Power Management for Cognitive Radio Platforms

Amin Khajeh, *Student member, IEEE*, Shih-Yang Cheng, *Student member, IEEE*, Ahmed M. Eltawil, *Member, IEEE*, and Fadi J. Kurdahi, *Fellow, IEEE*
University of California Irvine, Irvine, CA, USA
{akhajehd, shihyanc, aeltawil, kurdahi}@uci.edu

Abstract- This paper discusses how the cognitive radio concept can be extended to allow the system not only to manage shared resources such as spectrum, but to use this knowledge to optimize the overall system power consumption. We introduce a case study of video over wireless via a 3G WCDMA modem connected to an H.264 decoder. We show that by utilizing knowledge about the communication channel, a savings of more than 20% of the overall system power is possible while maintaining a required quality of service.

I. INTRODUCTION

Cognitive radios (CR) running on programmable software defined radios (SDR) platforms have been under vigorous research since their introduction in the late 90s [1]. However, the exact definition of cognitive radios differs depending on the context of discussion. In some forums, the focus is on spectrum utilization, where spectrum management becomes device centric rather than policy based. The goal is to develop adaptive radios that sense and share the spectrum with a focus on negotiations and radio protocols that leverage “spectrum holes” that are available in space, time and frequency domains. Other forums view cognition on multiple levels of which spectrum management is one very important facet of a multi dimensional space. For example, the SDR forum cognitive radio working group in conjunction with the IEEE P1900.1 defines (in their latest draft) cognitive radios as “A radio or system that utilizes a cognitive control mechanism that can sense and autonomously reason about the surrounding radio environment and adapt to it accordingly.” [2]. Even though the definition might be different, most forums agree that the combination of SDR and CR provide the most flexible solution to support adaptive and intelligent communication systems. However, the adoption and proliferation of SDR/CR platforms has been slow to materialize. There are multiple reasons for this; including technical, legal and in some cases even ethical reasons (how do you guarantee fairness between users in a cognitive system?).

On the technical side, the Achilles heel of SDR platforms is their power consumption. By definition, a platform designed for general purpose processing cannot compete with a custom crafted ASIC from the power and area perspectives. In this paper, we intend to discuss novel methods of power management for cognitive radios that expand the concept of cognition to address jointly, both environment and self cognition with the target of minimum power consumption for a given set of conditions. This goal is achieved by dynamically managing not only the shared resources (spectrum) but also the hardware resources in terms of power

consumption and subsequently increasing the battery life, while meeting the required system objective of reliable communication.

The essence of the idea is that current power management techniques tend to adapt the system to provide the optimum performance based on the operating conditions, with the universal underlying assumption that the adaptation technique used has to maintain 100% correctness of the computational engine of the system regardless of the application. In other words, the adaptation technique is not designed to be cognitive of the application requirements. While it is true that some applications cannot function unless the data processed is 100% correct (such as processor instruction code), there exists a broad family of applications that are inherently fault tolerant such as wireless and multimedia systems. This algorithmic resilience to errors can be utilized and co-designed with the hardware circuitry in mind to provide resilience not only to functional induced faults but also to hardware induced faults thus expanding the adaptation space to unexplored domains. To do so, we examine the relationship between (a) the constituent components of an architecture and their vulnerability in terms of power consumption and performance as a function of the operating conditions, and (b) the needs, assumptions and requirements of the algorithms executing on the design.

We consider a case study of a hypothetical SDR platform performing video over wireless. Without loss of generality we consider a sample point where the modem is a 3G compliant WCDMA system, while the video decoder is an H.264 system. The most effective means of managing the power consumption of such a system is supply voltage scaling. In current techniques the operating voltage is conservatively chosen to guarantee 100% error free hardware. Our hypothesis postulates that by aggressively scaling the voltage even further and allowing a controlled amount of hardware errors to occur, an optimum operation point can be achieved whereas the algorithmic tolerance of the algorithms can be used to compensate for hardware failures while achieving a target performance metric such as bit error rate (BER, modem) or peak signal to noise ratio (PSNR, H.264). By knowing the statistics of the channel, one can decide on an optimum supply voltage for each section independently, where the error correction capabilities of each section will allow it to recover from hardware induced failures. For example, the channel coding of the modem offers resilience to errors while the error concealment algorithms in the H.264 decoder provides another

layer of protection. Alternatively, one may decide to jointly change the supply voltage on both sections to achieve maximum power savings. The cognition cycle of the system thus becomes, **observe** the channel statistics, **identify** degrees of freedom to minimize power consumption (modem voltage scaling versus application voltage scaling), **modulate** operation conditions, and **monitor** performance metrics such as BER, PSNR.

The paper is organized as follows: Section II introduces the proposed adaptation technique and reviews the H.264 decoder and the WCDMA modem structure. Section III discusses the case study of video over wireless and quantifies the gains of the proposed technique. Finally conclusions are drawn in section IV.

II. PROPOSED TECHNIQUE

In general, hardware faults can be categorized into logic faults and memory faults. Tracking logic faults due to functional errors is an extremely difficult and currently unsolved problem. However, identifying memory faults presents a much more structured problem due to the structured nature of memories. In prior work [3], the authors have presented a complete analysis of memory faults under aggressive voltage scaling and methods of identifying and correcting for these faults in the context of wireless applications. This paper builds upon the fact that a significant (sometimes even dominant) fraction of embedded metrics (power, performance, area) is related to memories which form a large portion of most modern communication transceivers. These memories store raw soft bit values that have multiple levels of redundancy as follows:

1. At the algorithmic level, coding redundancy is inserted to protect against channel errors.
2. At the implementation levels, soft bit representations are available prior to any decisions being made.
3. The current approach to design specifies the architecture assuming operation at the minimum sensitivity of the device. However, the signal will be close to sensitivity only a small fraction of the time. The majority of the time, the receiver will be experiencing a relatively better SNR. At high SNR's an optimal design should target these extra "algorithmic dBs" to provide **runtime** flexibility to minimize power consumption.

To illustrate the last point, consider a Rayleigh flat fading signal with the Rayleigh probability distribution function given as:

$$p(r) = \frac{r}{\sigma^2} \exp\left(-\frac{r^2}{2\sigma^2}\right)$$

It is fairly straight forward to find that 90% of the time, the signal is within 9.8 dB of its mean expected value. Extending this concept to frequency selective signals, we constructed an impulse response consisting of 18 taps spanning a delay spread of 380 ns [4]. This channel model is representative of a dispersive indoor channel. Forty thousand channel samples were simulated and a histogram of the difference between the

mean and the deepest notch for each channel instance was recorded as shown in Figure 1(a). It is interesting to note that in this case the probability density function exhibits a Rayleigh-like distribution in the logarithmic domain with the most likely value centered around 12 dB. Based on this observation, it is safe to say that a high percentage of the time, the receiver will be experiencing a relatively higher SNR than the minimum required for demodulation and the "slack" can be used to allow some limited and controllable errors to occur in the hardware to relax the circuit specifications.

Fig. 1. (a) Histogram of difference between mean and peak attenuation for a frequency selective fading channel, (b) Effect of aggressive voltage scaling on 65nm SRAM

A. The Adaptive Voltage Biasing (AVB) Technique

The exponential relation of the leakage power with the supply voltage [5] and the square relation of the dynamic power with V_{dd} [6], point out the importance of voltage scaling as one of the most effective means of reducing power consumption. As a result, methods like Dynamic Voltage Scaling (DVS) [7] or Ultra Dynamic Voltage Scaling (UDVS) [8] are widely used for power saving. However these techniques are still required to maintain 100% correctness of the underlying hardware. Thus the power savings are inevitably accompanied by a reduction in performance, typically by running at a lower operating frequency which is set by the weakest performer in the overall system. Due to their dense structure, memories are typically the first circuit element to fail under aggressive voltage scaling. The failure mechanisms of memory circuits are well understood and typically have an exponential dependence on the supply voltage as shown in Figure 1(b). To achieve a low power solution, the designer must first decide on the acceptable level of error tolerance that is permissible by the application and the overall system design while still maintaining the required performance. Given that level, and a required performance level, the designer can select the appropriate V_{dd} .

The question now is given a specific SDR platform encompassing both the modem and the application, how to dynamically decide on the optimum voltage level for both modem and decoder sections assuming a-priori knowledge of the channel SNR? To better answer this question a brief review of H.264 and WCDMA is presented prior to analyzing the full end-to-end system.

B. H.264 Video Decoder Overview

H.264 [11] is emerging as one of the most promising video standards with higher quality, better compression and many other features. Figure 2 illustrates a typical configuration for an H.264 Video codec.

Fig. 2. Typical configuration of H.264 video codec

Similar to communication systems, multimedia systems also have inherent error resilience. In such systems, the quality of an output image or sequence of video frames is measured in terms of Peak Signal-to-Noise Ratio (PSNR) which compares the output image(s) to a reference set and computes the PSNR. Even in the case of ideal transmission conditions, residual errors occur due to quantization and/or filtering so systems typically have less-than-perfect PSNR values. In order to cope with imperfect signal recovery, some multimedia systems have explicit error resiliency built into the algorithms in two ways: (1) At the network level, error concealment strategies are embedded within the standards [13] as a means of maintaining resilience to transmission losses, and (2) at the implementation level (and during decoding) data redundancy is implicit in the fact that neighboring pixels are likely to have similar values. Indeed video compression schemes exploit that redundancy at all levels. Such redundancy is gradually reintroduced during the decoding process. It is important to highlight the differences between error concealment and implementation level data errors. In error concealment, errors occur in the encoded bitstream. When this happens, the errors affect several pixels and typically result in the loss of a complete macroblock (or slice) in which case a “substitute” macroblock (or slice) is generated locally using spatial or temporal redundancies to replace the lost one. In implementation errors, the errors occur in the middle of the decoding process on the actual image pixels. We will refer to these two types of errors as type 1 and type 2, respectively.

While type 1 errors can be fixed through well-known error concealment techniques, type 2 errors are more localized and can be fixed individually through exploring either spatial or temporal locality. The underlying assumption of motion compensation is that scene change is due mainly to object and camera motion and the difference between temporally adjacent pictures is so small that many parts of the current frame can be borrowed from previously decoded frames which are stored in a memory called the Decoded Picture Buffer, or DPB. In DPB, decoded YUV components which represent Luma (brightness) and Chroma (color) of the image are stored. In the 4:2:0 sampling format ('YV12'), U and V each have half the horizontal and vertical resolution of Y. Each YUV

component has 8 bit depth (in the simulated H.264 decoder) and a total of 12 bits are used to make one pixel because only one U and V components are necessary for every four Y components. Since DPB stores decoded images, it requires large memory space and can easily be the dominant memory in system design of H.264 or other similar standards such as MPEG-2 or MPEG-4 [12] which also require reference picture buffers as large as 16Mb. The new MB86H50 H.264 codec chip from Fujitsu [10] that integrates the DPB on-chip requires a total of 512Mbits (64Mbytes) of on-chip memory to reduce power for mobile applications. One can easily see that this huge amount of on-chip memory will dominate area, performance and power of the overall system. This motivates our proposed approach to focus on such memories in an attempt to reduce power. In a previous work [11] we showed that by aggressively reducing the supply voltage on the DPB, we can achieve significant savings in total system power while maintaining essentially the same output quality.

B. WCDMA Receiver Overview

Figure 3 depicts the top-level block diagram of a diversity enabled WCDMA modem. The system includes the modem section (RAKE receiver), the coding layer and the protocol layer of the standard. It is based on a dual embedded micro-controller architecture. This chip has been designed and fabricated by the authors and an extensive coverage of this modem has been presented in [9]. The combiner output from the modem are soft values with 10 bit precision that are available for all the data and control symbols transmitted. Control symbols are very important and thus must be stored in a protected memory with minimum loss. The SNR estimation is then used to determine a scaling factor for the physical data channel post de-interleaving. The factor is used to reduce the precision from 10 to 4 bits and to perform optimal scaling in fading conditions for soft Viterbi and/or Turbo decoding. Symbol level processing is then performed on the 4 bits soft values to find the transmitted data bits intended to higher layers.

In a previously published study, we have shown that, similarly to the H.264 decoder case, we can achieve significant power reduction by aggressively scaling V_{dd} for the data memories in both 0.18 μ m and 32nm technologies, respectively [3].

Fig. 3. Top level block diagram of the WCDMA transceiver System

C. Cross-Layer end-to-end system

In the proposed Adaptive Voltage Biasing technique, we separate the error-resilient memories (ERM) from the rest of the system. ERMs are memories that store data that is

inherently redundant such as raw data buffering memories in the modem case or the DPB memory in the H.264 case. The error-resilient memories are supplied from a variable source while the rest of the system is supplied in a traditional manner (i.e. nominal supply). During operation, the voltage to the ERM is aggressively lowered and errors are allowed to occur in the system, which will correct for them at the application layer. In this paper, we explore such an approach to a cross-layer system that merges both the modem and the source coding in one system. The proposed system diagram is illustrated in Figure 4. A variable supply (could be on chip or off-chip) supplies the ERM in both systems with separate V_{dd} s while a standard supply provides the rest of the system with nominal V_{dd} . The authors are aware that this is one example, and that the techniques of error resilience as well as the power saving are design dependent. However, the main point of the paper is that it is *possible to utilize the knowledge* of the requirements of the application using the modem to reduce power. In other words, the system becomes *self cognizant*.

Fig. 4. Proposed system structure

III. CASE STUDY

In order to explore the AVB technique on this system, we discuss three cases as shown in figure 5. These cases are: 1) Nominal- V_{dd} for the WCDMA modem and AVB for the H.264 decoder engine, 2) AVB for the W-CDMA modem and nominal- V_{dd} for the H.264 decoder and 3) AVB for the H.264 decoder and AVB for the WCDMA modem. For a given set of operating condition, the radio has to be smart enough to decide which approach to choose in order to make the optimum use of the resources. In our power calculations, the WCDMA modem consumes 72% of the total power [9] whereas the H.264 decoder consumes 28% of the total power in 65nm technology node [11]. Other systems may have different contributions of the two system components.

A. Nominal WCDMA and AVB H.264

This case represents the scenario that WCDMA receiver does not benefit from a high received SNR. In other words, the received signal's SNR in modem is very close to the minimum required SNR and the WCDMA modem can not tolerate any error in its data buffering memories.

The output of the nominal WCDMA modem which is running at its nominal voltage (for 65nm technology

$V_{nominal}=0.9v$) and at the nominal required SNR for AWGN ($SNR_{nominal}=7dB$) has $BER=10^{-6}$. In this case, there is almost zero type 1 errors (as defined in section II.B). Considering this error rate in the data stream, the H.264 decoder can utilize AVB technique. This will generate type 2 errors in the decoded picture frames. These errors can be fixed with a post filtering to achieve lower power consumption and desired quality. Figure 6(a) shows the effect of H.264 decoded picture buffer (DPB) voltage on the Y component PSNR. Edge and mean filters are used to improve the decoded image quality or in other words reduce the effect of the memory faults due to aggressive voltage scaling [11].

Fig. 5. Three different cases that the radio can choose based on the operating conditions.

Figure 6(a) shows that 24% reduction of the DPB voltage and applying edge and mean filters, the Y PSNR will degrade only by 0.01%. While this degradation of the Y PSNR does not have any effect on the visual quality of the decoded video, a power saving of the 35% in H.264 decoder can be achieved taking into account the overhead power of the filtering. It is important to note that for the power saving calculations we have only considered the dynamic power and even more power saving can be achieved by considering the leakage power.

B. Nominal H.264 and AVB WCDMA

As previously discussed in section II, a high percentage of the time the receiver will be experiencing a relatively higher SNR than the minimum required for demodulation. This "slack" can be used to allow some limited and controllable errors to occur in data buffering memories by reducing memory voltage aggressively to lower the power consumption of the modem [3]. Figure 1(b) shows that 28% reduction of voltage of the fault tolerant memories of the WCDMA modem leads to 10^{-3} error in the memory. From figure 7(a) one can easily find out that having 10^{-3} error in the fault tolerant memories of the WCDMA modem degrades the BER of the WCDMA system from 10^{-6} to 2.46×10^{-6} . These results were obtained by simulating an end to end WCDMA modem and injecting errors into the ERM. Moreover, Figure 7(b) shows that having 2.46×10^{-6} errors in the data stream generates type 1 errors which, even when error concealment schemes built into the H.264 decoder are used, would degrade the output Y PSNR by 0.7%. It is important to note that the fault tolerant memories (ERMs) in WCDMA modem consume almost 30% of total modem power. The 28% V_{dd} reduction of the fault

tolerant memories results in 48% saving in dynamic power and 58% saving in leakage power. Consequently 17% power saving can be achieved in WCDMA modem.

Since the effect of the data stream out of the WCDMA receiver errors are dominant compared to the errors in the DPB on the output video quality, this case represent the scenario that there is extra redundancy in the received signal that can be used to compensate for the effect of errors in the fault tolerant memories of the WCDMA modem.

Fig. 6. (a) Effect of AVB combined with filtering on H.264 decoder, (b) Expected power saving by utilizing AVB for H.264 and post filtering.

Fig. 7. (a) Effect of aggressive voltage scaling on WCDMA BER, (b) Effect of WCDMA BER on Y PSNR of a perfect H.264 system.

B. AVB H.264 and AVB WCDMA

This is the most general case where AVB is used on both the WCDMA receiver and the H.264 decoder. The assumption here is that the channel impairment is low and that it is possible to tolerate some minor degradation in the video output quality. There are, however some practical limitations to this approach. In this case, both type 1 and type 2 errors are present.

Figure 8(a) shows that the H.264 decoder can tolerate WCDMA BERs (i.e. type 1 errors) between 10^{-6} to 10^{-5} and WCDMA BER of 10^{-5} or greater degrade the quality of video drastically. This fact points out the limits on utilizing AVB for H.264 and WCDMA modem simultaneously. In other words by setting the voltage of the WCDMA data buffering memories to a voltage that leads to error rates less than or close to 10^{-5} in WCDMA (type 1 errors in the video stream) we can still benefit from utilizing AVB technique for the H.264 decoder.

Figure 8(b) shows the power savings that can be gained for the three different cases that we discussed above. Based on this figure the radio can decide on the optimum approach for the desired Y PSNR and the corresponding power saving. In other words, for a fixed Y PSNR the system will decide on the approach that results in the maximum power saving. From the graph, we observe that some points are inferior to others. In

other words, one case may yield higher power savings than another for the same target PSNR. Such points are considered as *pareto-optimal*. Case B points appear to be inferior to cases A and C. However, this is a direct conclusion of the ratio of power consumption of the receiver and video decoder. Since the receiver consumes more than 3x the power of the H.264 decoder, one would expect to get more power reduction by AVB of the receiver. This situation would be reversed in another system where the ratios are the opposite. As expected, case C yields the most pareto-optimal design points since it is a superset of cases A and B.

Fig. 8. (a) Utilizing AVB for H.264 and WCDMA, (b) Expected power savings for case A, B and C

IV. CONCLUSION

We have explored a new cross-layer design approach that proposes using aggressive Adaptive Voltage Biasing (AVB) to trade off system-level errors for lower power consumption with minimal effect on performance. Power savings of over 20% are observed when considering a typical system using this approach. In future research, we will consider design-time modifications that would make such a cross-layer system even more amenable to the proposed AVB approach.

REFERENCES

- [1] Joseph Mitola III and Gerald Q. Maguire Jr., "Cognitive Radio: Making Software Radios More Personal," IEEE PCS Magazine's Special Issue on Software Radio, August 1999.
- [2] "Standard Definitions and Concepts for Spectrum Management and Advanced Radio Technologies" IEEE P1900.1, Draft v0.13, March 2006.
- [3] Amin Khajeh Jahromi, Ahmed M. Eltawil, Fadi Kurdahi and Rouwaida Kanj, UC Irvine and IBM "Cross Layer Error Exploitation for Aggressive Voltage Scaling," ISQED 2007.
- [4] V. Erceg et al., "TGN channel models," IEEE 802.11-03/940r4, May, 2004..
- [5] An integrated cache/SRAM access time, cycle time, area, leakage, and dynamic power model CACTI, <http://quid.hpl.hp.com:9081/cacti/>
- [6] Y. Taur and T. H. Ning "Fundamentals of Modern VLSI Devices," Cambridge University Press 1998.
- [7] V. Gutnik and A. Chandrakasan, "Embedded power supply for low power DSP," IEEE Trans. VLSI Sys., Vol. 5, No. 4, pp. 425-435, Dec. 1997.
- [8] B. H. Calhoun, A. P. Chandrakasan, "Ultra-Dynamic Voltage Scaling (UDVS) Using Sub-Threshold Operation and Local Voltage Dithering," IEEE Journal of Solid State Circuits, Vol. 41, No. 1, Jan. 2006.
- [9] A. M. Eltawil, et. al, "Dual Antenna UMTS Mobile Station Transceiver ASIC for 2 Mbps Data Rate," in proceedings of ISSCC, Feb. 2003.
- [10] Fujitsu MB86H50 H.264 decoder chip, http://www.fujitsu.com.sg/news/pr/fmal_20061215.html
- [11] Kang Yi et. al. "Design and Analysis of Low Power Image Filters toward Defect-Resilient Embedded Memories for Multimedia SoCs," in Proceedings of the eleventh Asia-Pacific Computer Systems Architecture Conference, Shanghai, September 2006.
- [12] H. Arakida et. al. "A 160mW, 80nA Standby, MPEG-4 Audiovisual LSI with 16Mb Embedded DRAM and a 5GOPS Adaptive Post Filter," ISSCC 2003.
- [13] Y. Wang, S. Wenger, J. Wen, and A. K. Katsaggelos, "Error resilient video coding techniques-real time video communications over unreliable networks," IEEE Signal Process. Mag., vol. 17, no. 4, pp. 61-82, Jul. 2000.