SiFive: Driving Innovation In RISC-V Vector Processing

THE COMPANY APPEARS WELL POSITIONED TO CHALLENGE CPU INCUMBENTS WITH HIGH PERFORMANCE RISC-V CPUS AND VECTOR EXTENSIONS TO THE OPEN ISA ARCHITECTURE.

INTRODUCTION

The RISC-V CPU Instruction Set Architecture (ISA) is emerging as a serious challenger to current CPUs based on proprietary architectures, creating new opportunities for chip designers and investors alike. While RISC-V first gained traction in the low-end embedded market, where the open ISA model afforded more cost-effective designs, RISC-V is now getting more wind in its sails due to performance and power efficiency, especially with vector enhancements with advantages over alternative ISAs. Behind all the RISC-V buzz, SiFive is the company that makes many of the innovations of the open-source CPU architecture so appealing.

In addition to open-source and computing efficiency, RISC-V now offers a well-designed, highly efficient vector processing extension which can enable significant acceleration in applications where large data sets need to be manipulated in parallel. This research paper explores the potential of RISC-V, and how SiFive intends to drive further innovation with its vector extensions to penetrate high-value markets such as Artificial Intelligence (AI) in the data center, on the edge, in high end image processing, and in automotive infotainment. SiFive's latest release of the market leading X280 processor will lead this charge with enhanced capabilities including the newly announced SiFive Vector Coprocessor Interface Extension, coherent multicore support for up to 16 cores and WorldGuard Security.

BACKGROUND: RISC-V BENEFITS AND SiFive's ROLE

Silicon Valley startup SiFive has assumed the role of industry leadership and commercial IP innovation for the RISC-V movement, providing tested Intellectual Property (IP) and support for chip developers who incorporate RISC-V into their products.
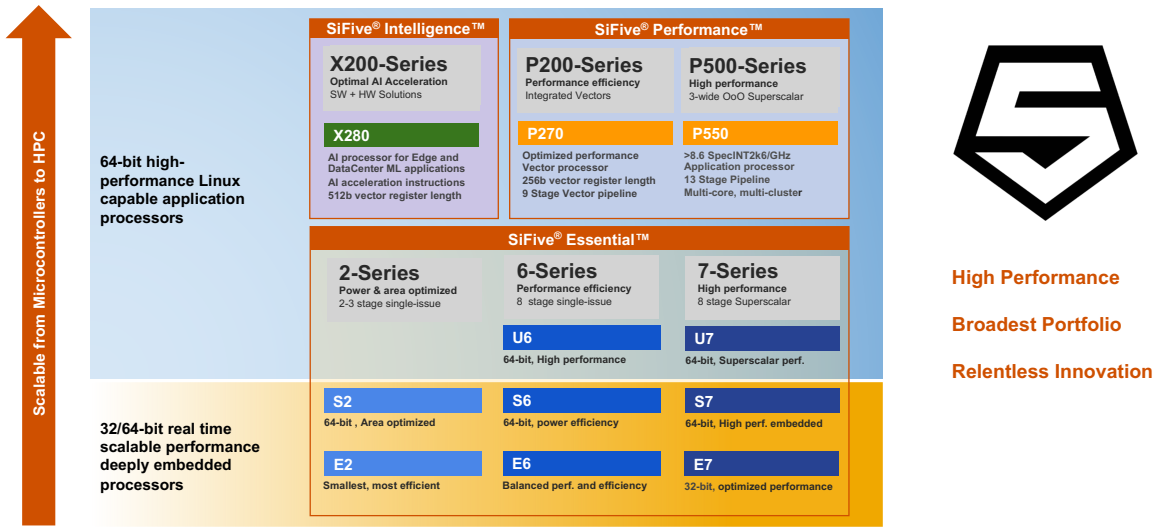
RISC-V portends to offer an alternative to proprietary processor cores in a user-friendly licensing and development environment. The raw performance of the latest SiFive RISC-V implementation is rapidly closing the gap with the incumbents, but with lower power and die area, and with no lock-in to a closed architecture. SiFive is further enhancing its portfolio with vector processing extensions that differentiate the ISA from other architectures.

SiFive is essentially the most visible and accomplished commercial steward of RISC-V, providing validated IP and support as well as open and proprietary enhancements to the RISC-V development community. With this open-standard approach and dependable tested IP, SiFive has garnered over 300 design wins with over 100 firms, including 8 of the top 10 semiconductor companies. With the addition of vector processing, we expect this trend to accelerate.

SiFive Strategy and Product Portfolio

In September 2020, SiFive announced it had hired CEO Patrick Little as the new President, CEO, and Chairman. Coming from Qualcomm where he led the company's successful foray into the automotive sector, Mr. Little has sharpened the company's business model on developing and licensing IP, selling the SiFive's OpenFive SoC design business to AlphaWave for $210 million. The company subsequently raised $175 million in a Series F funding round at a $2.5 billion post-money valuation. The latest round brings SiFive's total venture funding to over $350 million and was led by global investment firm Coatue Management LLC. Existing investors Intel Capital, Sutter Hill, and some others joined this latest round.

SiFive® RISC-V Processor IP Portfolio

	SiFive® Intelligence™	SiFive® Performance™	
	X200-Series Optimal AI Acceleration SW + HW Solutions	**P200-Series** Performance efficiency Integrated Vectors	**P500-Series** High performance 3-wide OoO Superscalar

64-bit high-performance Linux capable application processors

X280	P270	P550
AI processor for Edge and DataCenter ML applications AI acceleration instructions 512b vector register length	Optimized performance Vector processor 256b vector register length 9 Stage Vector pipeline	>8.6 SpecINT2k6/GHz Application processor 13 Stage Pipeline Multi-core, multi-cluster

SiFive® Essential™

2-Series Power & area optimized 2-3 stage single-issue	**6-Series** Performance efficiency 8 stage single-issue	**7-Series** High performance 8 stage Superscalar
	U6 64-bit, High performance	U7 64-bit, Superscalar perf.

32/64-bit real time scalable performance deeply embedded processors

S2 64-bit , Area optimized	S6 64-bit, power efficiency	S7 64-bit, High perf. embedded
E2 Smallest, most efficient	E6 Balanced perf. and efficiency	E7 32-bit, optimized performance

Scalable from Microcontrollers to HPC

High Performance

Broadest Portfolio

Relentless Innovation

Figure 1: The SiFive RISC-V portfolio.

VECTOR PROCESSING, RISC-V VECTOR EXTENSIONS AND SIFIVE INTELLIGENCE EXTENSIONS

In today's heterogeneous world of Domain-Specific Processors, parallel processing of large data sets is a critical adjunct to scalar processing. While accelerators such as GPU's and ASICs provide some incremental performance, they come at significant cost and generally require connectivity to CPU's along with the cost of data transfers from the CPU to and from the accelerator. And each accelerator requires its own distinct programming model. Now with RISC-V, general vector processing in the CPU cores offers an alternative approach,

Vector processing, where instructions manipulate data across a large dataset of numbers, has been a foundation of high-performance computing since the Cray 1 supercomputer in 1975. Other processors such as Intel Xeon and Arm Neoverse support Single-Instruction-Multiple-Data (SIMD) as a vector extension, however such SIMD implementations are rapidly losing favor as they are proprietary to the ISA, suffer from massive instruction set bloat as new data widths are added, and are unwieldy and cumbersome to program. In fact, the lack of generalization of

SIMD instructions across data types and operators has bloated the x86 and Arm instruction sets. Consequently, SIMD programs on Arm or x86 can require 10 to 20 times more instructions compared to RISC-V vector instructions.

RISC-V Vector extensions (RVV) enables RISC-V cores to process data arrays alongside traditional scalar operations to parallelize the computation of single instruction streams on large data sets. SiFive helped establish RVV as a part of the RISC-V standard and has now extended the concept in two dimensions.

SiFive Intelligence Extensions Accelerate End-to-End Models

~24x
Speedup

~6x
Speedup

CONV_2D

DepthwiseCONV_2D
other

RISC-V
Scalar ISA

CONV_2D

DepthwiseCONV_2D
other

RISC-V
Vector ISA

CONV_2D
DepthwiseCONV_2D
other

SiFive
Intelligence ISA
Extensions

Execution Time

MobileNet v1 Inference (batch size=1), VLEN=512
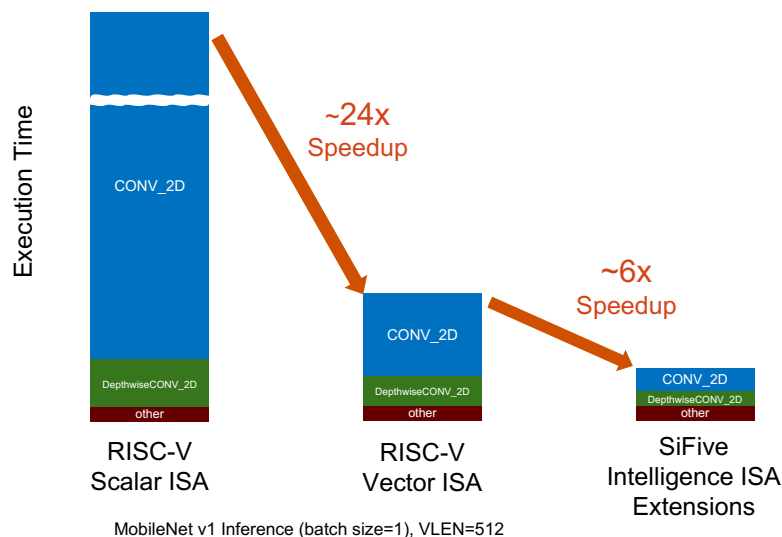
Figure 2: The SiFive extensions to the RISC-V vector capabilities can dramatically increase performance and efficiency.

First, the SiFive Intelligence Extensions add new operations such as matmuls for INT8, BF16 converts and compute operations, and enable vector instructions to operate on a broad range of AI/ML data types, including BFLOAT16. The SiFive Intelligence Extensions also add support for TensorFlow Lite for Machine Learning models, reducing the cost to port AI models to SiFive based designs.

Second, the company has recently announced the SiFive Vector Coprocessor Interface eXtensions (VCIX), enabling tight integration of the RISC-V CPU with dedicated accelerators. While a standalone x280 processor can perform very

SiFive: Driving Innovation In RISC-V
Copyright ©2022 Cambrian AI Research

well using its vector capabilities, some workloads require the additional performance of a dedicated special-purpose accelerator. The challenge here is how to minimize or even eliminate the costly movement of data from the CPU registers and DRAM to the accelerator, consuming power, and time. Also, traditional off-chip or off-package accelerators need to have their own memory subsystem, increasing costs and chip area. A tighter integration between the custom accelerator and the SiFive processors can help eliminate these disadvantages.

THE ADVANTAGES OF RISC-V VECTOR ENHANCEMENTS

The RVV extensions to RISC-V provide a single vector instruction set across a range of operations and data types, including Int-8, BFloat 16, and IEEE Floating Point 32 and 64-bit which greatly simplifies software development and reduces code size. The ISA is agnostic with respect to vector length, which provides full reuse and portability of libraries and applications. Instead of having a plethora of instructions to handle different vector lengths, RVV instructions operate on vectors whose lengths are specified by the VLEN variable, greatly simplifying code development and experimentation. Finally, RVV vector data resides in main memory, not a dedicated vector memory, reducing time and power needed to copy data before and after a vector operation.

VCIX REPRESENTS A STRATEGIC OPPORTUNITY FOR SIFIVE

In a world of increasing heterogeneity, there is a large opportunity to help SoC and System-on-Package (SoP) designers build tightly integrated solutions. The SiFive Vector Coprocessor Interface Extension (VCIX) is a direct interface between the X280 and a custom accelerator, enabling parallel instructions to be executed on the accelerator directly from the scalar pipeline. The custom instructions are executed from the standard software flow, utilizing the vector pipeline, and can access the full vector register set.

This new capability reduces both the design and test effort required when developing a customer accelerator, as the accelerator can share key processor resources, such as the vector register bank, main processor caching architecture, and memory system. The resulting system is both simpler to design, considerably more power and area efficient, and much easier to program.

SiFive: Driving Innovation In RISC-V   August 2022

THE SIFIVE PROCESSOR PORTFOLIO

The SiFive product portfolio is structured into three clearly differentiated product lines: the 32/64 bit Essential products (2-, 6-, and 7-Series) for embedded control/Linux applications, the SiFive Performance Series (the P200 and P500/P600 families) for high efficiency and higher performance, and the SiFive Intelligence Series (the X200 family) for parallelizable workloads such as Machine Learning at the edge and in data centers.

To capitalize on its advantage in vector processing, SiFive has built its vector capabilities into both the Performance P270 and the Intelligence X280 processors.
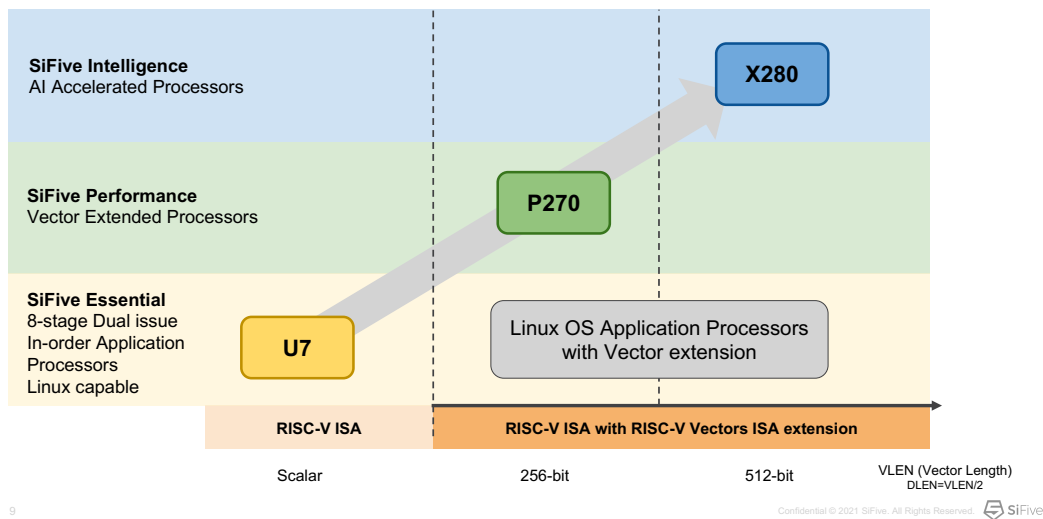
Figure 3: The portfolio includes the Essential, Performance, and Intelligence processors.

Let's take a deeper look into the P270 and X280. The Performance Series P270 is designed for high-performance applications with support for Linux and Hypervisors needed for data center applications. The 256-bit vector capability is complemented by a high-performance CPU, with an 8-stage dual-issue in-order (to keep it small and efficient) scalar micro-architecture. The cache-coherent architecture supports up to 8 cores.

The Intelligence Series X280 includes the SiFive Intelligence Extensions we discussed earlier, increasing performance and efficiency by up to 6-fold for Convolutional Neural Networks (Mobilenet with batch size =1) in addition to the 24-fold increase offered by RVV over scalar operations, and supports the TensorFlow Lite AI Framework with 512-bit vector registers. Taken together, SiFive promises nearly a 150-fold speedup.

SiFive Intelligence X280
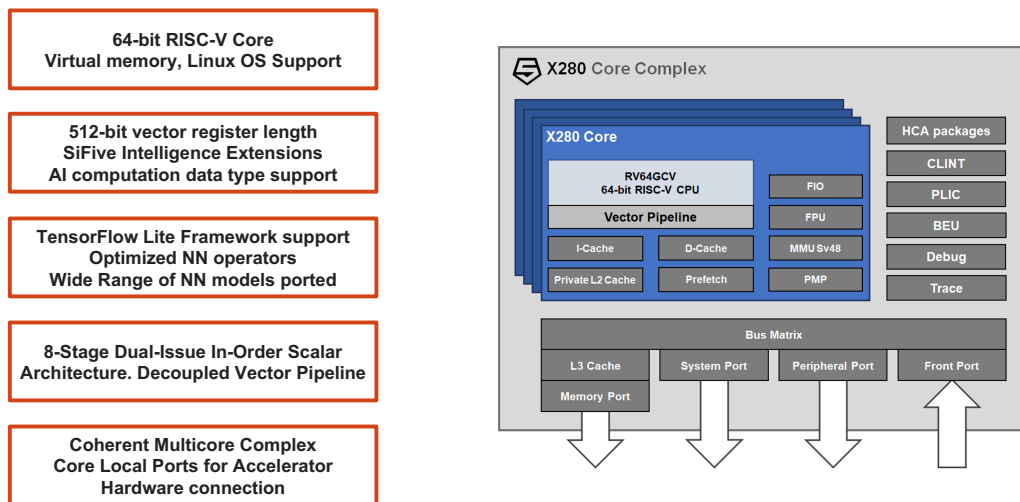
Figure 4: The flagship Intelligence X280 CPU is gaining traction in part due to its vector processing and Linux support.

EARLY ADOPTERS OF THE SIFIVE X280

SiFive X280 has already been adopted by several companies of note, including a Tier 1 semiconductor company and a US Federal Agency for a strategic initiative in the aerospace and defence sector. Another customer has selected the X280 for projects for its mobile devices and data center AI products. Similarly, a US company delivering autonomous self-driving platforms has selected the X280 for its next generation SoC. Of these opportunities, the last two could generate significant volumes, while the 1st could open more doors in the government sector.

On the startup front, we have already seen a number of SoC developers publicly announce their adoption of SiFive including Tenstorrent and Kinara (formerly known as DeepVision). Many are developing SoCs for AI acceleration, leveraging the vector processing of the X280 and complementing that with custom AI blocks. Tenstorrent tells us they are getting great support and that the cores are rock solid.

SiFive Development Tool Suite

Software for SIMD development and deployment has often been described as the Achilles Heel for vendors. The software for developing vector-accelerated applications on RISC-V is far more flexible. Starting with SiFive development tools which include the LLVM compiler with vector intrinsics and auto-vectorization support. For code already optimized to run SIMD, SiFive Recode translates SIMD instructions to RVV, simplifying migration of SIMD to RVV.

For AI applications, SiFive supports an Out-of-the-box software and processor hardware solution with TensorFlow Lite running under Linux OS to run NN models in the Object detention, Image Classification, Segmentation, Text, and Speech domains. Existing models can be run with little porting effort with a broad range of optimized NN operators in both 32-bit Float and Quantized 8-bit precisions.

Applications that can Benefit from Vector Processing

From our perspective, we believe that parallel processing is transitioning from the tool of a few to the norm for many applications, especially as AI and Machine Learning become pervasive. And as Moore's Law provides ever-diminishing returns, application developers still require more performance and Vector processing can provide the avenue for both higher levels of performance and better power efficiency especially with RISC-V. We see opportunities for RISC-V vector processing in multiple application domains including smart homes, telco, mobile devices, autonomous vehicles, industrial automation, robotic control, and health care. The simplicity and elegance of RVV and the performance gains are powerful selling points.

Figure 5: The X280 processor supports a wide range of use cases.

CONCLUSIONS

We are impressed by the progress that RISC-V and SiFive has made in the last few years. The new product line positioning makes a ton of sense, the processors are beefier, the software stack is getting much better and the vector extensions are impressive, both the open source RVV and the AI extensions the company has included in the Intelligence Series X280. The CPUs are relatively high performance with excellent scalability and power efficiency due to the simplicity that stems from the efficient RISC-V ISA and clever extensions. SiFive has also recently disclosed the intention of releasing an even higher performance P600 Series class Out-of-Order core with RISC-V vector compute in the near future.

Finally, the commitment to and leverage of the open-source community is perhaps RISC-V and SiFive's most important value they can offer as an alternative to Arm, especially for designers looking to build SoC solutions for Domain-Specific Architectures.