White Paper

FPGA

intel®

Empowering Semiconductor Manufacturing Equipment with FPGAs

Authors

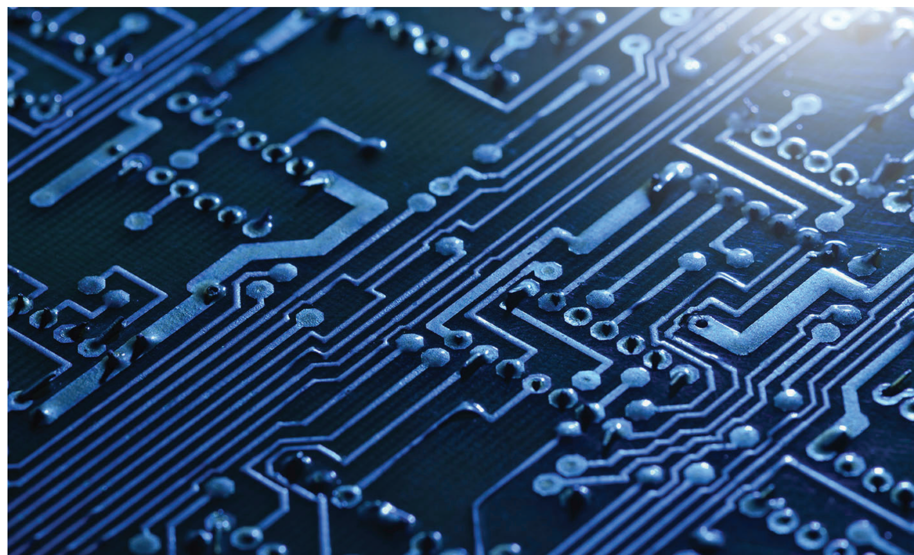
Dr. Lonna Edwards
Technical Specialist
Intel Programmable Solutions Group

Mr. Takayuki Ikushima
Director, Industrial Business Unit
Intel Programmable Solutions Group

Mr. Andreas Hagmann
Head of Digital Design
IMS Nanofabrication

Introduction

Many of the daily devices we use to simplify our lives are made possible by integrated circuits (ICs). Technologies, including phones, appliances, cars, and more, are powered by these components that begin their journeys as silicon wafers. As technology advances, multiple types of ICs are often combined within electronic devices to increase functionality. This increases demand for ICs, which is evident given the record number shipped by semiconductor manufacturers last year [1]. To meet the growing demand, many cleanroom facilities drive fabrication tools 24/7 [2]. Additionally, researchers at these semiconductor foundries continuously seek methods to optimize fabrication via process steps and manufacturing equipment improvements.



Past Meets Future Innovation

Intel has remained a leader in the semiconductor manufacturing industry since 1968, having been co-founded by an IC inventor and the inventor of Moore's Law. The company has an enduring record of producing competitive processors in addition to Ethernet, memory, wireless, and server products.

Given the extensive portfolio of ICs that Intel designs and manufactures, it is uniquely positioned to evaluate means of optimizing semiconductor manufacturing equipment. To improve machine performance and, thus, assist foundries worldwide in meeting chip demand, we propose integrating more FPGAs into cleanroom tools [3]. This paper examines FPGA advantages and use cases within two types of semiconductor manufacturing equipment: mask writers and surface inspection tools.

Table of Contents

Introduction 1
Past Meets Future Innovation 1
FPGA Overview 2
Semiconductor Manufacturing 3
FPGA Use Case #1: Mask Writing....3
Recommended FPGAs for Mask Writing and Wafer Surface Inspection 6
Conclusion............................6
Sources 7

FPGA Overview

I. FPGAs vs. ASICs, CPUs, and GPUs

FPGAs are gaining prevalence in numerous industries, including artificial intelligence, life sciences, automation, and quantum computing [4]. One can accredit this increase to the numerous advantages of FPGAs over ASICs, CPUs, and GPUs. Although ASICs are suitable for applications requiring high customization and high-volume production, FPGAs provide several exclusive advantages, including the following:

- Reconfigurability in the field
- Rapid time to market
- Low non-recurring engineering costs (NRE)
- ASIC prototyping capability
- Extensive hard and soft intellectual property (IP)

Compared to CPUs and GPUs, FPGAs offer lower latency, enhanced acceleration capability, and higher energy efficiency [5].

II. FPGA Architecture

The FPGA benefits previously described are inherently linked to architecture. For example, many FPGAs can be reprogrammed repeatedly because they are SRAM-based [6]. This feature is useful for ICs that require redesign after system integration, whether due to hardware bugs or changes in standards/protocols. The FPGA logic fabric enables rapid time to market for complete systems because the FPGA functionality is verified before it is offered to customers. Consequently, a customer can modify the logic fabric and verify the updated design without returning the device to the vendor. This ensures that NRE costs associated with the ASIC design cycle are averted. This significant cost reduction frequently causes a company to prototype a design in an FPGA, even if it will employ an ASIC in the final product [6].

FPGAs are characterized by their logic fabric, hard IP, and soft IP. The logic fabric contains reprogrammable lookup tables, shift registers, I/O, embedded memory, and interconnects [7]. These elements comprise the adaptive logic module (ALM), a fundamental FPGA building block. Although the logic fabric of FPGAs is consistent, some architecture can vary between product families. Each FPGA product family includes specialized hard IPs for customers to consider when implementing a design. IP refers to existing functional blocks designed to be efficient in power consumption, chip real estate, and performance [6]. For example, customers can select products with hard IP in the form of processor cores and digital signal processing blocks on a board. Intel also offers soft IPs that are implemented using the reprogrammable logic fabric.

Figure 1 showcases the Intel Agilex® 7 FPGA schematic with logic fabric and hard IP. All Intel® FPGAs contain ALMs, DSP blocks, embedded memory blocks, clocks, and routing as a standard. Customers can integrate additional hard IPs, including processor cores, transceivers, GPIO or DDR memory interfaces, embedded multi-die interconnect bridges (EMIB), and Secure Device Managers.

Due to their ability to incorporate soft IPs, FPGAs can serve as an interface between devices. In high-speed computing applications, they are now being utilized as accelerators, sometimes outperforming CPUs and GPUs. Implementing parallel processing directly within FPGA hardware rather than other accelerators results in low latency and high performance per watt. FPGAs also provide deterministic signal retrieval, ensuring all signals are captured with a consistently low jitter level. Although parallel processing is possible with ASICs, FPGAs are a noteworthy alternative for previously discussed reasons.

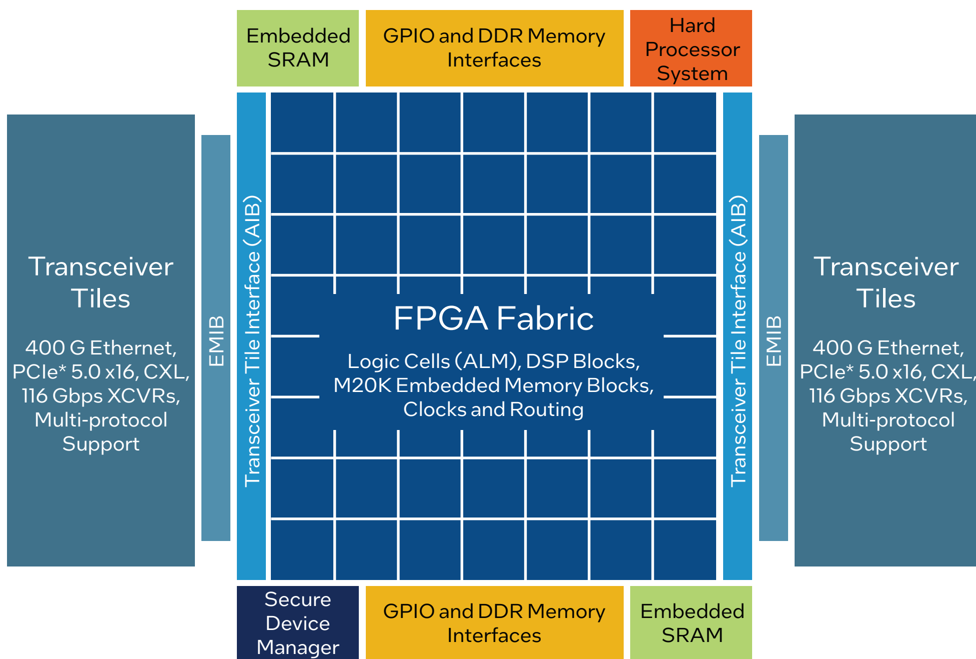


Figure 1. Intel Agilex® 7 FPGA Schematic

Semiconductor Manufacturing

I. Justification

The demand for ICs continues to increase, requiring cleanroom facilities to increase operation capacity to 100%. Nevertheless, there are constraints inhibiting growth in the semiconductor manufacturing equipment market. One factor includes the high costs of newer machines that perform complex operations. Another is the maintenance cost of older machines once parts become less available. FPGAs provide feasible solutions to both limitations since they can accelerate processors performing complex operations while maintaining long life cycles. Intel FPGAs, in particular, have life cycles of approximately 15 years.

FPGA integration could benefit many fabrication tools; however, this paper will focus on mask writers and inspection tools. First, we will examine the semiconductor manufacturing process at a high level. Then, we will discuss how mask writers and inspection tools operate to determine how functions can benefit from FPGA integration. Lastly, we will recommend suitable Intel FPGAs for mask writer and surface inspection applications.

II. Process Overview

ICs undergo hundreds of process steps before they are combined to form electronic devices. Each of these process steps corresponds to specific equipment. Fortunately, many steps are iterative, reducing the amount of equipment required in a cleanroom facility. General process steps for wafer fabrication include:

1) Cleaning
2) Thin film deposition
3) Lithography
4) Etching
5) Ion implantation
6) Establishing metal contacts
7) Testing
8) Packaging

Note that steps 1–5 can be repeated more than 20 times for a single wafer to create the layers of an IC [8-9]. Figure 2 shows a flow chart of the process.

Cleanliness is an extremely important requirement during fabrication. The facility must maintain an air quality standard, and the wafer must be free of contaminants to proceed to successive process steps. Wafer cleaning is performed at multiple stages and can include wet chemical etching or chemical mechanical planarization [10].

After a wafer is cleaned, a thin film is deposited onto the surface. This film comprises a dielectric or general protection layer for the wafer surface. Next, the entire surface is coated with photoresist, and the wafer is placed into a stepper or scanner system. A mask, or reticle, is then used to transfer a pattern onto the wafer surface via select area exposure to light. In other words, certain photoresist regions are removed after light exposure to reveal specific surface areas underneath. Whether the lithography mask is hard (reticle) or a design file, it must showcase the layout and dimensions of the desired 2D surface structures. The FPGA Use Case #1 section describes the procedure for fabricating a reticle.

In high-volume manufacturing, plasma (dry) etching is preferred for removing deposited material (i.e., a dielectric) in select areas to reveal the underlying silicon. During ion implantation, those newly revealed silicon areas are doped (i.e., introduced to free current carriers). Next, metal contacts are produced, allowing electricity to flow between desired components and devices. Before a semiconductor die reaches the end of the fabrication cycle (i.e., packaging), it undergoes rigorous inspection and testing [10].

FPGA Use Case #1: Mask Writing

I. Reticle Manufacturing

Although wafers can undergo direct writing during lithography (e.g., the system uses a design file rather than a physical mask to expose photoresist), they are traditionally fabricated with the help of hard masks (reticles) inserted into a stepper or scanner system. Numerous reticles are used to process a chip; however, each must undergo processing before transferring a pattern to a wafer. A reticle begins as a glass substrate coated in chrome (or another metal of choice) and a photoresist layer. It is then exposed to a design using a mask writer and later developed. The exposed chrome on the hard mask is etched, and the remaining photoresist is removed. Finally, a protective layer is added to the mask before shipping to a customer.

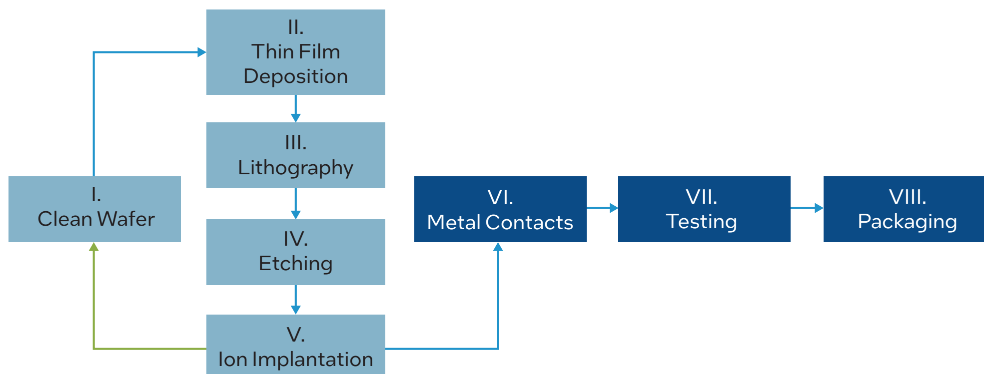


Figure 2. Semiconductor Fabrication Process

Figure 3 shows a flow chart of this process. Reticles undergo many of the same process steps as semiconductor wafers; however, once reticles are fabricated, they are used to manufacture semiconductor wafers. This relationship is never reversed. Reticles are used during lithography of semiconductor wafers. The finished reticle can then transfer a pattern layer onto a semiconductor wafer during the lithography step shown in Figure 4.

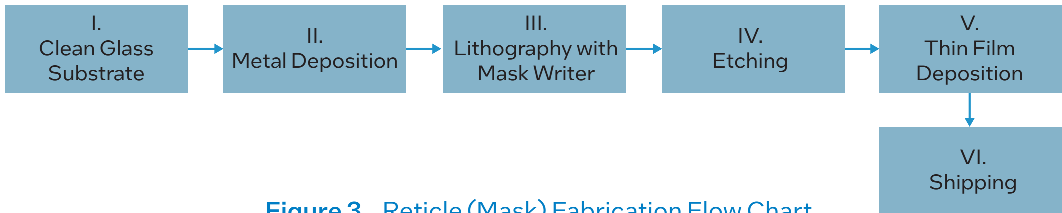
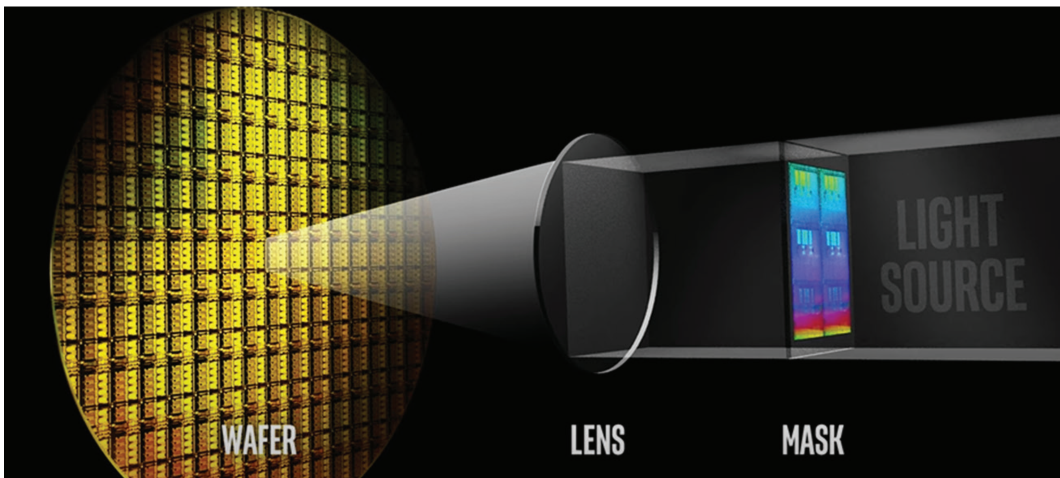


Figure 3. Reticle (Mask) Fabrication Flow Chart



Source: Reprinted from [11]

Figure 4. Reticle Use During Semiconductor Wafer Lithography

During the lithography of a wafer, a mask (reticle) is placed within a stepper or scanner system between a light source and a lens. The lens reduces the scale of mask features, which are projected onto a semiconductor wafer coated in photoresist. The scaled pattern is replicated on the entire wafer surface to produce numerous chips at the end of wafer processing. Note that a different mask is required for each layer of a single wafer, but masks are reused on numerous wafers.

II. FPGA Use within IMS Mask Writers

Mask writers are typically characterized by the type of beam used for exposure (i.e., a variable-shaped beam or a beam array) and exposure wavelength. Lithography systems, including mask writers, are continuously optimized to meet high throughput and high-resolution chip demands. The IMS MBMW-101 mask writer examined in this paper, shown in Figure 5, meets 7-nanometer technology node requirements [12]. The multi-beam tool can write a total area of 100 mm by 130 mm in less than 10 hours, surpassing the industry standard of less than 24 hours for a complex mask. It utilizes an array of programmable, homogenously shaped beams instead of a single variable-shaped beam, which is associated with traditional electron beam lithography systems. The multi-beam array enables consistent write times regardless of design pattern complexity. Additionally, the MBMW-101 can process data at a rate greater than 120 Gbps even when correcting the beam array for undesirable effects (i.e., the proximity effect, a defective beam, and more). Since the release of the MBMW-101, IMS has released two additional models with improved data transfer rate (150 Gbps) and software. The company is currently testing its newest model, the MBMW-301, which provides a high-speed data path of 800 Gbps.

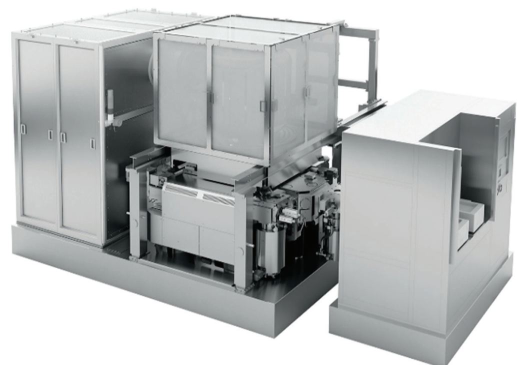


Figure 5. IMS MBMW-101 Mask Writer

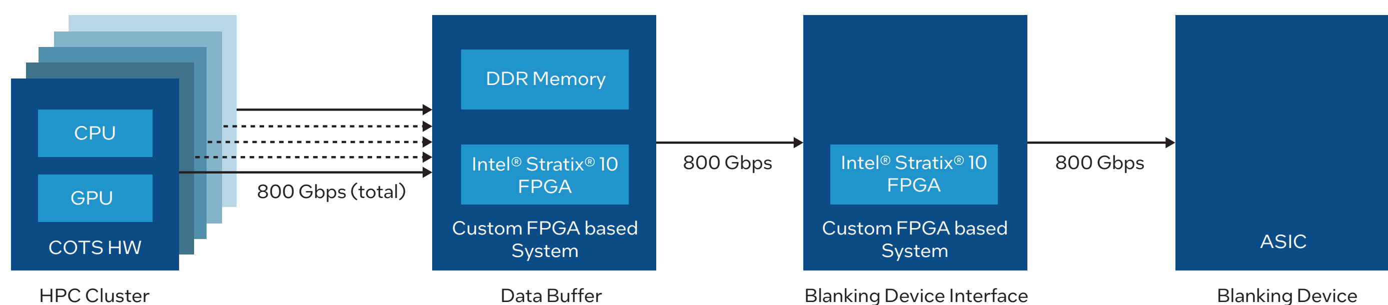


Figure 6. Data Path for IMS MBMW-301 Mask Writer

IMS ultimately integrated Intel FPGAs into all mask writer models due to their short development time and ability to meet latency requirements. Intel FPGAs also provide the high amount of memory and high-speed I/O channels necessary. IMS mask writers use the following Intel FPGAs: Stratix® V, Intel Stratix 10, Intel Arria® 10, Intel Arria 10 SoC, and Intel MAX® 10. These FPGAs perform several functions, including data buffering, high-speed data transfer, image rotation, error correction (i.e., cyclic redundancy checks), deterministic control, and real-time control. Time synchronization is in the nanosecond range, and position control is in the nanometer range.

Figure 6 showcases the data path for the MBMW-301. Commercial off-the-shelf hardware, including CPUs and GPUs, comprise the high-performance computing (HPC) cluster. Intel Stratix 10 FPGAs and DDR memory then buffer the data before it is sent to additional Intel Stratix 10 FPGAs that serve as an interface to a blanking device. A blanking device is required to write a desired pattern onto a mask when using a multi-beam system. It switches the numerous individual beams on and off by passing them through or deflecting them from apertures positioned before a mask substrate.

III. FPGA Use Case #2: Wafer Surface Inspection

For the second use case, the authors propose utilizing FPGAs within surface inspection tools to detect defects on semiconductor wafers in real time, which requires the consistent capturing and analysis of high-throughput data. Various image processing and machine learning algorithms are utilized during wafer surface inspection to locate defects such as foreign particles, scratches, pinholes, and stains [13-14]. Some researchers have even used algorithms to locate air gaps, cracks, and ink spots during packaging. Discovering such defects before product shipment can help manufacturers increase yield and discover which tools or processes negatively impact device quality.

To implement image processing or signal processing, in general, within a system, today's designers often consider the benefits and disadvantages of the following acceleration options: multi-core CPUs, digital signal processors (DSPs), GPUs, and FPGAs. Sometimes, two of the options mentioned above are combined to provide the best performance. Although it might seem counterintuitive, DSPs are not necessarily the optimal choice, especially when considering complex image-processing applications that utilize parallel algorithms. GPUs and FPGAs normally outperform competitors in processing speed because their architectures facilitate parallel computing.

Parallel computing has previously been suggested to increase system efficiency for automated visual inspection and, thus, IC yield in the semiconductor industry [13]. Parallel computing can be implemented in multicore DSPs but is more efficiently executed in FPGAs and GPUs [15-17].

Selecting between FPGAs and GPUs can be challenging since performance depends on the specific application and algorithms used. Nonetheless, FPGAs typically outperform other accelerators, including GPUs, in real-time image processing speed and power efficiency because FPGAs excel at rapid data capture and analysis for high-volume data [16,18]. Systems for optical inspection must handle large amounts of data and complex algorithms [19]. They must also perform multiple functions that filter images after capture, as well as correct image distortions, noise, and shading. FPGAs are suitable solutions for these tasks. As an added benefit, FPGAs can also support mathematical operations at arbitrary precisions (i.e., single-precision, half-precision, tensor floating point, and bfloat16 floating point), as opposed to GPUs that can only support operations at predetermined precisions [20].

Researchers have already demonstrated the use of FPGAs during an image preprocessing application associated with the visual inspection of an IC. In this case, an FPGA was used to improve picture quality, accelerate the CPU, and enhance the real-time capability of the system [21]. Although publications describing FPGA utilization in wafer surface inspection are limited, FPGAs have been used successfully in several image processing applications, including real-time 3D surface model reconstruction, real-time surface tracking, image boundary detection, real-time image denoising, and surface inspection of high-quality printing products such as banknotes and stamps [16, 19]. FPGAs have the added benefit of working well with sensors. For example, during one surface inspection application, FPGAs were used to generate line trigger signals for line-scan cameras before processing data [19].

FPGAs can receive and manipulate data from various inputs, including sensors, light sources, and actuators. As a result, their integration into surface inspection tools, which rely on such controls, should be feasible. FPGAs can also perform sensor signal conditioning, signal analysis, network interfacing, and functional safety tasks before outputting data to a network or display.

Figure 7 shows a block diagram applicable to FPGA use in surface inspection applications. FPGAs can receive data from various inputs associated with surface inspection tools and perform all functions shown in blue (i.e., sensor signal conditioning, signal/image analysis, network interfacing, and functional safety).

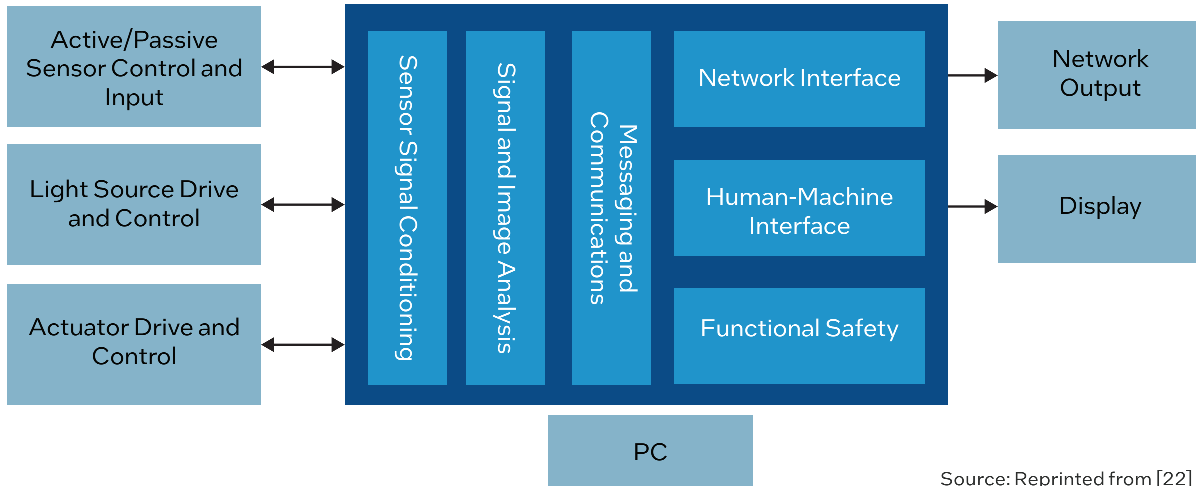


Figure 7. FPGA Use in Surface Inspection Applications Block Diagram

Recommended FPGAs for Mask Writing and Wafer Surface Inspection

Several Intel FPGAs utilized in IMS mask writers have previously been listed. Additionally, the authors recommend using Intel Agilex 7 devices in general mask writing and surface inspection applications for functional safety and image processing tasks. This product family leads the FPGA industry with the highest transceiver data rates (up to 116 Gbps), PCIe* 5.0 support, and DDR5 high-bandwidth memory support (over 1 TBps) [20]. Please refer to Figure 1 for further details. It is important to note that Intel Agilex FPGAs are enhanced for digital signal processing, showcasing a 38-TFLOPS performance for high-compute functions. This attribute is useful for mask writing and surface inspection, which require extensive data processing. Intel Agilex 7 FPGAs also allow end users the flexibility to use an integrated Arm processor or an Intel Xeon® Scalable processor via a Compute Express Link (CXL). Compared to Intel Stratix 10 FPGAs, Intel Agilex 7 devices demonstrate up to a 40% reduction in total power consumption or up to a 50% increase in performance [20].

Conclusion

As demand for ICs increases, foundries worldwide continue to assess ways to operate at 100% capacity efficiently. Advancing certain semiconductor manufacturing tools with FPGAs offers a promising solution. Although published research is limited regarding FPGA use in semiconductor manufacturing equipment (not for lack of use), this paper demonstrates their value in innovative IMS mask writers. Additionally, researchers and companies have successfully demonstrated FPGA use in image-processing applications. Market research also suggests that they have great potential for unmatched performance in surface inspection applications [3]. FPGAs demonstrate numerous advantages over competitors, including reconfigurability, rapid time to market, low latency, power efficiency, deterministic signal retrieval, parallel processing capability, and more. They are a feasible solution for any application involving high-throughput data processed at high speed or in real time.

Sources

[1] A. Kharpal, "Global semiconductor sales top half a trillion dollars for the first time as chip production gets boost," CNBC, 2022.

[2] Markets and Markets, "Semiconductor Equipment Manufacturing Market with COVID-19 Impact Analysis - Global Forecast to 2027," 2022.

[3] Markets and Markets, "Surface Inspection Market with COVID-19 Impact Analysis - Global Forecast to 2026," 2021, pp. 142-143.

[4] M. Nakajima, T. Ikushima and L. Edwards, "Case Study: Using Intel FPGAs for Real-Time Image Processing in High-Performance Industrial Cameras," Intel Corporation, 2023.

[5] J. Morss, "FPGAs vs. GPUs: A Tale of Two Accelerators," Dell Technologies, 2019.

[6] C. Maxfield, FPGAs: Instant Access, Elsevier Inc., 2008.

[7] Intel Corporation, "FPGA Optimization Guide for Intel® oneAPI Toolkits: FPGA Architecture Overview," [Online]. Available: https://www.intel.com/content/www/us/en/docs/oneapi-fpga-add-on/optimization-guide/2023-2/fpga-architecture-overview.html. [Accessed 25 10 2023].

[8] Intel Corporation, "The Journey Inside: Microprocessors, Intel Teach Program," 2010. [Online]. Available: https://www.intel.com/content/dam/www/program/education/us/en/documents/the-journey-inside/microprocessor/tji-microprocessors-back.pdf. [Accessed 25 10 2023].

[9] University of Illinois Urbana – Champaign (Department of Electrical and Computer Engineering), ECE 444 Lecture: Theory and Fabrication of Integrated Circuits, 2023.

[10] W. Kern, Overview and Evolution of Silicon Wafer Cleaning Technology, Elsevier Inc., 2018.

[11] Intel Newsroom, "Intel Mask Operation: An Inside Look at a Critical Manufacturing Step," YouTube, 2019.

[12] C. Klein and E. Platzgummer, "MBMW-101: World's 1st High-Throughput Multi-Beam Mask Writer," in Proceedings of SPIE, San Jose, 2016.

[13] S.H. Huang and Y.C. Pan, "Automated visual inspection in the semiconductor industry: A survey," Computers in Industry, vol. 66, pp. 1-10, 2015.

[14] M. Jariya et al., "Silicon wafer defect pattern detection using machine learning," in Materials Today: Proceedings, 2023.

[15] Texas Instruments, "Application Guide: Parallel Processing with the TMS320C4x," 1994.

[16] A. HajiRassouliha et al., "Suitability of recent hardware accelerators (DSPs, FPGAs, and GPUs) for computer vision and image processing algorithms," Signal Processing: Image Communication, vol. 68, pp. 101-119, 2018.

[17] S. W. Smith, "Chapter 28," in The Scientist & Engineer's Guide to Digital Signal Processing, California Technical Publishing, 1999.

[18] A. Boutros et al., "Beyond Peak Performance: Comparing the Real Performance of AI-Optimized FPGAs and GPUs," in International Conference on Field-Programmable Technology (ICFPT), Maui, HI, 2020.

[19] Y. Hasani et al., "FPGA based image processing for optical surface inspection with real-time constraints," in Proceedings of SPIE, Milan, Italy, 2014.

[20] Intel Corporation, "Product Brief: Intel® Agilex® 7 FPGAs and SoCs," 2023.

[21] W. Chen et al., "A High Speed Image Preprocessing Method for IC Wafer Inspection," in International Conference on Electronic Packaging Technology & High Density Packaging, Shanghai, China, 2008.

[22] Intel Corporation, "Intel FPGAs for Industrial Solutions Playbook," 2022. [Online]. Available: https://www.intel.com/content/dam/www/central-libraries/us/en/documents/2022-08/intel-fpga-industrial-solutions-playbook-2022.pdf. [Accessed 25 10 2023].

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software, or service activation. Performance varies depending on system configuration. Check with your system manufacturer or retailer or learn more at www.intel.com.

Intel reserves the right to make changes to any products and services at any time without notice. Intel assumes no responsibility or liability arising out of the application or use of any information, product, or service described herein except as expressly agreed to in writing by Intel. Intel customers are advised to obtain the latest version of device specifications before relying on any published information and before placing orders for products or services.

No product or component can be absolutely secure.

Your costs and results may vary.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. *Other names and brands may be claimed as the property of others.

WP-01332-1.0