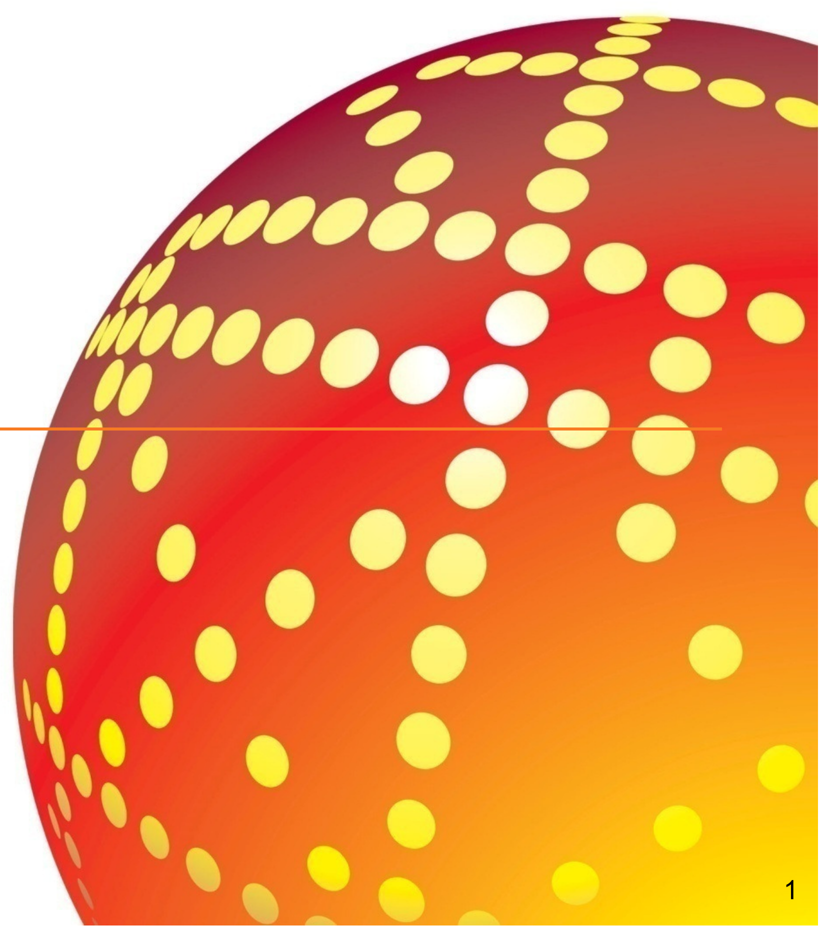
The Semiconductor Foundry Transition and it's Impact on the Mask Industry

Dr. Udo Nothelfer
Vice President and General Manager,
Fab1, GLOBALFOUNDRIES

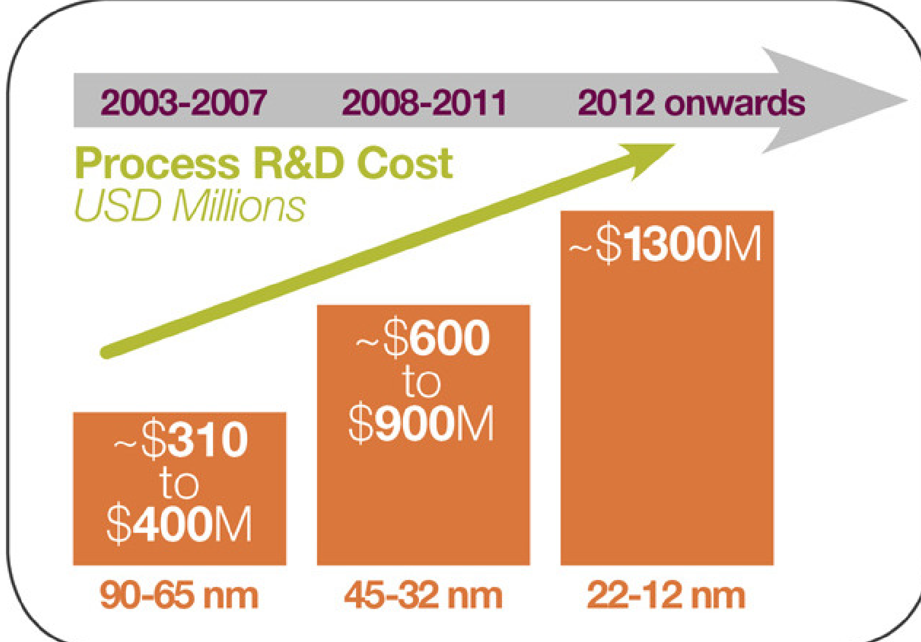


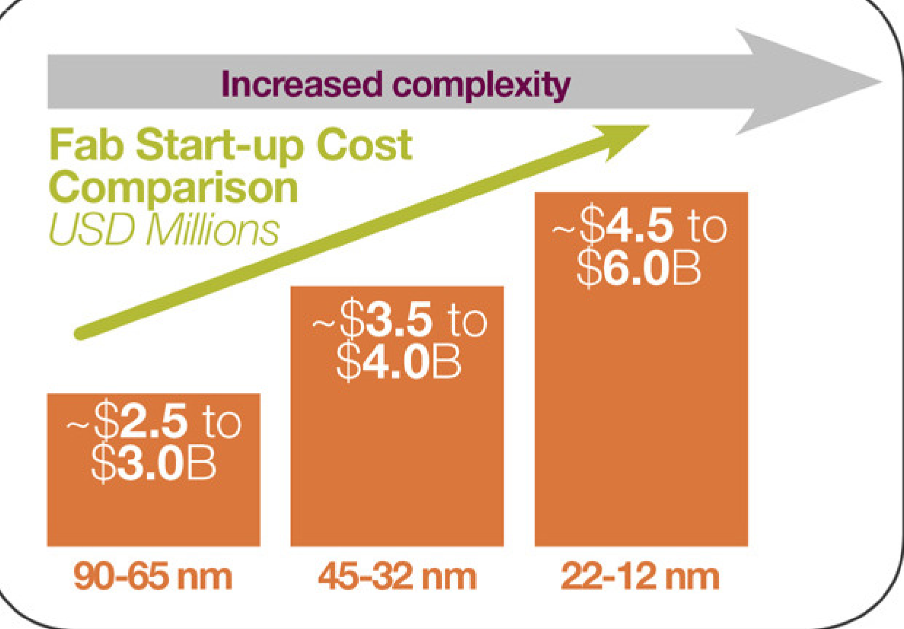
Outline

- The GLOBALFOUNDRIES story
- Transition to a foundry model: the mask cost topic
  - MLR as an answer on cost increase
  - The IP shuttle
- Transition to a foundry model: Technology challenges
  - GF MPU History
  - Tool and process control versus product specific control
  - Design flexibility and GF‘s solution
  - the role of high performance mask processes
- Ebeam Direct Write
- Summary

The Dilemma of the Semiconductor Industry

- Chip-makers need to keep pace with technology and focus on design
- …while chip manufacturing and technology R&D continue to grow in cost and complexity





January 2010: A New Foundry Leader is Launched

ATIC

Dresden, Germany

Munich, Germany

London, UK

Sunnyvale, CA

San Jose, CA

Saratoga County, NY

East Fishkill, NY

Austin, TX

Abu Dhabi, UAE

Yokohama, Japan

Shanghai, China

Hsinchu, Taiwan

Woodlands, Singapore

Tampines, Singapore

The New GLOBALFOUNDRIES

- Headquartered in Silicon Valley
  - Approximately 10,000 employees
  - Spanning three continents across 12 locations
  - 300mm Fabs in Singapore, Germany, New York
- One of the World's Largest Foundries
  - 2009 revenues in excess of $2B US
  - 150 Customers including many of the world's largest IC companies
- One of the Leaders in Foundry Technology and Service
  - Substantial Time-to-volume advantage for advanced technologies

Vision and Value Propositions

To be the first truly global semiconductor foundry, harnessing the world's resources to deliver maximum value to our customers and unlocking their potential to innovate.

GLOBALFOUNDRIES is:

A foundry technology leader.

- We ramp advanced technology ahead of all other foundries (CPU), and set the standard for new technologies

A leader in globally distributed capacity.

- Expanding to 1.6 million 300mm wafers annually with 2.2 million 200mm wafers supporting mainstream and advanced technologies

A foundry leader in customer-centric services.

- Flexibility and very close collaboration with customers in design enablement and technology delivery.

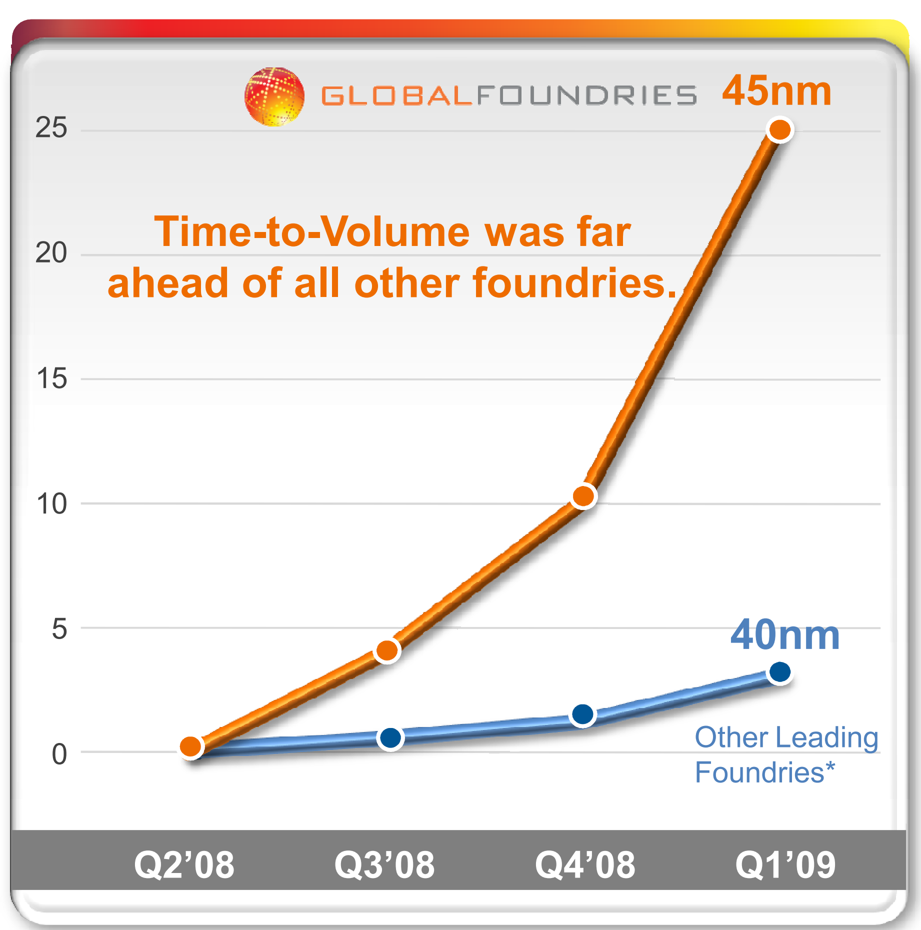
A foundry with focus and staying power

- $10B US financial commitment from investors

Leading Supplier of 45nm Technology

45/40nm Production Ramp

300mm Wafers Shipped per Quarter (1000's)



*Market data compiled by International Business Strategies

- Time-to-Volume is Unmatched in the Foundry Industry
  - High Volume, Large Complex x86 CPU Drives Yield Learning
- Initial 32nm HKMG production running in Dresden
- 32nm Ramp of Early Adopter will Precede 28nm Ramp of Other Customers

Technology Segmentation Strategy

	Technology	Target Audience/Applications
Super High Performance	45nm-SHP 32nm-SHP 22/20nm-SHP	Microprocessors, Game consoles, Graphics, Networking
Super Low Power, Low Power	65/45/40nm-LP 28nm-SLP 22/20nm-SLP	Smartphones, Portable consumer, Mobile gaming
High Performance, Generic	65/40nm-G 28nm-HP 22/20nm-HP	STB, DVD/DVR, Graphics, DTV, Servers / Storage

TRANSITION TO A FOUNDRY MODEL: THE MASK COST TOPIC

IDM versus Foundry mask related differences

- IN the IDM world
  - We drove technology and mask business to meet MPU needs
  - We strove to use up to 100% of the reticle field.
  - We ran thousands and thousands of wafers on one reticle
  - We used a technology-node for a limited time and moved quickly to the next
- IN the foundry world
  - We must drive to optimize cost and value each customer
  - We have to deal with both high and low wafer counts per reticle
  - We must continue to support older technologies for years
- Reticles are an important part regardless of IDM or Foundry
  - However, the mask related cost per wafer can vary considerably

Lithography cost analysis

- MPU is a high volume product! Wafer starts per mask >50.000 are typical
- In a foundry world, wafer counts as low as 1000w per reticle are possible!

Relative lithography costs at a 32nm technology

MPU, 50k wafer/reticle

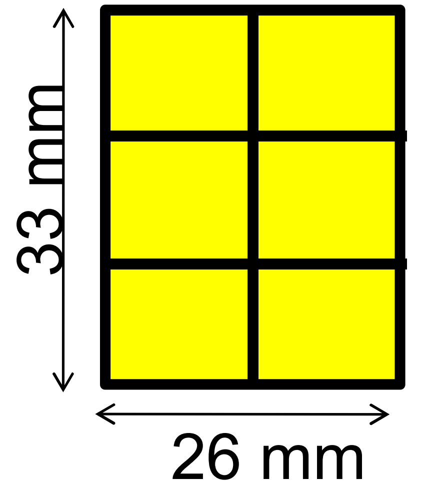
normalized costs
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
krF
immersion
mask
photochemicals
expsosure costs (laser)
variable costs (labor)
fixed costs (depreciation etc)

Foundry, 1k wafer/reticle

normalized costs
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
krF
immersion
mask
photochemicals
expsosure costs (laser)
variable costs (labor)
fixed costs (depreciation etc)

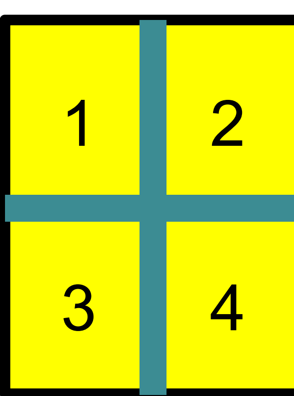
- Depreciation is clearly dominating for high volume manufacturing
- **Mask costs easily can become dominant in a foundry environment**

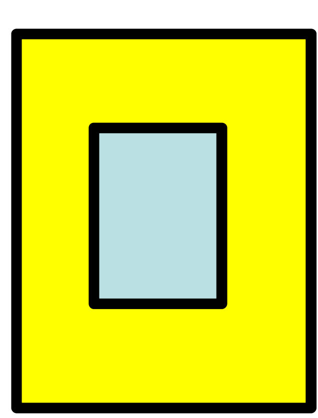
An answer toward the cost trap: Multi layer reticles (MLR) - Fix



1

2





Full Field
Many different Dice sizes possible but field filling is key

FF

(Full Field)

two masking levels or two products single level

2LR 2MP

four masking levels or four products single level

4LR 4MP

Small Field Vehicle
Field size is ~ 25% of field, can have many to only one dice

SFV

Mask type Cost samples

- Each Reticle Type will have an impact on Cost. One must compare both the set and any individual reticle to judge the Final approach

Generic 40 Mask Level Product

Most Critical Single

Full set Impact

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
FF
4LR
SFV
Other
Repair
Inspection
Process
Write
Data Prep
Blank Material

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
FF
4LR
SFV
14* Non Crit
12*Implants
M6
V5
M5
V4
M4
V3
M3
V2
M2
Via 1
Metal 1
Contact
Gate
STI

- **Mask approach allows choices to meet customer demands**

Consequences of Multiple Levels per Reticle

- Lower scanner throughput due to smaller field utilization
- Critical support structures (SLM, marks,…) must be accommodated for each layer with smaller exposure field
- Reduced area per reticle to accommodate product die
- Increased Data preparation cost due to multiple critical levels on a single reticle, mask build and Fab set-up time

Impact on Scanner TPT based on field size used

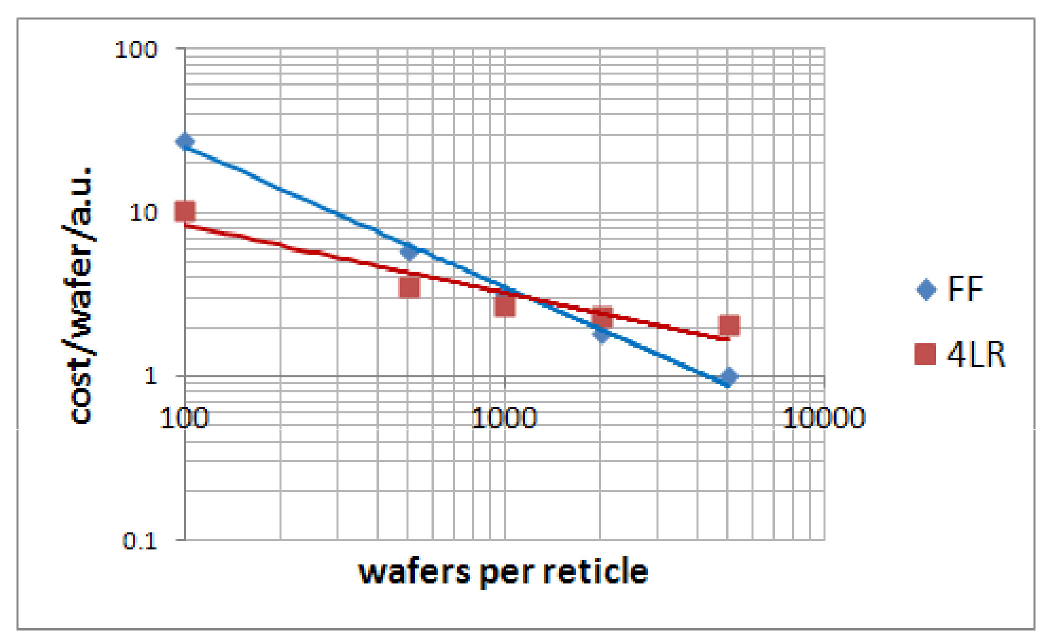
- Chiplet area includes the number of Dice and required frame
  - Max X and Max Y plus scribe lane determine Chiplet fill
- Flashes are the stepping pattern needed for each level exposed
  - Net Die per wafer determined by number of die in each Chiplet

Full field is 26 x 33mm

Layout / Product	Max X	Max Y	Reticle area	Effective Usage	Expected Flashes	THP (normalized)
4 Die Reticle	24.45	27.58	674.3	79%	98	**1**
1 x 2 - 2LR	24.64	15.4	379.5	44%	148	**0.6**
2 x 2 - 4LR **SFV**	12.32	15.4	189.7	22%	311	**0.31**
2 x 3 - 6LR	12.32	9.24	113.8	13%	534	**0.18**

Cost Comparison Next slide

Cost analysis for 4LR versus FF, most critical layer



- For the full field reticle, the cost increases exponentially at low wafer starts
- Depending on the assumption, the cost cross-over between FF and 4LR reticles is between 1000 and 2000 wafer exposures per reticle

MLR selection criteria

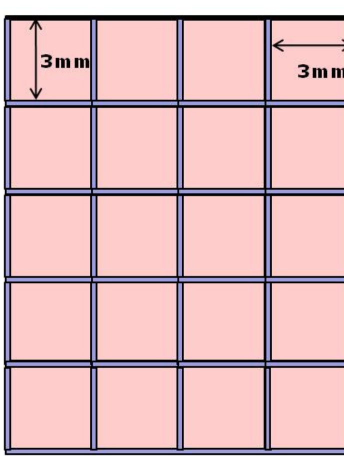
- Cost is very important, but it is not the only selection criteria!
- For every individual application, a complexity consideration will be done!
- Example for complexity rating of the individual mask field approaches:

Reticle Strategy and Factors	4LR - Product	4 MPW	2LR - Product	SFV	Full Field
Price (Single Most Critical)	4	10	6	4	8
Data Prep and OPC CT	5	10	4	1	3
Mask Manufacture Cycle time	10	10	8	2	6
Litho Impact	10	10	4	10	1
Revision Flexibility	8	10	6	1	2
Reticle Storage	4	7	5	7	7
Intangibles	5	2	3	2	1
Complexity	**46**	**59**	**36**	**27**	**28**

1: low
10: high

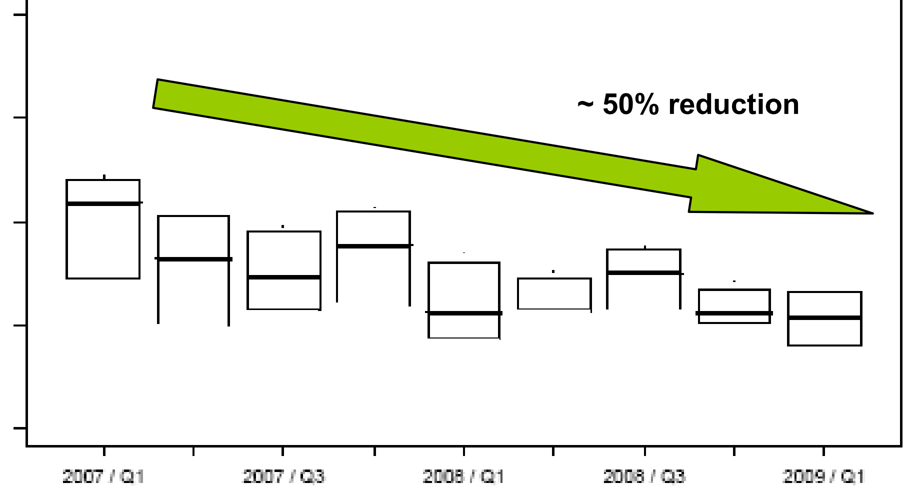
The IP shuttle: a typical MLR application

- Shuttles get customer IP and allow technology scoping on a frequent basis
  - Limited number of wafer starts per shuttle run
  - Predestinated application for MLR
  - Most foundries run shuttles on a regular basis for each technology
  - GLOBALFOUNDRIES will decide frequency based on business needs
- Customer loading and timing is not predictable
  - Most customers have different design cycles and timing
  - Shuttle format allows customers to test Foundry processes for a lower cost by sharing real estate but not IP with others foundry customers
- Providing standard “Tile” size for easier fab and Customer interactions
  - Frame and Scribe more standardized for integration and fab
  - Test structures and fab control standard on each technology shuttle

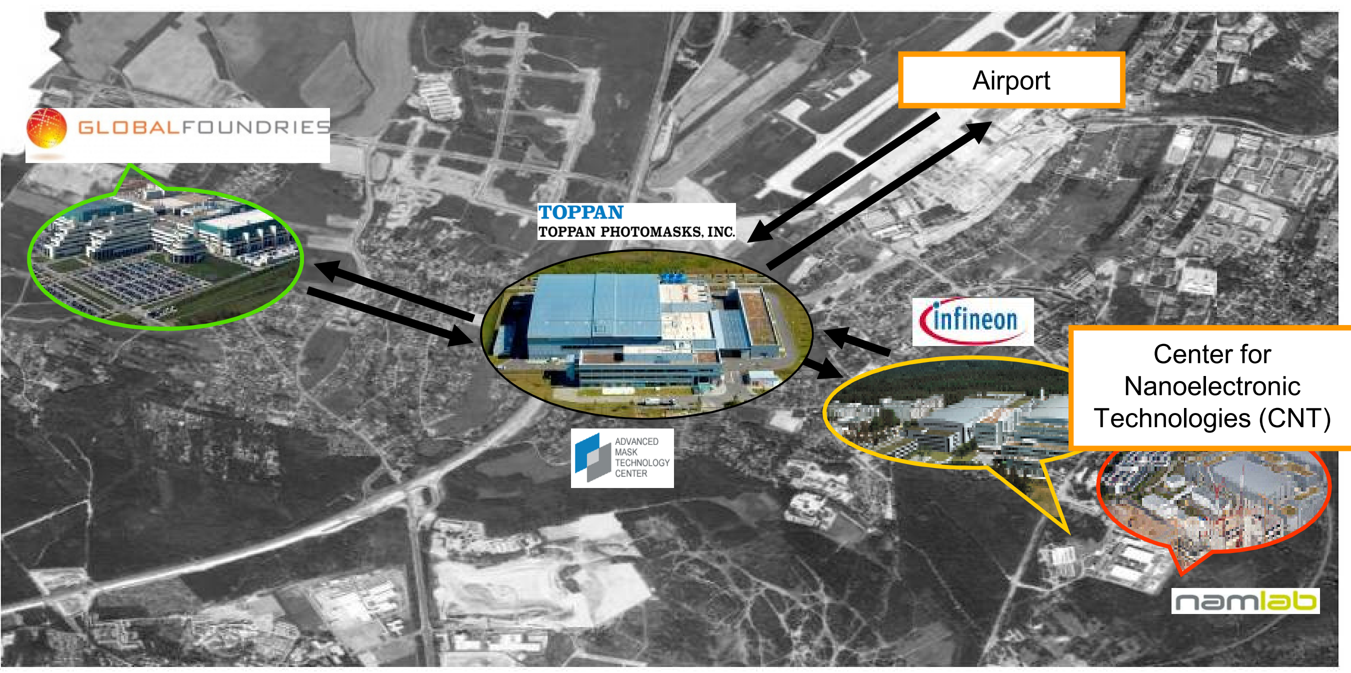


Delivery performance of masks

- Reliable and quick delivery of masks is essential to meet time to market requirements
  - Close interaction and communication between FAB's & Maskshop is needed to ensure deliveries meet lot and customer expectations
- Cycle time for the first layers are very critical and are the main focus
  - Cycle time is more critical in foundry space due to many overlapping products
- It is advantageous to have direct influence on the mask shop (captive?) !
- AMTC drives down cycle time to compete with commercial and captive capabilities



Synergies in Dresden – GLOBALFOUNDRIES and AMTC



- Maskshop close to the FAB helps to shorten shipping time in both directions
- Close interaction on all levels
  - F2F meetings and combined projects

TRANSITION TO A FOUNDRY MODEL: TECHNOLOGY CHALLENGES

Time Line for technology introductions IDM versus Foundry

Technology Development 1 | Pre T0 | Test reticles for Litho and transistor variants 1 → many years before T0

Technology Development 2 | T0 | Full Flow reticles are built to test technology

Product line, Fab and technology team technology and product interactions

Fab, technology team and customer technology and product interactions

Technology Validation

Final test Chips and Initial Product

Final test Chips, **Shuttles**, and Initial Product validation

Available for IDM products for 18 months to 3 years of volume

HVM

HVM N-1 (just 2 nodes overlap)

Available for ALL New Products on the New Node, many years

HVM – New Node

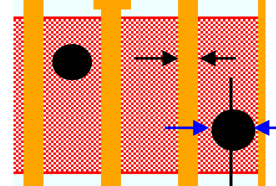
HVM N-1

HVM N-2 (long life cycle)

The fab keeps older technologies in place. This adds to the fab complexity but improves the yield learning and process sharing. This is the first big change from IDM to Foundry and virtual IDM → **The technology stays alive**

IDM versus Foudry: technology and manufacturing differences

- IN the IDM world
  - MPU is a high volume product, the number of products is limited
  - Tailoring of the technology toward special product needs is possible
  - Product specific control and improvement algorithms, in particular at the gate and contact layer, are possible and became sometime mandatory
  - The µP-speed and CD control and Contact to gate overlay always was of special importance



- IN the Foundry world
  - We have to deal with many more products.
  - Product specific control- and improvement algorithms must be the exception
  - The processes and tools, both for mask and wafer, must have built in stability!
  - Establish a good balance between process unification and process flexibility!

Example 1: Product specific overlay optimization for contact to gate

- Immersion puts new challenges on overlay (e.g.evaporation cooling)
- Chuck signatures can be different and can change over time !

Different/instable chucks in use

Same chuck /stable

6nm

8nm

reticle

Correction of higher order field term, lot history based

Due to our new correction scheme, we are able to run lots via mixed tools without violating 32nm overlay specs!

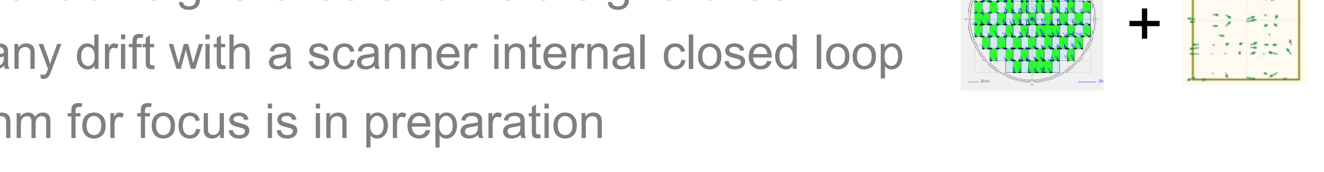
GLOBALFOUNDRIES dedicated correction and control system

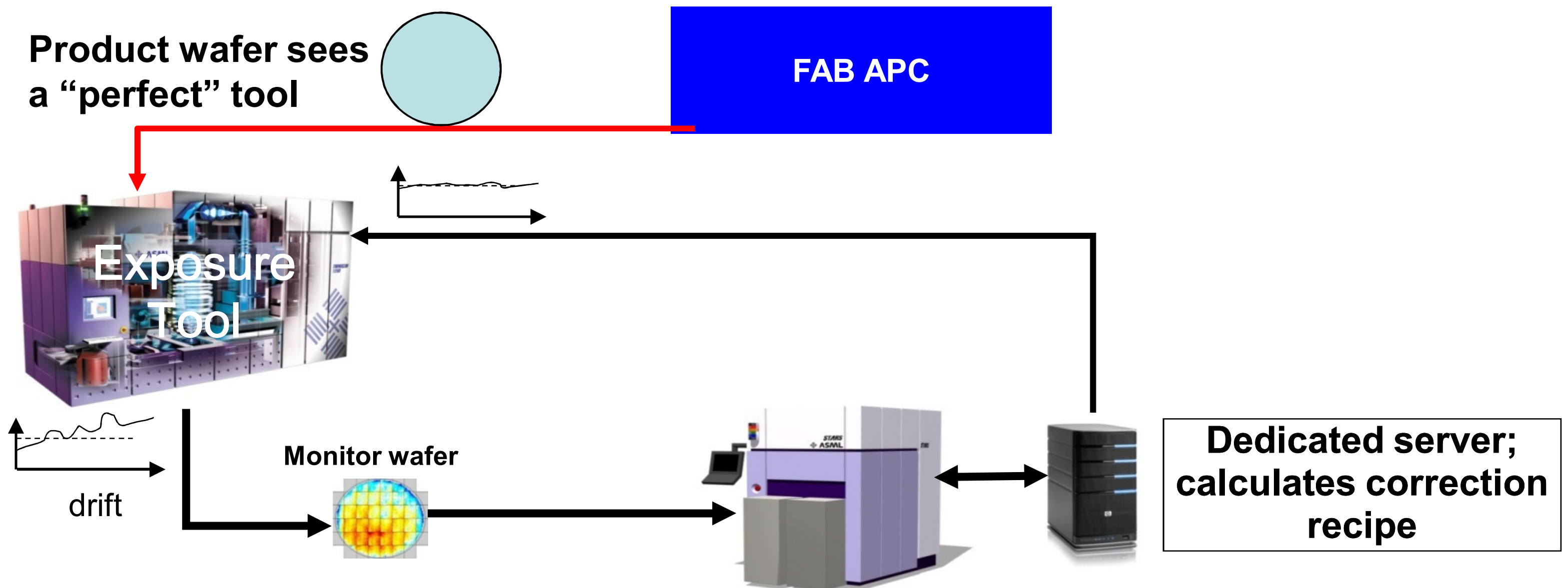
	Dedicated tool Dedicated chuck		Lot xxxx	mixed tool Dedicated chuck	
Layer 1	STPA ch1 ch2	STPB ch1 ch2	Time stamp	STPB ch1 ch2	STPA ch1 ch2
Layer 2	STPA ch1 ch2	STPB ch1 ch2	Correction map	STPB ch1 ch2	STPA ch1 ch2

Every wafer gets its „specific“ correction map depending on its history (tool, chuck, timing)

Overlay solution in a foundry world: generic TW based correction flow

- Test wafer with very dense sampling runs at a certain time interval
- Measure both chuck signatures and field signatures
- Correction of any drift with a scanner internal closed loop
- Similar algorithm for focus is in preparation

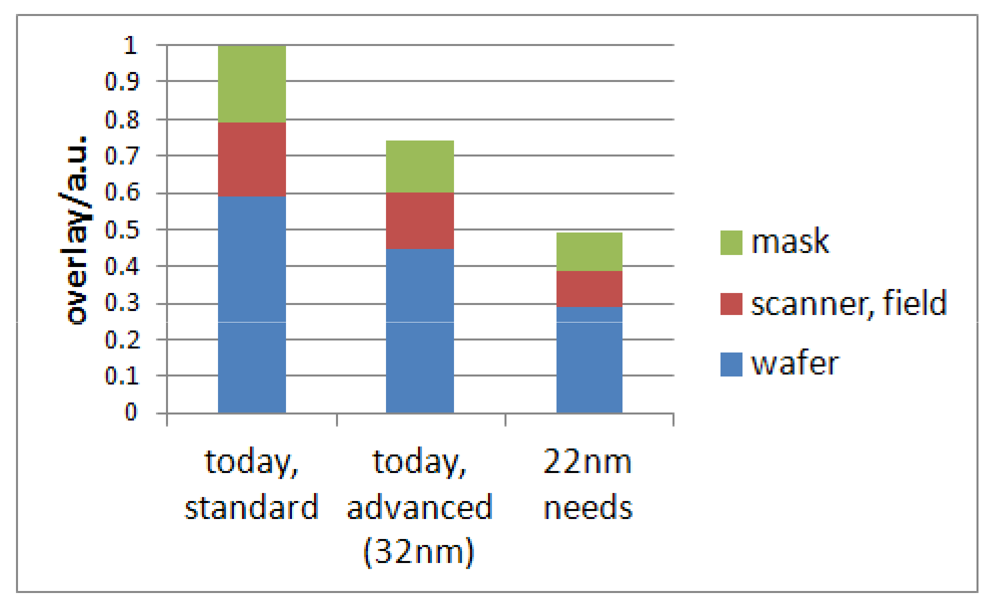




Please note: This correction does not address mask processes!

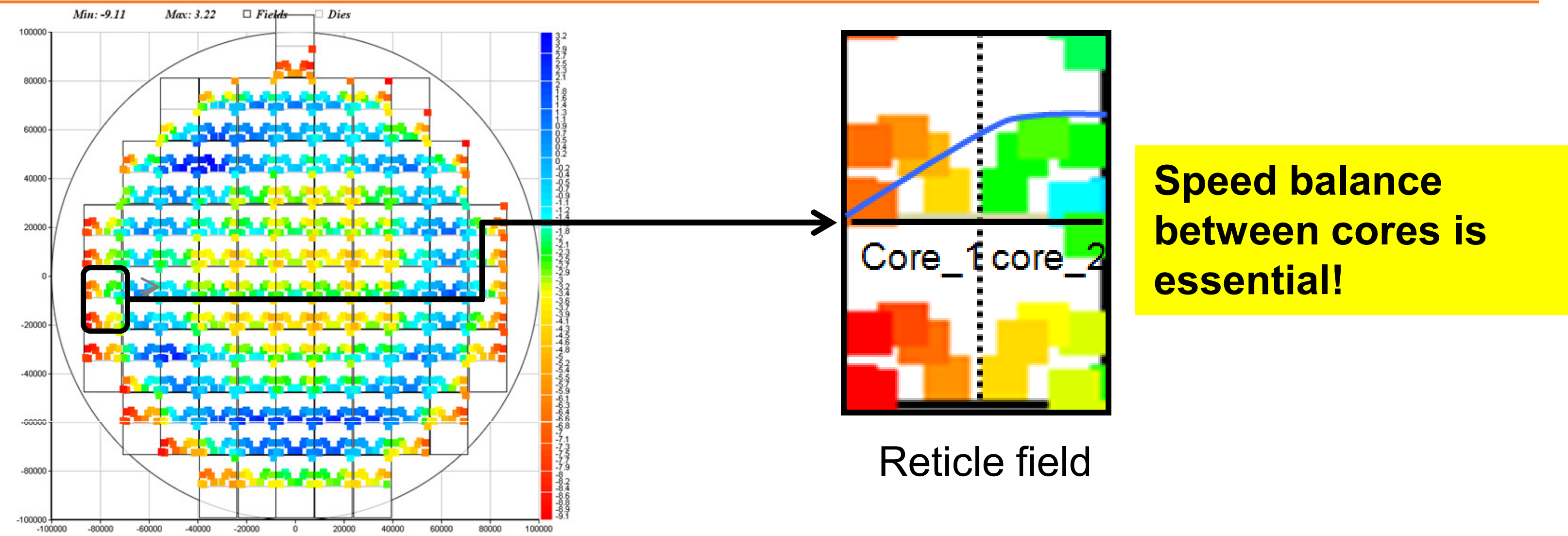
Reticle registration needs

Most important contributions to the overall overlay number



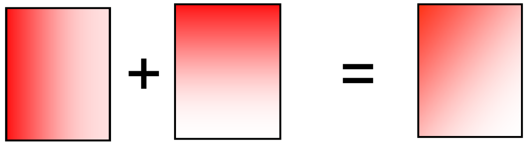
- Today, wafer terms are dominating over the reticle contribution
- With strong shrinking requirements toward 22nm, a further reduction of the mask contribution is mandatory
- By having an advanced mask shop like the AMTC, we have multiple techniques to drive the mask registration number toward the needs

Example 2: Product specific speed homogenization for the Opteron



Across wafer speed variation for the first dual core Opteron. Every dot represents one Ring oscillator; red means higher speed

DoseMapper JDP with ASML



Meanwhile, higher order correction profiles are possible both in x and y

Results of speed homogenization

Core_1 core_2

8660 STP90 Doserecipe AC1

Standard-sample

DoMa-sample

leakage

Fmax

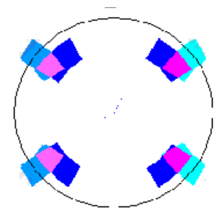
Package data of chips at the wafer edge; leakage versus speed

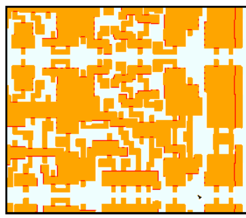
- Tremendous speed improvement due to the excellent correction of across chip speed variation by lithography
- **We can transfer that µP-learning to foundry products!**

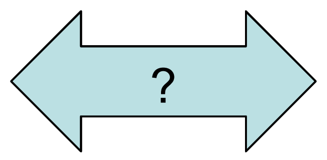
Design flexibility versus leakage (ACLV) performance

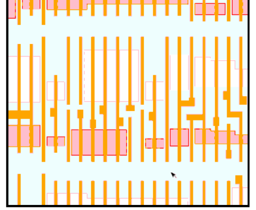
Gate design flexibility

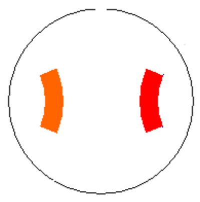
28nm, Quasar illumination







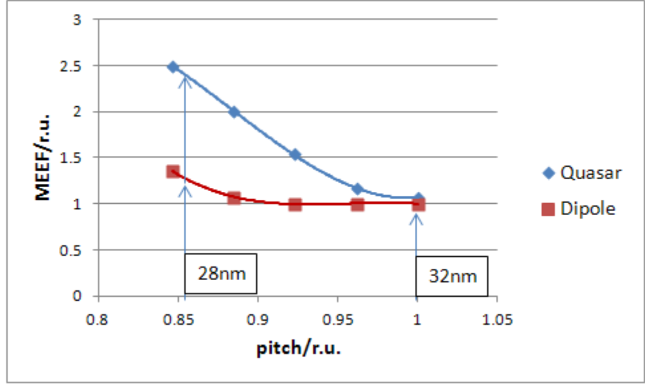




32/28nm, Dipole illumination

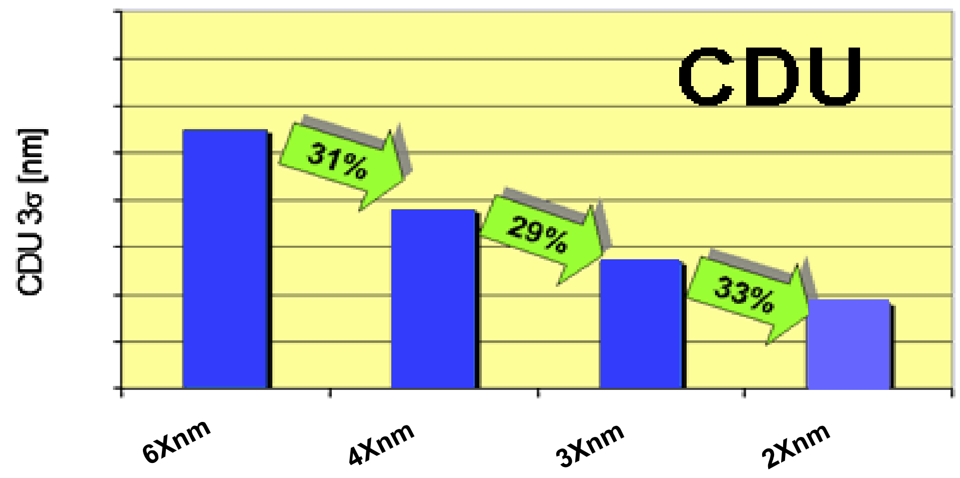
Some customers need design flexibility by requesting H and V gate orientation

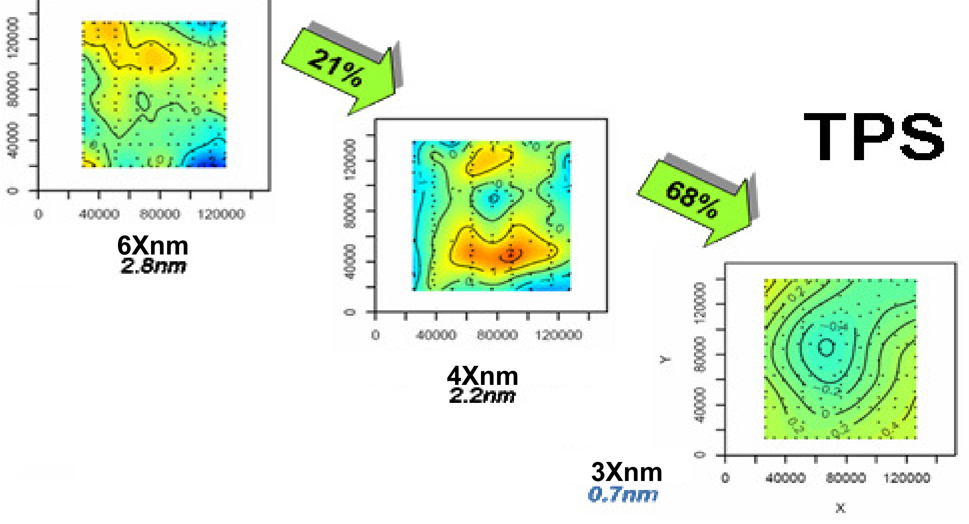
Some other customers accept restrictions for perfect leakage (ACLV) control



GF can offer both approaches by guaranteeing excellent mask CDU!

CDU Improvement Progress in Mask Making





Best CD performance on mask support across chip performance

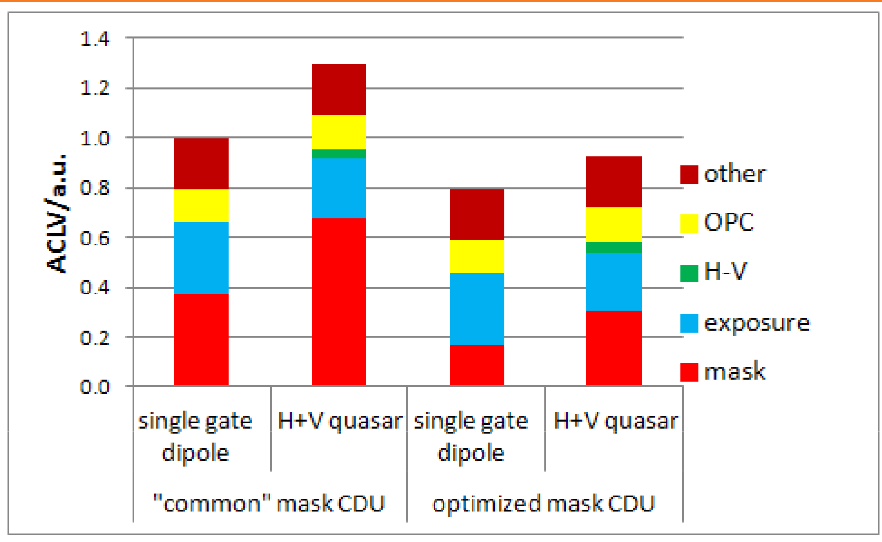
CDU signature correction on a small scale is essential during mask making process

CDU wafer level correction on mask can be very beneficial

Fab Litho process signature correction on mask will be a possible next step

Advantage of being able to drive reticle technology

Mask CDU improvement and it's impact on ACLV



- As long as we have the old „common“ mask process, ACLV is dominated by the mask contribution
  - Clear disadvantage for dual gate orientation
- With optimized mask process from AMTC, the mask contribution is in the order of other contributors
  - No big ACLV difference between dual and single gate orientation
- With post treatment of the mask, ACLV can be improved further!
- ACLV is just one parameter among others that contributes to device leakage!

E-beam-direct write: Is that an option?

- EBDW: a low-cost/high resolution lithography approach
  - maskless lithography
  - If the throughput is reasonable, it can be a cost effective alternative to conventional lithography
  - high potential particularly for low-volume manufacturing & prototyping
  - Currently, we don‘t see it as an alternative to EUV at high resolution / high volume manufacturing
- GLOBALFOUNDRIES promotes multi-eBeam tool & process development
  - Collaboration with multi-eBeam tool vendors
  - Process integration development
  - Assess progress from a potential end-user standpoint

EBeam Direct Write Lithography (II)

- EBDW challenges
  - **cost-effective direct write throughput**

    → needs to be significantly improved
  - beam control
  - overlay & butting
  - resist performance (resolution / LER / sensitivity)
  - **data preparation & data path**

    → considerably high data volumes
  - proximity correction
  - integration in manufacturing process / mix&match

Summary /conclusion

- The semiconductor industry currently compresses at the high-end technology space
- GLOBALFOUNDRIES is well positioned both due to it‘s high-end MPU volume manufacturing experience and it‘s foundry customer base
- Managing the increasing mask costs is a major challenge for future technology nodes. It is of particular importance for a foundry
  - By using MLR and SLV approaches we can control mask costs
  - IP shuttles enable customers to test their IP in an early technology phase
- The transition from an IDM to a Foundry world requires a concentration on process and tool control.
- Combining mass CPU and foundry manufacturing for flexibility and performance is a key success factor
- The Dresden semiconductor cluster with Fab1, AMTC and the CNT is essential within the world wide manufacturing and process development of GLOBALFOUNDRIES