Exploring the SoM Architecture and Applications of AMD Kria

Ayush Kumar Gupta, Kairatbek Davranbekov, Sai Mani Ritish Upadhyayula, Vidhiben Ashokbhai Vanani

Team 5 CS533 Computer Architecture, MSCS, City University of Seattle

[guptaayushkumar@cityuniversity.edu](mailto:guptaayushkumar@cityuniversity.edu)

[davranbekovkairatbe@cityuniversity.edu](mailto:davranbekovkairatbe@cityuniversity.edu)

[upadhyayulasaimanir@cityuniversity.edu](mailto:upadhyayulasaimanir@cityuniversity.edu)

[vananividhibenashok@cityuniversity.edu](mailto:vananividhibenashok@cityuniversity.edu)

**Abstract**

The AMD Kria System-on-Module (SoM) represents a modern approach to high-performance processing with remarkable adaptability. Designed for applications like AI acceleration, 5G, and edge computing, it integrates advanced components such as FPGA fabric, AI engines, and ARM processing cores into a compact, efficient module. Its architecture supports parallel processing and AI workloads while maintaining scalability and power efficiency. This paper explores the Kria’s unique features, including its memory management and connectivity options, as well as its relevance in industry use cases. By leveraging adaptive computing acceleration and flexible design, AMD Kria offers solutions for the ever-evolving needs of technology-driven industries.

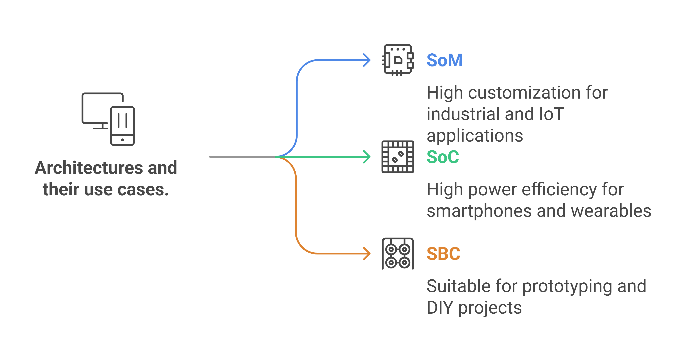
**Keywords:** Adaptive SoC, FPGA, AI engines, AMD Kria, Parallel Computing

**1.** **INTRODUCTION**

The fast growth of technology has increased the need for efficient, scalable, and high-performance computing solutions, especially in embedded systems and Internet of Things (IoT) applications. System-on-Modules (SoMs) have become popular by combining essential processing parts such as central processing units (CPUs), memory, power management, and connectivity features—into small, modular units than compared to completely integrated SoC architecture as shown in Figure (1). This combination makes development easier, speeds up the time it takes to bring products to market, and offers more flexibility for various industrial applications.

One of the leading SoM innovations is the AMD Kria, a mixed architecture that combines adaptable FPGA fabric, scalar processors, and AI engines. Designed to meet the demanding needs of modern high-performance applications, the

AMD Kria SoM is especially well-suited for areas like 5G technology, artificial intelligence (AI) acceleration, edge computing, and data center operations.

**Figure 1**: Illustration showcasing types of architectures and their applications.

This research aims to provide a thorough examination of the AMD Kria SoM's architecture, focusing on its instruction set, memory and I/O handling, specialized processing units, and overall performance. By examining these aspects, we aim to explain the SoM's capabilities and its potential impact on current and future technological applications.

**Figure 2:** AMD Kria Board

**2.** **DIFFERENT KRIA SOMs**

The Kria portfolio of system-on-modules (SOMs) includes the K24 and K26, which are designed for use with a carrier board as displayed in Figure (2) and share a common connector for easy migration. Both are based on a Zynq UltraScale+ MPSoC.

The K24 and K26 SOMs differ in the following ways:

|  |  |  |
| --- | --- | --- |
| **Feature** | **K24** | **K26** |
| **Size** | Half the size of a credit card | Larger than K24 |
| **Power Consumption** | 2.5W (base), up to 5W | Higher than K24 |
| **Logic Cells** | ~150,000 | ~250,000 |
| **DSP Slices** | Fewer | More |
| **Memory** | 2GB LPDDR4 (ECC-protected in industrial-grade) | 4GB DDR4 |
| **Operating System** | Ubuntu 22.04 Server | Ubuntu 22.04 Desktop |
| **I/O Availability** | Less I/O | More I/O |
| **Transceivers** | Not mentioned | 12.5 Gbit/s transceivers |
| **Cooling Requirement** | No active cooling required | May require active cooling |
| **Ideal Applications** | Smaller drones, motor control, DSP-intensive tasks, robotics, power generation, medical systems | Motion planning, navigation (KR260), vision processing (KV260), demanding robotics applications |
| **Primary Design Criteria** | Size, cost, power efficiency | Performance, higher processing needs |

**Table 1:** Comparison between K24 and K26

Both the K24 and K26:

* Use the same quad core Arm Cortex-A53 application processor and a dual core Arm Cortex-R5F real-time processor.
* Include DSP, programmable logic, a deep learning processor, a GPU, and multiple high-speed transceivers.
* Have a hardware root of trust for cybersecurity support.
* Offer a common connector for easy migration between the two.
* Are intended for production volumes.
* Come with starter kits that allow customers to quickly prototype their applications.
* Can be used in robotics applications.

The Kria SOMs are intended to speed up customers' time to market by enabling hardware developers to focus on their application differentiation and enable software developers to start sooner with pre-built hardware. They allow for the development of hardware-accelerated edge devices without requiring FPGA expertise. These SOMs support a variety of developer types and flows including RTL developers using Vivado, C/C++ developers using Vitus, MATLAB Simulink, and Python developers.

**3.** **INSTRUCTION SET ARCHITECTURE**

The AMD Kria SOMs use a custom-built Zynq UltraScale+ MPSoC .As seen in Figure (3) The processing system (PS) of the MPSoC includes an Arm Cortex-A53 based application processing unit (APU) and an Arm Cortex-R5F based real-time processing unit (RPU).

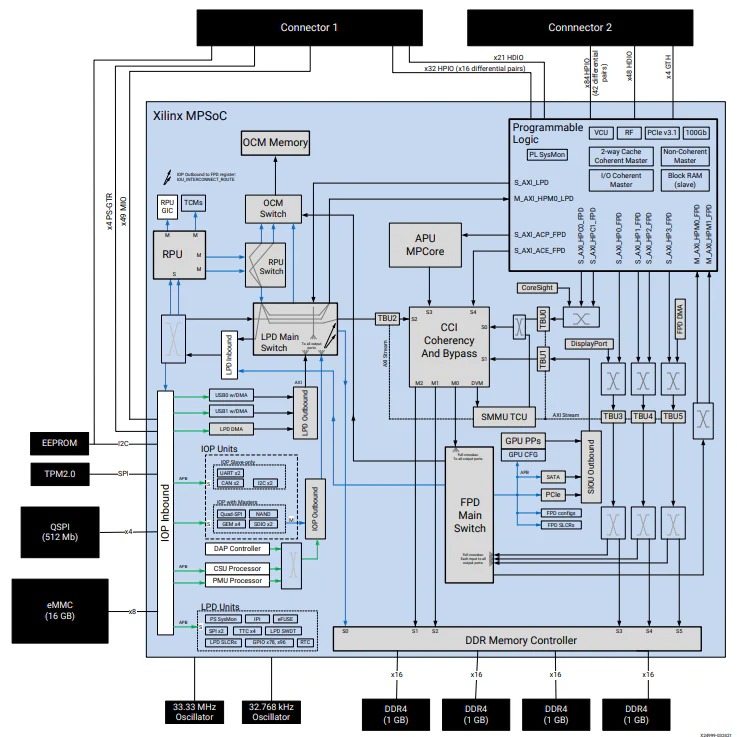
**Instruction Set:**

* The APU consists of quad-core Cortex-A53 processors with SIMD, VFP4 floating point, and cryptography extensions [4].
* The RPU consists of dual-core Cortex-R5F processors with floating point unit support [4].
* The Kria SOMs support the Arm instruction set architecture (ISA) [5].
* The programmable logic (PL) can be customized using Vitis, and/or Vivado Design Suite for custom instructions or acceleration.
* The Kria SOMs also include a Deep Learning Processing Unit (DPU) that can be used for accelerating AI inference.
* The Kria SOMs can be programmed with accelerated applications from the Xilinx App Store, which are pre-built for specific use cases and can be extended by the user.

**Memory:**

The Kria SOM has multiple types of memory:

* LPDDR4 Memory: The K24 SOM has 2 GB of 32-bit wide LPDDR4 memory. The industrial grade version has ECC support. The memory controller has configurable quality-of-service capabilities.
* QSPI Flash Memory: The K24 SOM has 512 Mb of QSPI flash memory. It can be used as the primary boot device for the MPSoC processing subsystem.
* eMMC Flash Memory: The K24 SOM has 32 GB of eMMC flash memory. It can be used as a primary or secondary boot device.
* EEPROM: The K24 SOM includes a 64 Kb EEPROM that stores device configuration, identification, and manufacturing data.
* Block RAM: The PL includes 216 blocks of 36 Kb Block RAM, totaling 7.6 Mb.
* Distributed RAM: The PL also has 1.8 Mb of distributed RAM.
* On-Chip Memory (OCM): The MPSoC has internal on-chip memory that can be used for secure boot.
* The SOM can use battery-backed RAM (BBRAM) for key storage, when battery backup is provided on the carrier card.
* In addition to the above, the Kria SOMs include a memory interface with 4 GB of 64-bit DDR4 memory.



**Figure 3:** AMD K26 Block Diagram

**4.** **SOM INPUT/OUTPUT INTERFACES**

The Kria SOMs provide a variety of input/output (I/O) interfaces through the Zynq UltraScale+ MPSoC, which includes both a processing system (PS) and programmable logic (PL). The I/O is highly configurable, offering flexibility in how the SOM interacts with external devices.

**Key aspects of I/O handling in Kria SOMs:**

1. Multiplexed I/O (MIO) Banks: The MPSoC has three MIO banks [4]. The Kria SOM uses the first bank for on-board peripherals, while the other two are customizable and available through the SOM connector interface [4]. All three MIO banks are powered by the SOM with a 1.8V power rail [4].

* MIO bank 501 contains MIO pins, with some pins reserved for power management [5].
* MIO bank 502 contains MIO pins [77:52] [5].
* The maximum data rate supported on MIO signals is 250 Mb/s [5].

1. High-Density I/O (HDIO) Bank: The HDIO bank (HDA, bank 26) is accessible through both the SOM240\_1 and SOM40 connectors [6].

* It supports 23 single-ended signals [6].
* Three signals are clock-capable inputs [6].
* The maximum data rate supported on HDIO signals is 250 Mb/s [7].
* The VCCO for the HDA bank is supplied by the carrier card [7].

1. High-Performance I/O (HPIO) Banks: The HPIO banks (HPA, banks 65 and 66) are accessible through both the SOM240\_1 and SOM40 connectors.

* Most signals in the HPIO banks are routed as differential pairs.
* The HPIO bank has a separate differential global clock input.
* The maximum data rate supported on HPIO signals is 2.5 Gb/s.
* The VCCO for the HPA bank is supplied by the carrier card.

1. PS-GTR Transceivers: Four dedicated PS-GTR receivers and transmitters with up to 6.0 Gb/s data rates supporting SGMII, tri-speed Ethernet, PCI Express® Gen2, serial ATA (SATA), USB3.0, and DisplayPort are accessible through the SOM240\_1 connector.

* The transceivers are configured via the AMD Vivado™ Design Suite.

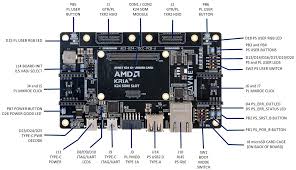
1. User-Defined I/O: The Kria SOM provides a large number of flexible user-defined I/O that can be configured for various I/O standards and voltage levels. Voltage levels for each HDIO and HPIO bank can be customized by the SOM carrier card design.
2. I/O Standards: The Kria SOM supports all I/O standards supported by the respective bank, except those requiring a reference voltage (VREF).

* Connectors: The SOM240\_1 connector provides access to MIO banks (501, 502), HPIO bank 65 (HPA), HDIO bank 26 (HDA), and the PS-GTR transceivers.
* The SOM40 connector provides access to HPIO banks 65 & 66, and HDIO bank 26.

1. Boot Mode Pins: The boot mode is selected by tying the MODE [3:0] pins on the carrier card.
2. Sideband Signals: These consist of power, processor, and configuration signals.

* JTAG signals connect to the SOM Zynq UltraScale+ MPSoC JTAG port.
* PS\_REF\_CLK input is connected to a 33.33 MHz oscillator.
* PS RTC inputs are connected to a 32.768 kHz crystal.
* I2C signals connect to an I2C master on MIO bank 500.
* PS\_MODE [3:0] pins connect to the SOM Zynq UltraScale+ MPSoC PS\_MODE pins.
* PS\_POR\_L signal drives the PS\_POR\_B signal on the Zynq UltraScale+ MPSoC.
* PWRGD\_LPD\_M2C indicates the power status of all SOM PS low-power domain rails.
* VCCOEN\_PS\_M2C enables the PS VCCO rails supplied by the carrier card.
* VCCOEN\_PL\_M2C enables the PL VCCO rails supplied by the carrier card.

1. Power Management: The K24 SOM has a PMU that controls power sequencing to the onboard PMICs.
2. Flexibility: The programmable I/Os and internal programmable logic can support transitions from one technology or interface to another, such as MIPI to SLVS-EC.

**Figure 4:** AMD K24 I/O Development Kit

The Kria SOM's I/O system is designed to be highly adaptable as we see in Figure (4), this allows developers to tailor the interfaces to specific application needs. The combination of configurable MIO, HDIO, and HPIO banks, along with high-speed transceivers, provides a versatile platform for various embedded applications.

**5. SOFTWARE DEVELOPMENT TOOLS**

The K24 SOM is supported by a variety of software and development tools, catering to different types of developers.

These tools and software flows include:

* Vivado for RTL developers.
* Vitus for C or C++ developers who want to code in a familiar language and implement that design in the programmable logic.
* MATLAB Simulink for control systems engineers.
* Python for Python developers.
* Ubuntu operating system, with the long-term support version 22.04. The K24 uses Ubuntu 22.04 Server.
* Docker Hub for deploying containerized apps supporting open Ubuntu Linux.

The Kria SOMs also support several AI frameworks:

* TensorFlow
* Cafe
* PyTorch

The K24 also supports the same software and development tools as previous Kria SOMs and kits.

The Kria App Store offers a variety of apps that can be downloaded directly to the target hardware. There are over 25 apps available for users to download for the KV260, KR260 and KD240. These apps cover applications ranging from communications, control, HMI and motor control. As we see in Figure (5). The apps are filled with both AMD and their ecosystem.

The Kria SOMs are designed to be a "welcome" sign for embedded and AI software developers who want to adopt the benefits of FPGA technologies but do not know where to start. The Kria SOMs speed up the time to market by enabling software developers to start sooner with pre-built hardware.

A black screen with blue and white text

AI-generated content may be incorrect.**Figure 5:** AMD K24 Developer Support

**6. InFO PACKAGING (PERFORMANCE and SIZE IMPACT)**

InFO (integrated fan out) packaging technology significantly impacts the K24's performance and size by removing the traditional substrate layer found in BGA (Ball Grid Array) construction.

Here's how it works and the resulting benefits:

* **Traditional BGA Packaging**: Typically includes the active silicon die, a substrate layer, and then the solder balls that connect to the printed circuit board (PCB).
* **InFO Packaging**: Eliminates the substrate layer, allowing the die to connect directly to the solder balls of the package.
* **Size Reduction**: By removing the substrate, InFO packaging enables a more compact footprint, reducing the size of the K24 SOM (System on Module) to roughly half the size of a credit card. This smaller size is beneficial for applications where space is a key constraint, such as small drones.
* **Thermal Conductivity**: The direct connection from the die to the PCB, enabled by InFO packaging, results in greater thermal conductivity. This means that heat generated by the die can be dissipated more efficiently onto the PCB.
* **Power Efficiency**: While not directly stated as reducing power consumption, more efficient heat dissipation can contribute to better overall power management, especially in applications with high processing loads. The K24 is designed for applications where power and size are key design criteria, typically using 2.5 to 5 watts.
* **Performance**: The K24, leveraging the Zynq UltraScale Plus adaptive SoC architecture, has a 2x latency advantage and enables high-performance and power-efficient DSP capabilities.
* **Scalability**: The K24 is a scalable option from the K26 where size, cost, and power efficiency are the primary design criteria. It complements the K26 SOM with connector compatibility.

In summary, InFO packaging allows the K24 to be significantly smaller, have improved thermal performance, and contribute to efficient power management, making it suitable for a range of applications with different requirements.

**7. COMPARING AMD KRIA WITH NVIDIA JETSON**

NVIDIA Jetson and AMD Kria are both leading System-on-Modules (SOMs) platforms used for edge AI and embedded computing, both have different applications. NVIDIA Jetson SOMs integrate ARM-based CPUs with powerful NVIDIA GPU, CUDA, Tensor Cores, and DeepStream used for AI inferencing, robotics, and vision-based applications. In contrast, AMD Kria SOMs combine ARM CPUs with FPGA-based acceleration, this offers real-time deterministic performance ideal for industrial automation and low-latency applications. While Jetson excels in AI model performance and deep learning, Kria provides a more customizable and power-efficient solution for real-time embedded systems.

|  |  |  |
| --- | --- | --- |
| Feature | NVIDIA Jetson | AMD Kria |
| Processing Unit | ARM Cortex-A + NVIDIA GPU (CUDA, Tensor Cores) | ARM Cortex-A53 + FPGA-based acceleration |
| AI Acceleration | GPU-accelerated AI (TensorRT, DeepStream) | FPGA-based AI acceleration (Vitis AI) |
| Edge AI Performance | High FPS inferencing for vision/robotics | Low-latency, real-time processing |
| Power Consumption | 5W–60W (varies by model) | 5W–15W (more power-efficient) |
| Memory | Up to 64GB LPDDR5 (AGX Orin) | Up to 4GB DDR4 (Kria K26) |
| Customization | Fixed GPU architecture | Fully customizable FPGA fabric |
| Development Tools | JetPack SDK, CUDA, TensorRT | Vitis AI, PYNQ, Vivado |
| Ease of Use | Large AI/ML developer ecosystem | Requires FPGA knowledge |
| Best Use Cases | AI vision, robotics, deep learning, autonomous vehicles | Industrial automation, real-time control, edge AI |
| Drawbacks | Higher power, less flexibility in hardware acceleration | Smaller ecosystem, requires FPGA programming |

**Table 2:** Nvidia Jetson vs AMD Kria

NVIDIA Jetson SOMs are mostly ideal for AI-powered applications that require high-performance inferencing, like robotics and autonomous vehicles. AMD Kria SOMs excel in industrial automation and real-time applications, where FPGA-based processing provides a deterministic performance with much lower power consumption. The choice depends on whether AI model acceleration (Jetson) or real-time adaptability (Kria) is the priority.

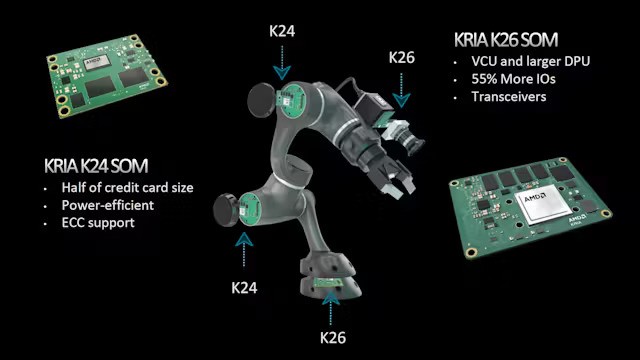
|  |  |  |
| --- | --- | --- |
| **Category** | **AMD Kria (e.g., Kria KV260)** | **NVIDIA Jetson (Jetson AGX Orin, Xavier, Nano)** |
| **Compute Performance** | ~1.4 TOPS (INT8) | Up to 275 TOPS (AGX Orin) |
| **Power Efficiency (TOPS/W)** | ~0.28 TOPS/W (FPGA-based efficiency) | ~4.58 TOPS/W (AGX Orin) |
| **Development Cost** | Higher (FPGA expertise, Vitis AI licenses) | Lower (CUDA ecosystem, easier software portability) |
| **Hardware Cost** | ~$300–$600 (KV260, edge SoMs) | $99 (Nano) to $2,000+ (AGX Orin) |
| **Total Cost of Ownership (TCO)** | Lower for long-term industrial applications (custom power optimization) | Higher due to power consumption, but lower initial development costs |
| **Performance Per Dollar (TOPS/$)** | ~2.3 TOPS per $100 | ~13.75 TOPS per $100 (AGX Orin) |

**Table 3**: Performance and cost-efficiency metrics of AMD Kria vs Nvidia Jetson

**8. APPLICATIONS AND INDUSTRY USAGE**

AMD Kria are highly suited SOMs for a wide array of applications, especially in domains that require high computational resources, real-time processing, and hardware-level customization. These domains include 5G networks, autonomous systems, industrial automation, and AI/ML acceleration. The FPGA-based architecture enables efficient data processing which makes them ideal for industries with low-latency, high-performance computing at the edge.

One of the advantages of AMD Kria is its suitability for edge computing, where processing is brought closer to the data source, significantly reducing latency. This particularly benefits applications like real-time video analytics, smart traffic management, and autonomous systems, where rapid decision-making is crucial. In industrial automation as shown in Figure (6), Kria SOMs are deployed in smart factories, optimizing machine vision, predictive maintenance, and robotics for enhanced efficiency and reduced downtime.

**Figure 6:** AMD K24 and K26 Usage

Furthermore, Kria SOMs are essential in data centers due to their focus on energy efficiency and scalability. Their design is tailored for extensive operations, efficiently handling power usage and heat management. Moreover, smart infrastructure solutions, including intelligent surveillance, IoT edge devices, and real-time control systems, utilize Kria’s adaptable FPGA architecture to accommodate changing technological demands.

Overall, AMD Kria excels in applications that demand real-time deterministic performance and hardware adaptability, making it a preferred choice for industries seeking innovation, efficiency, and scalability.

### **9. CONCLUSION**

In conclusion, the AMD Kria System-on-Module (SoM) represents a significant advancement in embedded computing, offering a versatile and high-performance solution for a wide range of applications. Its integration of adaptable FPGA fabric, scalar processors, and AI engines provides a flexible platform capable of meeting the demanding requirements of modern technologies such as 5G, artificial intelligence acceleration, edge computing, and data center operations.

The comprehensive examination of the AMD Kria SoM's architecture—including its instruction set, memory and I/O handling, specialized processing units, and overall performance—highlights its potential to drive innovation and efficiency across various industries. By leveraging the unique capabilities of the Kria SoM, developers and engineers can accelerate development timelines, reduce complexity, and deploy scalable solutions that are adaptable to evolving technological landscapes.  
  
As the demand for efficient, scalable, and high-performance computing solutions continues to grow, the AMD Kria SoM stands out as a compelling choice for those seeking to harness the power of advanced embedded systems in their applications.

**10. WORKLOAD ASSIGNMENT**

|  |  |
| --- | --- |
| **TEAM MEMBER** | **CONTRIBUTION** |
| **Kairatbek** | Handling the introduction, abstract, conclusion, formatting, and references. |
| **Ayush** | Focused on instruction set and covers memory, I/O handling. |
| **Vidhi** | Covered performance characteristics and applications. |
| **Ritish** | Analyzing special processing units, strategic analysis, and recommendations. |

**11.** **REFERENCES**

1. Dotterel Technologies. (2023). Dotterel Reinvents Audio Capture with AMD KriaTM SOM. In AMD + Dotterel Technologies Case Study [[Case study]](https://www.amd.com/content/dam/amd/en/documents/resources/case-studies/dotterel-case-study.pdf).
2. AMD Adaptive Computing. (2023). Kria K24 SOM Data Sheet. In Kria K24 SOM Data Sheet [[Report]](https://www.mouser.com/pdfDocs/ds985-k24-som.pdf?srsltid=AfmBOooj7dfN7deALZ_pAFbx1WsKJW9CHwe7E5UCo4XrkL9icwOqdQdi).
3. Radmantis, Local Coho, Bravo, A., & Huber, R. (n.d.). [[Case Study]](https://www.amd.com/content/dam/amd/en/documents/resources/case-studies/radmantis-case-study.pdf)
4. Xilinx, Inc. (2021). Achieving Embedded Design Simplicity with Kria SOMs. In WP528. [[Report]](https://docs.amd.com/v/u/en-US/wp528-simplified-som)
5. Electronic Design. (2023, September 19). A look inside AMD’s KRIA K24 System-on-Module [[Video]](https://www.youtube.com/watch?v=wCoo3KxBw1w). YouTube.
6. Component Search Engine. (2024, October 21). AMD KRIA Adaptive Motor Control SOMS [[Video]](https://www.youtube.com/watch?v=kuorJW4IHx0). YouTube.
7. Mouser Electronics. (2024, February 13). Introducing the AMD KriaTM️ K24 SOM and KD240 Drives starter kit: Tech Chats. Mouser Electronics [[Video]](https://www.youtube.com/watch?v=5d8Ez6vTkV8). YouTube.
8. AMD. (n.d.). Kria SOM Product Selection Guide. AMD. [[Guide]](https://docs.amd.com/v/u/en-US/kria-som-psg)
9. Antmicro. (2022, September). Kria UltraScale+ SOM baseboard. [[Article]](https://antmicro.com/blog/2022/09/kria-ultrascale-plus-som-baseboard/)
10. Newark. (n.d.). AMD Xilinx Kria K24 SOM. [[Article]](https://www.newark.com/amd-xilinx-kria-k24)
11. Reddit user. (2023, June 24). Pricing of Kria SOM vs. UltraScale SoC. [[Reddit]](https://www.reddit.com/r/FPGA/comments/14hw8yr/pricing_of_kria_som_vs_ultrascale_soc/?rdt=58605)
12. Reddit user. (2023, September 20). The AMD Kria K24 SOM: Save on power and reduce costs. [[Reddit]](https://www.reddit.com/r/AMD_Stock/comments/16nkjwt/the_amd_kria_k24_som_save_on_power_and_reduce/)
13. Crowd Supply. YouTube. (n.d.). [[Video]](https://youtu.be/PVQkpRfwP2I)