

Cell Coprocessor Acceleration (CPA) at IBM for Computational Lithography

**G. J. Dick ¹, J. Allen ¹, L. Biskup ²,
T. Dunham ¹, J. Rogich ¹, J. Russell ²**

¹ IBM Semiconductor Research and Development Center

² IBM Systems and Technology Group, Operations

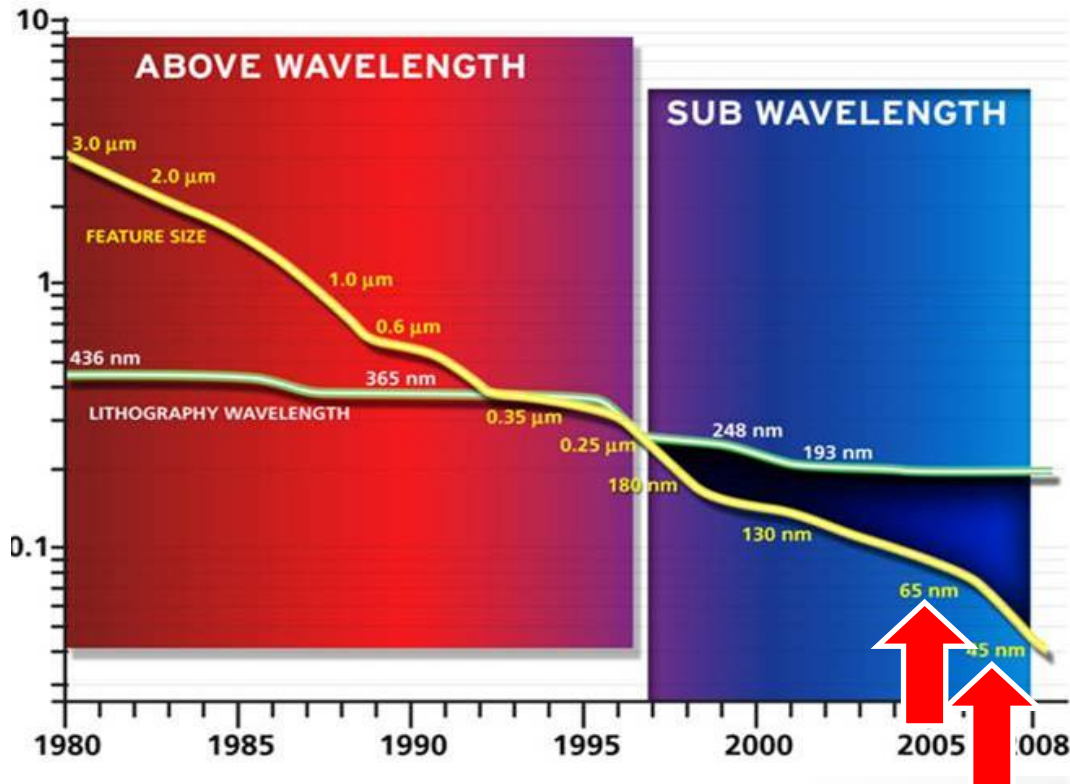
Computational Acceleration: Mentor Calibre™ nmOPC & Cell Broadband Engine™

Optical Proximity Correction (OPC)

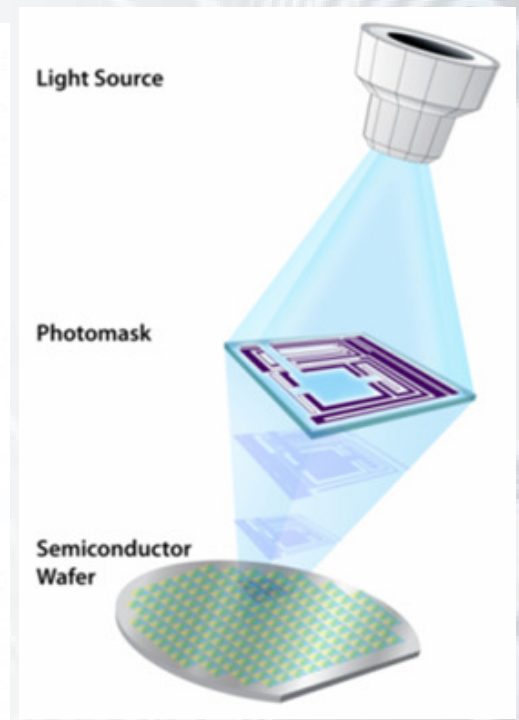
What Is It and Why Is It Needed?

Increasing Lithography Imaging Complexity

Optical Proximity Correction (OPC):
Anticipate and compensate for sub-wavelength optical lithography distortions in manufacturing.



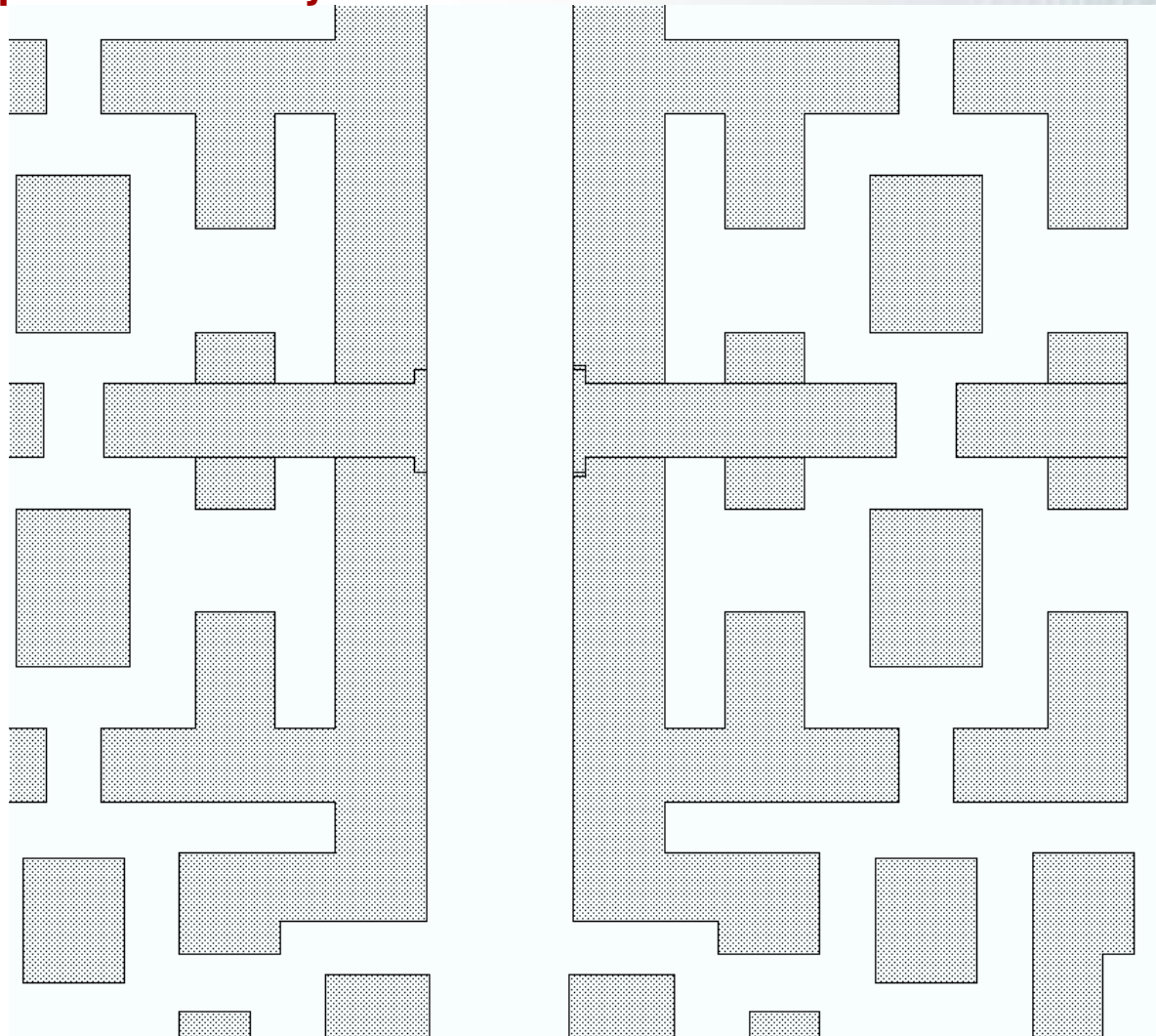
Lithography Imaging



Design

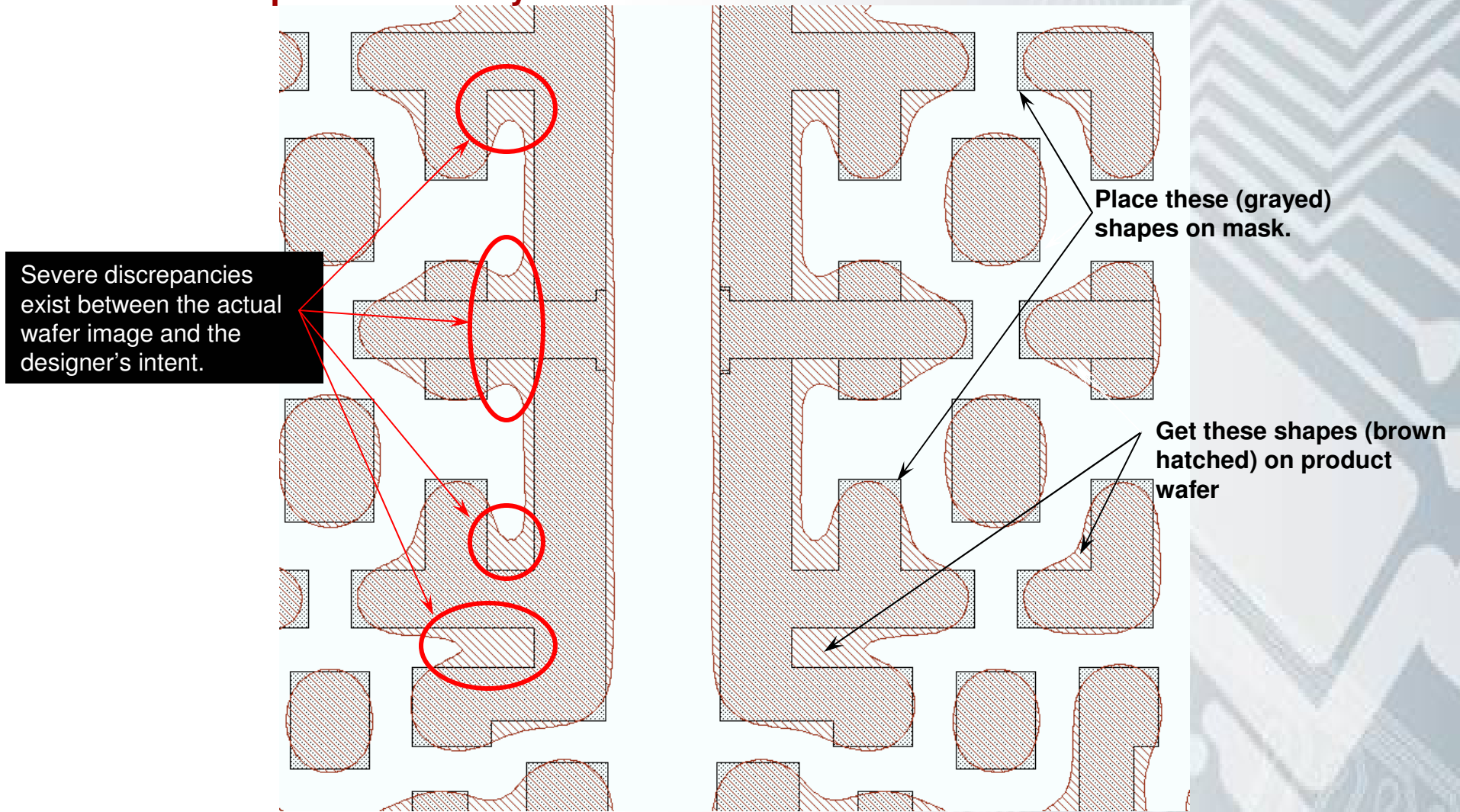


Need for Optical Proximity Correction



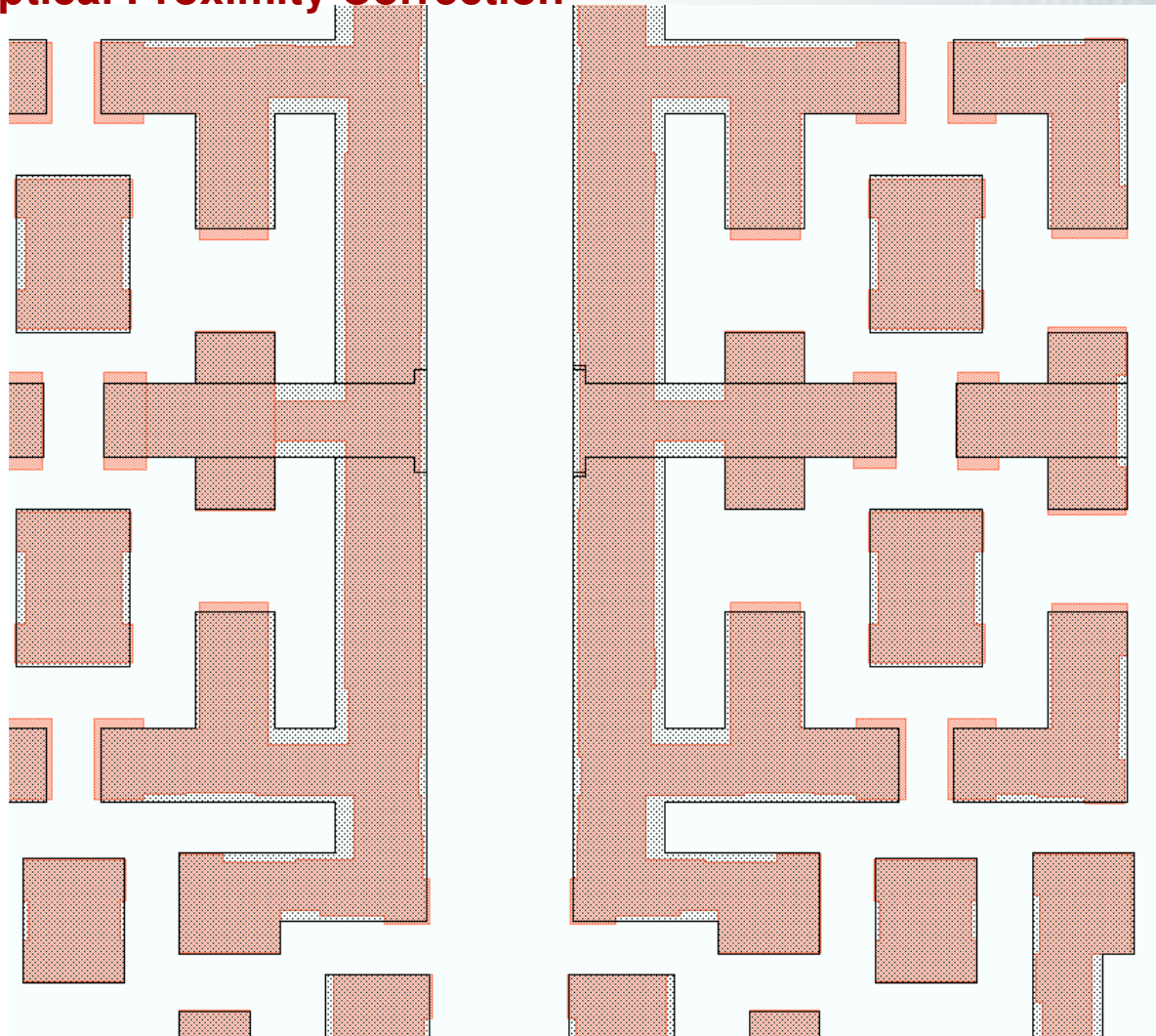
This is what the designer would like to see on the wafer

Need for Optical Proximity Correction



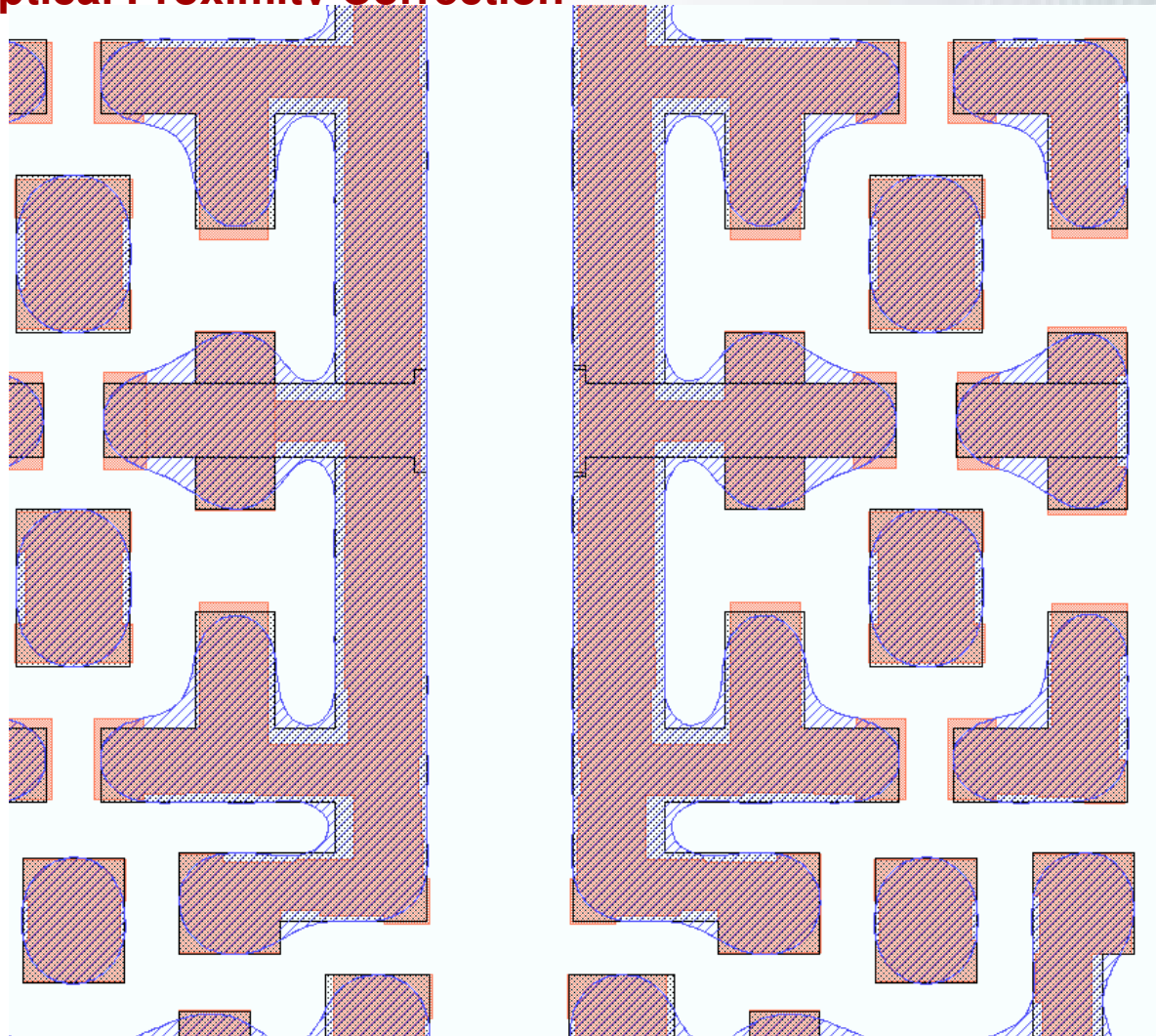
The brown hatched area is what would be on the wafer if the design is not corrected.

Need for Optical Proximity Correction



The solid shapes are the result of Optical Proximity Correction.

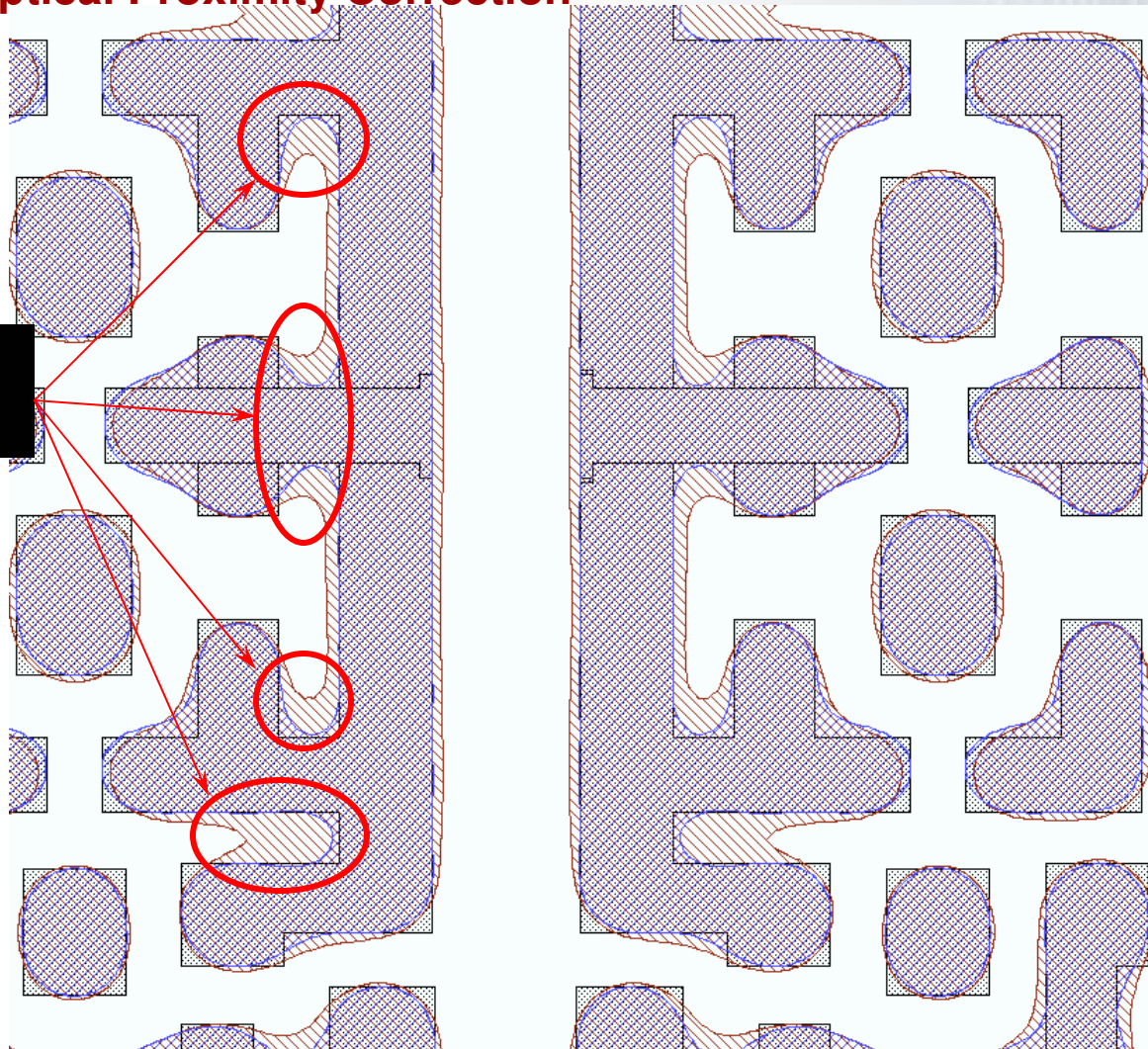
Need for Optical Proximity Correction



The blue hatched area shows what would be on the wafer when the Optical Proximity Corrected shapes are used to make the mask.

Need for Optical Proximity Correction

Wafer image now more closely matches the designer's intent.



Comparison of the pre-OPC (brown hatched), post-OPC (blue hatched) wafer images, and design intent (shaded gray).

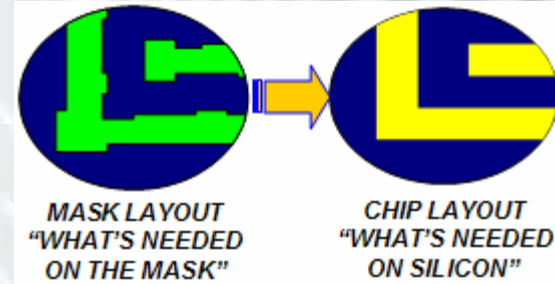
Hardware Acceleration of OPC

Mentor Graphics + Mercury Computer Systems + IBM Partnership

A comprehensive 3 company value proposition

The value of the whole is greater than the sum of the parts

- Calibre® nmOPC
 - Dense image simulation
 - Co-Processor Acceleration
 - Hierarchy engine
 - New resist process model
 - Process window correction
 - Design intent awareness
- Application software support



- MultiCore™ Plus middleware
- Software integration
- Algorithm optimization
- FFT library optimization
- CPA cluster integration and test
- Performance and tuning optimization
- IBM HW and HW support sale



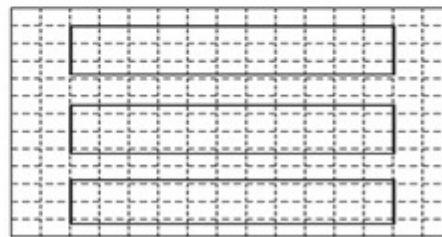
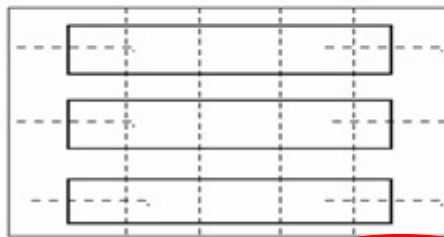
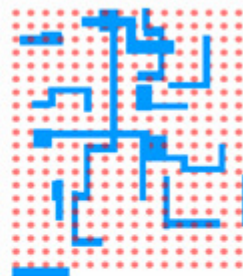
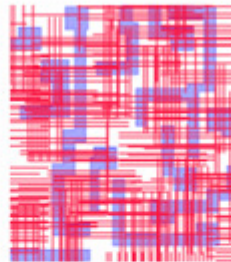
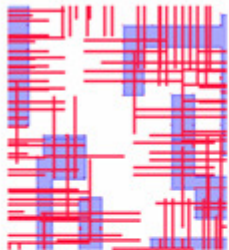
- Cell Broadband Engine™ Architecture
- HW warranty and single-point-of-service
 - ⇒ CPA standalone, or
 - ⇒ Hybrid x86 + CPA cluster system
- High density computing
- Data center services

Mentor Calibre® nmOPC - Convergence of Technologies: *Dense Imaging Simulation and Hardware Acceleration for OPC*

Why Transition from Sparse to Dense Imaging?

Grid-based simulation more efficient with increasing layout density

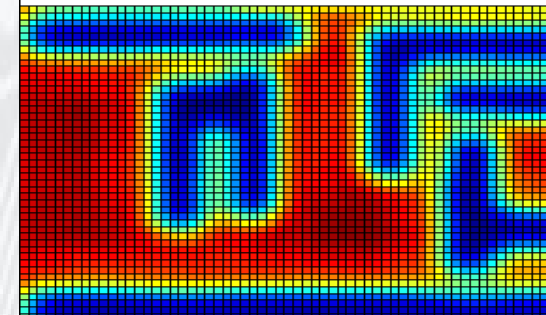
65nm sparse simulation 45nm sparse simulation 45nm dense simulation



10 Sites/Shape; 15 simulations/site: 450 simulations Grid-based simulation sites: 182 simulation sites

OPCpro

nmOPC

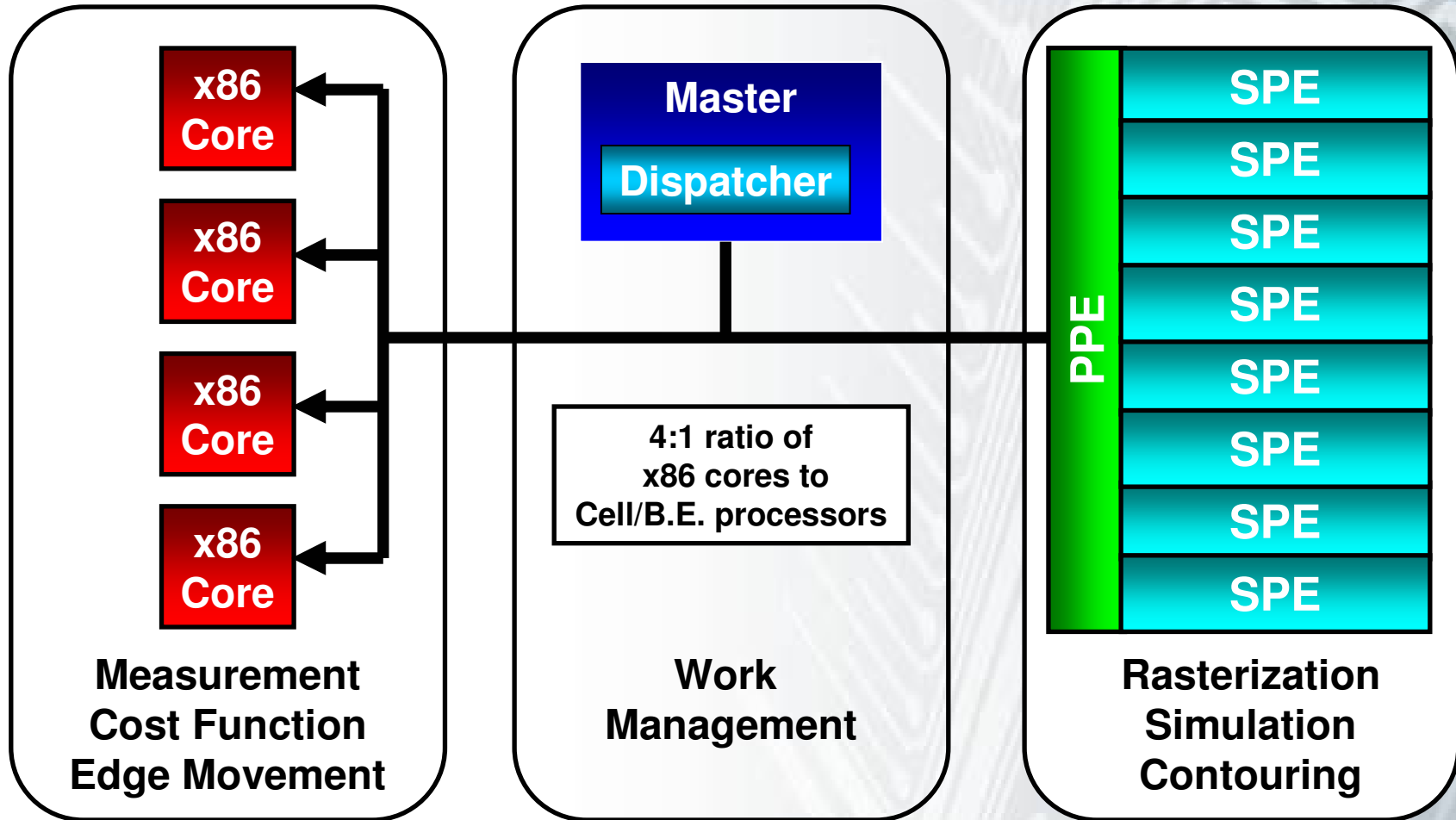


Dense vs. Sparse Imaging

- Computational efficiency
- Better model accuracy
- Support more complex shapes
- Designer intent

OPC Computational Acceleration: How it Works

Implementation by and

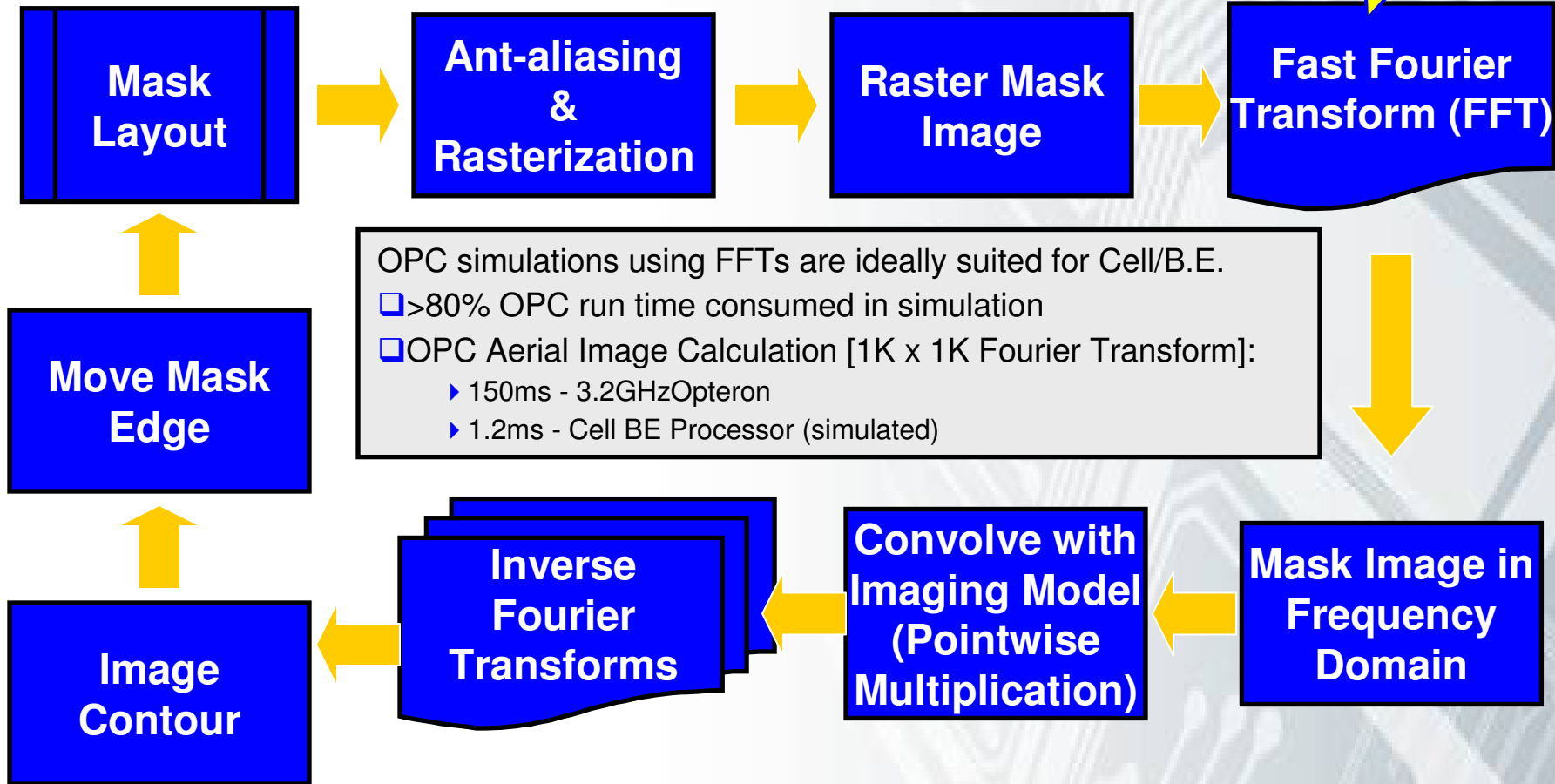


Enabled with Mercury's MultiCore Framework (*intra-Cell*) and Parallel Acceleration System (*inter-Cell*) middleware

How OPC Works in Dense Imaging

Iterative Model-based OPC: A Great Fit for Cell/B.E.

**Up to 100X
Speedup**



Example Higher Density Computing Performance / Sq. Ft. / Watt / \$

**Typical 4:1 ratio
of x86 cores to
Cell/B.E. processors**

Without CPA

7.4X GFLOPs/Watt

With CPA

7.2X GFLOPs/Sq. Ft.

1U

912 x86 Linux Cores
228 1U 2S Dual-Core Servers
5.43 Racks
90.5 KW = 16.7 KW/rack
(397 W/server)
~5,472 SP GFLOPS

256 x86 Linux Cores
64 1U 2S Dual-Core Servers
1.52 Racks
25.4 KW = 16.7 KW/rack
(397 W/server)
~1,536 SP GFLOPS

64 Cell BE Processors
32 Cell BE Blades
3 BladeCenter H Chassis
0.64 Rack
12.6 KW = 19.7 KW/rack
(5,527 W/chassis = 395 W/server)
~13,120 SP GFLOPS



	<u>From</u>	<u>To</u>	<u>Advantage</u>
SP GFLOPS	5,472	14,656	+9,184 (+168%)
KW	90.5	38.0	-52.5 (-58%)
Racks	5.43	2.16	-3.27 (-60%)
x86 Cores	912	256	656 Freed Up

Example Higher Density Computing Performance / Sq. Ft. / Watt / \$

**Typical 4:1 ratio
of x86 cores to
Cell/B.E. processors**

Without CPA

6.4X GFLOPs/Watt

7.1X GFLOPs/Sq. Ft.

With CPA

1U

104 x86 Linux Cores
26 1U 2S Dual-Core Servers
0.62 Racks
10.3 KW = 16.6 KW/rack
(397 W/server)
~624 SP GFLOPS

104 x86 Linux Cores
26 1U 2S Dual-Core Servers
0.62 Racks
10.3 KW = 16.6 KW/rack
(397 W/server)
~624 SP GFLOPS

26 Cell BE Processors
13 Cell BE Blades
1 BladeCenter H Chassis
0.21 Rack
5.1 KW = 24.3 KW/rack
(5,527 W/chassis = 395 W/server)
~5,330 SP GFLOPS

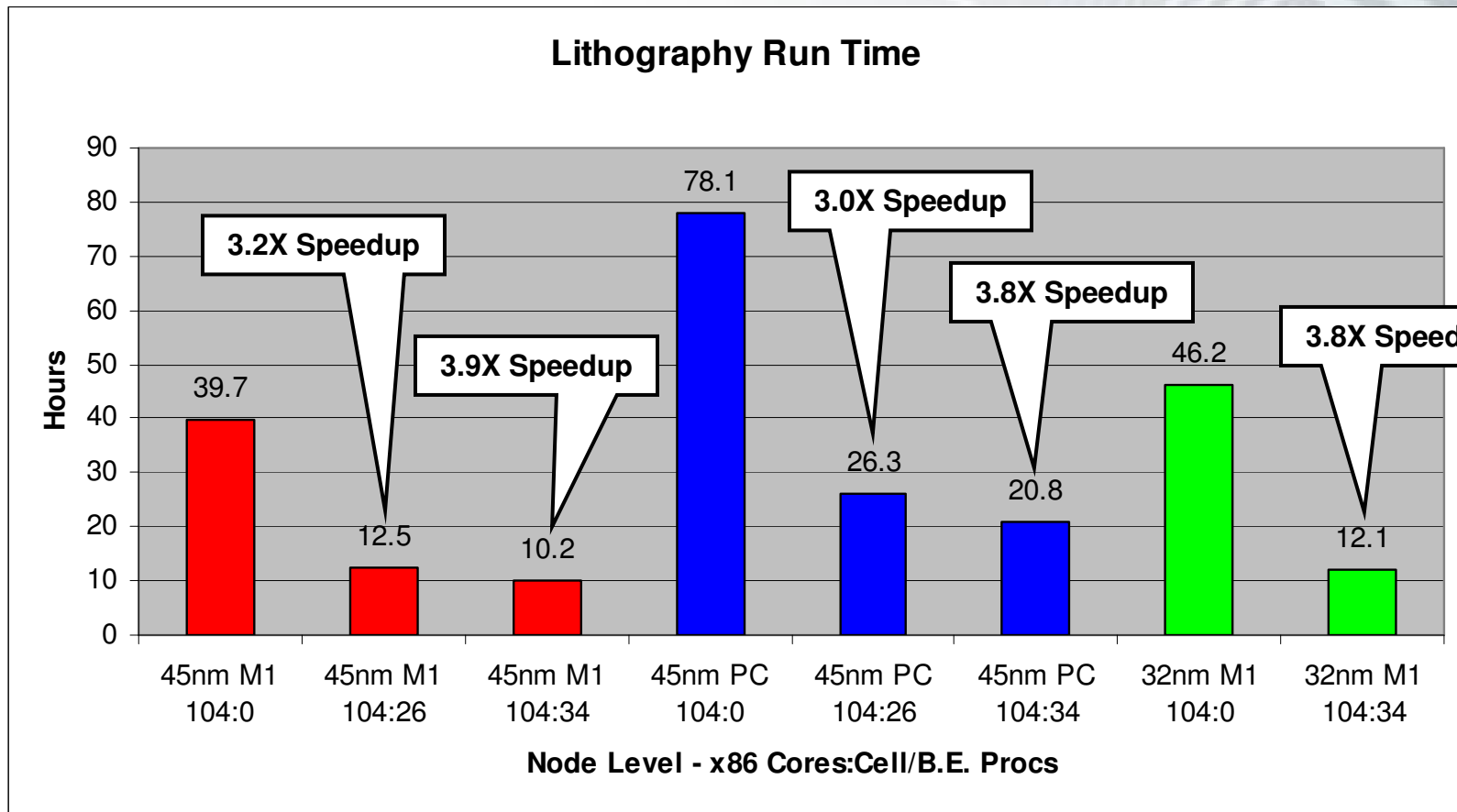


3.2X Speedup



	<u>From</u>	<u>To</u>	<u>Advantage</u>
SP GFLOPS	624	5,954	+5,527 (+854%)
KW	10.3	15.4	+5.1 (+49%)
Racks	0.62	0.83	+0.21 (+34%)

Calibre® nmOPC - Lithography Run Time and Speedup



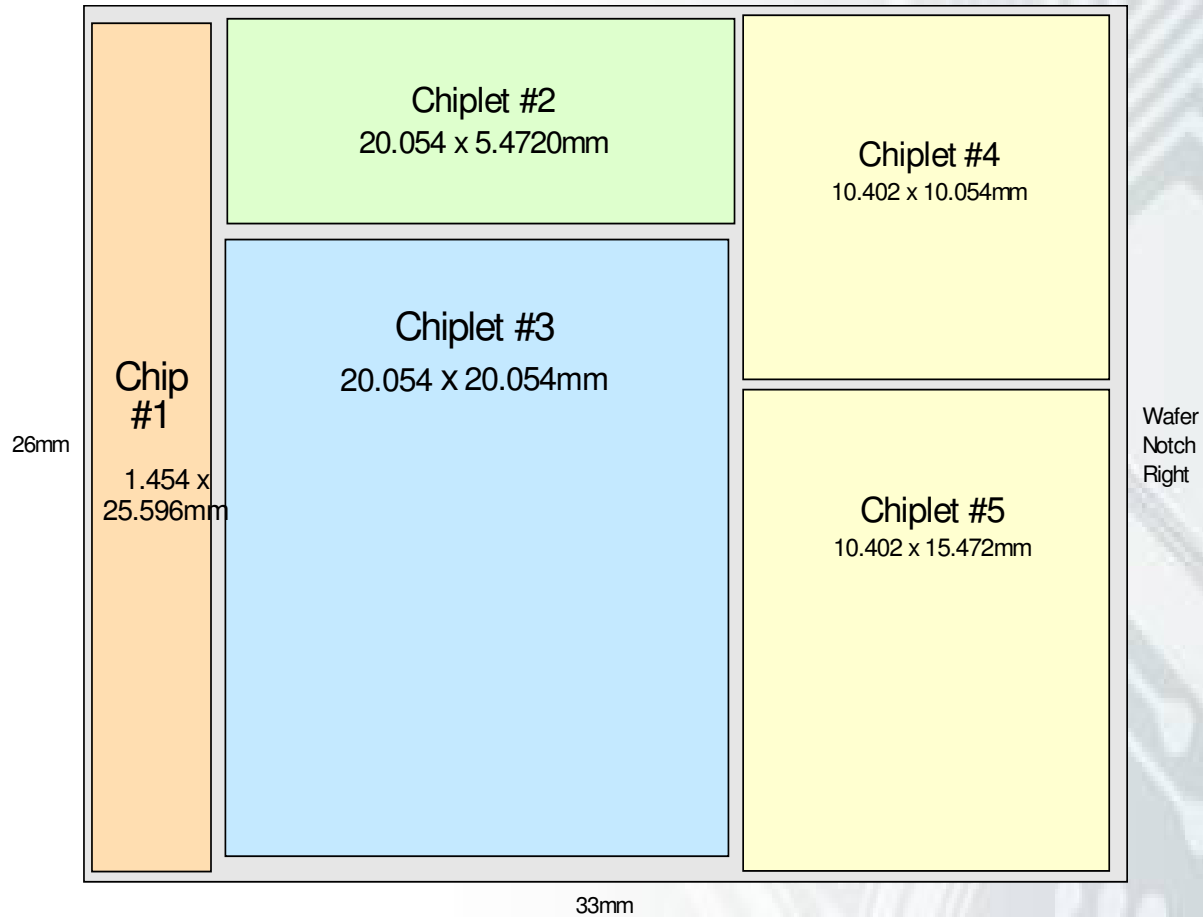
Some of our very first results. These were run at the Mentor Graphics Corporation facility in Marlborough MA while we were in the process of bringing up our installation at IBM Burlington.

IBM's Computational Lithography Experience Using Cell Coprocessor Acceleration (CPA)

Configuration of Accelerated Cluster

- QS21 Cell blades: 126 (Cell/B.E. processors: 252)
- x86 remote cores: Over 1000 available
 - ▶ A standard OPC run uses 120 x86 cores
- Standard ratio of x86 remote cores to Cell processors: 4:1
 - ▶ Standard accelerated run uses 30 Cell processors
 - ▶ The x86/Cell chip ratio can be varied.
- Memory per x86 remote core: 3GB
- Master System p P4 Regatta - 32-way with 256GB memory

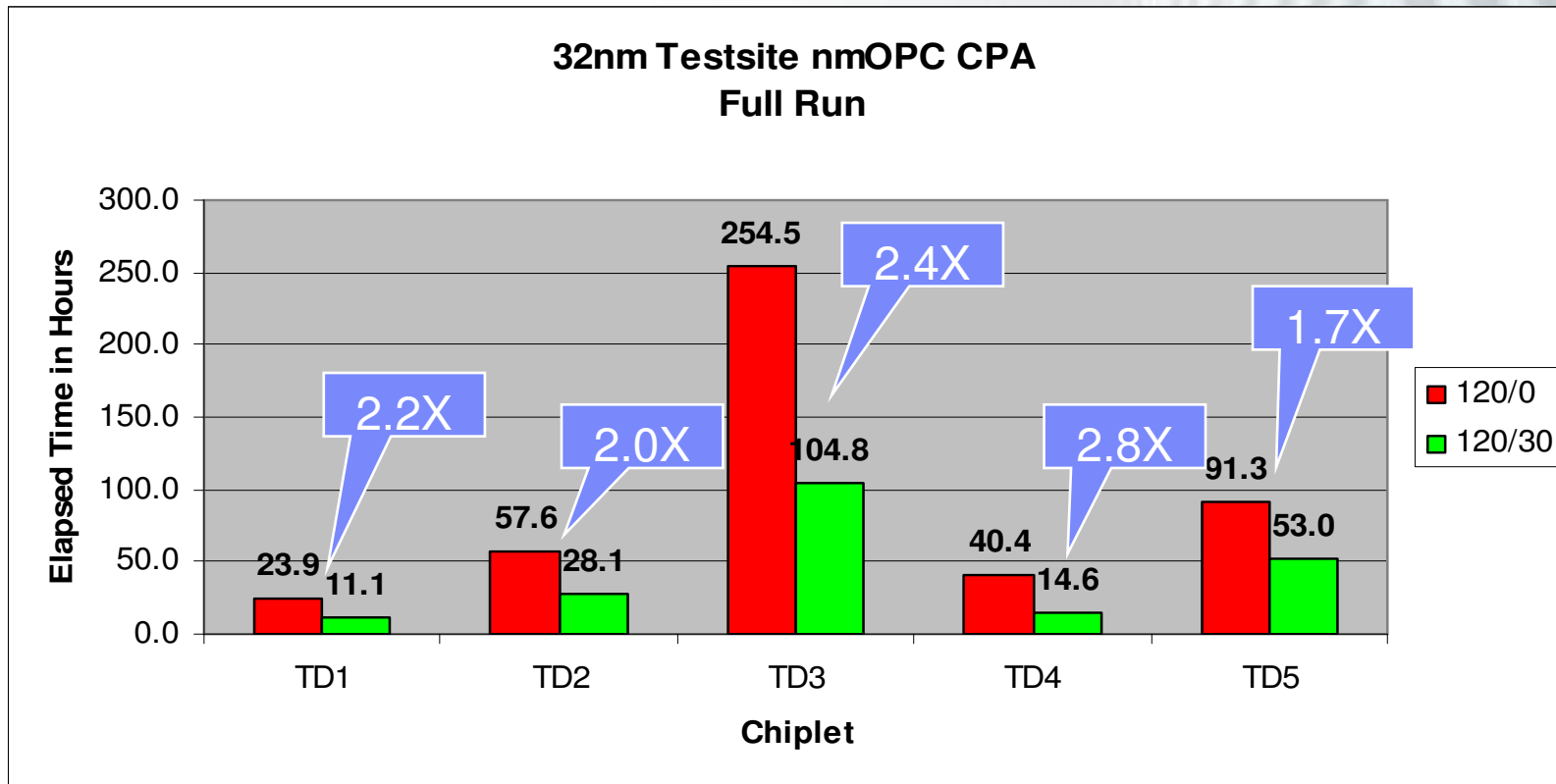
A 32nm Testsite Floor Plan



32nm Testsite's Chiplets' Total Elapsed Times

120/0 is 120 x86 cores and no Cell B.E. Chips (no acceleration)

120/30 is 120 x86 cores and 30 Cell B.E. Chips (4:1 ratio)

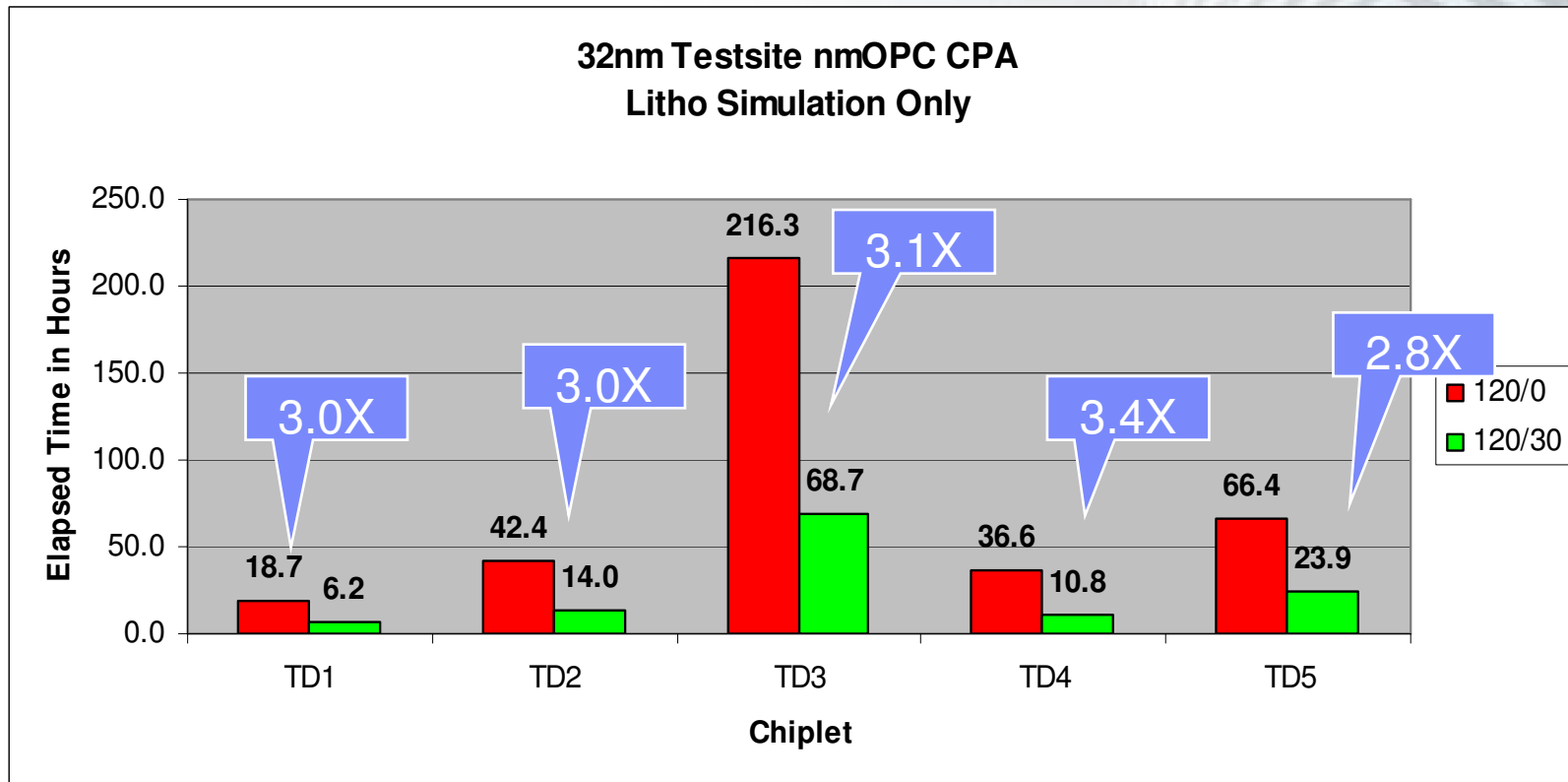


Size: ~1.5 x 26 ~20 x 5.5 ~20 x 20 ~10.4 x 10 ~10.4 x 15.5

32nm Testsite's Chipllets' Simulation Elapsed Times

120/0 is 120 x86 cores and no Cell B.E. Chips (no acceleration)

120/30 is 120 x86 cores and 30 Cell B.E. Chips (4:1 ratio)



Size: ~1.5 x 26 ~20 x 5.5 ~20 x 20 ~10.4 x 10 ~10.4 x 15.5



ERROR: stackunderflow
OFFENDING COMMAND: ~

STACK: