Trends in Low Power Digital Signal Processing

Anantha Chandrakasan, Rajeevan Amirtharajah, James Goodman, Wendi Rabiner
Department of EECS,
Massachusetts Institute of Technology, Cambridge

ABSTRACT

System level optimization will be required for the implementation of ultra low power portable devices. Often, system properties such as data distribution or network configuration can be exploited to minimize computational switching. Embedded power supply systems configured in performance feedback also minimize energy dissipation under varying temperature, process parameters and computational workload. If the power dissipation of digital processors can be scaled to low enough levels, self-powered techniques can be used to achieve "infinite" lifetime and low-maintenance wireless operation.

1. Introduction

Energy efficient system design requires systematic optimization at all levels of design abstraction ranging from process technology and logic design to architectures and algorithms [1]. Significant advances have been made in reducing the energy consumption of digital processors over the past few years [2], and this paper explores some of the emerging trends in energy efficient computing. As an application driver, we will focus on an ultra low power wireless camera (Figure 1). This camera transmits compressed video data over a wireless link (with a variable bandwidth up to 1Mbps) to a fixed base station. Many of the design issues faced in the context of our wireless camera are common to those in other wireless applications. Total system power (computation and communication) averaged over the normal operating conditions of the device should be minimized to maximize battery lifetime. The system should also be designed to service time varying data rates and quality of service requirements; embedded power supplies, which adapt supply voltages on demand, can save significant power in such systems.

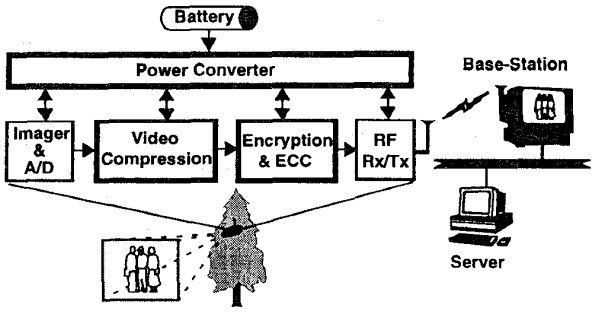

Figure 1. Portable video camera in a networked environment

The main application specific issues of our camera have to do with the asymmetry between a camera and the receiving base-station, which may be communicating with several cameras. The base-station is assumed to not be battery operated and this creates opportunities to design a system with skewed computational burdens, favoring the low power cameras.

2. Network Driven Optimization

Increasingly, portable devices are being operated within a networked environment, which gives them access to high powered compute servers through a combination of wired and wireless channels. This environment with distributed resources can be exploited to reduce the power dissipation in the portable by transferring large amounts of computation typically performed at the portable to the high powered servers on the network.

The camera described in the previous section must encode the video stream before it can be transmitted across the bandwidth-constrained wireless channel. Most video compression algorithms use some form of block-based scene motion estimation/compensation to remove the temporal correlation inherent in natural video sequences. The algorithms vary from low-complexity approaches such as conditional replenishment (i.e., simple frame differencing) to full search approaches. There is an inherent trade-off between the amount of computation and the rate of compression for conventional video compression algorithms-- the ones which achieve high rates of compression are computationally intense, whereas the computationally simple algorithms tend to not produce acceptable rates of compression.

An important observation, that can save significant power in the camera, is that the motion of objects is continuous from one frame to the next in natural sequences. Thus, knowing the location of an object in a few previous frames, it is possible to predict its location in the current frame (i.e., it is possible to predict the motion vectors of the current frame based on the motion vectors of the previous frames). It therefore is possible to remove the motion estimation computation from the portable encoder and perform it at the receiving base-station (or at another server on the network). Since the base-station only has access to the previous reconstructed frames, it must per-

0-7803-4455-3/98/$10.00 © 1998 IEEE

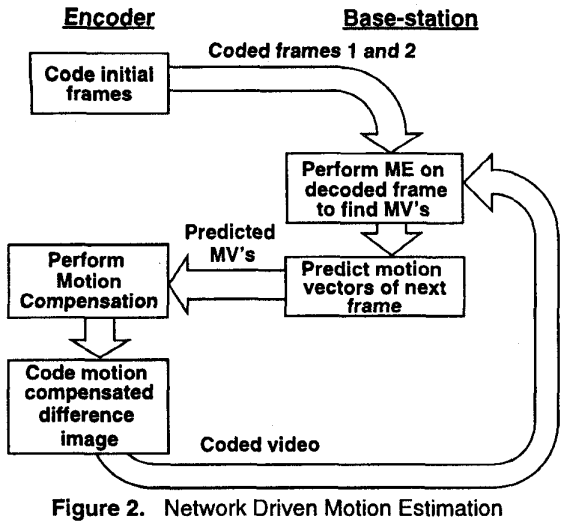

Figure 2. Network Driven Motion Estimation

form motion estimation on these previous frames and predict the motion vectors of the current frame from these motion vectors of the previous reconstructed frames. These predicted motion vectors are transmitted through a low-bandwidth reverse wireless link to the encoder, where the remainder of the compression algorithm is performed. This algorithm, shown in the flow graph in Figure 2, is called network-driven motion estimation and illustrates the effective reworking of a conventional video compression algorithm so that the computationally demanding task of motion estimation can be performed at a high powered network server [3].

A video compression system which employs network-driven motion estimation performs nearly as well as one which performs the motion estimation computations at the encoder using full search computation. This is illustrated in Figure 3, which shows the number of bits required to code each frame of a video sequence using encoder-based motion estimation, conditional replenishment, and network-driven motion estimation. Network-driven motion estimation compresses the images much more than conditional replenishment, while using the same amount of encoder power. Alternatively, network-driven motion estimation achieves nearly the same compression rates as encoder-based motion estimation while using over *two orders of magnitude* less computation at the encoder.

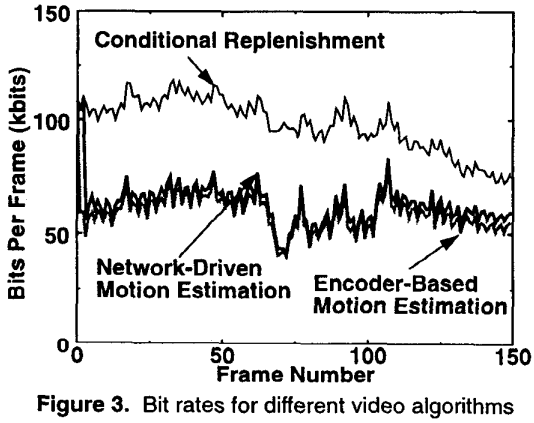

Figure 3. Bit rates for different video algorithms

3. Energy Scalable Computing

In many applications, such as the wireless camera, it is desirable to design digital processors that allow a trade-off between the quality of service (QoS) provided and the energy consumed to process a sample. This allows the user to evaluate the application's requirements and set the desired quality while minimizing the energy consumption. We have developed an energy scalable encryption processor where the level of security (i.e., quality) and energy consumed to encrypt a bit can be traded-off dynamically based on demand. Since transmitted data streams can often be partitioned into different priority levels, an energy scalable processor ensures that important information is adequately protected, while sacrificing some security for low priority data in order to reduce the total system energy.

The energy scalable encryption processor in this work is based on a variable-width quadratic residue generator (QRG). The QRG is a cryptographically-secure pseudo-random bit generator that is based upon the work in [4]. The QRG operates by performing repeated modular squarings. The modular squaring is performed using an algorithm based on Takagi's iterated radix-4 algorithm [5] which requires $(\log_2 Q)/2$ iterations to compute the result $P = X \cdot Y \bmod Q$. The least significant $\log_2\log_2 Q$ bits of each result can be extracted and used as a strong reproducible pseudo-random source for applications such as a stream cipher or key generator.

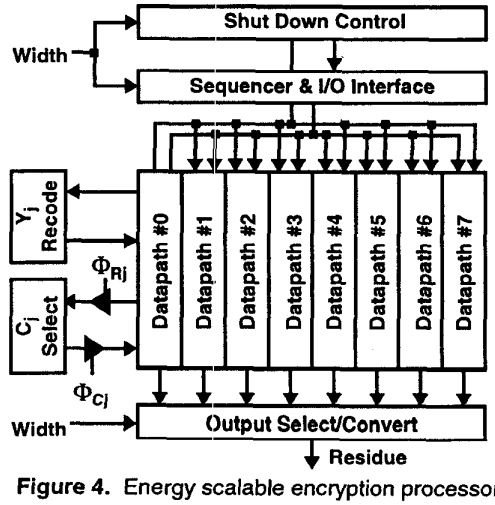

Figure 4. Energy scalable encryption processor

Energy scalable computing requires dynamically reconfigurable architectures that allow the energy consumption per input sample to be varied with respect to quality. In the case of the QRG the quality scales sub-exponentially with the modulus length, while the energy consumption scales polynomially. A fully scalable QRG architecture was developed where the width ($w = \log_2 Q$) can be reconfigured on the fly to range from 64 to 512 bits in 64 bit increments (Figure 4) [6]. The design makes extensive use of clock gating to disable unused portions of the QRG. Hence the switched capacitance of the QRG is minimized and energy scalability is achieved.

Further energy/security scalability can be achieved through the use of an adaptive supply [7]. Rather than designing a system with a static supply to meet a specific timing constraint under worst case conditions (i.e., establishing the feedback around the power converter to fix the output voltage), it is more energy efficient to allow the voltage to vary such that the timing constraints are just met at any given temperature and operating conditions; this is accomplished by establishing the feedback around a fixed processing rate or delay. In this example, when operating at a reduced width, the number of cycles required per multiplication is reduced and therefore the supply voltage can be reduced for a given throughput. The supply is varied using an embedded custom DC/DC converter. The use of an adaptive supply enables us to substantially reduce the energy consumption as the multiplier width is varied (Figure 5). Figure 6 shows a die photo of the scalable encryption processor with embedded power supply.

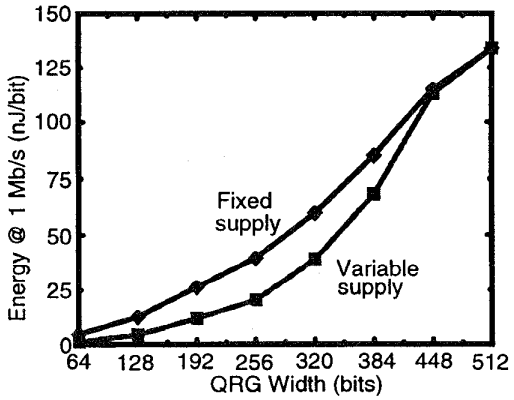

Figure 5. Power dissipation for fixed and variable supply

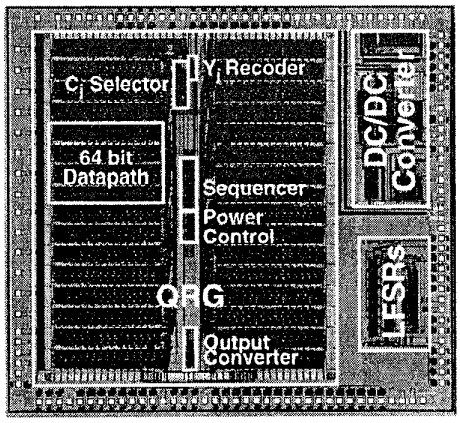

Figure 6. Encryption processor with embedded supply

4. Self-powered Computing

Based on continued advances in power management techniques, it is projected that the power consumption of future low to medium throughput DSPs will be scaled to 10's to 100's of μW. At these low power levels, an interesting question arises: can we use ambient energy sources to power electronic systems? Ambient energy is energy that is in the environment of the system and is not stored explicitly, for example in a battery. The sources of ambient energy available to the system depend on the application. This includes solar power (e.g., used in electronic calculators), electromagnetic fields (used in RF powered ID tags, inductively powered smart cards, or non-invasive pacemaker battery recharging), thermal gradients, fluid flow, and mechanical vibration.

Figure 7 is a detailed block diagram of our self-powered system [8]. This approach involves transduction of mechanical vibration to electrical energy. A moving coil generator is used which consists of a mass attached to a spring, which is attached to a rigid housing. The generator and rectifier subsystem is shown at the top. Transformer X1 (with a 1:10 turns ratio) converts the output voltage of the generator V_{gen} to a higher voltage that can be rectified by the half-wave rectifier formed by diode D1 and capacitor C1. Note that with proper electromechanical design, the transformer can be eliminated. Voltage V_{in} is the time-varying input voltage to the regulator.

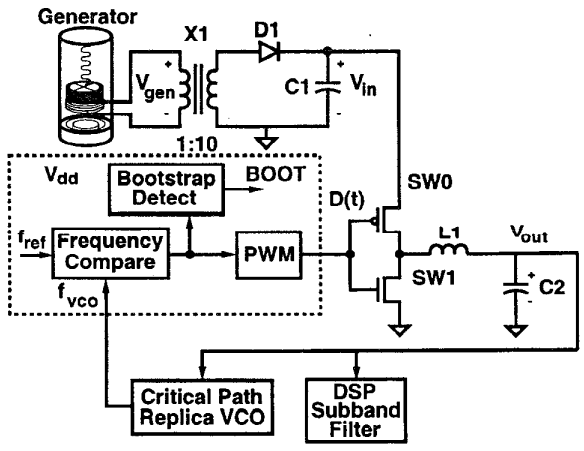

Figure 7. Vibration based self-powered system

The regulator consists of five main subsystems: a VCO, frequency comparator, pulse-width modulated (PWM) waveform generator, bootstrap detection circuit, and a Buck converter. An external input provides the performance constraint: the rate at which the load circuit must produce results. The load circuit in this case is an 8 bit FIR filter. The DSP rate command is delivered in the form of a clock, f_{ref}, the rate at which new input samples are fed to the DSP. Its period therefore corresponds to the total delay between valid output samples of the DSP. To achieve the lowest possible power consumption, the converter downconverts V_{in} to the lowest voltage at which the DSP can run and still produce correct results at the rate set by f_{ref}.

The overall control scheme is similar to a PLL. The rate f_{ref} is compared to the output of a voltage-controlled oscillator, f_{vco}. The VCO is a ring oscillator consisting of the DSP critical path padded with a few inverters. It is powered from the regulated output voltage V_{out}. Thus it is a replica of the circuit whose power supply is being controlled with some delay margin to account for processing mismatches. The controller adjusts V_{out} until the period of f_{vco} is short enough to satisfy the performance demand but not so short as to waste power.

An important difference between the self-powered system and a battery-powered system is that the former requires a backup power source. This is necessary since at startup the voltage regulator must derive its power from some source and the generator output is too uncontrolled to be used. The source could be a very small battery or a previously charged large capacitor, but it need not provide much energy since it is only used during the startup transient of the system. The bootstrap detect block switches the controller to V_{out} when the output voltage is deemed stable.

A prototype regulator and low-power DSP load have been fabricated and tested with this generator. One excitation of the generator produces 23ms of DSP operations, corresponding to 2,340 *free* operations. The clock rate is adjustable from 500kHz to 1MHz for the supply voltage varying from 0.9V to 1.1V. Figure 8 shows the closed loop regulation of the generator output.

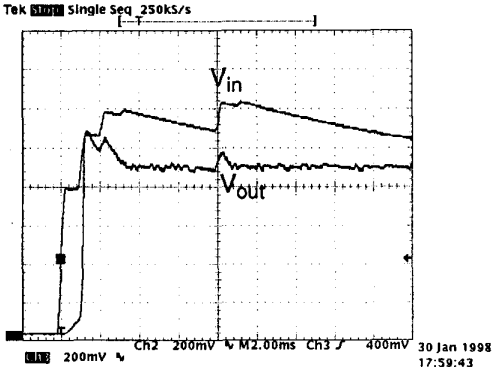

Figure 8. Closed loop regulation of generator output

5. Conclusion

Extremely low power operation can be achieved for digital circuits by aggressively reducing the supply voltage. To maximize energy savings, power supply circuits should be designed to meet a fixed processor performance and *not* a fixed voltage specification. If the power dissipation of digital processors can be scaled to a low enough level, the concept of embedded power supplies can be extended to self-powered systems. Design of efficient generators and regulators can enable low-maintenance wireless operation. Minimizing computational switching is another approach to reduce power and techniques exist to accomplish this at many levels of design abstraction. At a system level, networking can be exploited to save power by optimum partitioning of the computation. These emerging techniques can save orders of magnitude of power dissipation.

Acknowledgments

This work is sponsored in parts by DARPA, under contract #DAAL-01-95-K3526 and by the ARL Advanced Sensors Federated Lab program under contract #DAAL01-96-2-0001. We thank Abram Dancy for the design of the DC-DC converter circuitry.

References

[1] A. Chandrakasan, R. Brodersen, *Low Power Digital CMOS Design*, Kluwer Academic Publishers, July 1995.

[2] A. Chandrakasan, R. Brodersen, *Low-Power CMOS Design*, IEEE reprint book, Jan. 1998.

[3] W. Rabiner and A. Chandrakasan, "Network-Driven Motion Estimation for Wireless Video Terminals," IEEE Transactions on Circuits and Systems for Video Technology, pp. 644-653, Aug. 1997.

[4] L. Blum, M. Blum, and M. Shub, "A Simple Unpredictable Pseudo-Random Number Generator," SIAM Journal on Computing, v. 15, n. 2, pp. 364-383, 1986.

[5] N. Takagi, "A Radix-4 Modular Multiplication Hardware Algorithm for Modular Exponentiation", IEEE Transactions on Computers, pp. 949 - 956, Aug. 1992.

[6] J. Goodman, A. P. Chandrakasan, "A 1Mb/s Energy/Security Scalable Encryption Processing Using Adaptive Width and Supply," IEEE International Solid-state Circuits Conference, pp. 110-111, Feb. 1998.

[7] V. Gutnik, A. Chandrakasan, "Embedded Power Supply for Low-Power DSP," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, pp. 425-435, Dec. 1997.

[8] R. Amirtharajah, A. P. Chandrakasan, "Self Powered Low Power Signal Processing," IEEE Symposium on VLSI Circuits, pp.25-26, June 1997.