Then, 4 of our technology experts will join me and provide a glimpse into the roadmaps for logic, DRAM and heterogeneous integration that are key to enabling the AI inflection.



- Let's begin with the customer perspective.
- In a recent article, TSMC illustrated that energy-efficient performance has been improving every year, and they forecast it improving 3X every 2 years.
- TSMC suggests 4 drivers for energy-efficient performance.
- First, new transistors and materials; second EUV and DTCO; third, circuit architecture innovations; and fourth, more advanced packaging and STCO.



- As Mark Liu from TSMC described, there are 2 ways to scale.
- First, intrinsic scaling through lithography, which was the dominant technique through 10nm.
- Second, Design Technology Co-Optimization – DTCO – which is the rearrangement of elements of the transistor and logic cell to fit a smaller area.
- DTCO is becoming increasingly important, contributing 50% of scaling at 3nm.
- DTCO is enabled by materials engineering.



- Second is STCO, which stands for System Technology Co-Optimization.
- Mark Fuselier talked about 2.5D and 3D packaging which drives “volumetric scaling.”
- STCO can increase transistor density by orders of magnitude.
- New innovations like hybrid bonding enables shorter interconnects which can reduce power by orders of magnitude.
- STCO is enabled by materials engineering.

- Applied uses a unique, connected materials engineering portfolio for better logic and memory chips and systems.
- We co-innovate with our customers to achieve these advances.
- Six years ago, we formed our IMS group — dedicated teams that work with customers to enable device inflections.
- This innovation happens at 3 major centers: Maydan Technology Center in Sunnyvale, CA, META Center in Albany NY and our Advanced Packaging Development Center in Singapore.
- Additionally, last year we launched EPIC, which is designed to be the largest collaborative innovation center for semiconductor process technology and manufacturing equipment R&D.
- EPIC will provide early access to next-generation processes and equipment to accelerate innovation and commercialization of foundational semiconductor technologies.

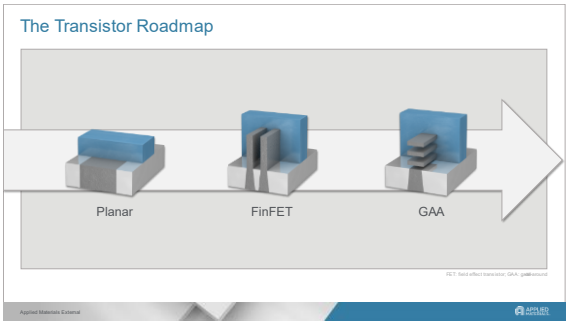
SEMICON West 2024 Technology Breakfast

KEY POINTS | July 9, 2024

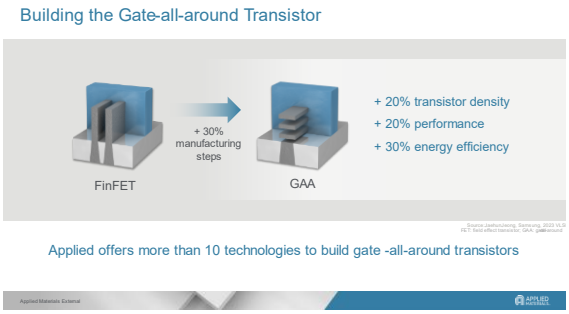
- Applied has strong partnerships with multiple customers.
- Here are comments from Naga Chandrasekaran who leads technology development at Micron.
- We work closely with his team to develop new manufacturing technologies for DRAM transistor architectures and advanced packaging for high-bandwidth memory.

- Next, I'd like to invite four of our technology experts to join me.
- Dr. Mehul Naik spoke at our May 2022 master class on wiring, where he introduced Backside Power Delivery. Today, he'll cover the Gate-All-Around transistor, and frontside and backside wiring.
- Dr. Sony Varghese participated in our master class on memory in May 2021. Today's he'll cover the DRAM roadmap.
- Next, Dr. Jinho An will discuss high-bandwidth memory.
- Finally, Dr. Sarah Wozny will cover the heterogeneous integration roadmap.

- Thank you for joining us, Mehul.

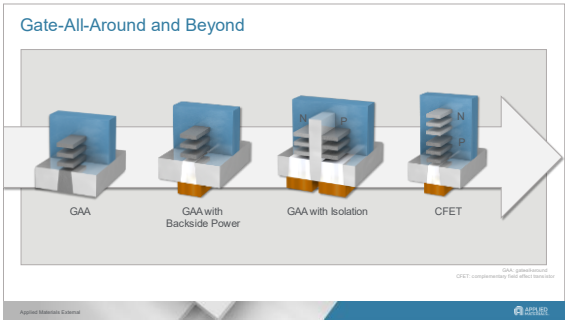


- A short time ago, Mark described AMD's AI computing technology vision, and both he and TSMC have discussed the need for new transistors.
- What's on the transistor roadmap, and why?
- Let me start with some background.
- The FinFET transistor was introduced about 10 years ago.
- Since then we have had 5 generations of FinFET improvements, mainly in making the fins narrower and taller to continue scaling.
- This approach has reached its limit at the 3nm node, and the industry is now implementing a new transistor architecture called Gate-All-Around.

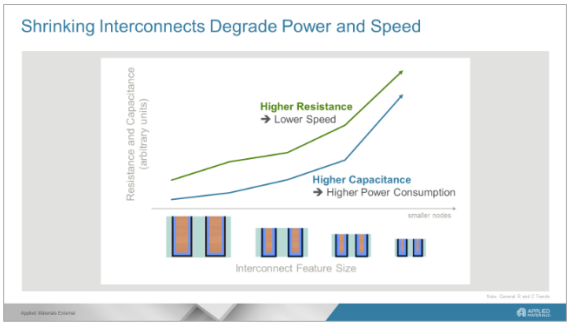


- How is the new Gate-All-Around transistor different, and what are its benefits?
- In the FinFET architecture, the silicon channels are vertical and surrounded by a gate on three sides.
- With Gate-All-Around, the silicon channel is horizontal and stacked which allows us to fit 20% more transistors in each chip.
- We surrounded each of the channels on all 4 sides, which gives us better leakage control and the opportunity to reduce power consumption by up to 30%.
- The Gate-All-Around architecture also allows us to use more channels, which means we can increase drive current up to 20% higher than with FinFET.
- Is Gate-All-Around more complicated to manufacture, and how does Applied help?

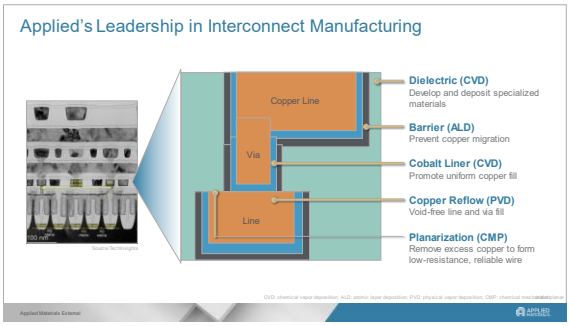
- Yes, it's significantly more complicated with about 30% more process steps.
- From an equipment perspective, new innovative unit process technologies are required.
- As Prabu showed us earlier, more co-optimization is needed, and Integrated Materials Solutions will play an even larger role.
- Applied offers 10 major technologies for Gate-All-Around manufacturing, and the number of Integrated Materials Solutions is growing.



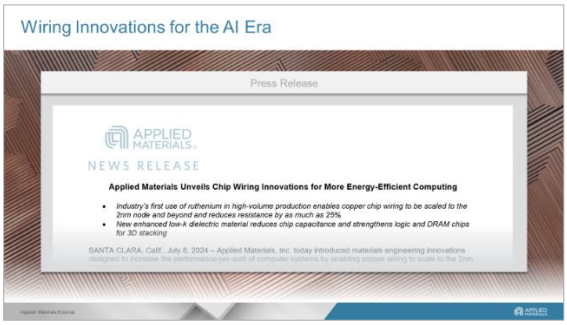
- You mentioned the FinFET architecture was extended for around 5 generations, primarily by increasing the fin heights. How far can we take Gate-All-Around, and how?
- I believe Gate-All-Around can be extended for 2-3 generations.
- One area of progress will be combining Gate-All-Around transistors with backside power.
- Another idea is called Gate-All-Around with isolation, which is designed to bring the NMOS and PMOS transistors closer together to increase density and performance.
- Does the industry see another transistor inflection behind Gate-All-Around, and when?
- Yes, we do. It is called CFET which stands for Complementary Field Effect Transistor.
- In Gate-All-Around, the N and P transistors are laid out next to one another.
- In CFET, we will stack the N and P transistors on top of one another, which in theory will double the transistor density.
- CFET will be even more complex than Gate-All-Around, and we expect them to appear by early in the next decade.
- So, the transistor roadmap for the next 10 years is fairly clear.



- Next comes wiring. Chipmakers want to keep shrinking with EUV. How does this impact wiring?
- First, shrinking brings the wires closer to one another which gives less space for the insulating low-K dielectric material, causing crosstalk, signal delays and distortion.
- Second, shrinking gives smaller spaces for copper wiring which increases resistance.
- This reduces performance and energy efficiency.



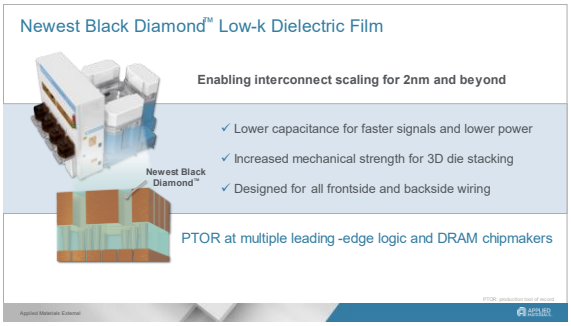
- So what is Applied's role in this?
- We have a comprehensive role. We not only develop and deposit specialized CVD dielectrics that insulate the wires from each other, but we also have metals that lower resistance in smaller features.
- We deposit the ALD barriers to prevent copper migration and we deposit CVD liners to promote copper fill followed by PVD copper fill, and finally, we use CMP to remove the excess copper to form a low-resistance, high-yielding, reliable wire.



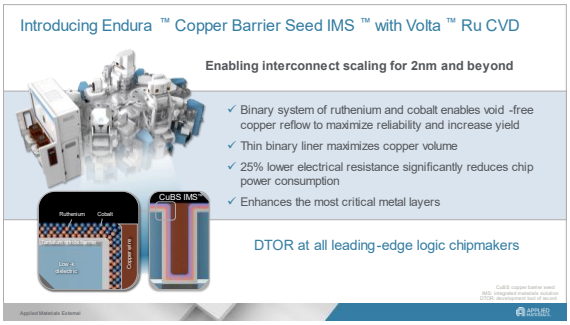
SEMICON West 2024 Technology Breakfast

KEY POINTS | July 9, 2024

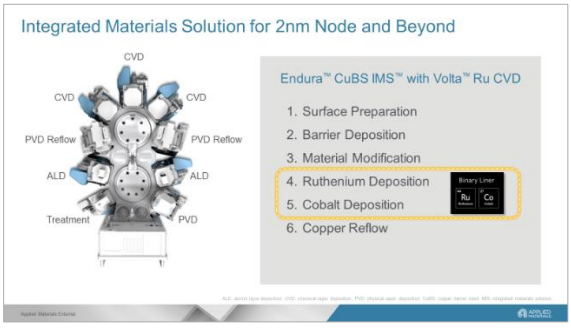
- Yesterday Applied introduced two new products that allow us to keep shrinking and extend copper wiring to the 2nm node and below. How are we doing that?



- First, we introduced an enhanced version of our Black Diamond low-k dielectric material.
- It maintains low capacitance even as we shrink.
- In addition, it has 20% higher mechanical strength than its predecessor, which allows customers to take 3D logic and DRAM stacking to new heights.

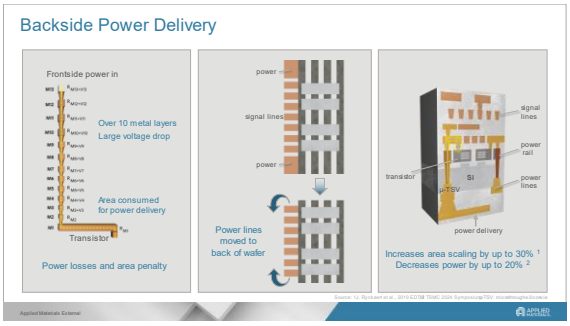


- Great. And what is the second innovation?
- We invented a new binary liner for the copper wiring that combines ruthenium and cobalt for the first time.
- This new liner is 33% thinner, which reduces the liner width from 3nm to 2nm.
- This allows us to deposit pure, void-free copper wires and reduces resistance by up to 25%.

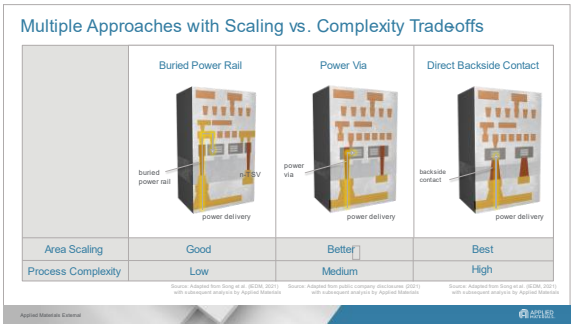


- Can we make this new wiring with unit process tools?

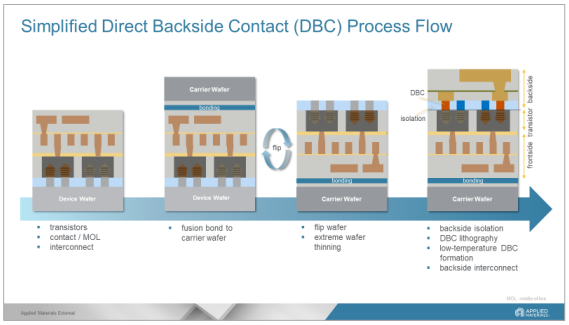
- Absolutely not.
- The only way to engineer the ruthenium, cobalt and copper is in an Integrated Materials Solution which combines 6 sequential steps in high vacuum.



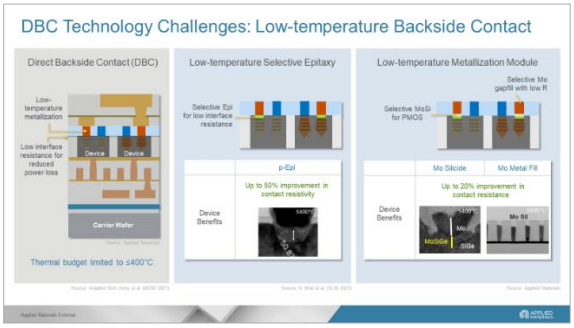
- Finally, let's talk about backside power delivery. What is the motivation?
- Moving the power wires to the back has many benefits.
- First, it reduces wiring complexity on the front side of the chip and gives customers new kinds of design flexibility.
- Second, it provides area savings of 6-30 percent.
- Third, designers can choose between increasing chip performance by up to 10% or reducing power consumption by up to 20%.



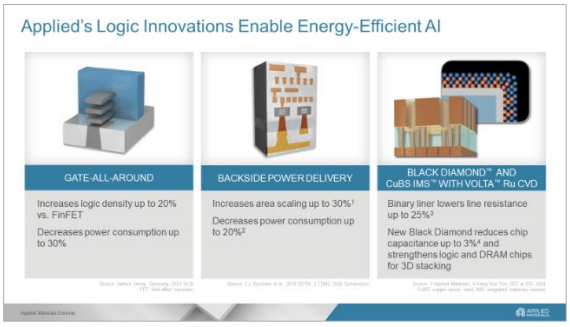
- You gave a wide range of 6% to 30% for the area savings. What drives that?
- It really depends on how aggressive and complex the design is.
- The simplest, buried power rail scheme saves the least area.
- The power via approach offers greater scaling.
- The most complex approach is providing direct power connections to the transistor source drains which also provides the maximum scaling.



- How is the manufacturing done?
- First, we build all of the transistors and frontside wiring.
- Then we bond a carrier to the frontside of the wafer for stability.
- Next we flip the wafer over and grind it and use CMP to get close to the transistors and logic cells.
- Then we use brand new steps to build the backside isolation, contacts and wiring.

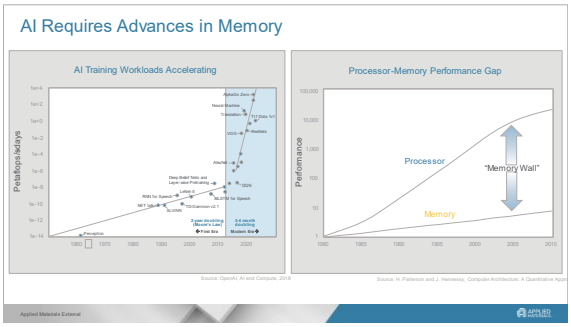


- What are the special challenges, and how can Applied help accelerate this?
- A major challenge is that all of the backside steps have to be low-temperature to avoid degrading the frontside transistors and wiring.
- Because of Applied's leadership in frontside wiring, we have the broadest portfolio for backside wiring.
- We are also developing new technologies and Integrated Materials Solutions to efficiently deliver power to the transistors for the most complex schemes.
- For example, we have a special epitaxy system for the backside transistor contact that can be integrated with a new material called moly to reduce backside contact resistance.
- We aim to reduce transistor interface resistivity by up to 50% and metal contact resistance by up to 20%.

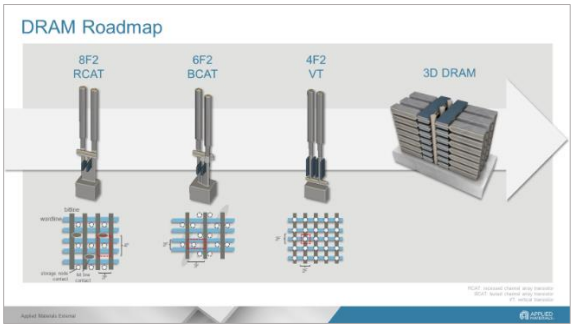


- Thank you, Mehul for giving us the roadmap insights for better transistors, frontside and backside wiring.
- I have no doubt these innovations are going to be key to more energy-efficient computing.

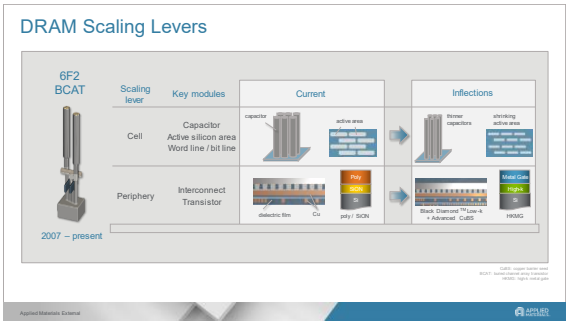
- Now, I'd like to invite Sony Varghese to join me and share our DRAM roadmap.



- Sony, we're here to talk about the roadmap for better AI. What role does DRAM play in this?
- If you look at where we are today, AI models are doubling in complexity every 3-4 months. But compute capability is only doubling every year at best.
- There are two key levers to increasing compute performance: the processor, which performs the computations on the data, and the DRAM, which keeps the data and the model close by.
- Processor performance is increasing at a rapid pace, so the main bottleneck was the memory – and that's why high-bandwidth memory DRAM is the workhorse for AI today.
- We need to focus on making better DRAM chips that can improve memory capacity, bandwidth and latency – all while lowering power consumption.

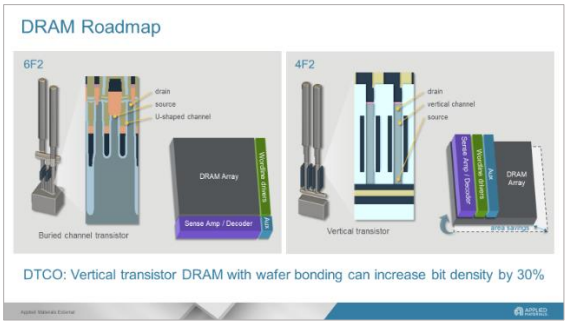


- Let's start with the basic DRAM architecture. A lot of people are interested in when planar DRAM scaling comes to a halt and forces a change to a 3D architecture, similar to what happened in NAND. Is this the next big change?
- Actually, no.
- Over time, DRAM has gone through a number of innovations.
- The architecture went from 8F squared to now 6F squared.
- This nomenclature refers to the area needed by the DRAM transistor and capacitor in relation to the smallest lithographic features.
- Interestingly, between today's 6F squared architecture and 3D DRAM, we believe there will be a new architecture called the vertical transistor DRAM.

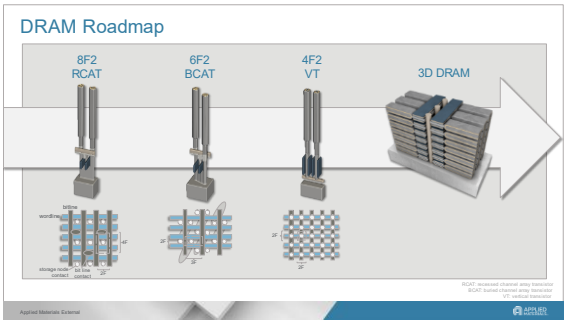


- OK, before we go there, how much further can we scale the 6F squared DRAM, and what challenges will we need to solve along the way?
- We think we can continue to scale 6F squared for 2 to 3 more generations.
- But this will create new challenges such as in wiring as we heard from Mehul.
- We expect that new materials like the new Black Diamond low-K dielectric film will need to be adopted in DRAM just as in logic.
- Also, the DRAM roadmap is following the logic roadmap in transitioning to high-K metal gate transistors.
- Faster transistors and I/O are great for DRAM latency, bandwidth and power, and all of the leading chipmakers are moving to them.

- I believe 6F squared DRAM will have a hard time scaling beyond 10nm due to lithography costs and also due to scaling-related reliability issues.

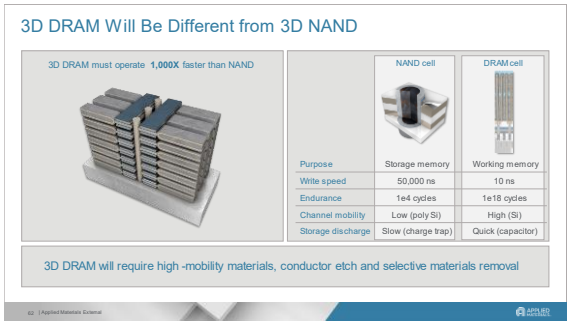


- What changes with the vertical transistor DRAM?
- The 6F squared DRAM uses a U-shaped transistor that occupies a large area.
- The vertical transistor DRAM will take up less area, so we should be able scale chip area by around 30%.
- Another big change with the vertical transistor DRAM is that companies may take the opportunity to build the chips using two wafers and performing wafer bonding.
- By building the memory cells on one wafer, and the peripheral logic on the other, companies may be able to reduce some of the patterning and integration complexity of scaling both of these elements on the same wafer.

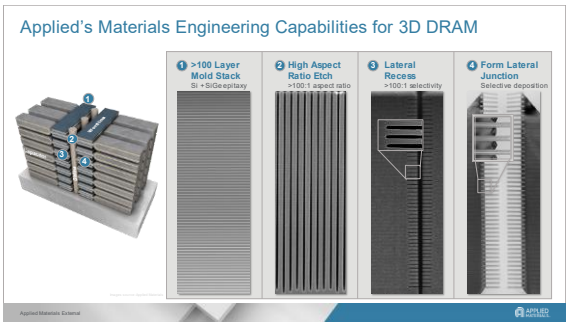


- When will the new vertical transistor DRAM reach the market?
- We expect to see it in the market within the next 4 years.
- You mentioned patterning complexity. How expensive is it to pattern the vertical transistors?
- Actually, the patterning of the vertical transistors will be more relaxed and will most likely require fewer advanced lithography steps.
- So after the vertical cell transistor DRAM comes 3D DRAM. The big question is, when?
- The sooner the better.
- And because it's a big change, we're already working on it.

- I imagine we'll see high-volume 3D DRAM production by the end of the decade, but we have to meet a number of milestones before then.

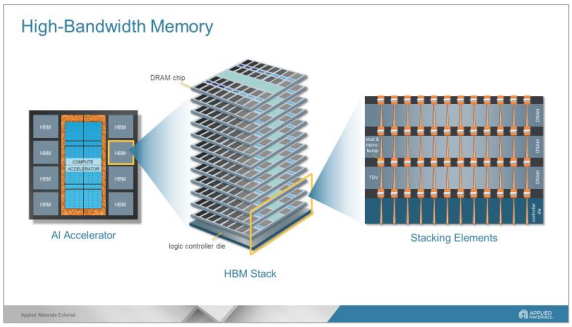


- Will 3D DRAM be a lot like 3D NAND?
- Actually, no.
- 3D DRAM will need to be around 1,000 times faster than 3D NAND.
- It will also require a different set of materials and processing steps.

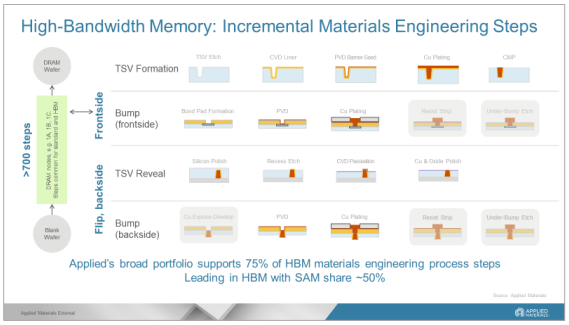


- 3D NAND uses a lot of dielectric deposition and etch.
- What kinds of equipment will be needed for 3D DRAM?
- 3D NAND used dielectric layers to form the stack, and dielectric etch is used.
- 3D DRAM uses high-mobility silicon and conductor etch.
- We are developing new epitaxy systems to deposit the silicon and silicon germanium stacks.
- We are also using tools from our conductor etch and selective materials removal portfolio to shape and remove the materials to create the memory cells.
- You mentioned that vertical cell DRAM could adopt wafer-to-wafer bonding. What about 3D DRAM?
- It would make perfect sense for companies to design the 3D DRAM memory stacks on one wafer and the peripheral logic on another and use wafer-to-wafer bonding.
- Thank you, Sony for being here with us.

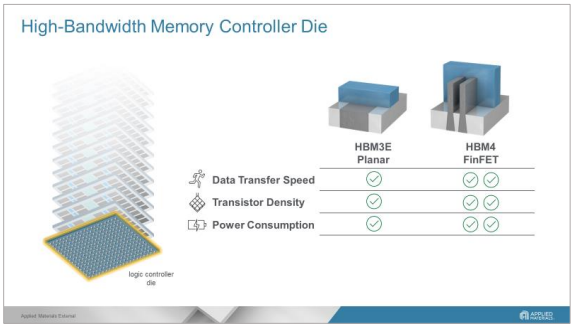
- And now I'd like to invite Jinho An to talk about combining DRAM chips to make high-bandwidth memory.



- So to begin with, Jinho, what are the differences between high-bandwidth memory DRAMs and a standard DRAMs?
- High-bandwidth memory DRAM is designed for low latency, high bandwidth and low power, which make it great for AI. If you look at its structure, it has a logic controller die on the bottom, and you can currently stack as many as 12 DRAM dies on top.
- For the DRAM die, the industry uses the fastest DRAM technology available.
- As Sony mentioned, this includes many of Applied's logic innovations, including high-k metal gate and high-performance wiring solutions.
- Within the DRAM layouts are spaces for thousands of through-silicon vias. This allows data to be transferred from every chip in the stack to the bottom of the stack where the logic controller resides.
- All this space needed for TSVs however means high-bandwidth memory DRAMs have only about half the bit density per area.

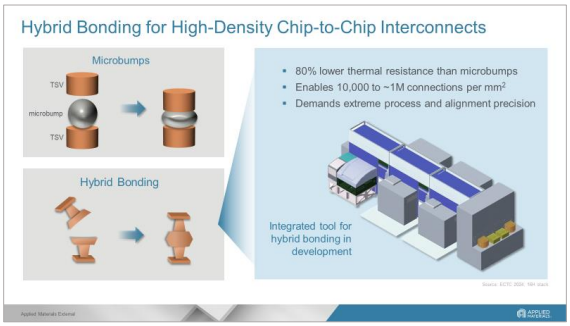


- What are the manufacturing differences?
- The high-bandwidth memory DRAMs go through the same, 700 or so process steps as regular DRAMs.
- But connecting the dies vertically adds nearly 20 more front-end materials engineering steps.
- Because of our industry leadership in wiring, Applied has products for around 75% of these steps, and we have the leading share position in high-bandwidth memory front-end equipment.
- Stacking the dies involves etching high-aspect-ratio through-silicon vias and using CVD oxide film to electrically isolate the metal from the silicon.
- PVD then deposits a barrier that prevents copper from migrating into the silicon, and a copper seed layer is deposited for electroplating the copper. Then we use CMP on the wafer to create a flat surface.
- In addition, these wafers require a frontside micro-bump and additional backside pad formation steps to create the metal connections to stack the dies.
- After dicing the wafer, the high-bandwidth memory dies are sequentially stacked on top of the logic controller die which manages the data transfer.



- How important is it to advance the logic controller at the bottom of the stack?
- It's very important.
- The logic controller is key to fast data transfer, and it can consume 40% of high-bandwidth memory power.

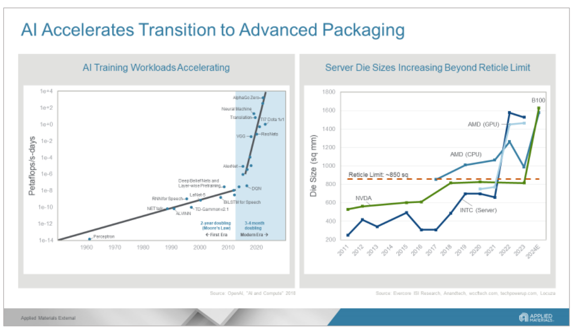
- As high-bandwidth memory scales to higher densities and improved performance, we see the logic die moving to FinFET technology. This will include the more advanced wiring technologies Mehul described earlier.



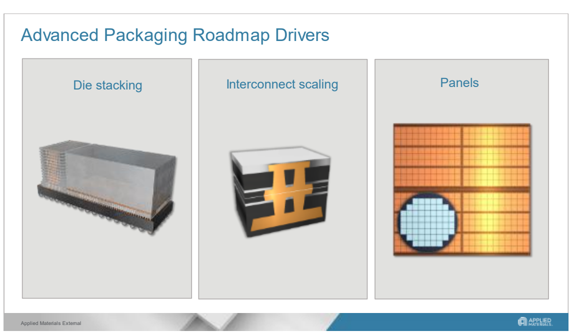
- What else do you see coming in the HBM roadmap?
- The biggest change in the roadmap is how the stacked dies are going to be connected.
- Currently, stacked dies are spaced apart and connected using the micro-bumps I mentioned earlier.
- Within the next 2-3 years, we see high-bandwidth memory adopting a more advanced technology called hybrid bonding.
- Hybrid bonding does away with the micro-bumps and instead directly connects the chips to one another.
- Why is it called hybrid bonding?
- We use the term hybrid because we are connecting both the copper pads and the surrounding dielectric materials to each other in the same process sequence.
- Compared to micro-bumps, hybrid bonding eliminates the space between the dies. This allows an increased number of dies that can be stacked in the available height. And this will help increase the overall memory capacity and speed.
- But the greatest benefit in using hybrid bonding will be the thermal performance, which could improve by 20%.
- This will have a significant impact on energy-efficient AI.
- And what is the readiness of hybrid bonding technology?
- It’s already here in advanced logic, such as in AMD’s high-performance chips for servers and even PCs.
- And all of the leading-edge foundry-logic and memory companies have hybrid bonding on their roadmaps.
- Applied Materials and our partner Besi are developing an integrated hybrid bonding tool. This has everything our customers need to take the technology to very high-volume manufacturing over the next several years. This includes high-bandwidth memory.

- With the capability to do hybrid bonding, Applied is also co-optimizing the CVD, copper electroplating and CMP materials engineering steps to improve and expedite hybrid bonding technology.
- Thank you for joining us, Jinho.

- Now I'd like to welcome Sarah Wozny to share the advanced packaging roadmap.

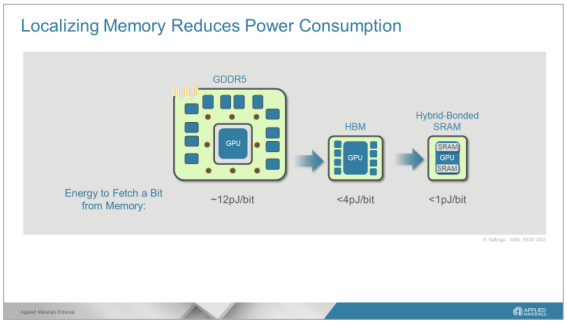


- Why is the advanced packaging roadmap just as important?
- There is a dilemma in the computing industry because on one hand, AI datasets and models have been growing exponentially, but at the same time, Moore's Law has been slowing.
- As a result, server die sizes have been growing and now exceed the reticle limit.
- The only way to get all the transistors and on-chip memory we need for AI is to shift from system-on-a-chip to system-in-a-package.

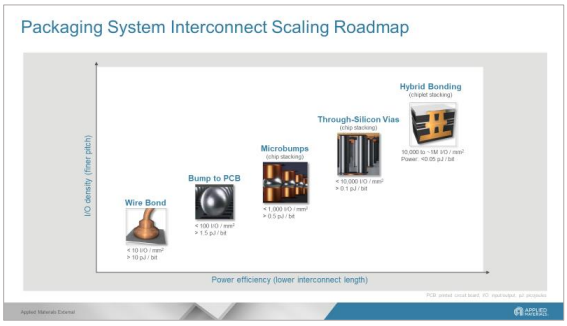


- What are the key drivers of the advanced packaging roadmap?

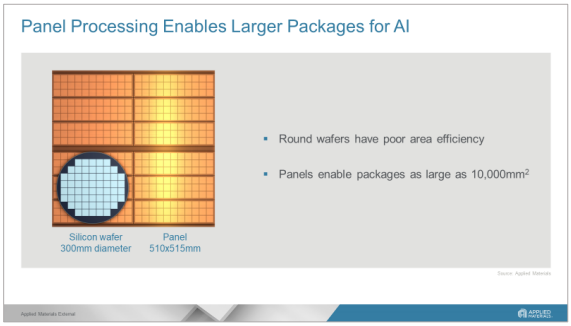
- Classic Moore's Law is about 2D scaling and getting more transistors and memory bits onto a single chip.
- Advanced packaging is 3D, volumetric scaling.
- We're trying to get as many transistors and memory bits as possible into a cubic area.
- And there are three main drivers.
- One, we need to stack more chips vertically.
- Two, we need to shrink the interconnects between the chips.
- And Three, to accommodate more chips, we need to move beyond wafer-level packaging to larger substrates called panels.



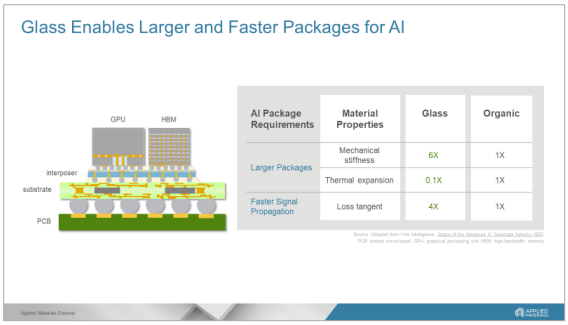
- Is higher performance the main goal of advanced packaging?
- Actually, no.
- Today, the biggest constraint is power.
- And what you'll notice is that power consumption goes up exponentially the further we have to go to fetch the data we need to process.
- Fetching data from a conventional DRAM consumes 10 to 100 times the energy of having the same data available on chip.
- Also, shifting from conventional DRAM to advanced packaging technologies like high-bandwidth memory and hybrid bonding dramatically reduce power consumption.



- What are the benefits of interconnect scaling?
- It's absolutely critical to both performance and power.
- First, interconnect scaling increases the number of I/Os per area, and this increases bandwidth which increases performance.
- The major interconnect scaling innovations shown on the slide increased I/O density by a factor of 10, and in the case of hybrid bonding, by as much as a factor of 100.
- Power also drops dramatically as we scale the interconnects, with hybrid bonding giving us a 10x improvement in energy efficiency when compared to micro-bumps.

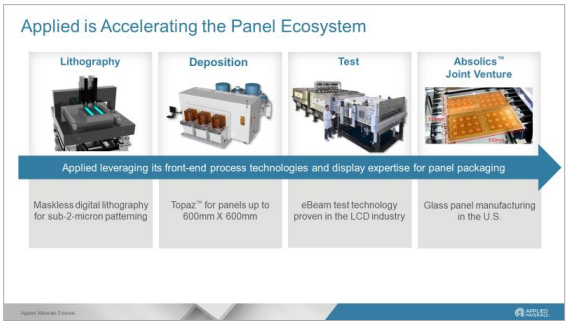


- Can you tell us why everyone is talking about panels?
- If you've been following the latest AI superchip introductions, you know that the chips are getting much bigger.
- Companies are integrating multiple, reticle-sized GPUs.
- This requires larger and larger substrates.
- Silicon is a great substrate material, but wafers are round, so you can only make a small number of rectangular AI superchips per substrate.

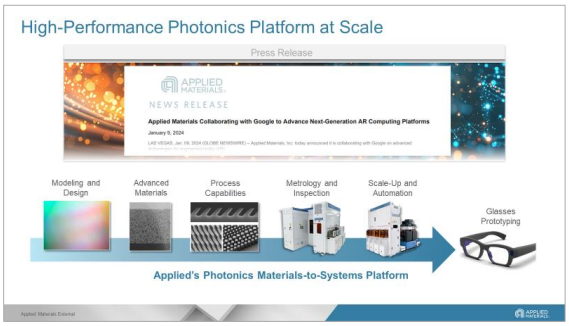


- We're also hearing about glass as a new material for panels. What are the benefits?
- Compared to organic substrates, glass is flatter and more rigid.
- It also has better thermal stability and signal insulating properties.

- Glass would give us larger AI superchips with faster signal propagation which means higher bandwidth and performance.



- Does it make sense for a company like Applied to be involved in the new panel ecosystem?
- Absolutely.
- We have decades of experience in the display industry, so we know a lot about materials engineering on large glass substrates.
- We know about glass handling, high-volume production, cleanliness and yields.
- We already have major products for panels including digital lithography, deposition systems and eBeam test.
- In addition, we entered into a joint venture called Absolics which is preparing to manufacture glass panels here in the United States, in Georgia.



- The last trend we're hearing more about, Sarah, is system interconnect moving to optical I/O. Does Applied have any plans for that?
- It's a little too early to go into details, but one interesting thing about heterogeneous integration is that you can integrate silicon photonics as well.
- In fact, Applied's CTO office has been working with companies like Google to do materials engineering on glass to develop technology for augmented reality computing platforms.
- We are also exploring how to use photonic materials engineering technologies to enable faster and lower power I/O for AI systems in the data center.
- Thank you for joining us, Sarah.

SEMICON West 2024 Technology Breakfast

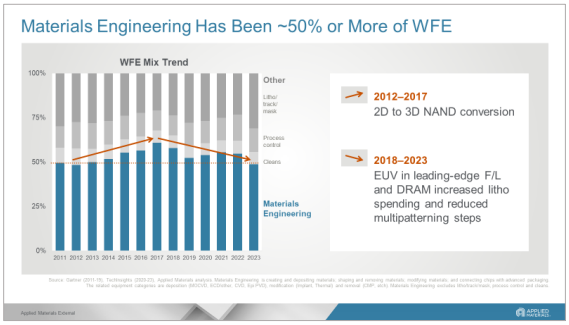
KEY POINTS | July 9, 2024

- In summary, Applied has been working on our AI-enablement strategy for a number of years.
- When Gary Dickerson spoke at our AI Design Forum in 2018, he talked about a 1,000x increase in performance per watt.
- And he introduced the need for a new playbook for PPACt.

- Today, we are aiming to help our customers generate a 10,000x improvement by 2040.
- The new playbook we described is going to drive this.
- The playbook talked about new materials: Mehul showed our new binary liner to scale copper interconnects.
- The playbook talked about 3D structures: Mehul talked about Gate-All-Around transistors, and Sony showed us new DRAM architectures.
- The new playbook promised new ways to shrink: I described DTCO, Mehul talked about backside power and Sony described DRAM vertical transistor.
- Finally, the new playbook promised advanced packaging: Jinho talked about higher density stacked DRAMs, as in high-bandwidth memory, and Sarah talked about improving system performance through heterogeneous integration: Hybrid bonding, panel packaging and our photonics platform.

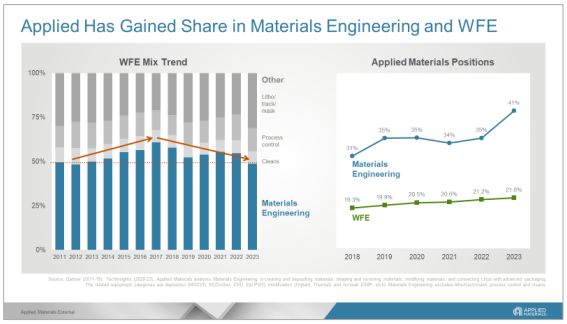
- I hope you enjoyed hearing how Applied is co-innovating the inflection roadmap with our customers and to see the progress we are all making together.
- And now I'd like to invite Prabu Raja to summarize what the new roadmap means for our business.

- Today, we talked about a race for AI leadership fueled by key architecture inflections enabled by materials engineering.
- Applied Materials fuels this race through our unique, connected portfolio.

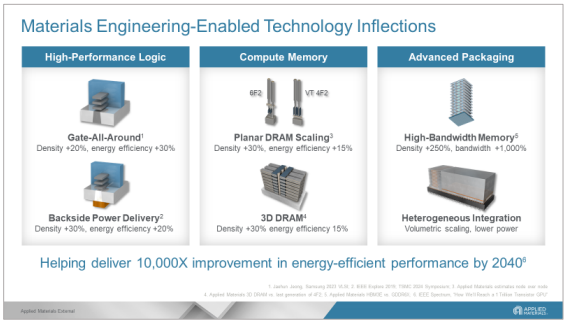


- WFE mix can be partitioned as materials engineering and other, which consists of lithography, track and mask systems, process control, and cleans.
- Over many years, materials engineering has been around half of WFE.
- Various inflections can change this mix.
- Between 2012-2017, the 3D NAND inflection increased materials engineering.

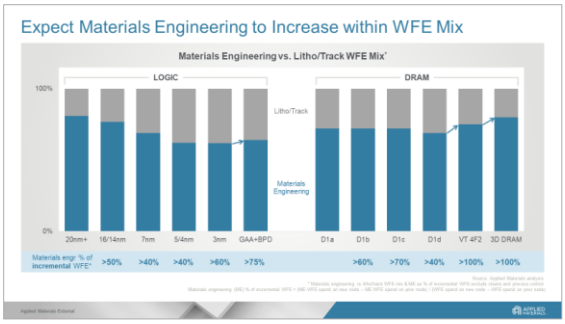
- From 2018-2023, EUV was adopted in logic and DRAM, reducing materials engineering with lower double and quad patterning.



- Despite lithography headwinds, Applied has grown WFE share.
- While materials engineering trended down, Applied gained share in materials engineering from inflections.
- We are in the early innings of these materials-driven inflections.

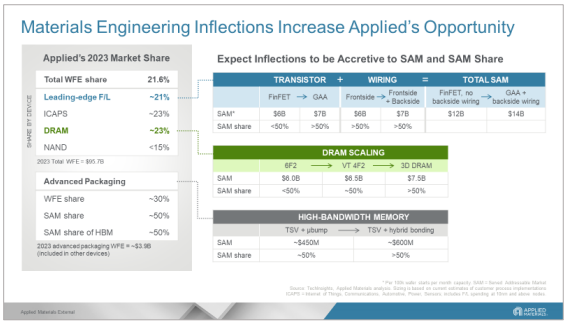


- Device architecture inflections across high-performance logic, DRAM and advanced packaging will help the industry reach the goal of a 10,000x increase in energy-efficient performance by 2040.
- Also, we believe each of these inflections will grow Applied's share opportunity in materials engineering.



- This chart shows the WFE mix of lithography and materials engineering node over node.
- The trend in logic starts to reverse, with EUV step growth slowing and materials engineering content growing. We are in the early innings.

- In DRAM, materials engineering grows more rapidly with the adoption of vertical transistor DRAM and 3D DRAM.
- The row on the bottom shows the materials engineering percentage of incremental WFE spending. Note that materials engineering has a greater proportion.



- We believe these inflections will increase materials engineering as a percentage of WFE, and this will increase Applied's opportunity.
- With Gate-All-Around and backside power, we expect our SAM opportunity to increase from $12 billion to $14 billion for 100,000 wafer starts per month of capacity, and we are gaining share.
- We expect our DRAM SAM opportunity to grow from $6 billion to $6.5 billion for vertical transistor DRAM and by another $1 billion for 3D DRAM.
- In high-bandwidth memory, as hybrid bonding is adopted, we expect our SAM opportunity to grow from $450 million to $600 million.
- We expect our SAM share in all of these applications to be at or above 50%.

- I'll conclude by connecting this to our company's growth thesis.
- We believe semiconductors will significantly outgrow GDP.
- Because of increasing complexity, WFE will grow as fast or faster than semiconductors, with materials engineering gaining share within the mix.
- With our unique, connected materials engineering portfolio, Applied's Semiconductor Systems business will outgrow WFE.

- And, with better connected technologies we can grow our services business as fast or faster than our equipment business.

- Thank you for joining us.