Tolerance-Based Wafer Verification Methodologies with a Die-to-Database Inspection System

Kohji Hashimoto, Satoshi Usui[1], Kenji Yoshida[1], Satoshi Tanaka, Toshiya Kotani, Shigeki Nojima, Ichirota Nagahama, Osamu Nagano, Yasuo Matsuoka, Yuuichiro Yamazaki, and Soichi Inoue

Process and Manufacturing Engineering Center, Semiconductor Company, Toshiba Corporation, 8 Shinsugita-cho, Isogo-ku, Yokohama 235-8522, Japan
[1]*Process and Manufacturing Engineering Center, Semiconductor Company, Toshiba Corporation, 3500 Matsuoka, Oita 870-0197, Japan*

Received August 23, 2008; accepted April 6, 2009; published online July 21, 2009

With a die-to-database inspection system using electron beam, we have constructed state-of-the-art verification methodologies for the design for manufacturability (DfM), process proximity correction (PPC), minimization of process errors and process fluctuations, and so on. The experimental methodologies make it possible to extract exact hotspots and result in short development turnaround time (TAT) in low k_1 lithography. In the methodologies, the die-to-database inspection system, NGR-2100, has remarkable features for the full-chip inspection within reasonable operating time. This system is equipped with tolerance-based "verifiers" and provides higher hotspot extraction accuracy than the conventional optical inspection tool. As a result, hotspots extracted using the system included all killer hotspots extracted by electrical and physical analyses. In addition, the new methodologies are highly advantageous in that they shorten the development TAT by two to four months. In the application to 65-nm-node complimentary metal oxide semiconductor (CMOS) devices, we verified yield improvement using the proposed methodologies. © 2009 The Japan Society of Applied Physics

DOI: 10.1143/JJAP.48.076502

1. Introduction

Turnaround time (TAT) in ultra large scale integrated (ULSI) development is one of the most important factors with regard to the acceleration of the shift toward much smaller complimentary metal oxide semiconductor (CMOS) geometries, because a short development TAT brings about steep ramp-up of LSI volume production. As shown in Fig. 1, however, the shift to low k_1 $(= \mathrm{HP} \cdot \mathrm{NA}/\lambda)$ lithography could decrease patterning fidelity and result in the generation of many hotspots; a hotspot[1)] is a device pattern that has relatively large critical dimension (CD) and patterning feature errors with respect to the targets on wafers. Hotspots could arise owing to uncertain design for manufacturability (DfM) and process proximity correction (PPC), process errors and process fluctuations, and so on. Verifications of drawn LSI layouts, PPC and LSI fabrication processes with hotspots constitute one of the key issues concerning the achievement of shorter development TAT in low k_1 lithography, because hotspot eradiation is time-consuming. Hotspot extraction in the verifications, therefore, has required accurate evolutional methodologies with short operating time. In this paper, an innovative approach is described for the verifications. By contributing to the minimization of ULSI development lead time, the methodologies could help accomplish the maximum process-related yield.

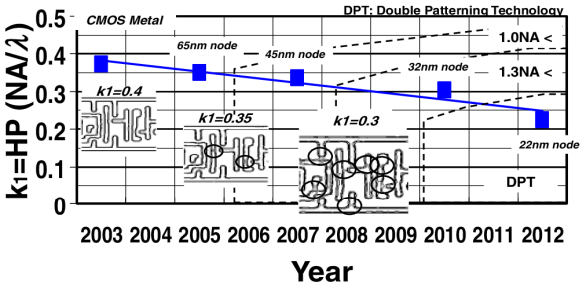


Fig. 1. (Color online) Trends of k_1 factor provided by minimum half pitch (HP), wavelength (λ), and numerical aperture (NA) of an exposure tool for each CMOS technology node. Lower k_1 induces many hotspots.

2. Theory

2.1 Tandem PPC

For the achievement of both accurate linewidth control and rapid correction remodeling, we proposed a tandem process proximity correction[2)] (Tandem PPC) at the SPIE Symposium on Optical Microlithography in 2002, as shown in Fig. 2. The flow on the left side is Tandem PPC data-processing flow and that on the right side is pattern fabrication process flow on wafers. Mask fabrication, lithography and etching on wafers are described in the tandem PPC flow. In this method, the proximity corrections for unit processes are carried out in reverse order of process events as follows. Firstly, linewidths and top-down images of the design data are process-proximity-corrected for wafer etching (etching PPC) to ensure good agreement with those on the final patterns on wafers, and the etching PPC generates data for resist linewidths and top-down images. The generated data for resists are the intended target for the lithography PPC. Secondly, the data generated by the etching PPC are process-proximity-corrected for wafer lithography (lithography PPC) to ensure good agreement with those on the resist patterns on wafers, and the lithography PPC generates data for masks linewidths and top-down images. The generated data for mask are the intended target for the mask PPC. Finally, the data generated by the lithography PPC are process-proximity-corrected for mask fabrication (mask PPC) to ensure good agreement with those on the mask patterns. These three PPCs, namely, etching PPC, lithography PPC, and mask PPC, are carried out in tandem at EB data processing. Also, each correction model is obtained individually with traces of experimental unit process proximity effect (PPE).

 © 2009 The Japan Society of Applied Physics

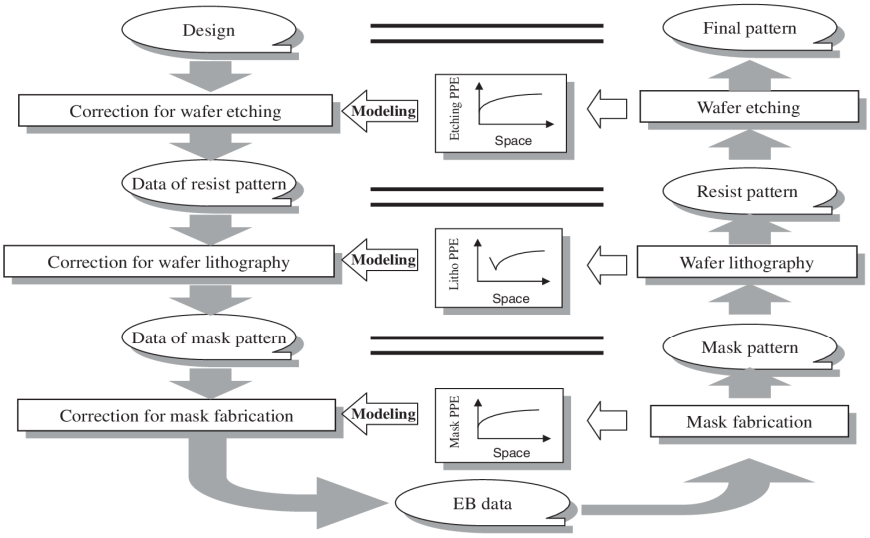


Fig. 2. Tandem PPC flow. In this method, the proximity corrections for unit processes are carried out in reverse order of process events. Mask fabrication, lithography on wafers, and etching on wafers are described.

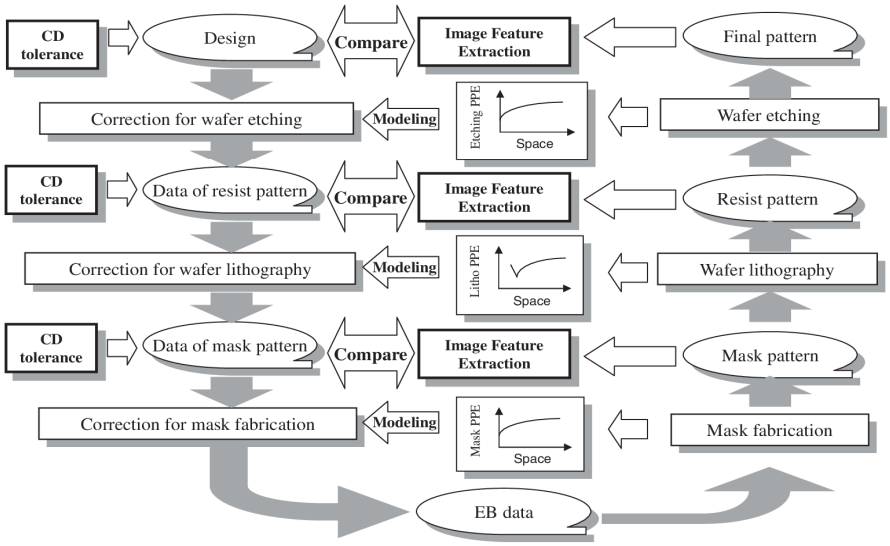


Fig. 3. Tolerance-based PPC verification flow in Tandem PPC. The flow on the left side is Tandem PPC data-processing flow and that on the right side is pattern fabrication process flow on wafers.

2.2 Tolerance-based verification

At the SPIE Symposium on Optical Microlithography in 2003, we proposed the Tandem PPC verification methodology with image feature extraction.[3)] Tandem PPC accuracy is verified from the agreement between extracted image features and target pattern features in computer aided design (CAD) data at each unit process. Figure 3 shows the tolerance-based PPC verification flow[4)] in Tandem PPC reported at Photomask Japan in 2004. The flow presents the tolerance-based PPC verification methodology with target features including CD tolerance for the judgment criteria in each process step.

Figure 4 shows the concept and procedure of the tolerance-based PPC verification methodology. First, we prepare the target feature at each unit process. Next, we generate CD tolerance data (inner and outer tolerance data) for each unit process step in CAD data, and input the extracted data to the system. These tolerances vary with pattern features (e.g., linewidth and line end), process steps, and layers. Then, we extract image features from scanning electron microscope (SEM) and overlap them into the target features in CAD data on the system. Finally, we compare the extracted pattern image features and the target features in designed CAD data including CD tolerance at any hotspots in the system.

© 2009 The Japan Society of Applied Physics

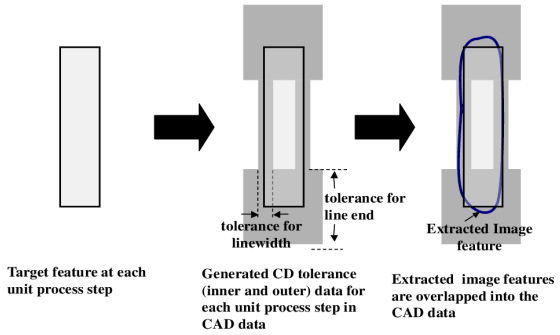


Fig. 4. (Color online) Concept and procedure of tolerance-based PPC verification.

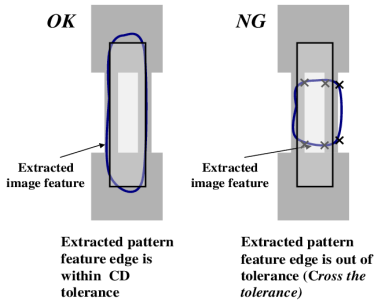


Fig. 5. (Color online) How to verify the PPC.

Figure 5 shows how the PPC accuracy is verified by this method. If the extracted pattern feature edge is within CD tolerance, this PPC is judged to be sufficiently accurate. On the other hand, if the extracted pattern feature edge exceeds the tolerance, PPC is judged to be insufficiently accurate; this means that the edge crosses the tolerance lines. The concept of tolerance-based PPC verification is also available for the verification of design for manufacturability (DfM), process errors, and process fluctuations, not only PPC.

2.3 Die-to-database wafer verification methodologies with patterned feature tolerances

Figure 6 shows an example of die-to-database inspection. Target features such as GDS-formatted design data are overlapped to processed patterning features on wafers. The patterning features whose deviations from target features are out of tolerance are extracted as hotspots. This methodology is reasonable for the hotspot extraction, and minimizing the deviations allows hotspot reduction in drawn LSI layouts, PPCs, LSI fabrication, and so on. Figure 7 shows how the die-to-database inspection system is used to identify hotspots. To avoid omissions in hotspot extraction, full-chip inspection within reasonable operating time is necessary. Therefore, both short operating time and accuracy for hotspot extraction are important factors. For the short operating time and extraction accuracy, high-speed image processing and high-resolution EB inspection are required, respectively. The die-to-database inspection system, NGR-2100,[5,6] has features satisfying the above requirements. Hardware and software for the NGR-2100, which is shown

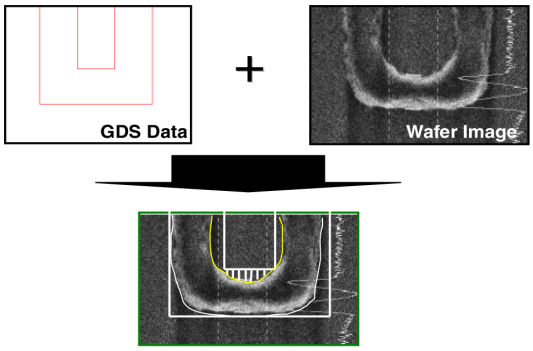


Fig. 6. (Color online) An example of die-to-database inspection.

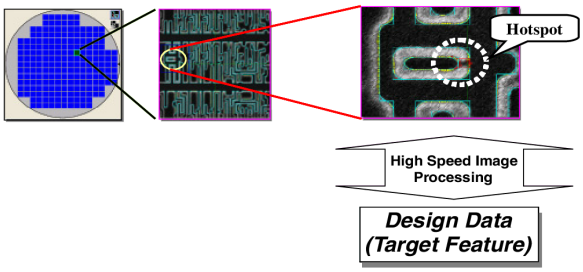


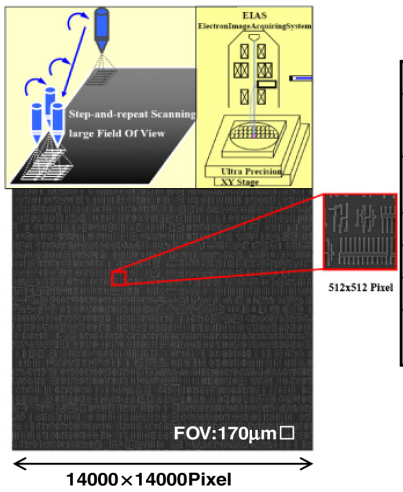
Fig. 7. (Color online) How the die-to-database inspection system is used to identify hotspots.

Fig. 8. (Color online) Photograph of NGR-2100.

in Fig. 8, are produced by NanoGeometry Research and Topcon, respectively.

Figure 9 summarizes the hardware configurations compared with CD-SEM. To accomplish high inspection throughput, a larger field of view (FOV) and higher probe current than those of CD-SEM are implemented. The inspection time of 3.5 to 6.5 h for a 100 mm^2 chip is realized with the hardware condition shown in Fig. 9. Figure 10 shows the "verifier" software function for extracting hot-

© 2009 The Japan Society of Applied Physics



	CD–SEM	NGR2100
FOV	1 μm (512pixel)	60–170 μm (14000pixel)
Landing Energy	~1000eV	~2600eV
Resolution	2nm	3–12nm
Probe Current	10pA	1–10nA
Scan Rate	7.5MHz	100MHz
Accumulation Image	16–32Times	1–2Times

Fig. 9. (Color online) Hardware configurations compared with CD-SEM.

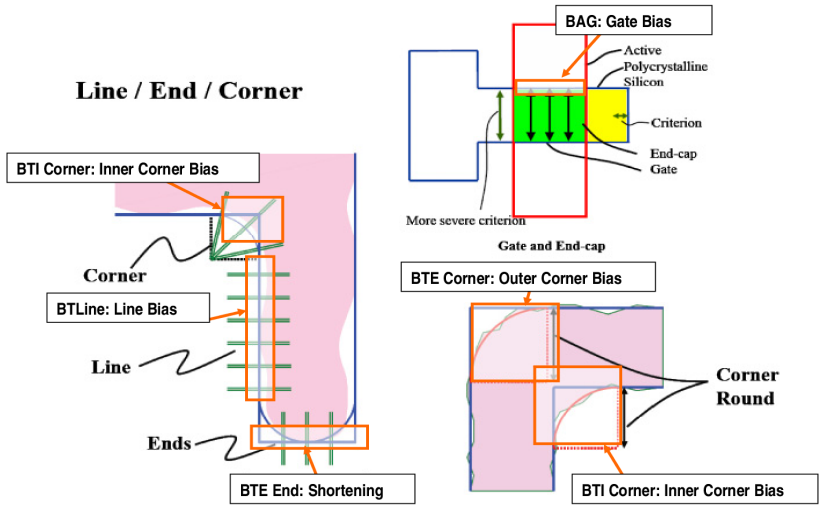


Fig. 10. (Color online) Tolerance-based verification "verifiers".

spots. Five main verifiers are currently available for hotspot extraction: inner corner bias, outer corner bias, line bias, shortening, and gate bias. Set biases in verifiers can be varied freely in accordance with tolerance in each process step.

Figure 11 shows the state-of-the-art wafer verification flow in LSI fabrication with the die-to-database inspection system. Hotspots extracted experimentally using the die-to-database inspection system are used to judge the adopted patterning processes: design, mask data processing (MDP), mask fabrication, and wafer processes (lithography and etching). If the extracted hotspots are considered unsuitable to proceed to the next process steps, some unit processes for pattering are reoptimized until the hotspots are eliminated.

3. Experimental Results and Discussion

3.1 Hotspot extraction accuracy and development turnaround time

Before the discussion on hotspot extraction accuracy, we

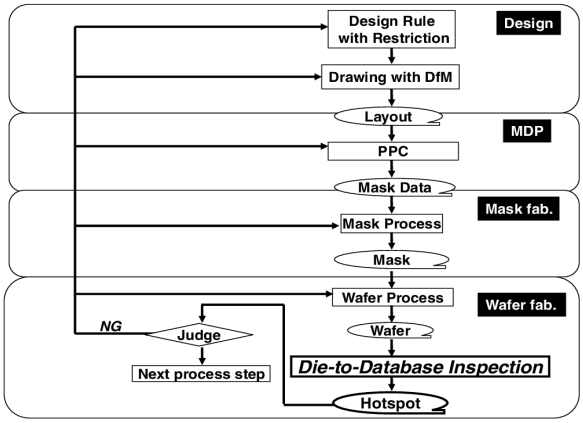


Fig. 11. State-of-the-art verification flow in actual LSI fabrication with the die-to-database inspection system.

© 2009 The Japan Society of Applied Physics

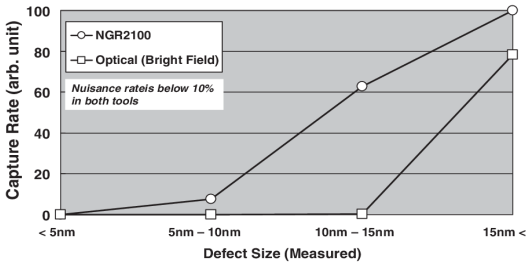


Fig. 12. Defect capture rates for the NGR-2100 and the conventional optical inspection tool. Nuisance rate is below 10% in both tools.

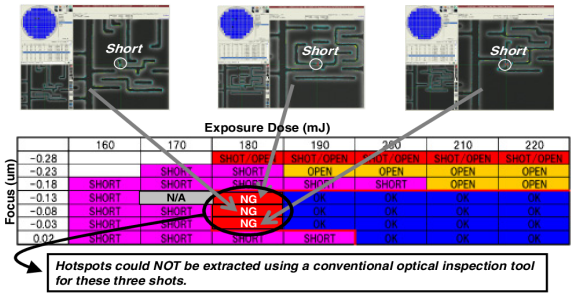
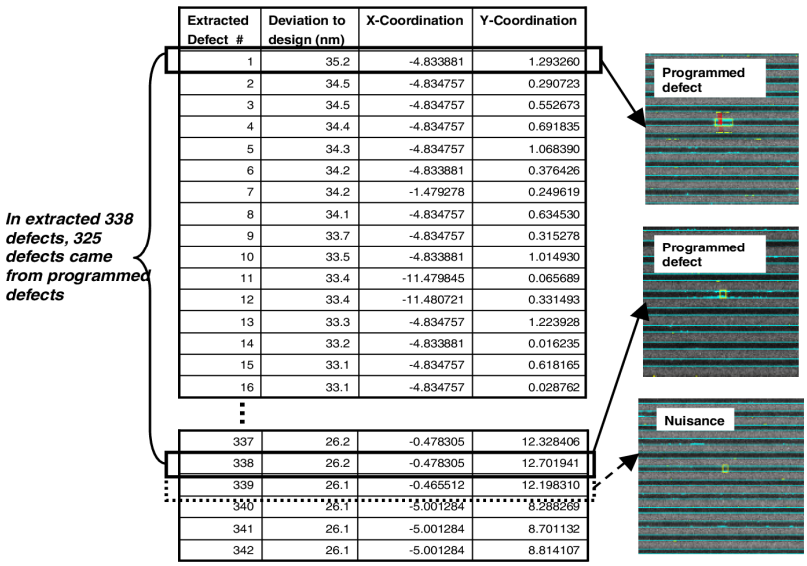


Fig. 14. (Color online) Verifications in hotspot extraction accuracy of the NGR-2100 compared with the conventional optical inspection tool on 65-nm-node CMOS local metal.

compared defect detection abilities between NGR-2100 and the conventional optical inspection tool. For this investigation, an experimental test mask involving the programmed defects was fabricated. The programmed defects of various sizes, including both opaque and clear defects, were placed at the pattern edges of 32 nm line-and-space (L/S) patterns on the mask. The formation of 32 nm L/S patterns on the wafer was performed using a spacer patterning technology with an immersion ArF scanner with 1.0 numerical aperture (NA). Figure 12 shows the defect capture rate comparison (arb. unit) for both inspection tools. Each inspection recipe was optimized for real defect detection with low nuisance rate (below 10%). The data provides that the NGR-2100 has demonstrated a higher capture rate than that of the conventional optical inspection tool. Figure 13 shows the list of defects extracted defects with the NGR-2100 using the programmed defect mask. In the 338 extracted defects, 325 defects came from programmed real defects and 13 defects were other real defects. The extracted defect #339 was a nuisance defect. Therefore, the nuisance defect detection rate could be further reduced with reasonable criteria such as the threshold between defects #338 and #339. From both Figs. 12 and 13, it can be concluded that the NGR-2100 has remarkable defect detection ability compared with the conventional optical inspection tool.

Using high-resolution EB inspection, we demonstrated the accuracy of hotspot extraction. Figure 14 shows verifications in hotspot extraction accuracy of the NGR-2100 compared with the conventional optical inspection tool. Comparison was carried out using focus exposure matrix wafers (FEM) on the first metal layer (Cu damascene metal) in 65-nm-node CMOS device. From the results, hotspots arise markedly on defocused and underdosed chips, and the boundary between usable processed chips (with "OK" script) and failed chips (with "OPEN" or "SHORT" script) is critical for extracting hotspots. On three underdosed chips (with "NG" script), hotspots (patterning short) were not extracted using the conventional optical inspection tool. However, the NGR-2100 could accurately extract hotspots.

We compared the results of hotspot extraction using NGR-2100 with the results of electrical and physical analyses on the first metal layer (Cu damascene metal) in 65-nm-node CMOS device as shown in Fig. 15. Shown on the left side in Fig. 15 are killer hotspots that directly affect



Extracted Defect #	Deviation to design (nm)	X-Coordination	Y-Coordination
1	35.2	-4.833881	1.293260
2	34.5	-4.834757	0.290723
3	34.5	-4.834757	0.552673
4	34.4	-4.834757	0.691835
5	34.3	-4.834757	1.068390
6	34.2	-4.833881	0.376426
7	34.2	-1.479278	0.249619
8	34.1	-4.834757	0.634530
9	33.7	-4.834757	0.315278
10	33.5	-4.833881	1.014930
11	33.4	-11.479845	0.065689
12	33.4	-11.480721	0.331493
13	33.3	-4.834757	1.223928
14	33.2	-4.833881	0.016235
15	33.1	-4.834757	0.618165
16	33.1	-4.834757	0.028762
⋮			
337	26.2	-0.478305	12.328406
338	26.2	-0.478305	12.701941
339	26.1	-0.465512	12.198310
340	26.1	-5.001284	8.288269
341	26.1	-5.001284	8.701132
342	26.1	-5.001284	8.814107

Fig. 13. (Color online) Extracted defects list obtained with the NGR-2100 using the programmed defect mask.

© 2009 The Japan Society of Applied Physics

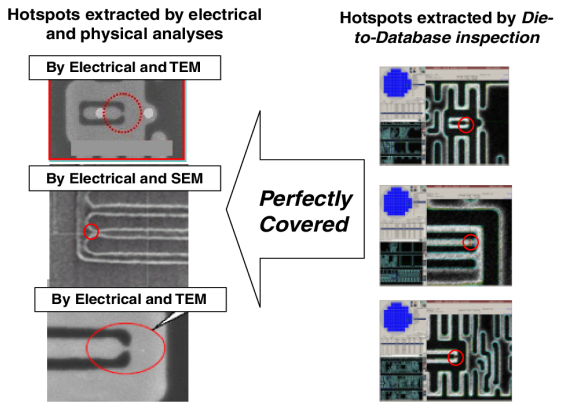


Fig. 15. (Color online) Results of hotspot extraction using NGR-2100 compared with the results of electrical and physical analyses on the first metal layer (Cu damascene metal) in 65-nm-node CMOS device.

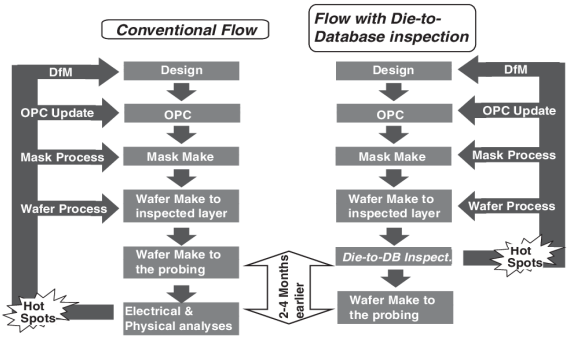


Fig. 16. Development TAT comparison between conventional flow and flow with die-to-database inspection.

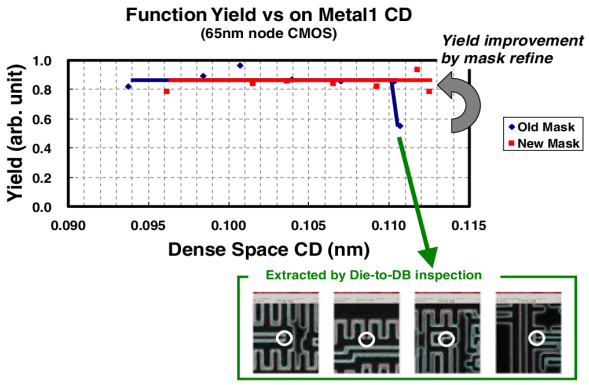


Fig. 17. (Color online) Device yields to process window with respect to variable exposure dose condition.

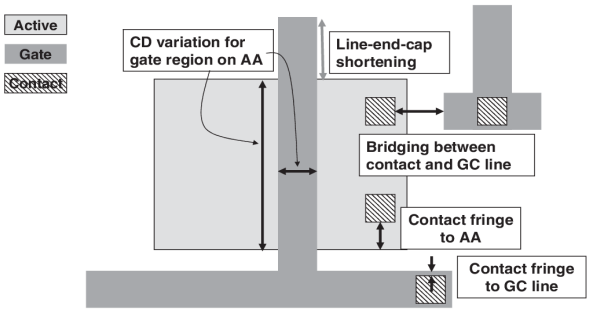


Fig. 18. Examples of interlayer hotspots among active area (AA), gate conductor (GC), and contact levels at front-end-of-line (FEOL).

device characteristics. After full wafer process flow, those hotspots were extracted by electrical or physical analysis such as SEM or transmission electron microscopy (TEM). It took two to four months to indicate these hotspots after this layer patterning. Shown on the right in Fig. 15 are hotspots extracted using NGR-2100. Hotspots extracted using NGR-2100 included all the killer hotspots extracted by electrical or physical analysis. These results prove that NGR-2100 is a promising system for detecting killer hotspots. Also, it should be noted that hotspot extraction using NGR-2100 only requires patterning wafers for inspecting layers, not fully processed wafers. This means that the exact hotspots could be detected two to four months earlier than in the case of using the conventional electrical or physical methodologies, as shown in Fig. 16. Therefore, the hotspot extraction using NGR-2100 is highly advantageous for shortening development TAT.

3.2 Application 1: Feed to design refinement

We applied the above flow to the first metal layers (Cu damascene metal) on 65-nm-node CMOS device. Device yields to process window were verified with respect to variable exposure dose condition as shown in Fig. 17. In the first trial (old mask), the yield decreased under overexposure dose condition because several killer hotspots considerably contribute to this degradation. These killer defects were all extracted using the NGR-2100 under this overdose condition, and the extracted hotspots were reduced on the basis of feedback of design refinements to the new mask fabrication. With the new mask, the function yield could be improved under overexposure dose condition.

3.3 Application 2: Interlayer hotspot extraction

In addition to short operating time and accuracy for hotspot extraction, the implementation of interlayer hotspot extraction may be desirable from the viewpoint of LSI fabrication. Figure 18 shows examples of interlayer hotspots among active area (AA), gate conductor (GC), and contact levels at front-end-of-line (FEOL). In the relationship between AA and GC, CD variation for gate region on AA and line-end-cap shortening should be investigated in terms of device performance. Regarding contact layers, moreover, contact fringe to AA, contact fringe to GC line, and bridging between contact and GC line should be monitored to obtain higher process yield. Because GDS data is used, die-to-database inspection can easily extract interlayer hotspots. Figure 19 shows the flow for interlayer hotspot extraction using die-to-database inspection. Firstly, wafer image (first layer) extraction was carried out by die-to-database inspection. Secondly, the extracted wafer image was transferred to

© 2009 The Japan Society of Applied Physics

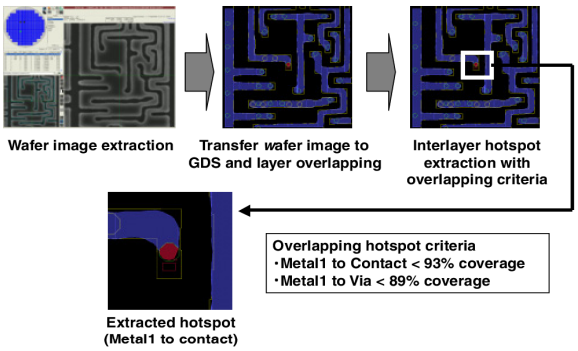


Fig. 19. (Color online) Flow for interlayer hotspots extraction using die-to-database inspection for 65-nm-node CMOS local metal to contact or via.

GDS data, and the second layer (GDS) overlapped the first layer. Finally, interlayer hotspots were extracted with overlapping criteria. Figure 19 also shows the interlayer hotspot between local metal and contact or via in 65-nm-node CMOS. In this example, overlapping hotspot criteria were below 93% and below 89% for contact and for via, respectively.

4. Conclusions

We constructed tolerance-based wafer verification methodologies with a die-to-database inspection system that is required for both hotspot extraction accuracy and short development TAT. The NGR-2100 is a suitable EB inspection system that, compared with conventional optical tools, offers wider FOV, higher landing energy, higher scan rate, and higher probe current to achieve both accuracy and short operating time. The evolutional methodologies with die-to-database have been applied to 65-nm-node CMOS device. As a result, we verified the function yield improvement with the new methodologies. Also, we verified that the die-to-database inspection system is highly advantageous in the sense that it shortens development TAT by two to four months compared with the conventional electrical or physical analysis flow. Also, the die-to-database inspection system demonstrates excellent interlayer hotspot extraction from the viewpoint of LSI fabrication.

Acknowledgments

The authors would like to express their thanks to the following people: Mr. M. Yamamoto, Mr. T. Kitamura, Mr. Y. Okubo, Mr. H. Sakaizawa, Mr. J. Ueda, and Mr. T. Hasebe of NanoGeometry Research Inc. for fruitful discussion and experimental support; Mr. M. Inoue of Topcon Corporation, for hardware support; Mr. K. Ehara, Mr. K. Takahata, and Mr. Y. Hagio of Process and Manufacturing Engineering Center Semiconductor Company, Toshiba Corporation, for experimental data analysis; Mr. K. Okuda of Corporate Manufacturing Engineering Center, Toshiba Corporation, for supporting in software development; and Ms. R. Sakai of Meitec Corporation for experimental assistance.

1) S. Nojima, S. Mimotogi, M. Itoh, O. Ikenaga, S. Hasebe, K. Hashimoto, S. Inoue, M. Goto, and I. Mori: Proc. SPIE **4889** (2002) 187.
2) K. Hashimoto, T. Kuji, S. Tokutome, T. Kotani, S. Tanaka, and S. Inoue: Proc. SPIE **4691** (2002) 1070.
3) K. Hashimoto, T. Ito, T. Ikeda, S. Nojima, and S. Inoue: Proc. SPIE **5040** (2003) 1156.
4) K. Hashimoto, H. Fujise, S. Nojima, T. Ito, and T. Ikeda: Proc. SPIE **5446** (2004) 471.
5) T. Kitamura, K. Kubota, T. Hasebe, F. Sakai, S. Nakazawa, N. Vohra, M. Yamamoto, and M. Inoue: Proc. SPIE **5756** (2005) 73.
6) T. Kitamura, T. Hasebe, K. Kubota, F. Sakai, S. Nakazawa, D. Lin, M. Hoffman, M. Yamamoto, and M. Inoue: Proc. SPIE **6518** (2007) 651834.

© 2009 The Japan Society of Applied Physics