Memory Yield Improvement - SoC Design Perspective

Jitendra B. Khare
Ample Communications, Inc., Fremont, CA USA.

Abstract

On-chip memories are a major source of yield loss in SoC designs. Currently, redundancy is the only available option to improve memory yield. However, other techniques - e.g., DFM-based bit-cell design, flexibility in bit-cell choice, and ability to choose the number of metal layers - can be more effective. The availability of such techniques would allow designers to tailor memories to the specific SoC architecture. Such strategies would reduce die cost, but would require close collaboration between the foundry, IP companies and customers.

The shift in IC technology to 0.13um has resulted in a substantial increase in the number of large SoC designs. A significant portion of the die on such designs is typically dedicated to on-chip memory. This memory can consist of a few very large RAMs for data storage, and/or a large number of smaller RAMs/register files used mainly for control and data processing. In both cases, chip yield is dominated by these on-chip memories.

The usual practice to improve memory yield is to apply redundancy. Memories with built-in redundancy are readily available from IP companies. However, redundancy works well only for large memories (e.g. on-chip caches). For designs with numerous small memories, (e.g. communications chips) redundancy is wasteful and expensive. These designs are unable to optimize their memory yield with existing solutions. Since most SoCs have a mixture of large and small memories, a smaller portion of their memory yield loss can be recovered through redundancy.

Another issue – which affects small design houses - is the increase in IP costs due to redundancy. IP companies charge 50-100% more for memory compilers with redundancy. Small design houses tend to circumvent these costs by adding extra rows/columns and configuring them through BISR circuitry. This results in some inefficiency, and hence, a further reduction in yield gains.

Besides redundancy, other techniques could be used to improve memory yield. First, DFM techniques can be used to create yield-optimized bit-cell layouts. With the current focus on redundancy, bit-cells are designed to be as compact as possible. However, by relaxing design rules, memory yield can be improved with a small area penalty. Such bit-cells can be used in smaller memories without redundancy, resulting in decreased die costs.

A second technique is to give designers the ability to choose the bit-cell for each instance of the compiler-generated memory. The optimal bit-cell for each memory depends on the overall memory in the design and the number of instances of the specific memory. For example, a design may have one instance of a 4096x64 memory, but 32 instances of a 1024x64 memory. Clearly it would be more beneficial to choose the DFM bit-cell for the latter. Depending on cost analysis, the first memory could use either the DFM bit-cell or the compact one.

Another useful option in a memory compiler is to have the ability to choose the number of metal layers in the memory. Currently, metal layers used for a given memory type (single-port RAM, two-port RAM, two-port register file, etc.) are fixed within the compiler. However, for cost sensitive designs, this can be a problem. By reducing metal layers in memories at the expense of area, designers may be able to save one or two routing layers, which can result in a large reduction in wafer cost. Thus, in spite of increased area, overall die cost may actually reduce.

The development of the above techniques would require a close collaboration between the foundry and IP companies. Clearly, defect data would have to be shared, bit-cells would need joint development, and more test chips would be necessary. In addition, inputs would be needed from customers regarding typical design characteristics and the various cost-yield-area trade-offs. However, a successful deployment should result in significant savings in die costs.

A compiler that incorporates all the above options would cost more than a traditional one. However, this additional cost will have to be worked out within the overall cost analysis framework – designers would be willing to pay if substantial die cost savings can be achieved.