



Evaluating ASIC, DSP, and RISC Architectures for Embedded Applications

Marc E. Campbell
Northrop Grumman Corporation

Abstract

Mathematical analysis and empirical evaluation of the solid state equation $Power_{CMOS} = P = C \cdot V^2 \cdot f \cdot N \cdot \%N$ is presented in this paper which identifies a measurable metric for evaluating relative advantages of ASIC, DSP, and RISC architectures for embedded applications. Relationships are examined which can help predict relative future architecture performance as new generations of CMOS solid state technology become available. In particular, Performance/Watt is shown to be an Architecture-Technology Metric which can be used to

- calibrate ASIC, DSP, & RISC performance density potential relative to a solid state technology generations,
- measure & evaluate architectural changes, and
- project a architecture performance density roadmap.

1. Architecture Metric Basis

Intuitively, more efficient architectures will use less energy to complete the same task on the same generation CMOS solid state technology. Beyond this intuition,

(1) How can the relative advantages between ASIC, DSP, and RISC architectures be quantified?

(2) How can the relative advantages between different ASIC, DSP, and RISC options be projected into future embedded system design plans?

The key concept is the power consumption behavior of a CMOS gate.

1.1. CMOS device power consumption

Power consumption is an externally measurable physical property of CMOS ASIC, DSP, and RISC devices which are used to build larger computational systems. The power consumption for a CMOS device is shown in equation 1-1.

$$Power_{CPU} = P = C \cdot V^2 \cdot f \cdot N \cdot \%N \quad (1-1)$$

where,

C = Capacitance

V = CPU Core Voltage

f = CPU Core Clock Frequency

N = Number of Gates

$\%N$ = Percentage of Gates which change state on a given clock cycle.

Some primary architectural and technology components in the CMOS power consumption equation 1-1 can be noted.

1.2. Architecture

First, the architecture is a function of the number of gates N and how gates toggle to perform a task, $\%N$. Physical CMOS gate real estate is quantified by N . How the software interacts with the CMOS gates is quantified by how the gates toggle, $\%N$. Thus software architecture is also reflected in $\%N$ for software programmable devices. The system architecture of combined hardware and software is largely represented by $N\%N$.

$$Architecture = N \cdot \%N \quad (1-2)$$

More complex architectures require more gates. More specialized architectures have fewer gate toggles per unit task.

Thus, an Architecture Specialization metric can be defined based on how many clock cycles and gates are required to complete a given functional task. The fewer gates and cycles used to get the task done the more specialized, efficient, or “tuned” the processor is for the required task.

1.3. CMOS technology

Second, the advances in solid state process technologies can be observed as a function of gate geometry, voltage, and frequency. Newer technologies systems are design with smaller gate geometry, lower voltages, and higher frequencies. Changes in gate geometry affect changes in capacitance.

$$SolidStateTechnology\left(\frac{1}{C}, \frac{1}{V^2}, f\right) \quad (1-3)$$

2. Architecture Metric Observations

At this point, several observations based on sustained performance per watt can be made:

2.1. Clock frequency

Observation 1. Higher performance by only increasing the clock rate f has no net performance per watt benefit.

Let,

$$\begin{aligned} Performance &= \frac{task}{unit_time} = \frac{operations}{second} \\ &= \left(\frac{operations}{cycle}\right) \left(\frac{cycles}{second}\right) \\ &= O_c f \end{aligned} \quad (2-1)$$

where,

O_c = sustained operation per cycle, and

f = clock frequency.

Then it follows that,

$$\begin{aligned} \frac{Performance}{Power_{CPU}} &= \frac{\left(\frac{operations}{second}\right)}{watt} \\ &= \frac{O_c}{C \cdot V^2 \cdot N \cdot \%N} \end{aligned} \quad (2-2)$$

The observation that clock frequency f does not affect performance per watt for the same CPU architecture N is also illustrated in Figure 1-1 based on information available from

Digital Equipment Corporation (DEC) [2]. Figure 2-1 does not include the power consumption required by support logic.

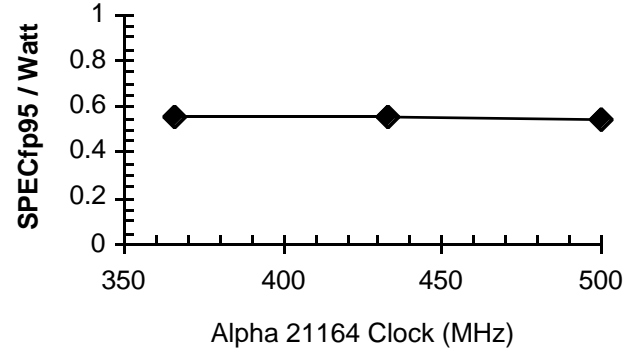


Figure 2-1. Alpha Performance/Watt vs. Clock

2.2. Voltage

Observation 2. A voltage (V^2) decrease significantly improves performance per watt for the same architecture.

For the same architecture $N\%N$ with corresponding O_c task definition, the relative performance per watt improvement can be shown by equation 2-3,

$$\begin{aligned} VoltageChangeBenefit &= \frac{\left(\frac{Performance}{Power_{CPU}}\right)_{new}}{\left(\frac{Performance}{Power_{CPU}}\right)_{old}} \\ &= \frac{\frac{O_c}{C \cdot V_{new}^2 \cdot N \cdot \%N}}{\frac{O_c}{C \cdot V_{old}^2 \cdot N \cdot \%N}} \\ &= \frac{CV_{old}^2}{CV_{new}^2} \end{aligned} \quad (2-3)$$

For the cases where C is near constant or the change in C is small with respect to the change in V^2 the performance per watt benefit is approximated as,

$$VoltageChangeBenefit \approx \frac{V_{old}^2}{V_{new}^2} \quad (2-4)$$

The approximate performance per watt benefit as an architecture voltage changes from 3.5 volts toward 2.0 volts is calculated in equation 2-5.

$$\frac{V_{3.3}^2}{V_{2.0}^2} = \frac{3.3^2}{2.0^2} = 2.7 \quad (2-5)$$

Note also that,

$$\frac{\partial P}{\partial f} \propto V^2 \Rightarrow \frac{\frac{\partial P_{new}}{\partial f}}{\frac{\partial P_{old}}{\partial f}} = \frac{V_{new}^2}{V_{old}^2} \quad (2-6)$$

Thus, the 2x improvement in $\partial P / \partial f$ slope improvement shown in figure 2-2, and the associated 2x improvement in performance / watt, for the DEC Alpha 21164 [2] can be mostly attributed to the change in V^2 rather than by a change in architecture. The data in figure 2-2 does not include power consumed by support logic.

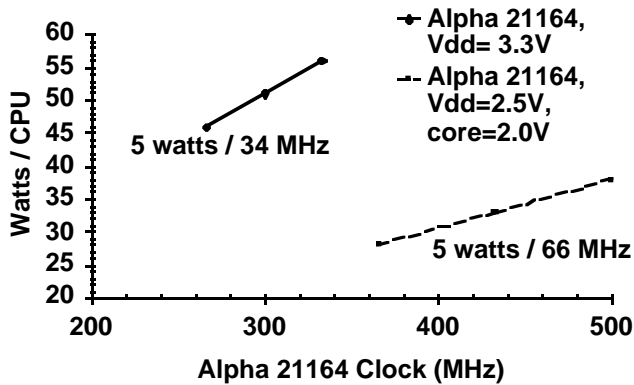


Figure 2-2. DEC Alpha Watt vs. Clock

2.3. Architecture width

Observation 3. Smaller architecture widths will have lower power consumption per completed task.

Since 64-bit floating point register uses 2x more gates N than a 32-bit floating point register, the 64-bit architecture can be expected to consume 2x more power for the same solid state CMOS generation.

The difference between 32-bit and 64-bit performance / watt can be observed in figure 2-3. The 32-bit MIPS and 32-bit PowerPC show a 2x SPECfp92 performance per watt compared with the 64-bit DEC Alpha and 64-bit Hewlett Packard PA-RISC processors in the 1995 and 1996 release years as the respective architectures matured in the market.

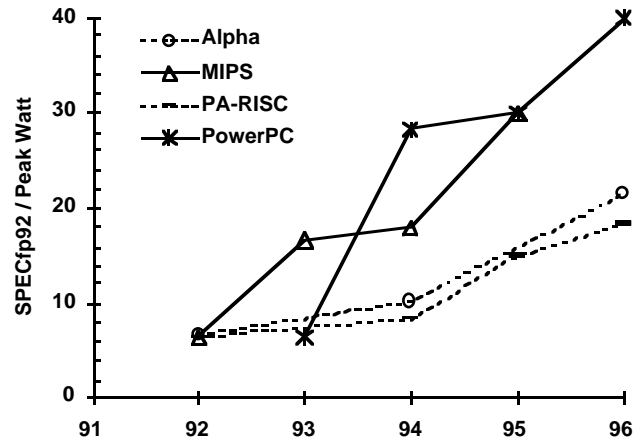


Figure 2-3. SPECfp92 Performance/Watt vs. Release Year

Using a 64-bit processor, where only 32-bits is required, will potentially consume on the order of 2x more energy on large scale systems. Likewise, Field Programmable Gate Array (FPGA) processors would be an energy efficient choice for bit level template matching algorithms.

2.4. Architecture specialization

Observation 4. For given operational task O_c and a given technology generation, Performance / Watt provides a metric of Architecture Specialization.

The Architecture Specialization metric is concisely stated in equation 2-7.

$$ArchitectureSpecialization_{RISC}^{DSP} = \frac{\left(\frac{Performance}{Watt}\right)_{DSP}}{\left(\frac{Performance}{Watt}\right)_{RISC}} \quad (2-7)$$

The Architecture Specialization metric has the nice feature of being directly measurable. In contrast, an $N\%N$ count determination would require a gate level simulation which is not as widely available an actual production device.

The Architecture Specialization metric of a 32-bit SHARC DSP with respect to a 64-bit 21064A and 21164 Alpha RISC for sustained FFT throughput / watt is shown in figure 2-4.

The 18x Architecture Specialization measurement is based on measured sustained throughput / watt for 1D complex FFTs averaged over vector lengths ranging from 64 to 32K for the same voltage system. The SHARC 21060 is designed for efficient FFT performance. The Architecture Specialization metric will vary significantly for other tasks.

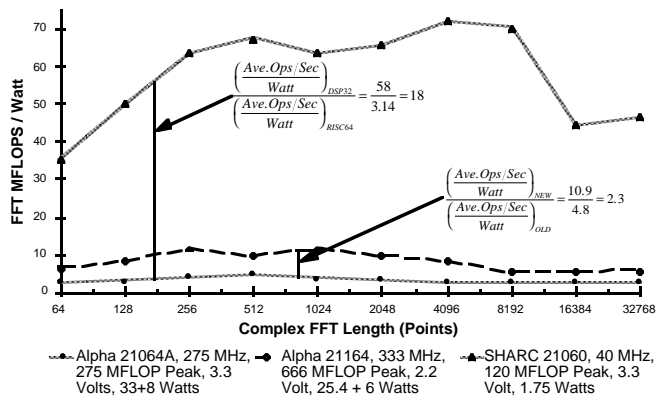


Figure 2-4 FFT MFLOPS / Watt

Notice that the 2.6x Alpha performance per watt increase can be largely attributed to a V^2 core voltage change and not to a fundamental shift to a more DSP-specialized architecture.

An aside observation is that more specialized architectures inherently require more specialized programming techniques and tools. FPGAs use a Hardware Description Language (HDL). Floating point DSP chipsets primarily have assembly and C support. General purpose RISC processors have a wide variety of programming languages including niche languages such as Fortran, Prolog, and SmallTalk.

2.5. Solid State Technology

Observation 5. Solid state advances will uniformly benefit each distinct architecture specialization.

Architectures tuned to specific tasks will continue to outperform architectures which are not tuned to the same tasks for the same voltage and same CMOS technology generation. The relative Performance/Watt is shown in figure 2-5.

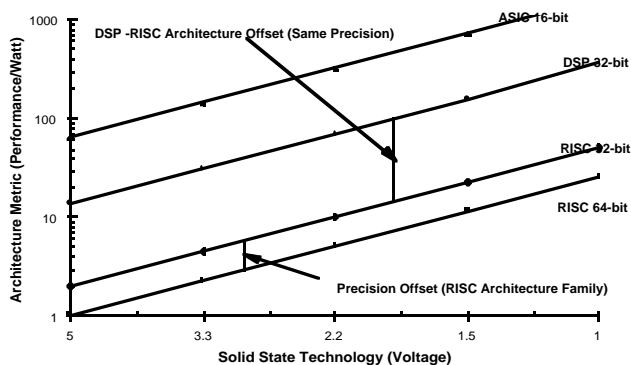


Figure 2-5 Persistent Offset

The values are based on measured and estimated 1D complex FFT throughput. The values are normalized to

RISC 64-bit @ 5 volts being given the value of 1. CMOS technology generation axis shows the corresponding minimum and/or specified device voltage. Notice that the DSP-RISC architecture offset is greater than the precision offset within a RISC Architecture Family.

Figure 2-6 illustrates how a lag of 2 CMOS generations can diminish the specialization advantage that an architecture might have.

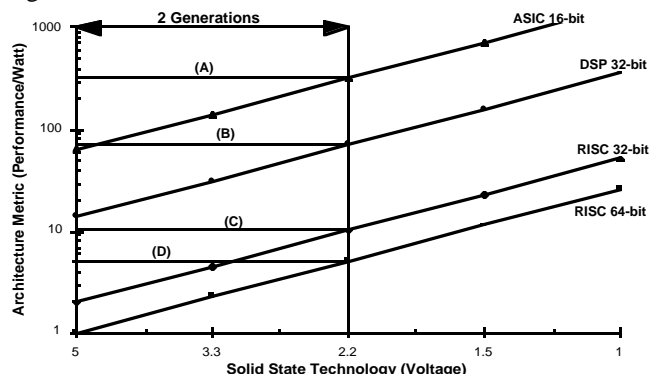


Figure 2-6 Two Generation Lag

The performance/watt offset between FPGA, Integer DSP, Floating Point DSP, and RISC will be persistent, as each architecture takes advantage of the same advancements in solid state technology.

3. Summary

The paper provided an analytical basis with empirical supporting data that,

- (1) Based on $P = CV^2 f N \%N$, Performance Per Watt is useful as an Architecture - Technology Metric.
- (2) Performance Per Watt can be used to calibrate ASIC, DSP, & RISC architectural performance density potential independent of solid state technology.
- (3) Performance Per Watt can be used to evaluate generation to generation architectural changes.
- (4) Performance Per Watt can be used to project a device performance density roadmap once calibrated to a given CMOS technology generation.

REFERENCES

[1] Analog Devices, DSP/MSP Products Reference Manual, 1994
 [2] Digital Equipment Corporation. www.digital.com
 [3] Device Electronics for Integrated Circuits. Muller, 1997