Application Paper

Industrial Mask Inspection

Description of Application

The "inspection of masks used in semiconductor (integrated circuit) manufacturing" is a difficult application due to the resolution required to detect defects that are significant to the process. The small feature size (less than 1 micron) used in today's processing exacerbates the problem because of the squared relationship between resolution and the number of pixels to be processed, which is in direct relationship with the time needed to complete the inspection.

The equipment used encompasses three precision elements; (a) a digital line-scan camera, which in this application is 4096 pixels long, (b) a high quality lens able to image over the length of the line scan sensor at the diffraction limit of the light being used, and finally (c) a positioning stage which moves the mask under the sensor in successive passes. This last step is performed with enough precision to be sure that every part of the mask is imaged at the required resolution.

Lighting and lens selection are not trivial as 0.5 microns approaches ultra-violet wavelengths.

Imaging the Mask

The mask is imaged by moving it under the line scan camera in a back and forth pattern, so that each pass overlaps the last pass by 1% (40 pixels, 20 microns). This assures that no portion of the mask is left unexamined due to positioning errors of the stage. It also gives the processing system enough data to determine if a defect is bridging two passes, or if two defects are near the edge of the two passes.

The lighting of the mask is designed so that the background is dark and any defects will show as bright spots (i.e. Dark Field Lighting). This is typically accomplished by lighting from the sides so that any defect will scatter light into the camera.
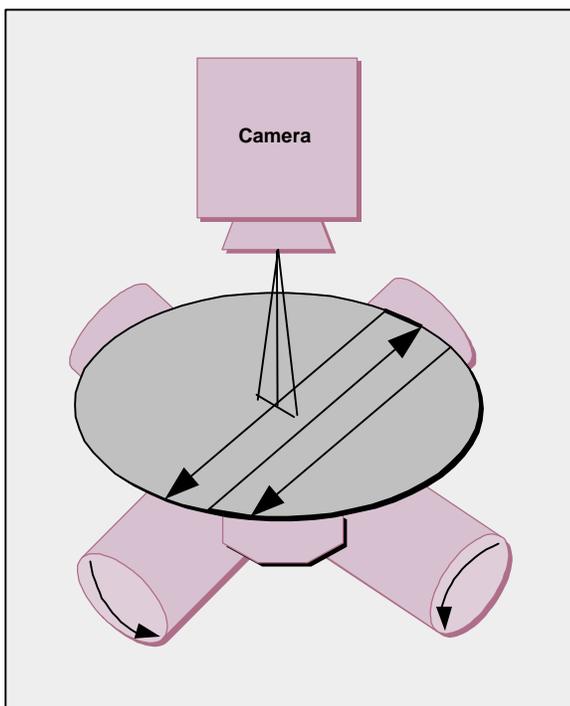

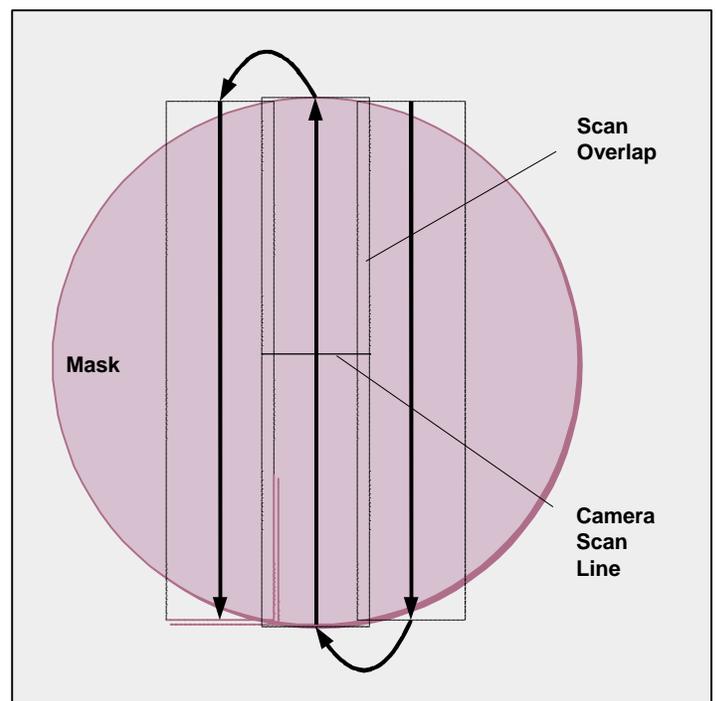Figure 1: Diagram of inspection apparatus.

Figure 2: Diagram of imaging path.

As each pass is executed, image data is processed by the computer system that retains images of the defects, throwing out normal background data. This discarding of background data is a cost savings method, as it eliminates the need to store 40 GB of image data, most of which would be uninteresting. In addition, the data is being collected at a high data rate (from 50 to 200 Mbytes/s), which would require expensive hardware to capture and store (striping raid disk systems are used typically).

Image Processing

Image processing consists of two parts, one very simple step which must be performed on every pixel, and a more complex image processing step to be performed on the defect images only.

The first part of the processing corrects the pixels for variation in the line scan detector (gain and dark current correction), and then compares the corrected pixel to a threshold. The correction step allows the threshold to be set very low so that a greater portion of the defect is detected for later processing. Also correcting the sensor for a 'flat field' of view will reduce systematic errors, which might show up as errors that are correlated to the position of the sage and camera, rather than true defects in the mask.

The second part of the processing collects the image data for the defect by collecting pixels in a rectangular region that are slightly larger then the defect image. These regions of interest (ROIs) are collected and further processed by a blob detection algorithm, and then measured. The measurements taken are position, area of the convex hull of the defect, radius of the smallest circle that will encompass the defect, average density of the defect (brightness), and the perimeter of the defect. This data will be processed by the host computer to determine if the mask should be discarded. The defects are utilized by later processes, that use the mask, to reject circuits that are produced by portions of the defective mask, or trigger a cleaning step should the type of defect indicated suggest contamination.

Performance

The processors shown in the table below were compared in the implementation of this application. In each case the central loop of the first processing step dictated the number of processors needed to 'keep up' with processing the pixels as they came in. Additional processing power is used by the defect analysis, however this step is performed after the imaging is complete and is insignificant compared to the first imaging step, unless the mask is loaded with defects.

The application is parallelized by data partitioning so that each processor gets a portion of the data, a vertical slice along the motion of the stage. Each vertical slice is taken to overlap (40 pixels) so that bridging defects can be resolved. Errors are collected in memory and processed at the end of the scan.

There is no physical process that dictates the speed at which the image should be collected. It is a trade off between the cost of the system required to inspect the mask, and the cost of the time it takes in the process. The first step in determining the cost is determining the number of processors required to inspect the mask in a given amount of time. For the purposes of this article, 60 minutes down to 15 minutes is considered. At higher performances (>200 Mbyte/s) two cameras are required as the data rate exceeds that obtainable from the faster line scan cameras (Dalsa CT-F3-4096 8 tap camera).

As can be seen in the table below, the number of processors required is best for the TriMedia TM1300 processor. The TI processors came in second however they do not fare too well when cost is considered. The ADI processors do not stack up as well due to the limited number of processing units available in each processor.

Data Rate	200MB/s	160MB/s	100MB/s	80 MB/s	50 MB/s
Processor	15 min	20 Min	30 Min	40 Min	60 Min
ADI ADSP21160	8	6	4	3	2
Philips TM1300	3	2	1	1	1
TI TMS320C620X	4	3	2	2	1
Intel-PIII-450	NP	NP	1	1	1

Table 1: The number of processors required.

The PIII-450 processor almost outperforms the other processors, but it is limited by the mother board's I/O capability (PCI) of less than 132 MB/s. If one were to build a private memory multiprocessor PIII based product, it would perform quite well in the kind of application requiring high memory bandwidths as its memory system is currently the fastest (800 MB/s peak). When cost is considered, the PIII processor does not fare as well – the low cost versions (<330 MHz) do not perform as well as other processors in the table, while the higher performance parts (>400MHz) are too expensive. In addition, the physical size of the high-performance Pentium processors make them difficult to use.

The mask inspection application is an example of a memory performance limited application, which is solved by scaling the memory bandwidth to deliver the highest performance solution with tolerable cost. Real-time constrained applications exhibit other characteristics as the next example shows.

ALACRON

71 Spit Brook Road, Suite 200
Nashua, NH 03060
Tel: 603.891.2750   Fax: 603.891.2745
Web: www.alacron.com   E-mail: sales@alacron.com