# Characterization of Contact Module Failure Mechanisms for SOI Technology using E-beam Inspection and In-line TEM

Xing J. Zhou[1], Oliver D. Patterson[1], Woo-hyeong Lee[1], Hyoung H. Kang[1], Roland Hahn[2]
[1]IBM Semiconductor Research and Development Center
2070 Route 52, Mail Stop: 46H, Hopewell Junction, NY 12533 USA
[2] KLA-Tencor Corp
20 Corporate Park Road, Suite C, Hopewell Junction, NY 12533 USA

*Abstract* – **Electron-beam inspection (eBI) of the contact (CA) module for silicon-on-insulator (SOI) technology is discussed in this paper. Voltage contrast is used to detect CA opens in the SRAM and both CA opens and shorts in special test structures. The inspection is performed after the tungsten chemical mechanical planarization (W CMP) step. In-line transmission electron microscope (TEM) samples at select defect sites are then prepared and imaged to determine the failure mechanism. This methodology greatly enhances yield learning and provides quick feedback for lithography and etch process adjustments.**

## I. INTRODUCTION

Electron beam (e-beam) inspection is widely used to detect electrical defects such as opens or shorts underneath the wafer surface and tiny physical defects that are beyond the resolution limit of optical systems [1-5]. The most common application for eBI is at the CA module, where the wafer is scanned right after the W CMP process. This is because CA is one of the most challenging elements to implement for each new technology and because the CA module has a number of characteristics that make it a perfect application for eBI. Defects are generally hidden under the surface so bright-field inspection has limited usefulness. The wafer surface is planarized, removing topography constraints which would otherwise add complexity to eBI. Furthermore for bulk technologies, the large N well underneath SRAM PFET regions provides excellent grounding, resulting in a very strong signal for PFET CA opens and NFET CA shorts. In contrast to bulk technology, silicon-on-insulator (SOI) technology does not enjoy this third advantage for eBI; it does not contain natural grounding. Almost all the circuitry is floating over a buried oxide layer. This of course provides major performance benefits [6,7], but makes eBI much more difficult.

In this paper, we will discuss application of eBI to the CA module for SOI technology. Examples will be drawn from a recent SOI technology. With each new technology, the CA critical dimension (CD) shrinks proportionally to the gate length design rule and the depth to diameter aspect ratio increases. Often the CAs are printed a little bit larger than designed dimension to gain robustness [8]. If the CAs are printed too big, however, they will short to the poly gate. On the other hand, if they are too small, they will etch-stop, causing open problems. Locating a process window for defect-free CA formation is thus difficult.

Opens for SOI technology may still be detected with an array inspection of the SRAM, but special beam conditions must be used. These will be discussed in Section II-B. Shorts on the other hand have proven to be much more difficult to detect with an SRAM inspection. Therefore, we have developed special test structures for this purpose. These, along with test structures for CA opens detection, will be discussed in Section II-C.

The eBI methodology that IBM and its partners use involves detecting failure sites with eBI and then determining the root cause with in-line TEM. The mechanics of the in-line TEM portion will be discussed in Section II-D. Results collected using this methodology will be shared in Section III. The failure mechanisms for CA opens are mainly blocked or smaller size CAs. Elemental analysis can help to determine blocking materials that are located at the bottom of the tungsten. For shorts, the failure mechanisms are more complicated. Compiled together with the critical size measurements, the root-cause for the shorts can be classified as various types of misalignment, CA size variability or other reasons. The eBI methodology described in this paper, where defects are detected and characterized in-line for very fast feedback of CA module experiments has proven very efficient and valuable for SOI technology development.

## II. METHODOLOGY

### A. Equipment

A KLA-Tencor eS35 e-beam inspection system was the primary inspection tool used for the work in this paper. This is the latest eBI tool from KLA-Tencor [9]. In-line TEM in this paper means that lamellas appropriate for TEM analysis are prepared and then plucked from wafers in-line. Two tools are used for this: an FEI CLM in-line dual-beam system and a TEMLink for sample lift out [10]. The samples are transported

to a TEM outside the fab in an environment with less vibration and noise. The turn-around time for images is generally less than 24 hours.

### B. Inspection of SRAM

Inspecting SRAM for an SOI technology at W CMP is quite difficult. The challenge is that there is not much difference in the virtual grounding of a good CA versus an open CA as shown in Fig. 1. The virtual grounding is essentially proportional to the size of the electrical node. Both electrical nodes include the W plug, but the good CA also attaches to a small active area labeled source/drain (s/d) in the figure. For pFET CAs, the active area also forward biases to a small channel area. The SRAM gate CAs land on the gate so that the virtual grounding is also small, similar to the nFET and pFET CA's. In contrast, pFET CAs for bulk technologies forward bias to the N Well, which is very large and provides a great deal of virtual grounding. The weak virtual grounding of SOI CAs makes it difficult to see a difference in voltage contrast signal between good and bad CAs. For earlier technologies with a previous generation eBI tool, pre-charging the CAs was a must to maximize the difference in electrons that could be extracted from good vs. bad CAs. With our current eBI tool, however, this no longer is necessary.

The following beam conditions are used for this application: Wehnelt voltage 500 V, landing energy 750 eV, pixel size 0.035 nm, beam current 25 nA, and scan frequency 200 MHz. A strong extraction voltage is needed to pull out as many electrons as possible and to minimize charging. A smaller beam current is used because there just aren't that many electrons available to extract. It also generates a small spot size for better resolution. The pixel size is much smaller than would be used for a bulk technology of similar proportions to compensate for the poor signal to noise. Theoretically, these conditions should work. We also verified that they are optimal experimentally.

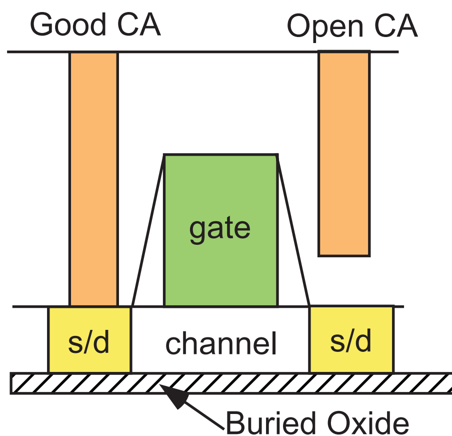



Fig.1. Schematic of good and bad CAs for SOI technology.

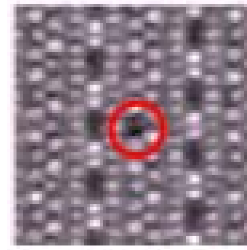


Fig.2. High resolution image of an SRAM open. The red circle marks the defect location.

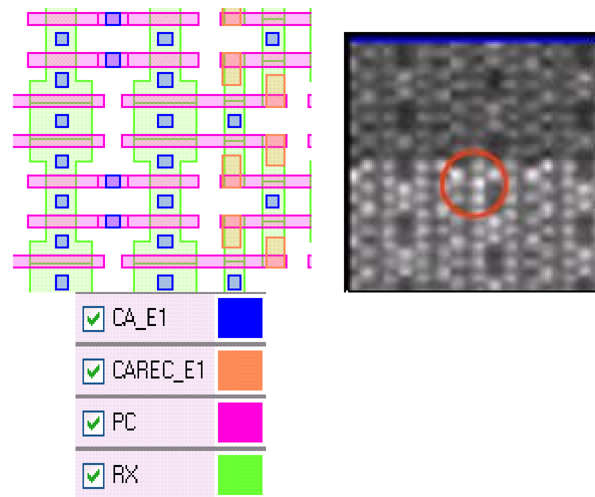



Fig.3. Left: SRAM layout; Right: eS35 patch image of a bright nFET CA. This is the expected signal for CA to gate shorting.

There is a bright side to the challenge of this poor grounding situation. Since nFET CAs, pFET CAs and gate CAs are similarly grounded, they give a similarly strong signal when an open occurs. This means the sensitivity to each of these mechanisms is equivalent. In contrast, for bulk technologies, the signal for pFET CA opens is much stronger than for nFET CA or gate CA opens. If the latter two CA opens need to be detected, the inspection has to be optimized with a much slower speed. Figure 2 shows an eS35 high resolution image of a CA open using these beam conditions.

It would also be desirable to detect CA to gate shorts in the SRAM. A CA landing on gate is typically a little brighter than nFET and pFET CAs. Presumably if an active area CA shorted to the gate, the active area CA would become the same brightness as the gate CA. Fig. 3 shows a possible example of this along with the SRAM layout for comparison. In practice, this has proven quite difficult to verify. To submit the failure analysis of a short signal is often complex because the failure analysis (FA) tools are not able to redetect the bright gate CA. Work-arounds such as giving directions relative to a FIB mark or die corner are necessary. To this point, we have not developed a confidence that these bright CAs are yield limiting shorts. All or a large portion of them could be natural

variation in the brightness. Limited failure analysis of these bright CAs has not shown CA to gate shorting. As a result, special voltage contrast test structures have been developed to monitor for CA to gate shorts. These structures use hard grounding to the substrate and therefore the signal to noise is much, much better that for an SOI SRAM scan.

*C. Test Structures*

Special test chips designed specifically for eBI are included on most development mask sets for SOI technology. Fig. 4 shows one of the structures designed to detect CA to gate shorts. All the gate lines (PC) are grounded and all the CAs are floating (Fig. 4). The CA to PC space is the same as for the SRAM. Every tenth CA lands on a PC, creating a clearly visible pattern under voltage contrast inspection conditions to help the eBI tool compare the reference to defect images. If there is a short between CA and PC, that CA will appear bright because the short provides a path to ground. Fig. 5 shows a typical CA to PC short high resolution image. The CAs land on STI rather than active area. This simplifies the design, eliminating the chance of nuisance defects due to PC to active area shorts. The impact on CA size is considered negligible. The PC lines are spaced evenly apart. The throughput for a 300 mm wafer is 104 million CAs in 180 seconds.

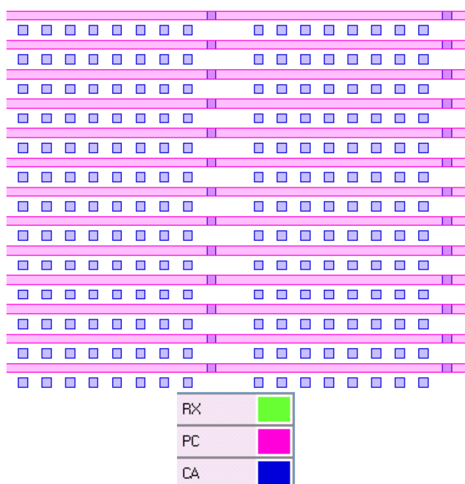



Fig. 4. CA array test structure for CA to poly (PC) shorts structure. All poly lines are grounded. Every tenth CA lands on grounded poly line, creating a clearly visible pattern under voltage contrast inspection conditions.

The eBI test chip for many SOI technologies contains a family of these structures to monitor the CA to gate shorts process window [11]. For a recent technology, these have CA to gate spaces of +1, 0, -1, -2 and -3nm relative to the dimension used in the SRAM layout.

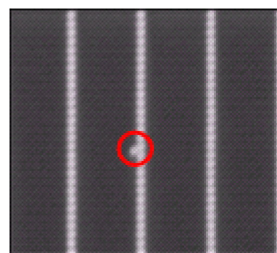


Fig.5. Typical eBI patch images for short (red circle).

The SOI test chip for this recent technology also contains CA opens test structures. While CA opens are already monitored directly from the SRAM inspection, these arrays offer a number of unique advantages. First, the basic CA array design is at a more relaxed pitch than the SRAM layout. This pitch is more representative of logic circuitry. Second, these structures can be scanned much more quickly because they have a hard ground and defects really stand out. Third, many variations of the basic CA open structures are included in the chip design enabling the testing of a large variety of possible issues.

Fig. 6 shows the basic CA open test structure. The CAs land on a grounded plane of active area. Under extraction mode, good CAs will appear bright. Open CAs, on the other hand, will appear dark since they are not grounded. Fig. 7 shows an example of a CA open in this structure. Similar to CA to gate shorts, the CA opens process window may be monitored. A family of structures with a range of CA sizes is monitored. Often eBI test chips also contain structures where the CAs are bounded by PC. Landing CAs in between PC is more difficult than landing them on a plane of active area.

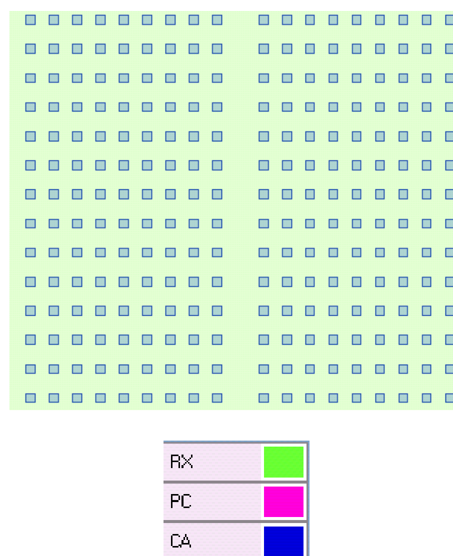



Fig.6. CA array test structure for opens. Grounding is provided by a plane of active underneath the CAs. Every tenth CA is missing, creating a clearly visible pattern under voltage contrast inspection conditions.

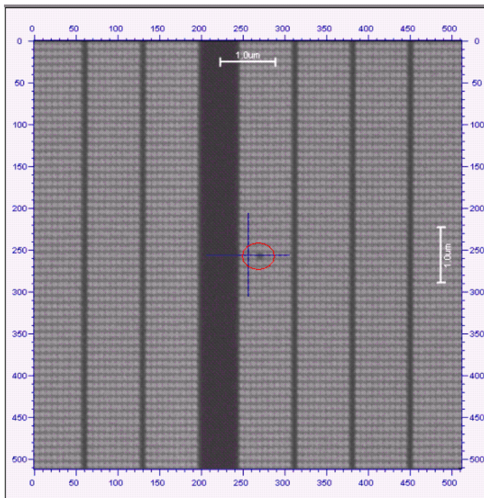



Fig.7. Typical image for CA opens in the array test structure (red circle shows the CA array open location).

Since there are so many test structure arrays that might be monitored, inspection efficiency is very important. For each test, only one cell size can be used, thus test structures with different x pitch cannot be monitored in the same test. This degrades inspection efficiency. The solution, explained in detail in ref. [11], is to design all structures with a common super-pitch, 10 um. Then they all can be scanned with a single test. Classification software automatically classifies the defects for different test structures to different bins based on their location within the die. For instance 35 nm CA opens would be binned A1; 36 nm CA opens would be binned A2 and so on.

Defect limited yield (DLY) is our preferred way to report eBI defects. The defect limited yield is the percentage of good die (i.e. zero defects for a particular bin). We find DLY preferable to reporting defect count or defect density because often there are single die with large counts of defects which make the random defect density look much worse than it actually is. Possibly we could use clustering to minimize this effect, but we have found DLY works well for our purposes. Supplementing the DLY numbers, we often include wafer maps for the key bins in our reports.

It is important to note that only a fraction of test structures in the test chip are routinely monitored. These are the ones that are problematic. Structures that are routinely good are of no interest and will not be scanned. A goal is to avoid reporting 40 different bins which dilutes the message to the integration and process engineering teams.

## D. In-line TEM Analysis

After eBI, the selected wafers are sent for in-line TEM. The failure mechanism generally is not obvious from top down imaging, but with a TEM cross section, the root cause is usually very clear. Detailed measurements of features in the TEM can be made to help determine the cause. A sample request is prepared specifying which defects should be analyzed. Failure analysis engineers re-isolate each defect using the inspection output ("KLARF") file from the eS35 and appropriate beam conditions. They then prepare an approximately 80 nm thin lamella of the defect region that can be "lifted out" with the TEMLink [10]. A set of 10 TEM samples may be attached to a grid which is taken to the TEM lab for images. The whole turn around time is usually within 24 hours. A major benefit of the in-line TEM sample preparation is that wafers can continue to be processed and do not need to be scrapped.

After a period of time, a root cause Pareto emerges and different root causes can be linked to wafer spatial signatures or even top down appearance. At this point, the number of samples submitted for in-line TEM can be reduced.

# III. APPLICATION

## A. Opens

E-beam inspection is used both to trend key CA module metrics by wave of yield learning hardware and to evaluate CA module split experiments. The DLY for SRAM opens and test structure opens and shorts detected by the eS35 are the key metrics of wafer health. Figs 8 and 9 show the trend charts for these metrics over a period of around six months for a recent SOI technology. The Y-axis is the eS35 defect limited yield for the inspected area. For the SRAM, the inspected area is generally a faction of the SRAM size. These charts reflect improvements made with each new wave.

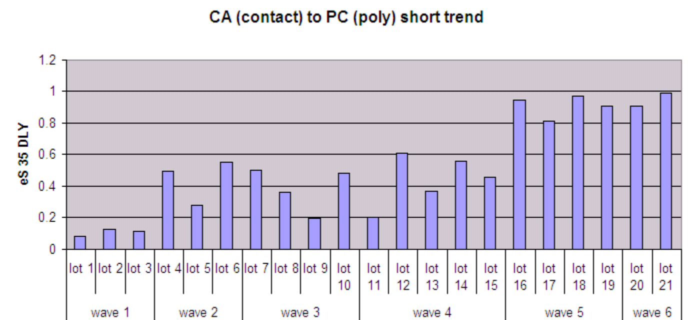



Fig.8. eS35 shorts trend chart for multiple development waves.

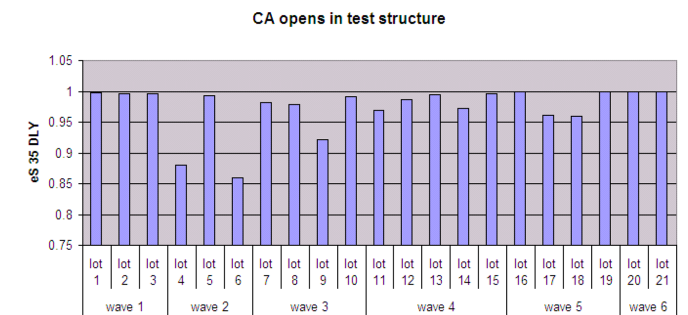
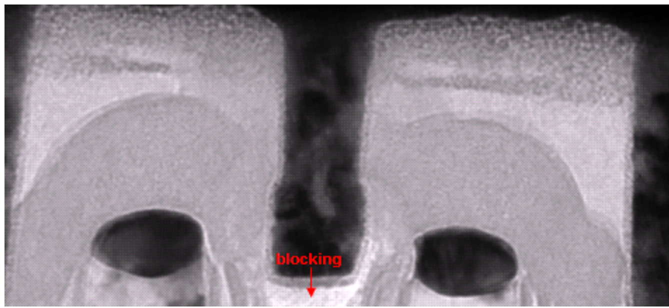
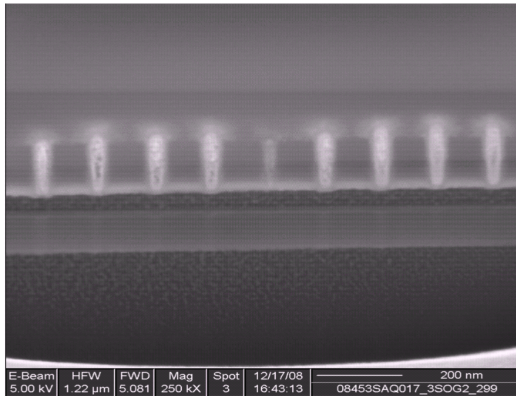



Fig.9. eS35 opens trend chart for multiple development waves.

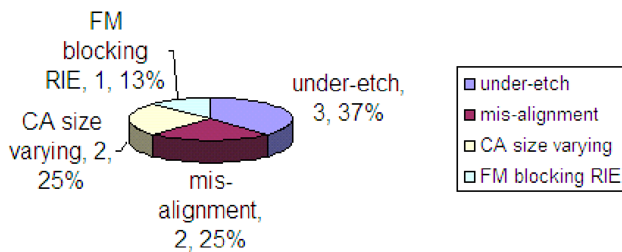



(a)




(b)

Fig.10. (a) In-line TEM image of SRAM open due to blocking in Fig. 2 (sensitive gate stacks have been deleted due to censorship), (b) SEM cross-section of CA opens in test structure due to smaller CA.



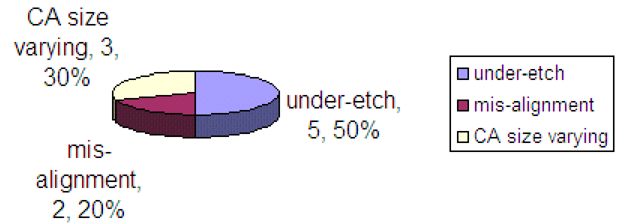
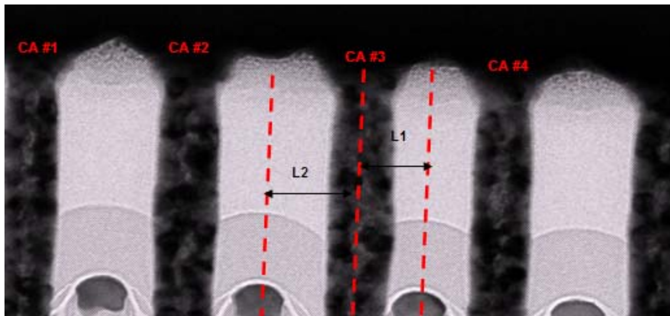




Fig.11. Defect Pareto for CA open failures for a wave of hardware.

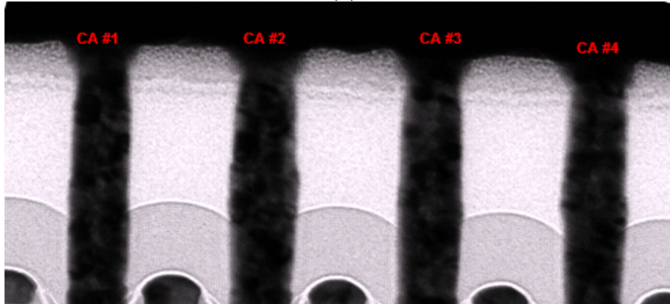
Fig. 10 shows two examples of CA opens. Fig. 10 (a) is an in-line TEM image of a CA open in the SRAM. It shows foreign material blocked the CA reactive ion etch (RIE) step. The subsequent elemental analysis found NiSi underneath the CA. Fig. 10 (b) is a SEM cross-section image of a CA open on a test structure. The open is because the CA was printed smaller and therefore could not reach the bottom. Fig. 11 shows the CA opens defect Pareto for a wave of yield hardware.

*B. Shorts*

Figs 12 and 13 show in-line TEM images of CA to PC shorts. Table 1 was used to determine the failure mechanism in Fig. 12. Offset is calculated as the difference of distance between neighboring CA to poly line. For example, in Fig. 12(a), L1 is the distance between CA #3 and the metal gate on the right. L2 is the distance between CA #3 and the metal gate on the left. The CA-PC offset is (L1-L2)/2. Offsets were calculated for each CA shown in Fig. 12(a) and (b) and are summarized in Tables 1 (a) and (b) respectively. We can clearly see that CA #3 has a much larger CA-PC offset than the other CAs in the same image. The big offset for CA #3 indicates that the misalignment occurs locally. On the other hand, in Table 1 (b), the offsets for the other CAs are comparable. All are in the range of 7 to 9 nm. This indicates global “shifting” of the alignment. So this failure is categorized as global misalignment, which is different from local misalignment.




**(a)**




(b)

Fig. 12. In-line TEM images for a CA to PC short. These defects were captured by eBI. (a) local misalignment; (b) global misalignment.

**TABLE 1: Offset for the CAs in Fig. 12 (a)**

| | CA #1 | CA #2 | CA #3 | CA #4 |
|---|---|---|---|---|
| CA-PC offset | n/a | -0.7 | -12.9 | 0.3 |

**Offset for the CAs in Fig. 12 (b)**

| | CA #1 | CA #2 | CA #3 | CA #4 |
|---|---|---|---|---|
| CA-PC offset | -8.4 | -8.5 | -7.2 | -8.2 |

**Note: CA-PC offset = (L1-L2)/2**

Fig. 13 shows another failure mechanism for CA to PC shorting. In this case, we can see that the spacer was etched off and it is completely gone. The CA is not uniform in size from the top to the bottom: some "fat" parts touch the poly line (PC) and cause the short.

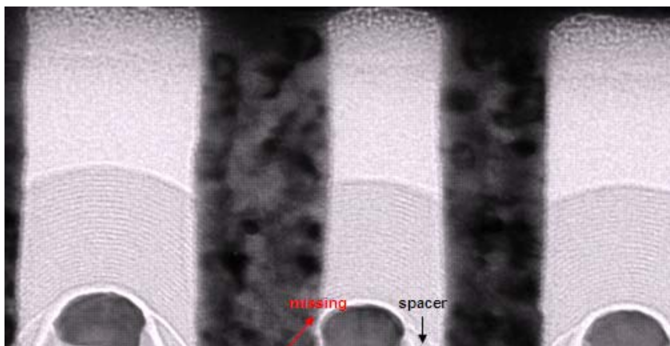



Fig. 13: Contact to poly short caused by missing spacer (red arrow shows where the spacer is missing)

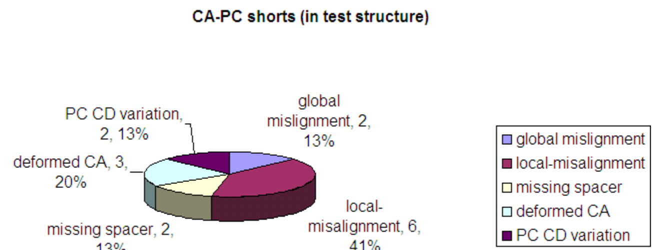



Fig. 14: Defect Pareto for CA short failures in the same DYC wave

Fig. 14 shows the Pareto for shorts for a DYC wave. This helps to determine the main failure mechanisms for a particular process condition.

*C. Split experiment*

Fig. 15 shows results from a split experiment designed to explore the appropriate RIE conditions for two different tools/chambers. There are three wafers for each split. S1 through S7 are combinations of chamber and RIE parameters. It is obvious that S4 is the best for short yield and S7 is the worst scenario of all. This is a fast way for CA module team to optimize process parameters.

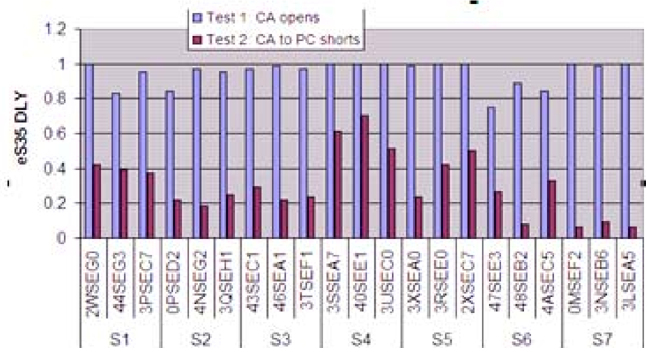



Fig. 15: Results of split experiment for CA open and short. (Split details were eliminated due to classification)

IV. SUMMARY

A methodology for fast yield learning for the CA module for SOI technologies is described. It involves e-beam inspection and in-line TEM analysis. The SRAM can be inspected for CA opens using special beam conditions. Special eBI test structures can be used to monitor CA to gate shorts and CA opens. Top down images are not sufficient to determine the root cause of failure; in-line TEM imaging is used for this purpose. This combination provides fast and accurate feedback to the integration team, accelerating yield learning.

ACKNOWLEDGMENTS

This work was performed at the IBM Microelectronics, Semiconductor Research & Development Center, in E. Fishkill, NY. We thank Bachir Dirahoui, Tso-Hui Ting, Aimin Xing, Florian Fesquet and others for stimulating discussions related to this work. We thank Jin Wallner and Jay Strane for their suggestions in designing CA structures for eBI test chips. We thank John King for his TEM imaging support.

REFERENCES

[1] O. D. Patterson, K. Wu, D. Mocuta, K. Nafisi, "Test structure and e-beam inspection methodology for in-line detection of (non-visual) missing spacer defects," *IEEE/SEMI Advanced Semiconductor Manufacturing Conference Proceedings*, pp. 48-53, June 2007.

[2] R. Guldi, J. Shaw, J. Ritchison, S. Oestreich, K. Davis, R. Fiordalice, "Characterization of Copper Voids in Dual Damascene Processes", *IEEE/SEMI Advanced Semiconductor Manufacturing Conference Proceedings*,, pp. 597-602, Nov. 2004.

[3] O. Moreau, A. Kang, V. Mantovani, I. Mica, M.L.Polignano, L. Avaro, C. Pastore, G. Pavia, "Early detection of crystal defects in the device process flow by electron beam inspection", *IEEE/SEMI Advanced Semiconductor Manufacturing Conference Proceedings*,, pp. 334-339, May 2006.

[4] J. Jau, W. Fang, H. Xiao, "A Novel Method for In-line Process Monitoring by Measuring the Gray Level Values of SEM Images", *Proceedings of ISSM*, pp. 143-146, 2005.

[5] A. Ache, K. Wu, N. Rowand, "E-beam inspection of SOI wafers for the evaluation of barrier nitride RIE etching," *Advanced Semiconductor Manufacturing Conference and Workshop*, IEEE/SEMI issue, pp. 124-126, April 2005

[6] M. Horstmann, "Advanced SOI CMOS transistor technology for high performance microprocessors," *Solid-State Electronics*, vol. 53, no. 12, pp. 1212-1219, Dec. 2009.

[7] S. Cristoloveanu, F. Balestra, "Silicon on insulator: Technology and Devices," *Advanced Semiconductor and Organic Nano-Techniques*, pp. 325-365, 2003

[8] Y. H. Liu, Y. L. Tu, W. Y. Lain, B. W. Chan, M. Chi, "Contact size dependence of highly selective self-aligned contact etching with polymer formation and its mechanism," *Advanced Semiconductor Manufacturing Conference and Workshop*, IEEE/SEMI issue, pp. 153-156, Sep. 2000

[9] KLA-Tencor, "Electron Beam Wafer Inspection System", online (http://www.kla-tencor.com/front-end-defect-inspection/es35.html)

[10] H. H. Kang, M. A. Gribelyuk, O. D. Patterson, S. B. Herschbein, "A new method of wafer level plan view TEM sample preparation by DualBeam," *Proceedings of ITSTFA*, pp. 168-171, 2008

[11] O. D. Patterson, K. Wu, H. H. Kang, P. Feichtinger, "In-line process window monitoring using voltage contrast inspection," *IEEE/SEMI Advanced Semiconductor Manufacturing Conference Proceedings*,, pp.19-24, May 2008.