Breaking Analysis: Arm Lays Down the Gauntlet at Intel’s Feet  by, David Vellante

April 3rd, 2021

Contributing Author: David Floyer

Exactly one week after Pat Gelsinger unveiled plans to reinvent Intel, Arm announced version 9 of its architecture and put forth its vision for the next decade. We believe Arm’s direction is strong and compelling as it combines an end-to-end capability, from edge to cloud to the datacenter to the home and [...]
Contributing Author: David Floyer

Exactly one week after Pat Gelsinger unveiled plans to reinvent Intel, Arm announced version 9 of its architecture and put forth its vision for the next decade. We believe Arm’s direction is strong and compelling as it combines an end-to-end capability, from edge to cloud to the datacenter to the home and everything in between. Moreover it doubles down on Arm’s model of enabling ecosystem partners to add significant value while at the same time maintaining software compatibility with previous generations. We see this as extremely important because the variety of use cases requiring specialized silicon is rapidly expanding in the marketplace, and the Arm architecture is by far in our view the best positioned to capitalize on this coming wave.

In this Breaking Analysis we’ll explain why we think this announcement is so important and what it means for Intel and the broader technology landscape. We’ll also share with you some feedback we received from the CUBE community on last week’s episode and a little inside baseball on how Intel, IBM, Samsung, TSMC and the U.S. government might be thinking about the shifting landscape of semiconductor technology.

Competitive Responses to Intel’s IDM 2.0

Intel’s announcement comprised three broad pillars: 1) sticking to its internal factory network strategy; 2) increasing its use of outside foundries and 3) launching a new quasi-independent foundry business unit.

This past week there were two notable announcements directly related to Intel’s new direction:

<table>
<thead>
<tr>
<th>Arm (and TSMC) Respond to Intel</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Arm unveils Armv9 architecture" /></td>
</tr>
</tbody>
</table>

The Armv9 news and TSMC’s plans to invest $100B in chip manufacturing and development over the next three years. That’s a big number and appears to trump Intel’s planned $20B investment to launch two new fabs in the U.S. starting in 2024. You may remember back in 2019, Samsung pledged to invest $116B to diversify its production beyond memory chips.

Why the Rush to Spend and Will it Cause an Oversupply?

Why are all these companies getting so aggressive and won’t this cause a glut in chips? First, China looms
large and aims to dominate its local markets, which in turn will confer advantages globally. Second, there’s a huge chip shortage right now and the belief is that it will continue through the decade and possibly beyond. We are seeing a new inflection point in demand as we discussed last week, stemming from digital, IoT, cloud, autos and new use cases in the home.

As to the glut, these manufacturers believe that demand will outstrip supply indefinitely and they understand that a lack of manufacturing capacity is more injurious than an oversupply. If there’s overcapacity in the market, manufacturers can cut production and take the financial hit. Whereas capacity constraints mean you can miss entire cycles of growth, complete miss markets and the resulting benefits of volume learning curves and cost reductions. So they’re all going for it, hard.

What Arm Announced & why it’s so Compelling

A key element of Intel’s new strategy was Pat Gelsinger’s vision of system on package. This is an attempt to leapfrog system on chip (SoC). In a bold architectural move, Arm is taking a similar systems approach. However in our view it’s even broader than the vision laid out by Pat. Arm is targeting a wide variety of use cases as shown below.

---

Arm’s fundamental philosophy is that the future will require highly specialized chips. And Intel as you recall from Pat’s announcement would agree. The difference is that Arm has been executing on a strategy that allows for flexible customization and value add by ecosystem partners for decades. Intel is just now finally acting to take advantage of the future – but the future is here now and Intel won’t be ready for years.

Comparing the Arm & Intel Models

Arm historically has taken an ecosystem approach that is far different from Intel’s model. Arm is all about enabling the production of specialized chips to fit the application. For example – think about the amount of AI going on in iPhones moving from fingerprint to face recognition. This required specialized neural processing units (NPUs) that are designed by Apple for the specific use case. Arm is facilitating the creation of these specialized chips to be designed and produced by the ecosystem.
Intel on the other hand has historically taken a one size fits all approach, built around the x86. Intel’s design has always been about improving the processor (e.g. in terms of speed, density, adding vector processing to accommodate AI, etc.). And Intel does all the design and manufacturing and any specialization required by ecosystem. Intel has been a value sucking machine, putting as much function into its chips as it saw profitable, leaving less value add for ecosystem partners.

For example, think about PC’s and servers, Intel’s biggest markets. Much of the value add from customers like Dell, Lenovo, HP and HPE has frankly been bending metal or adding displays or other features at the margin. It’s true that storage and networking vendors have been able to add more value but those markets are much smaller than PCs and servers. For the most part, Intel has been the big winner in hardware and left its OEM customers to live in a world of relentless consolidation and margin pressure.

Nonetheless, the big advantage Intel and its ecosystem has enjoyed is that the x86 architecture is well understood, reliable and most enterprise software runs on x86. So this has been particularly lucrative for software companies like Microsoft and VMware who realize most of their value add writing code that works extremely well on x86. And having a consistent, stable processor platform is a bonus that has supported massive growth in software markets.

So as you can see, Arm and Intel have historically had very different models, which we heard Gelsinger say last week they were going to change with their new trusted foundry strategy.

The problem is that the vast majority of Intel’s operations are serving a business model that is losing momentum at the macro level.

Arm’s Approach is Well-Suited for the Future

Let’s go through an example that will help further explain the power of Arm’s model.

Take AWS for example, developing its own Graviton chips. Or think about Apple designing the M1 chip. Or Tesla designing its own chip. Or any other company in one of these use cases shown in the graphic above.

The Tesla case study is instructive. In order to optimize for video processing Tesla needed to add firmware in the NPU for its specific use case that involved processing video. It was happy to take off-the-shelf CPU or GPU and components of the Arm architecture -as is- and leverage Arm’s standards. But Tesla saw an opportunity to add its own value in the NPU. The advantage of this model is Tesla could get to tape out in less than a year, versus taking many years with traditional models.

Arm is like customizable lego blocks that enable unique value add by ecosystem partners with much faster time to market. Tesla goes from tape out (logical tape out if you will) to Samsung and says ‘ok run this against your manufacturing process’ and it should all work as advertised. Tesla interestingly chose the 14nm process to keep costs down – it didn’t need the latest and greatest density. It was all about the flexibility to fit the application and time to market.

Based on Pat’s announcement, Intel is vectoring toward the Arm model for its foundry business. It has to do so. But as we pointed out last week, foundry is a very different business with different margin models, relationships, go to market; and very importantly, volume requirements.

The Armv9

What’s so special about Armv9?
Armv9 is backwards compatible with previous generations. Perhaps Arm learned from Intel’s failed Itanium effort where it had no backward compatibility and floundered. As well, Arm adds some additional capabilities. In this post we’re going to focus on the two areas highlighted in red, the machine learning piece and security.

Take note of the call out—300 billion chips. That’s a lot and we’ve said Arm wafer volumes are 10X those of x86. Volume = cost reduction and is critical to the foundry business. We’ll come back to that later.

**Betting on AI/ML: The Trend is Your Friend**

A key component of the Armv9 architecture is its Scalable Vector Extension 2 (SVE2), designed to deliver more granular data parallelism and enable more work per instruction. This will support increasingly robust machine learning applications across a variety of use cases such as language processing, visualization, dynamic pricing, fraud detection, etc.

Arm is betting on AI/ML as are many others and the chart below shows why.
The above graphic shows ETR spending data from more than 1,200 respondents. Note that ML/AI gets the top spot on the vertical axis which is Net Score – a measure of spending velocity. The horizontal axis is Market Share or presence in the data set. We give this sector four stars to signify its consistent lead in the survey data each quarter.

So improving on specific machine learning functionality in its architecture is a pretty safe bet by Arm.

**Security is Front and Center**

At its Vision Day, Arm talked about Confidential Compute Architecture & Realms.
Arm Confidential Compute & the Concept of Realms

Note the left hand side showing data traveling all over and different use cases. And the call out from the CISO at a large public airline in an ETR Venn roundtable who noted that the shifting endpoints increase the threat vectors.

Arm said something at its announcement that really resonated. Paraphrasing Arm’s commentary...Today there is too much trust on the OS and hypervisor that run applications. And their broad access to data creates exposures and threats from hackers. Arm’s concept of Realms as shown in the right hand side of the chart underscores the company’s strategy to remove the assumption that privileged software like the hypervisor needs to be able to see the data. So by creating realms in a virtualized, multi-tenant environment, data can be more protected from memory leaks – a major opportunity for hackers.

This is an elegant concept and a way for the system to isolate a tenant’s data from other users.

From the Community - Intel Didn’t Miss Mobile, it Missed Smartphones

We want to share some feedback from our last post summarizing our thoughts on Intel.
A tech exec from Citi pointed out that Intel really didn't miss mobile – it missed smartphones. In fact, while that's a minor distinction it's important to recognize. Intel facilitated WiFi with Centrino under the direction of Paul Otellini – who was not an engineer, he was a marketing person by background. Ironically, Intel's work in WiFi connectivity, designed to make laptops more useful leveraging hotspots in airports and coffee shops, enabled the smartphone revolution. Maybe that makes the smartphone miss by Intel all that more egregious because their heads were buried so deep into PCs that it couldn't see the future.

The other piece of feedback we received related to our IBM scenario and our 3-way joint venture prediction bringing together Intel, IBM and Samsung.

The Role of IBM, Samsung, GlobalFoundries & the U.S. Government

Remember we said that IBM with Power 10, has the best technology in terms of disaggregating compute from memory and sharing memory in a pool across different processor types. We'll come back to that in a moment.

IBM, when it restructured its microelectronics business under Ginni Rometty, catalyzed a partnership with GlobalFoundries which took over IBM’s East Fishkill, NY semiconductor manufacturing plant. We believe this exit from a vertically integrated IBM was actually architected by Arvind Krishna. GlobalFoundries by the way, announced plans to sell off this plant in 2019– the plan just didn't pan out.

For context, GlobalFoundies was created in 2009 when AMD got out of the business of manufacturing its own chips. In 2009 it broke ground on a new fab in Malta, NY, outside of Albany. Malta would be where GF would produce some of its most advanced technologies, including eventually 7nm products.

The inset picture above shows the GF facility in Malta. AMD and IBM were committed to have their 7nm chips built by GF, but GlobalFoundries backed away from 7nm forcing IBM and AMD to look elsewhere for the latest and greatest technology. For a decade or more, Samsung, IBM and GF have been collaborating on a common design spec for advanced technologies.

But GF was losing money and its private backers were likely eager to recoup some of the $21B invested in fabs over the last several years. GlobalFoundries’ CEO Tom Caulfield has signaled an impending IPO. This
all has caused a bit of a rift between the companies and leaves a hole with respect to the original vision of the Malta factory. It also leaves the United States in a weaker position relative to China.

Intel, IBM and the U.S. Government

Intel for its part has great FinFet technology. FinFet is tech that goes beyond CMOS – you old mainframes will remember when IBM burned the boat on ECL and moved to CMOS...well this is the next generation beyond and could give Intel a leg up on AMD’s Chiplet IP, especially as it relates to latency. And there could be some benefits for IBM.

Now while Arvind was the face of the Intel announcement and clearly has deep knowledge of IBM’s semiconductor strategy, Dario Gil we think is a key player in the mix. He’s the SVP and Director of IBM Research, the part of IBM that was mentioned extensively in Intel’s announcement, and is in a position to affect some knowledge sharing with Intel – possibly as it relates to disaggregated architectures.

Why would IBM do that? Well it wants to compete more effectively with VMware who has done a great job leveraging x86..and that’s the biggest competitor to OpenShift. So Arvind needs Intel chips to execute on its cloud strategy because almost all of IBM customers are running apps on x86.

Where it gets really interesting is NY Senator Chuck Schumer is keen on building up an alternative to Silicon Valley in NY – known as Silicon Alley.

So is it possible that Intel, who has great process technology, with the U.S. government and IBM and Samsung could make a play for the NY foundry as part of its trusted foundry strategy and reshuffle the deck in Albany?

Sounds like Game of Thrones.

Volume Remains the Linchpin for Foundry

There’s one more piece to this puzzle. TSMC has been so consumed servicing Apple for 5nm and eventually 4nm that it has dropped the ball on some of its other customers, namely Nvidia. Remember, long term competitiveness and cost reductions come down to volume. And we think Intel can’t get to volume without Arm.

Maybe the JV strategy we laid out is a stretch and Samsung may not be willing to play ball given its huge investments in South Korea – although we think that’s still a viable scenario. But if Intel can actually become a trusted foundry with the help of IBM and the U.S. government, maybe it can compete on volume by attracting some disaffected TSMC customers.

How would that work? Well let’s say that Nvidia isn’t too happy with TSMC because it has to get in line behind Apple. Why wouldn’t Nvidia entertain Intel’s foundry as a second source. What about AWS, Google and Facebook. Maybe this is a way to placate the U.S. government and call off the anti-trust dogs. Give Intel foundry our business to secure America’s semiconductor leadership and perhaps the U.S. government is less contentious. And Microsoft, even though it’s not getting as much anti-trust scrutiny, would likely be happy to play ball as well.

And why wouldn’t Apple want to have an on shore source of advanced technology supply. Gelsinger is just the person to pull something like this off as he’ll do well with the public policy folks.

Would this give Intel a competitive volume posture? Yes it would. For sure if it can gain the trust of these companies the volume would be there.

But as we’ve said – currently this is a long shot and will take years to play out. But the dots are connecting in this scenario and the stakes are exceedingly high.
Intel is in a Fight for its Life…and a Nation’s Competitiveness

Make no mistake, the stakes are higher than they’ve ever been for Intel because it transcends survival for a company. We’re talking the survival of an industry and a nation’s competitiveness.

The Bottom Line - Intel is in a Fight for its Life

- China, China, China...
- Volume is the name of the game
- Arm (& TSMC) have responded to Intel’s moves
- Arm going beyond SoC
- Intel must (and will) secure the high end
- Puts immense pressure on Intel to be the trusted foundry
- U.S. government remains a key - Too strategic to fail

As we emphasized with the Clay Christensen steel industry example, the threat from China can not be overstated. And a key to long term competitiveness absolutely lies in volume manufacturing.

Both Arm and TSMC have responded to Intel’s moves. Nvidia’s acquisition of Arm remains precarious as regulators and competitors object to the union, however if Nvidia pulls it off Arm gets even stronger.

Arm announced its intentions to go beyond SoC, keeping pace with Intel, albeit with a purely outsourced strategy. Intel must, and we think will, maintain its high end position in the market but Arm will continue to gain footholds in the enterprise.

Intel’s big differentiator will be its IDM strategy. Vertical integration is a long game and brings both quality and cost advantages. But without volume, Intel will not be able to compete on cost and could literally go bankrupt trying.

As such this puts major pressure on Intel to successfully execute on its foundry strategy, which remains a high risk/high return venture. The role of the U.S. government will be to reduce that risk and increase the probability of Intel’s survival and the continued leadership of the country’s semiconductor industry.

Keep in Touch

Remember these episodes are all available as podcasts wherever you listen.
Email david.vellante@siliconangle.com | DM @dvellante on Twitter | Comment on our LinkedIn posts.
Also, check out this ETR Tutorial we created, which explains the spending methodology in more detail.
Watch the full video analysis:

Image credit: Suttipun

Note: ETR is a separate company from Wikibon/SiliconANGLE. If you would like to cite or republish any of the company’s data, or inquire about its services, please contact ETR at legal@etr.ai.
David Vellante is co-CEO of SiliconANGLE Media, as well as co-founder and Chief Analyst at The Wikibon Project, the world’s leading open source technology research community. Dave is a long-time tech industry analyst, entrepreneur, writer and speaker. As co-host of theCUBE – “The ESPN of Tech,” Vellante has interviewed over 5,000 experts since 2010. He is also a co-founder of CrowdChat, an angel funded startup based in Palo Alto using big data techniques to extract business value from social data. Prior to these exploits, Dave founded a CIO consultancy and spent a decade growing and managing IDC’s largest business unit. He lives in Massachusetts with his wife and four children where he is active in town activities including serving as the president of his town’s local “Kiddie Sports” association. Dave holds a B.S. in Applied Mathematics from Union College.