Intel Unveils More 10nm Details

By Dick James

On March 28, Intel held a Technology and Manufacturing Day in San Francisco, not surprisingly focusing on the work of the Technology and Manufacturing Group (TMG) within the company. This event was an exposition of the 10nm process, a new 22FFL ultra-low-power process, a quick mention of EMIB packaging, a plug for the enhanced 14nm technologies, some more marketing of Intel foundry, and all within the context of "Moore's Law is alive and well at Intel!"

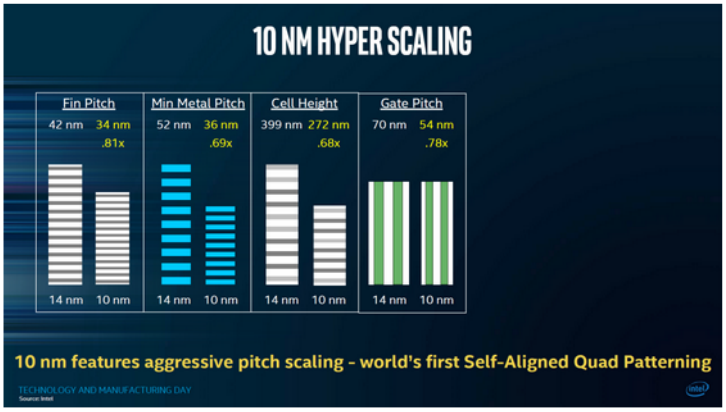
Agenda

Title	Speaker
Welcome	Laura Anderson
Strategy Overview	Stacy Smith
Moore's Law	Mark Bohr
14nm Leadership	Ruth Brain
10nm Leadership	Kaizad Mistry
22FFL Technology	Mark Bohr
IDM Advantage	Murthy Renduchintala
Q&A	Various Speakers
Custom Foundry Panel	Various Speakers
Investor Roundtable	Various Speakers

Source: Intel

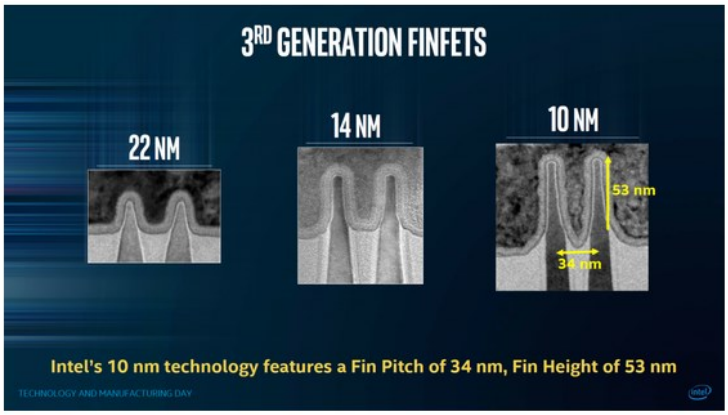
Getting straight to the 10-nm presentation, which was the fourth of the day, Kaizad Mistry unveiled some of the mysteries:

He put the numbers up remarkably quickly:



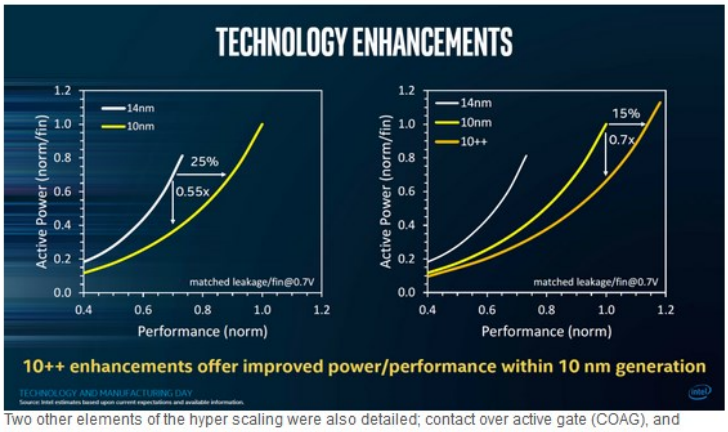
We had speculated in the last blog that self-aligned quadruple patterning (SAQP) would be used for the fins, and it would be "really ambitious" for metal definition; and Intel have surprised us by being really ambitious! Going below 40nm takes us into the realm of SAQP, or alternatively LELELE (litho-etch, litho-etch, litho-etch), adding to the complexity and cost of the process; Intel obviously considers the extra bump in density worth the cost. It also likely sets them up for the next node; take the pain now, and hopefully reduce the time to 7nm!

The gate pitch was announced at IDF as 54nm, and we now know the fin pitch is 34nm, the metal pitch is 36nm, and the cell height is 272nm. Taller fins were also mentioned, and indeed we have that, with an increase from 42 to 53nm:



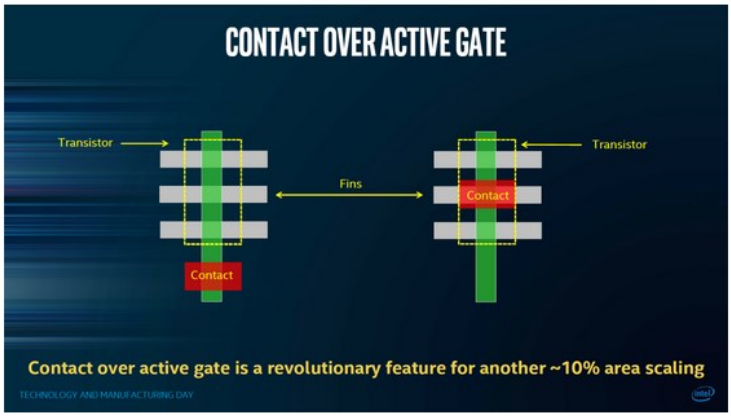
Putting a ruler on the fins, we come up with a fin width of 5 – 15nm, and ~7nm at half height. Gate width is ~110nm, compared with the ~85nm of the previous generation. Gate length is still an unknown, but we can speculate that it will be in the 18 – 20nm range, assuming dielectric thickness of ~8nm between the gate and contacts.

Kaizad claimed a 25% performance improvement, with another 15% to come from the 10++ version in a couple of years, and corresponding power reduction to 0.55x, and 0.7x for the 10++ sub-generation.

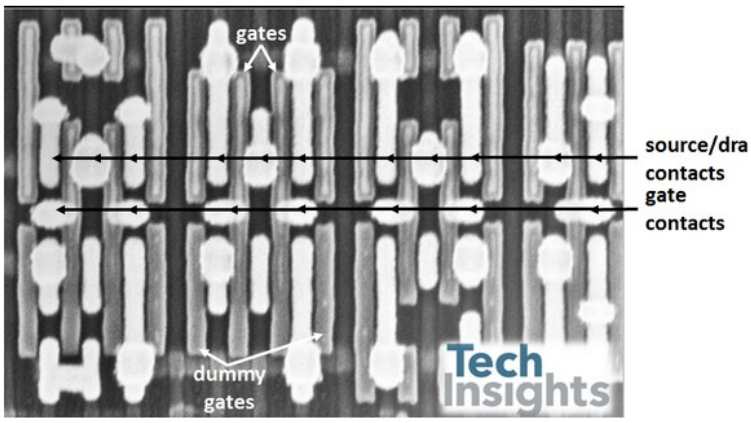


Two other elements of the hyper scaling were also detailed; contact over active gate (COAG), and single dummy gate (SDG), which in total are claimed to add another 30% transistor density improvement.

COAG means moving the contact from a position away from the fins to directly over the active part of the gate:

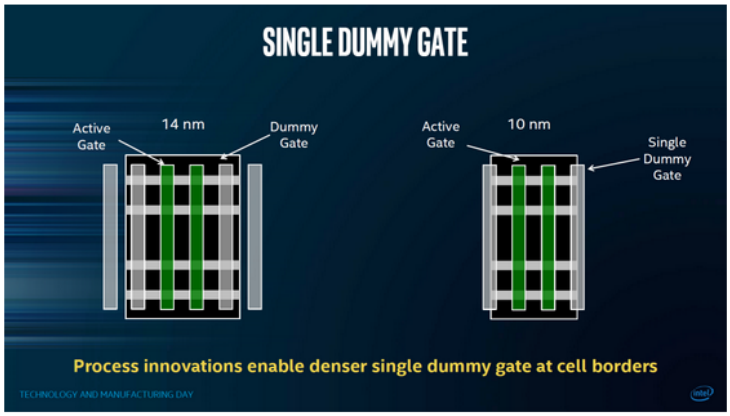


We can perhaps see the concept a bit more clearly in this (somewhat fuzzy) image of 14nm transistors:

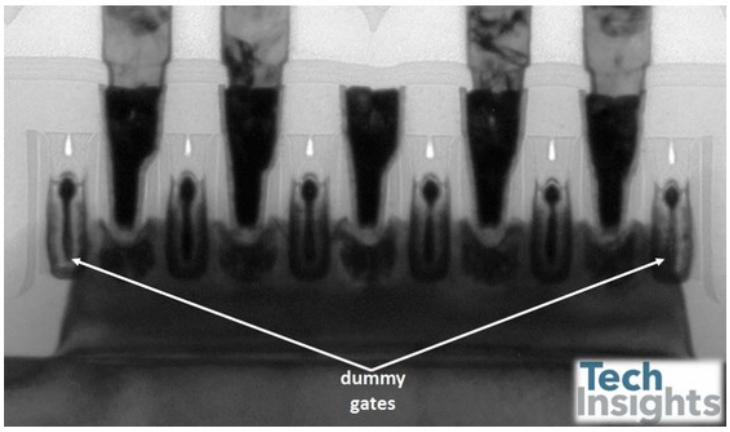


We can see here that if we can somehow squeeze the gate contacts in between the source/drain contacts, then we don't need the vertical space for them, and we can shrink the cell height; the vertical distance between the gates is reduced to the gate tip/tip spacing. However, this strikes me as quite a challenge, likely requiring at least self-aligned gate contacts, and it's no wonder that Kaizad commented that "there are a number of technology attributes and innovations that we needed to introduce to allow the contact to be placed directly above the active transistor." This extra complexity may mean that the cell architecture could have also changed, and the routing metal may not be at the minimum pitch; the minimum metal may start at the M3 level.

For the single dummy gate, he also said that Intel "introduced unique innovations to overcome the difficulties of single dummy gate," to ensure that the performance of centre transistors and the edge transistors are closely matched.



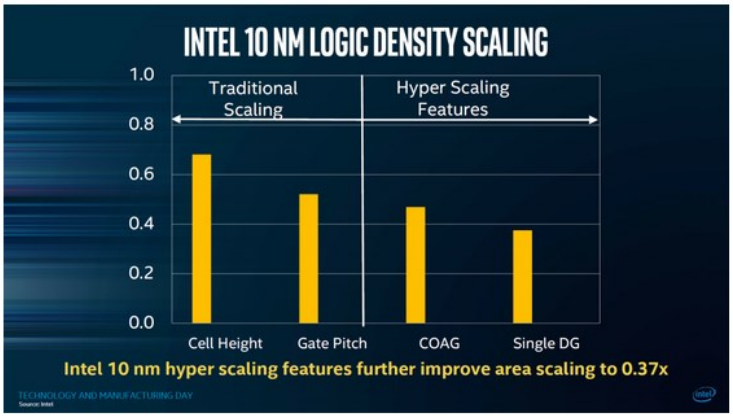
The dummy gates are shown in the plan-view image above, but what can't be seen is that he dummy gates usually overlap the end of the fin, unlike the left-hand schematic above. Here's a shot of 22nm PMOS gates on a fin:



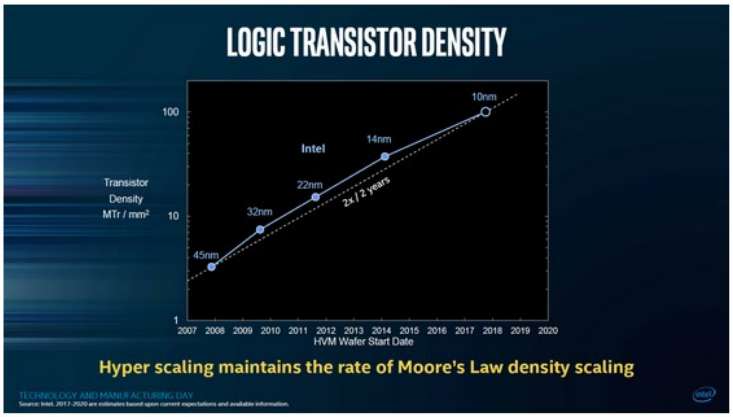
This clearly adds to the width of the cell, but it has the advantage of matching the performance of all the transistors on the fin. If we have a single dummy gate on the STI between the fins, as in the Samsung 14nm and earlier 20nm devices, then the source/drain cavity etch and the contacts are on or near the ends of the fin. They are then subject to overlay errors, increasing variation in both the contact resistance and the source/drain epi growth.

The simplest structure to get rid of this problem would be to do without the STI between fins, making the dummy gate some sort of isolation structure, but presumably that would mean having contacts to those gates to keep them tied to the correct potential – adding more complexity to the structure and design rules. We'll see what those "unique innovations" are when the part comes out!

In total these changes make for a shrink beyond the usual 50% to 37% of the 14nm technology:

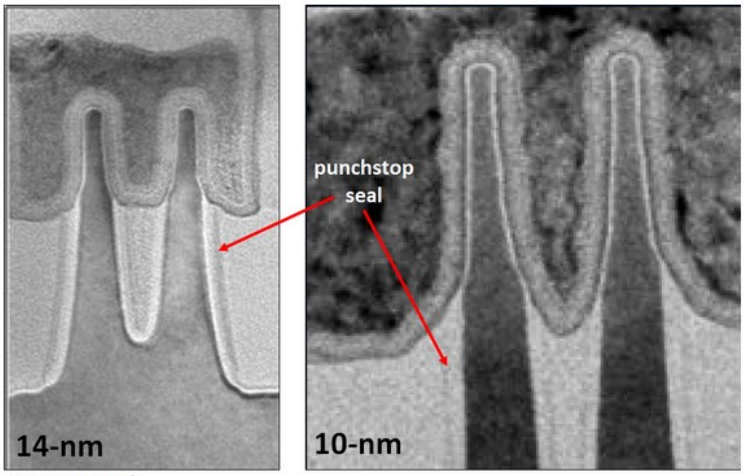


Intel claims that this hyper-shrink actually brings them back on to a two-year cadence from the 45nm node, assuming high-volume production as of the second half of this year.



The SRAM cells are scaled by a factor of ~0.6, so that the low-voltage 1:2:1 (fins in Pull-Up:Pass-Gate:Pull-Down transistors) cell goes from ~0.059 μm^2 to ~0.037 μm^2, and the high-density 1:1:1 cell shrinks from ~0.050 μm^2 to ~0.031 μm^2. (The TSMC and GF/IBM/Samsung 7-nm cells announced at IEDM, presumably 1:1:1 cells, were 0.027 μm^2.)

If we look at the transistor image, there are features in common with the 14-nm device. Comparing at the two cross-sections, it appears that the solid-source punch-stop diffusions introduced at 14-nm are present, since we can see the seal layer(s).



Source: Intel

Looking at the gate stacks, we do not seem to have any significant change, so we now have the fifth generation of Intel's HKMG technology, and of course it's their third-generation finFET.

Things are not getting any easier, not the least getting the gate and contact materials into ever smaller spaces. With a smaller fin pitch the implant angle needed for doping is also shrinking; I measured it as less than 30°, compared with the 52° and 41° of the 22- and 14nm processes, but I am told that if the implant has a twist (i.e. angled with respect to the fin orientation), then it is till feasible to get implants into the right location.

Taller fins with higher aspect ratios will have their own mechanical challenges, implying tighter stress control to avoid fin bending and distortion, which was quite noticeable in some samples of the 14nm product.

We could go on detailing more problems, but suffice it to say that I don't have much sympathy with some of the media criticism that I see of the slow-down in process generations. It wasn't easy in the days of Grove, Kilby, and Moore, given the technologies of the time, and now that we are counting atoms it certainly isn't any easier – don't forget that a 7nm fin is actually less than 25 atoms wide!