FormFactor™

Probe in the Spotlight

Enabling advanced packaging, chiplets, and heterogenous integration

Mike Slessor

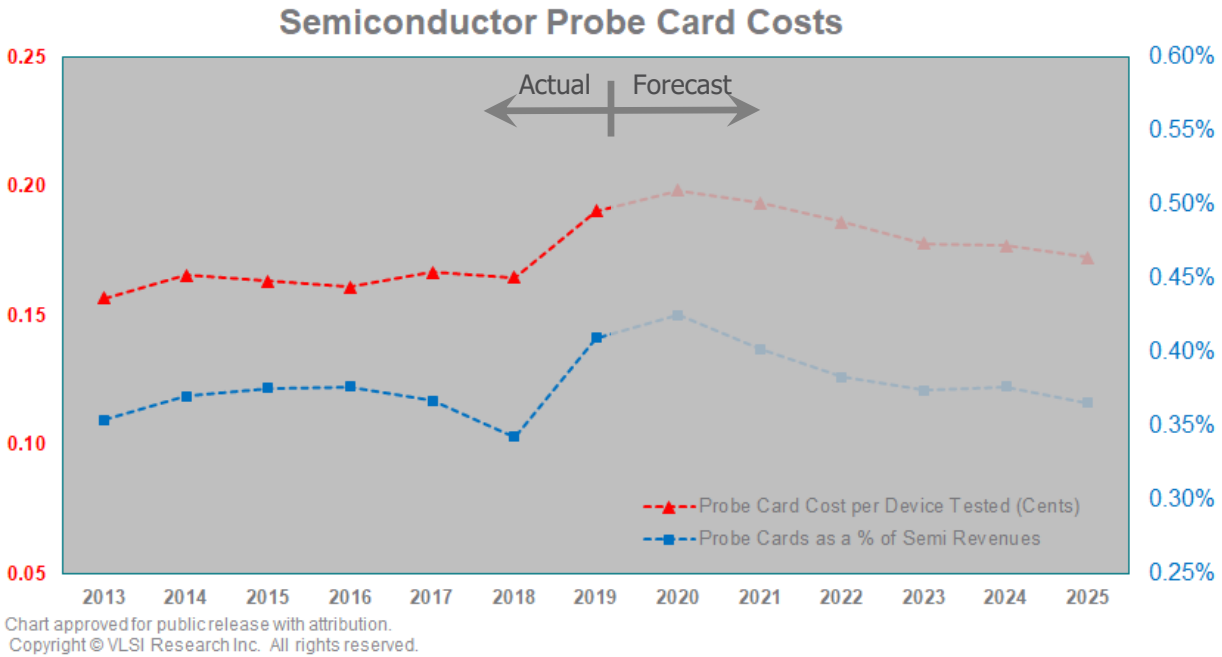
President and CEO, FormFactor

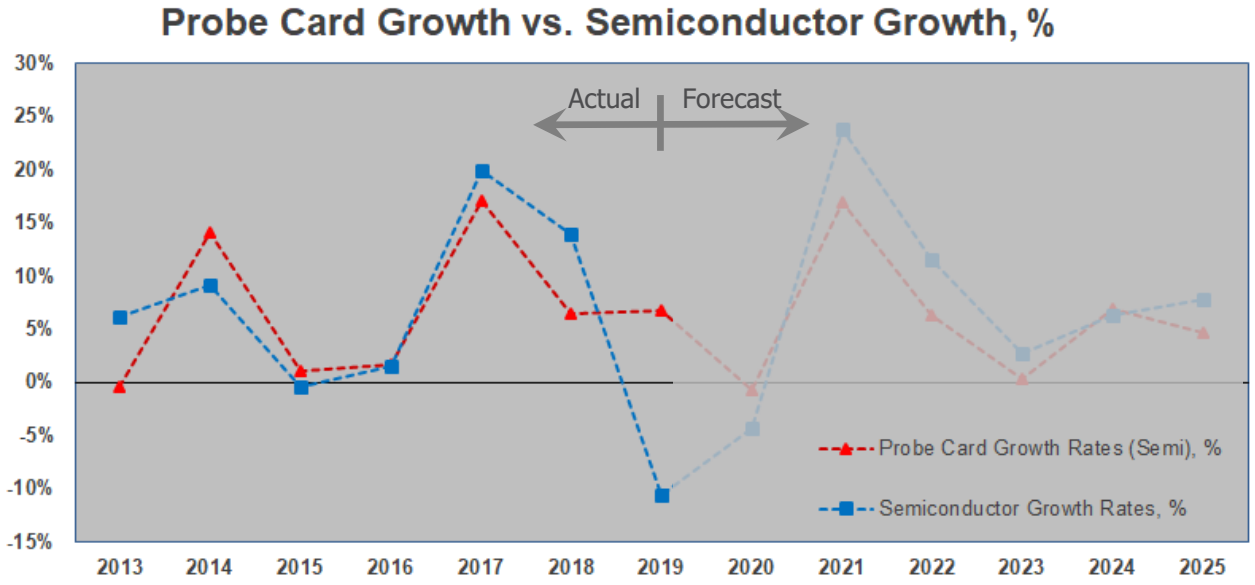
A Roadmap for Our Time Together

- An observation about the probe card market
  - It's growing faster than historical macro correlations would suggest
- A problem statement (and solution) for the semiconductor industry
  - The historical cost and performance driver is slowing/sputtering/dead
  - Advanced packaging offers a new path forward
- Are the observation and the problem statement (solution) connected?
  - Yes, please refer to the title of this talk
- What does probe need to do to enable the shift to advanced packaging?
  - Technically, and (more importantly) economically
- Q&A

An Observation: Something is Going On in the Probe Market

From VLSI Research's "The Probe Card Report 2020"*





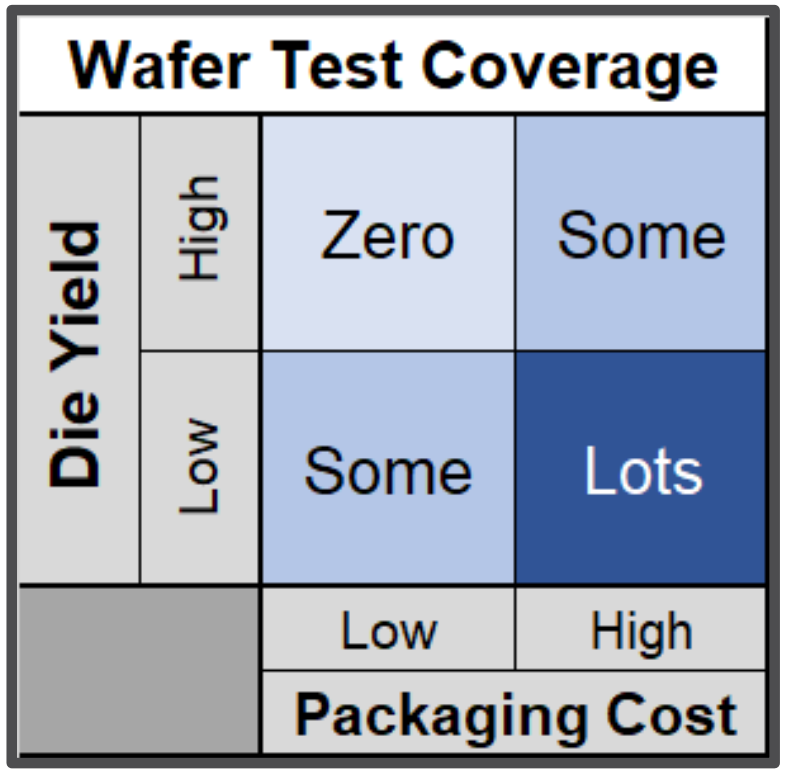
- In 2019, probe-card spending diverged from historical correlations
  - Intensity of probe-card spending grew on both a unit and dollar basis (chart on left)
  - Probe card market grew by 7%, while semiconductors overall shrank by 11% (chart on right)
- In 2020, early evidence suggests continued divergence
  - FormFactor 2020 YTD actuals plus Q4 guidance (10/28 earnings call) midpoint implies ~15% growth in '20

*Reproduced with permission of John West, VLSI Research

FormFactor™

Why Does Anyone Spend Any Money On Probe?

- Inline test (eg, probe) reduces overall chip manufacturing cost
  - Generally true for any inline measurement/inspection
  - There are exceptions, but relatively minor
    - Inform an adjustment/trim/change, eg, DRAM redundancy
    - Outgoing QC for product title transfer, eg, bare-die sales
- Easy to assess value in single-die integration schemes
  - Need (Cost of Probe) < (Cost of Packaging Bad Die)
  - If yield low and packaging cost high – probe is a good idea
    - Example: large flip-chip die fabbed on leading-edge logic nodes
      - Apps processors, high-performance compute, etc
  - If yield high and packaging cost low – probe is a bad idea
    - Example: small wire-bond die fabbed on mature nodes
      - Analog jellybeans, MOSFETs, etc



- Assessment for multi-die integration schemes gets more complex (and interesting)
  - Advanced packaging, heterogenous integration, and chiplets

Shifting Gears: A Problem Statement (and Potential Solution Path)

Moore's Law (at least the front-end transistor-reliant version) is slowing, or perhaps even worse…

Source: Su (AMD), IEDM 2017

COMPONENTS

CES 2019: Moore's Law is dead, says Nvidia's CEO

The long-held notion that the processing power of computers increases exponentially every couple of years has hit its limit, according to Jensen Huang.

BY SHARA TIBKEN | JANUARY 9, 2019 11:46 AM PST

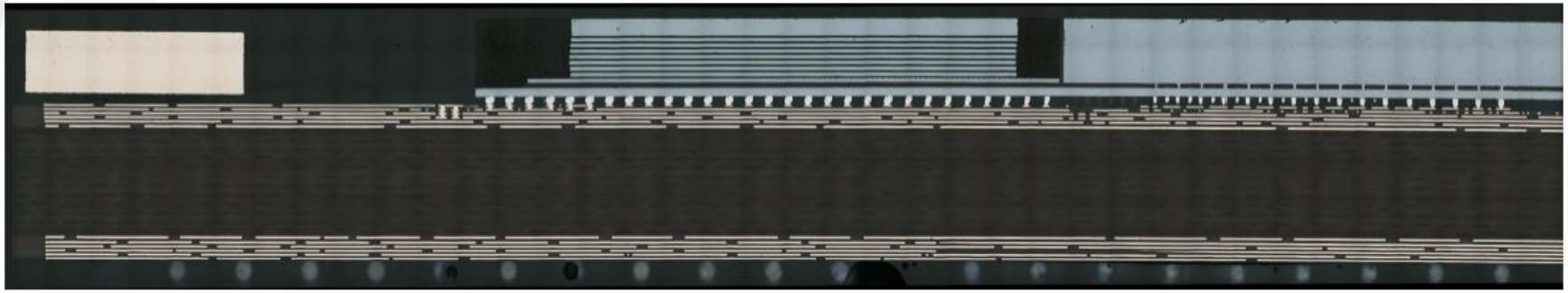
However, all is not lost - there are options to continue our industry's 50+ year trajectory of innovation

"Heterogeneous integration of best-in-class technology is a way to continue Moore's Law performance trends"

Nagisetty (Intel), IEEE Spectrum 2019

A Real Example of Heterogenous Integration (aka Advanced Packaging)

NVIDIA's GPU + HBM



- **NVIDIA's GPU with 4 HBMs (8 high stack + logic layer) mounted on Si interposer**
  - HBM with wide bus (1,024 I/Os, ~4,000 bumps, 55µm micro bump pitch)
  - Silicon interposer is 34 mm x 43 mm
- **NVIDIA's latest A100 uses GPU + 6 HBMs**

Source: NVIDIA.

techsearchinc.com

TechSearch INTERNATIONAL

From "The New Era of Heterogeneous Integration: Promises and Pitfalls", Vardaman CISES 2020. Reproduced with permission of Jan Vardaman, TechSearch International

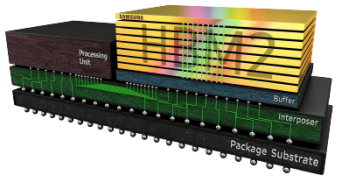
FormFactor™

A Complex Set of Choices Must be Made in Testing this Device

Observed probe-card growth is a result of test coverage requirements to reduce cost of advanced packaging processes

Test Challenges for Heterogeneous Integration

- **Known Good Die (KGD) required**
  - BIST and redundancy
- **Known good HBM stacks!**
- **Known good substrate needed**
  - AOI used for inspection
- **Know good interconnect (assumed)**
- **Need more comprehensive test content that can be run at wafer-level**
- **Need new methods to probe fine pitch bumps or test coverage without touching µbumps**



Source: Samsung.

"Stack and Pray" = accumulated yield loss

1 chip	2 chips	3 chips	4 chips
90%	81%	73%	64%

techsearchinc.com

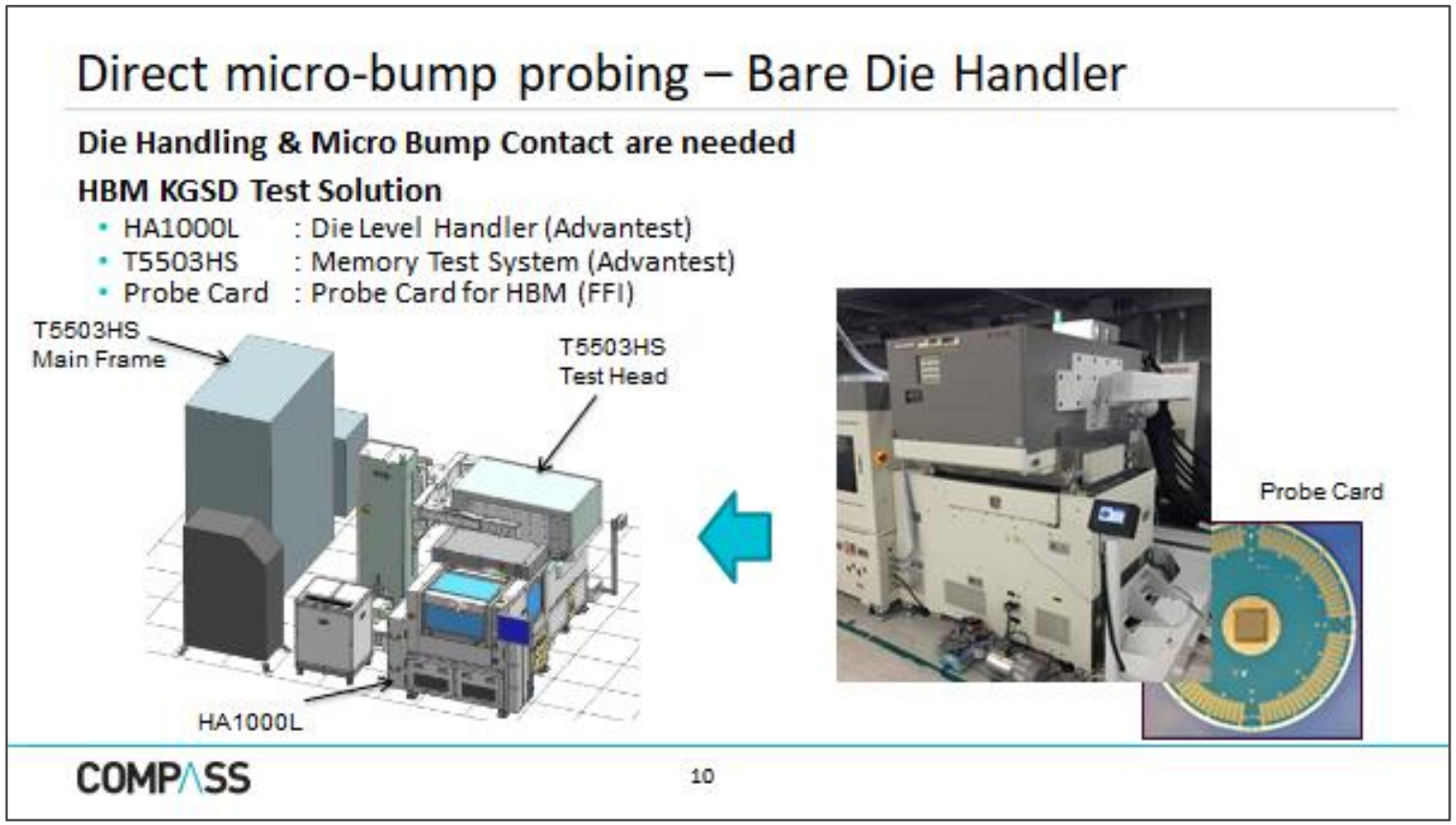
TechSearch INTERNATIONAL

From "The New Era of Heterogeneous Integration: Promises and Pitfalls", Vardaman CISES 2020. Reproduced with permission of Jan Vardaman, TechSearch International

- Ideally, each component known good before integration
  - This has spawned calls for Known Good Die (KGD)
- Caveat #1: "Known Good" within redundancy/repair
  - Interposers with redundant vias, HBM sub-die repairability
- Caveat #2: Economics (always?) dictates less-than-KGD
  - And as Dr. Tadayon noted earlier, it's all about economics
  - Analogous to quad chart one 1-die scrap-cost avoidance
- Yield & cost asymmetries* create test-flow choices
  - "Stack and Pray" quickly becomes economically unfeasible
  - GPU: $500 cost, 75% raw die yield
  - HBM: $100 cost, >99.9% raw unit yield
    - Final tested from DRAM vendor (will discuss flow later)
  - Interposer: $10 cost, 95% raw unit yield
- Almost 20% of BoM is parts other than the GPU
  - In monolithic GPU cases, package cost <<20% of BoM
  - Integration scheme driving GPU test coverage towards KGD
- Interposer only 2% of BoM, so not-KGD is expensive
  - A brand-new test insertion, with its own probe cards
- More unique probe cards & higher unit test intensity

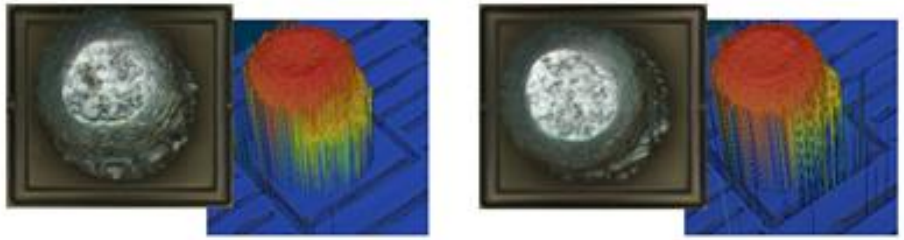
*I made these numbers up to illustrate a point

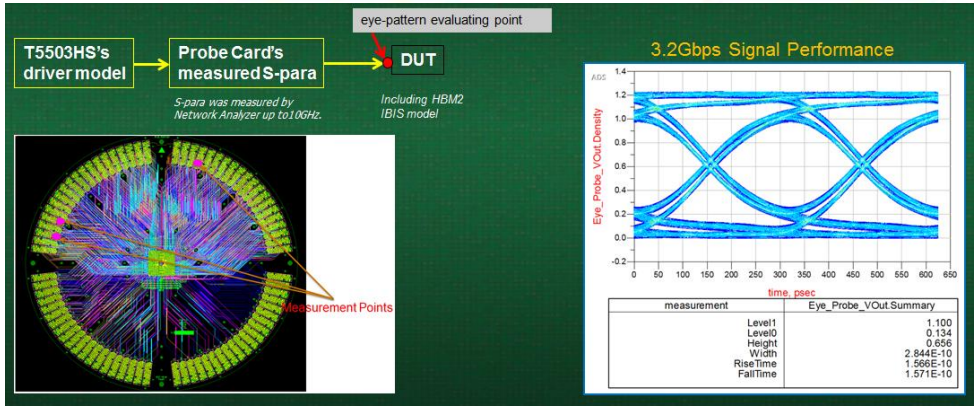
An HBM Case Study: KGD DRAM Test Through the Micro-Bumps



From Kiyokawa (Advantest) and Nhin (FormFactor), Compass 2019

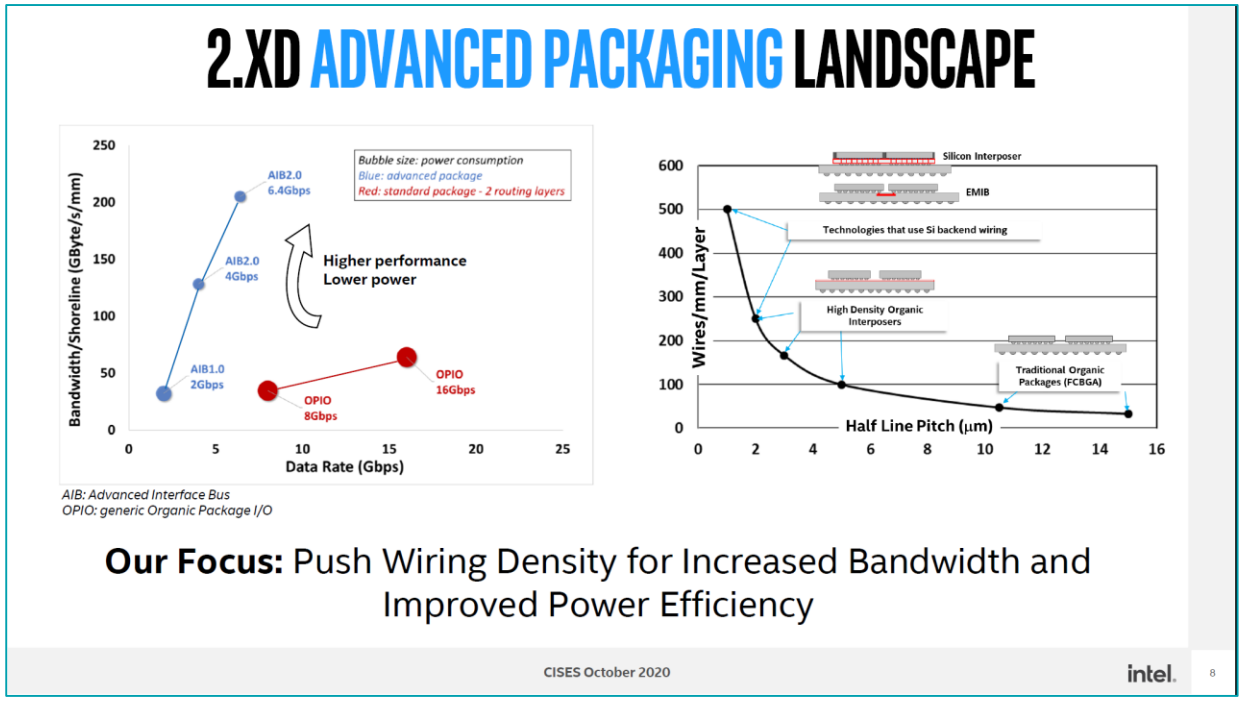
Condition	T.T:600sec 1 time	T.T:600sec 2 times
Scrub depth[um]	2.61	2.99
Scrub diameter[um]	14.81	15.04

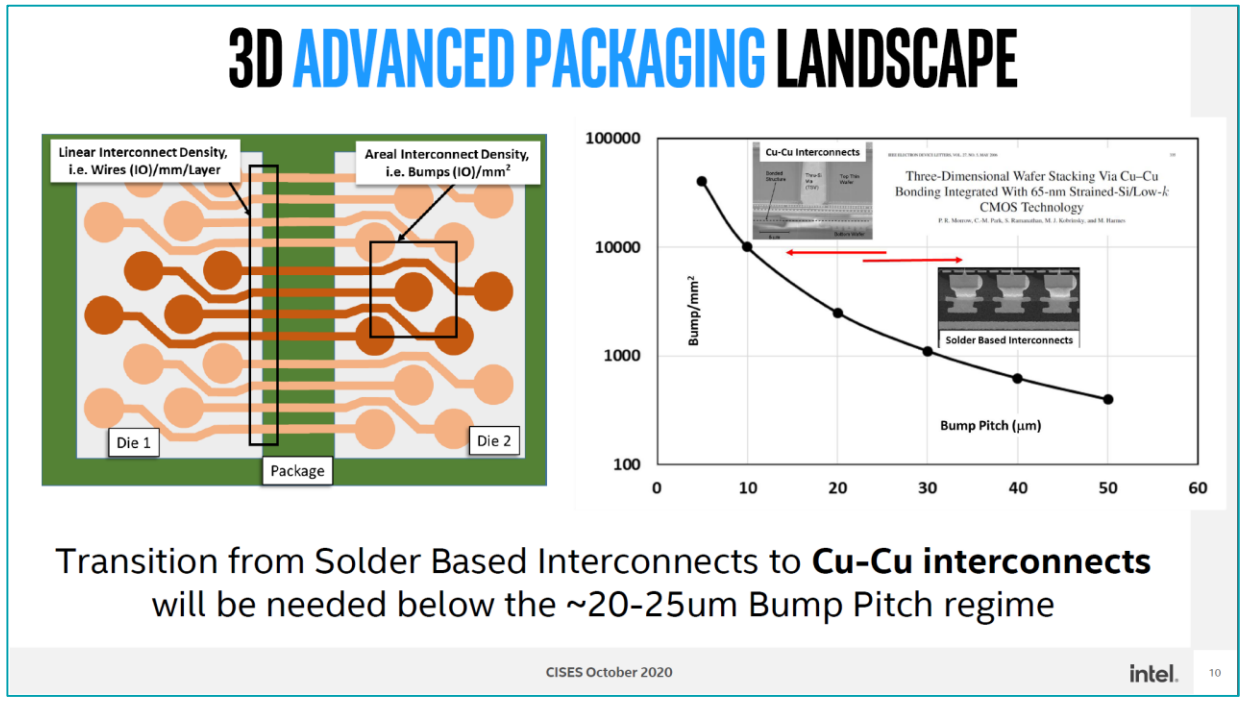




- Low-volume production KGD test (3.2Gbps) demonstrated for HBM component die
  - 25um SnAg micro-bumps on 55um full-grid-array pitch
- Technically feasible to perform KGD-level test through the packaging contacts/elements

Probing of Advanced Packaging Structures Requires Technical Advances

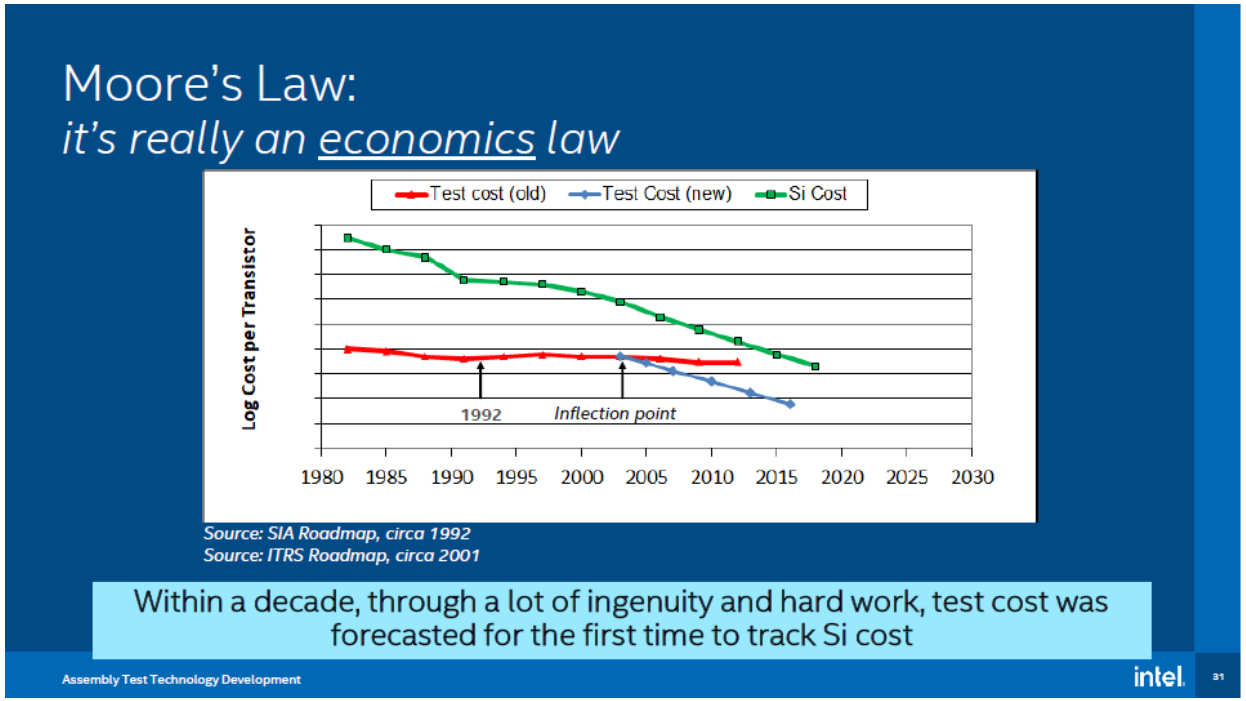


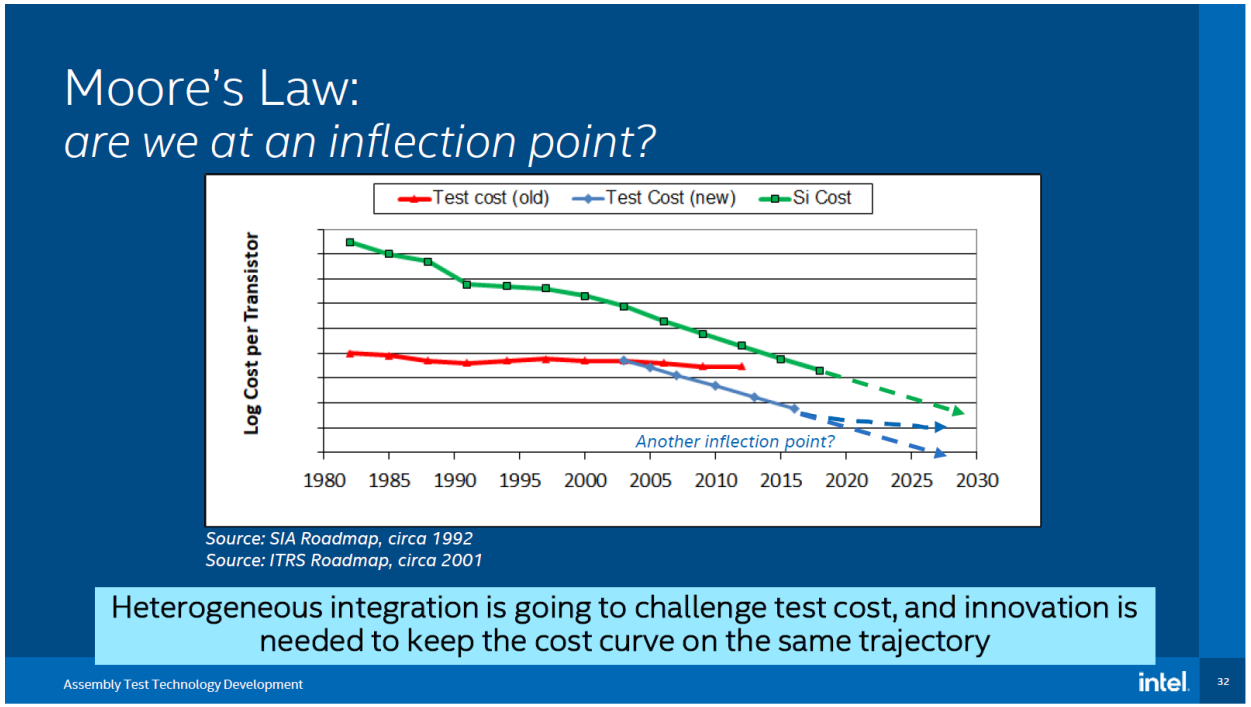


From "Advanced Packaging Architectures: Scaling for a Heterogenous World", Sabi CISES 2020. Reproduced with permission of Babak Sabi, Intel Corporation.

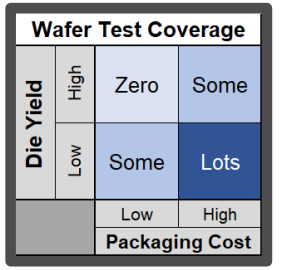
- Significant advancements for three basic capabilities are required to probe these structures:
  - Mechanical/geometrical – pitch, alignment, probe count
    - A few years ago, this was almost all you had to worry about in the probe card business
  - Electrical – frequency/bandwidth/data-rates and power (CCC, PI/SI)
  - Composition/materials – probing copper at dimensions of a few 10s of microns
- The good news is that the industry can meet most of these individual requirements today
  - The bad news is we can't do it all at once in one probe card with usable parallelism/probecount (today)

Technical Solutions Must Support the Economic Realities



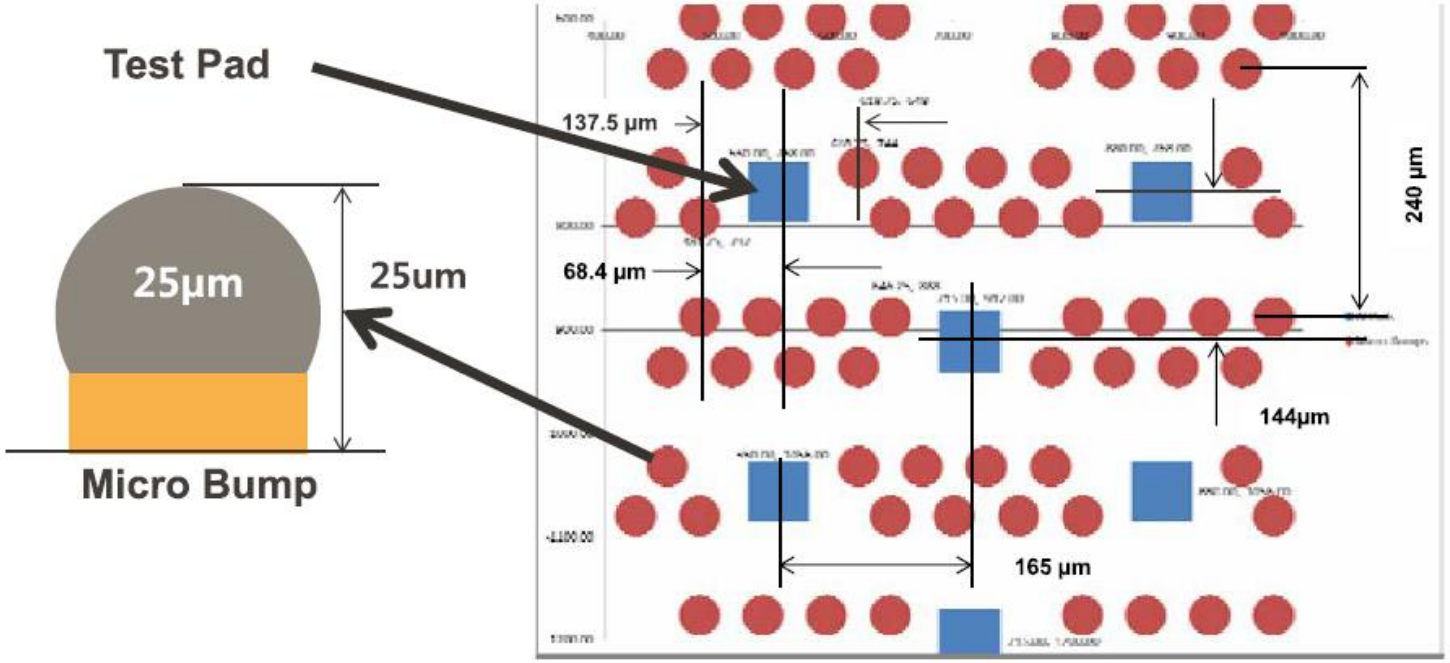


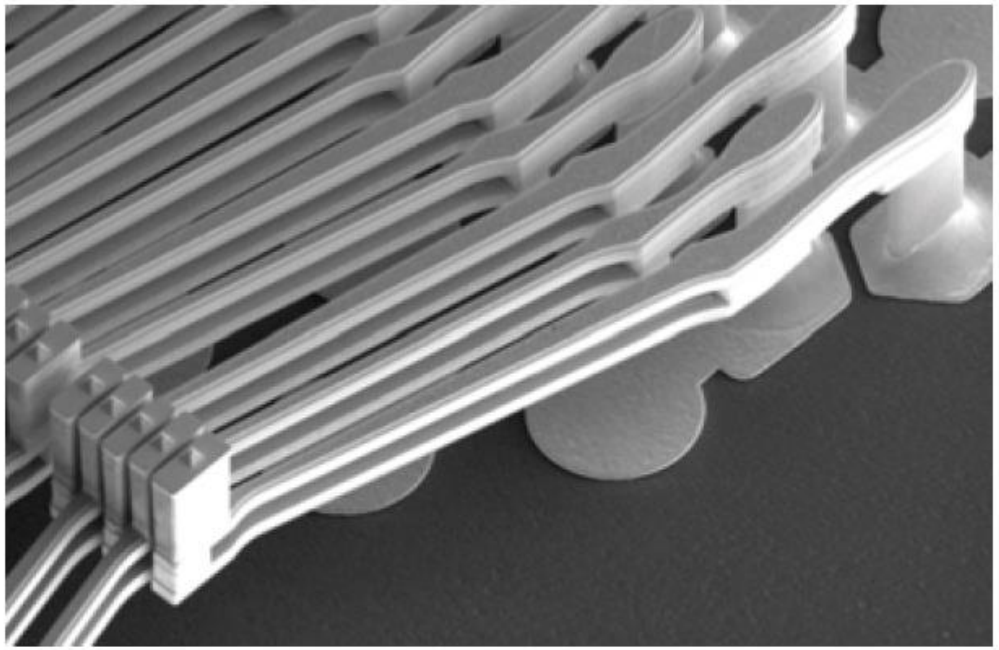
From "Moore's Law and the Future of Test", Tadayon SW Test 2020. Reproduced with permission of Pooya Tadayon, Intel Corporation.



- A key message from both keynotes at SWTW 2020 – it's all about cost and economics
  - Probe exists (solely) to prevent a more expensive downstream scrap event
- Technical advances to meet requirements not useful if probe cost > scrap cost
- Likely that optimization (read compromise) will need to occur at the IC and test-cell system level
  - There will be more probe cards that are more complex (expensive), but probe cards can't do all the lifting

A Potential Compromise (Back to HBM again): Don't Probe the Micro-Bumps if You Don't Need To

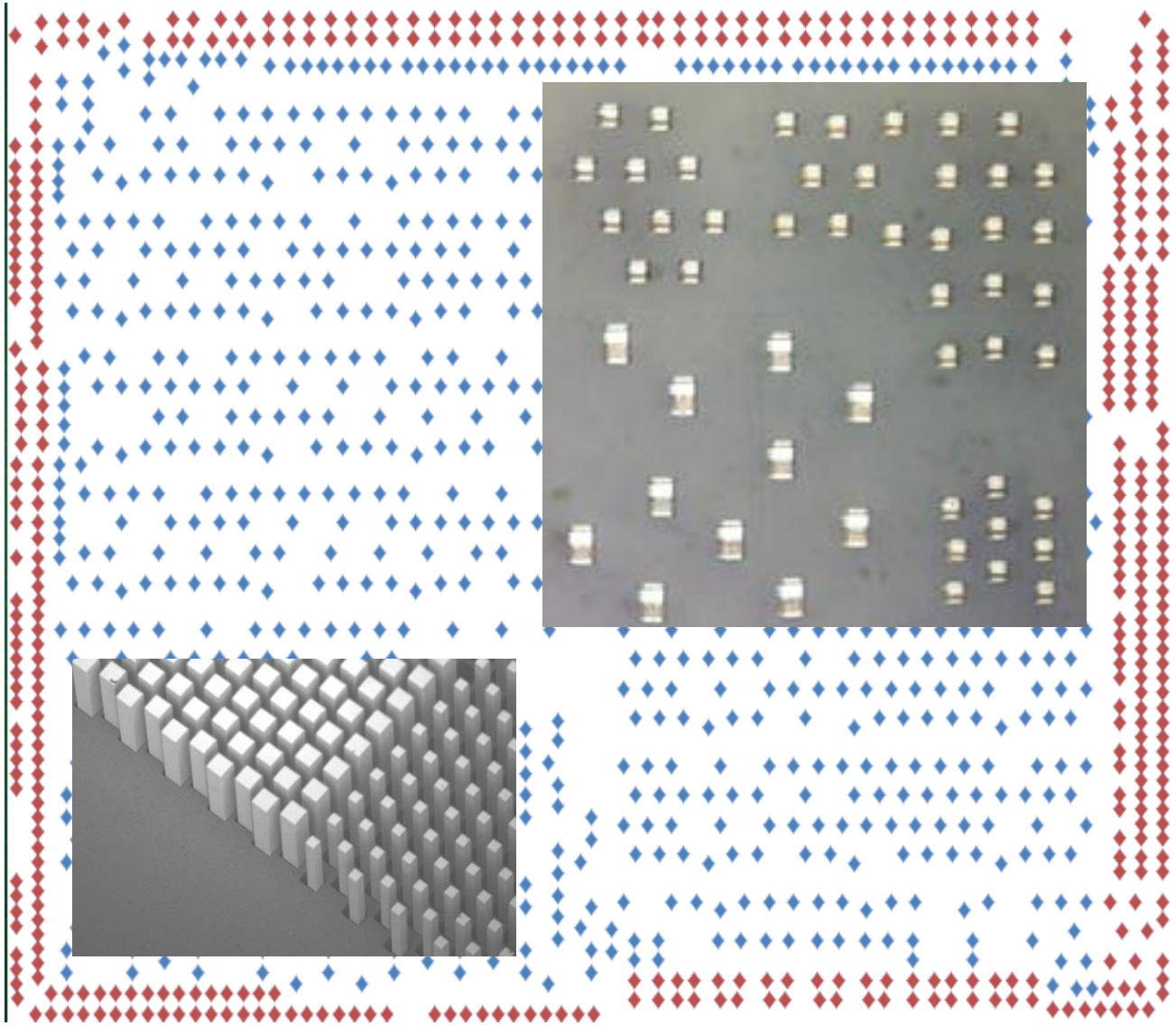




From Loranger+Yaglioglu (FormFactor) and Oonk (Teradyne), IEEE Design & Test 2016

- HBM DRAM die are sparsely populated with micro-bumps, real estate available for dummy test pads
- Some significant advantages to this approach, similar to "regular" DRAM sort
  - No micro-bump damage to worry/argue about
  - One-touchdown (whole wafer) parallelism provides a tremendous (~100x) cost advantage over single-die
- For high-volume HBM manufacturing, this has become the standard probe methodology
- Still more (high-speed) probe cards required, since one HBM stack is 8+ DRAM die and 1 SoC die

Another Potential Compromise: Hybrid Bump and Probe Layouts



- Approach: Use different probes in different areas of the die to do different jobs
  - Enables decoupling of different requirements
- An example use for application-processor test:
  - I/Os at fine pitch with low current requirements
  - Powers/Grounds at larger pitch with higher current
- Higher uptime from reduced probe burn events
  - 40% improvement in power impedance
- Many permutations of this general approach
  - Ex: One probe contacting multiple bumps
- Other examples for compromise/choices:
  - On-die test capability – BIST and DFT
  - Reduced compliance/overtravel requirements
    - Does a 40um pitch probe really need 100um max overtravel?

Summary and Conclusions – Probe In the Spotlight

- Advanced packaging is taking over from a slowing front-end-driven Moore's Law
- This is shifting the enabling burden from the front end to the back end
  - Where litho+etch once exclusively enabled the industry, now assembly and test more important
  - Probe is becoming (legitimately) more valuable to the industry
- Consistent with this shift, industry spending on probe is outpacing historical norms
  - Up 7% in a down 2019, likely continuing to accelerate in 2020
- Significant challenges emerging with increasing test coverage and complexity
  - Increased coverage means more probe cards: composite yields of component die
  - Increased complexity means more capable probe cards: higher densities, faster speeds, etc.
  - KGD with probe cards that do it all are feasible (and comforting), but likely too expensive for ROI
- Viable solutions likely need technical compromises made at the system level
  - Multi-supplier and customer collaboration to intelligently make these choices and compromises
- Thank you to Babak, Jan, John, & Pooya for permission to reproduce their work