

An Energy Efficient Rate Selection Algorithm for Voltage Quantized Dynamic Voltage Scaling

Lama H. Chandrasena
Dept. of E & E Engineering
Adelaide University
Adelaide, SA 5005, Australia
lama@eleceng.adelaide.edu.au

Priyadarshana Chandrasena
Dept. of E & E Engineering
Adelaide University
Adelaide, SA 5005, Australia
vpc@eleceng.adelaide.edu.au

Michael J. Liebelt
Dept. of E & E Engineering
Adelaide University
Adelaide, SA 5005, Australia
mike@eleceng.adelaide.edu.au

Abstract— This paper presents a highly energy efficient alternative algorithm to the conventional workload averaging technique for voltage quantized dynamic voltage scaling. This algorithm incorporates the strengths of the conventional workload averaging technique and our previously proposed Rate Selection Algorithm, resulting in higher energy savings while minimizing the buffer size requirement and improving the overall system stability by minimizing the number of voltage transitions. Our experimental work using the Forward Mapped Inverse Discrete Cosine Transform computation (FMIDCT) as the variable workload computation, nine 300-frame MPEG-2 video sequences as the test data, and a 4-level voltage quantization shows that our algorithm produces better energy savings in all test cases when compared to the workload averaging technique, and the maximum energy saving for the test cases was 23%.

1. INTRODUCTION

In CMOS circuits energy dissipation for processing a data sample is given by [2]:

$$E = nCV_{dd}^2 \quad (1)$$

where, n is the number of clock cycles per sample period, C is the averaged switched capacitance per clock period, and V_{dd} is the supply voltage. Due to this quadratic dependence of energy on voltage, supply voltage reduction is the most effective method of energy minimization. However, a reduction in supply voltage increases the circuit delays as shown by Equation 2[6]:

$$T_d = \frac{C_L V_{dd}}{\mu C_{ox} (W/L) (V_{dd} - V_t)^2} \quad (2)$$

where, T_d is the delay, V_{dd} is the supply voltage, C_L is the total node capacitance, μ is the mobility, C_{ox} is the oxide capacitance, V_t is the threshold voltage, and W/L is the width to length ratio of transistors.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISSS '01 October 1-3, 2001, Montréal, Québec, Canada.
Copyright 2001 ACM 1-58113-418-5/01/0010 ...\$5.00.

In highly sequential data-dependent computations, such as iterative variable workload computations, an increase in delay is acceptable provided the worst-case time constraint is not violated [6]. For such computations, the data samples with sub-optimal workload values can be processed at dynamically scaled down voltages without incurring a performance loss. As a result, the overall energy dissipation for the computation can be significantly reduced.

Even though the minimum energy dissipation in dynamic voltage scaling (DVS) is achieved with continuous voltage levels, it is impossible to prepare a stable supply voltage in such systems [3]; as an alternative, a small number of voltage quantizations has been proposed [2]. If the voltage can be dithered during a sample period, a 4-level voltage quantization produces a very good approximation to the continuous voltage level DVS. However, voltage dithering technique introduces additional voltage transitions, approximately doubling the total number of voltage transitions in the system.

The use of buffers and workload averaging has been shown to significantly improve energy savings in DVS [2]. The *Rate Selection Algorithm (RSA)* [1] demonstrates an alternative technique to the conventional workload averaging technique. This algorithm is specifically developed to minimize energy dissipation in voltage quantized DVS. This algorithm uses a single buffer, and where possible selects rates that are equal to quantized rate values. At quantized rate values the energy dissipation of voltage quantized DVS equals the energy dissipation of continuous voltage level DVS. However, the experimental results of [1] indicate that *RSA* only produces better energy savings than the conventional workload averaging technique if voltage is *not* dithered.

In this paper we present an improved rate selection algorithm named *Average-based Quantized Rate Selection (AQRS)* that produces better energy savings than the conventional workload averaging technique [2]. This algorithm uses the strengths of both the workload averaging technique and *RSA* in determining the rate. Moreover, this algorithm takes advantage of the convex nature of the energy versus rate relationship to improve energy savings by giving a higher priority to smaller quantized rates in its rate selection process. As a result the algorithm produces better energy savings, uses half the buffer size requirement of the conventional workload averaging technique, and improves overall system stability by eliminating the need for voltage dithering. To illustrate the performance of our algorithm we used nine 300-frame MPEG-2 video sequences as test data, the Forward Mapped Inverse Discrete Cosine Transform [4] as the variable workload computation, and a 4-level voltage quantization.

The rest of the paper is organized as follows. Section 2 presents the energy model, and section 3 describes the proposed *AQRS* algorithm. The experimental setup is given in section 4, and section 5 presents the results. Finally, section 6 presents the concluding remarks.

2. ENERGY MODEL

Combining Equations 1 and 2, energy per data sample is given by equation 3 [2]:

$$E(r) = E_0 r \left[\frac{V_t}{V_0} + \frac{r}{2} + \sqrt{r \frac{V_t}{V_0} + \left(\frac{r}{2}\right)^2} \right]^2 \quad (3)$$

where, E_0 is a scaling factor with units of energy, $r = f/f_{ref}$ is the normalized processing rate ($0 \leq r \leq 1$), $V_0 = (V_{ref} - V_t)^2/V_{ref}$, V_t is the threshold voltage, and f_{ref} is the frequency at maximum voltage V_{ref} .

Using Equation 3, the normalized energy and rate relationship for the continuous voltage levels model and voltage quantized model using a 4-level voltage quantization corresponding to normalized rates of 0.25, 0.5, 0.75, and 1.0 is shown in Figure 1.

The ideal continuous voltage levels model requires the availability of unique supply voltages for all rate values. However, in voltage quantized DVS only a fixed number of quantized rates (voltage levels) are available. Therefore processing samples with rate values that do not equal to quantized rates can be handled by two different techniques:

The first technique is to process the sample at the next highest quantized rate. For example, if a sample has a normalized rate value of $r = 0.6$, the sample will be processed at $r = 0.75$ using the quantization shown in Figure 1. This technique is known as DVS *without* voltage dithering, and the energy dissipation for this model is represented by the *stair-step* curve in Figure 1.

The second technique processes part of the sample at one quantization and the remainder at another quantization. For example, the same sample above with the normalized rate value of $r = 0.6$ using this model would have 40% of the sample period processed at $r = 0.75$ and the remainder (60%) at $r = 0.5$. This technique is known as DVS *with* voltage dithering, and the energy dissipation for this model is represented by the *straight-line* curve in Figure 1.

As Figure 1 shows, the voltage dithering technique significantly improves the energy savings. However, these energy savings are achieved at the cost of increased voltage transitions in the system (approximately one additional transition per sample period), degrading the overall stability of the DVS system.

Finally, Figure 2 shows the energy difference between the voltage quantized DVS *with* voltage dithering and *without* voltage dithering, and the continuous voltage levels DVS.

3. AVERAGE-BASED QUANTIZED RATE SELECTION

3.1 AQRS Algorithm

The *RSA* algorithm in [1] uses buffered workload to where possible select rates that are equal to quantized rates. As Figure 2 demonstrates, the energy difference between the voltage quantized model and the continuous voltage levels model approaches zero at quantized rates. Although

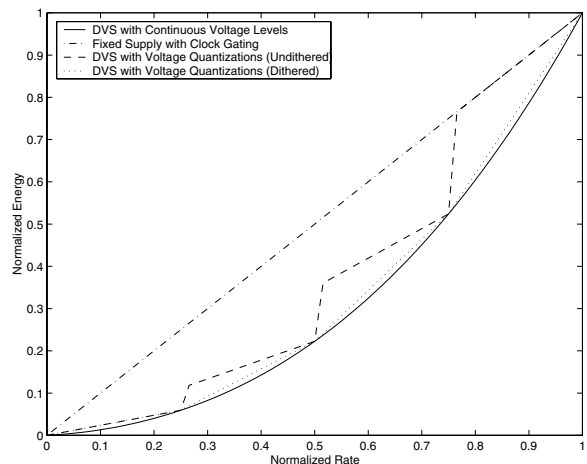


Figure 1: Energy vs. Rate

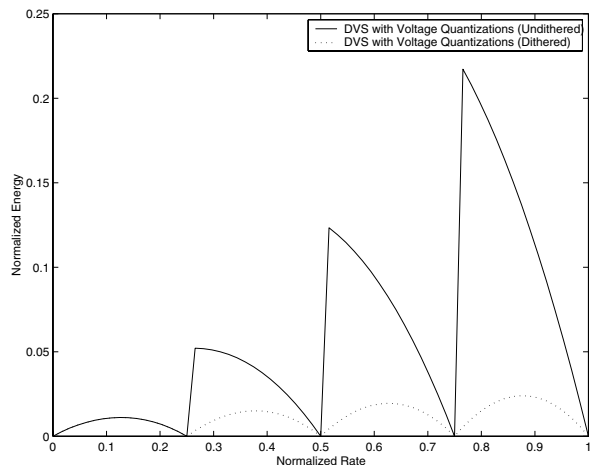


Figure 2: Energy Difference in Voltage Quantized Dynamic Voltage Scaling

the *RSA* algorithm effectively selects the smallest quantized rate that always clears at least one buffer location, the experimental results demonstrate that the algorithm always produces a small percentage of rates corresponding to the highest quantized rate ($r = 1.0$ normalized). Because of the convex nature of the energy versus rate curve, even a small percentage of rates at the highest quantization can be detrimental to the overall energy savings. For example, the number of samples that can be processed at each quantization of Figure 1 for the same energy dissipation is given by Equation 4:

$$1 \cdot E_{r=1} = \begin{cases} 16.87 \cdot E_{r=0.25} \\ 4.49 \cdot E_{r=0.5} \\ 1.91 \cdot E_{r=0.75} \end{cases} \quad (4)$$

where, $n \cdot E_{r=i}$ represents the number of samples n and the normalized energy at normalized rate value of $r = i$. As this equation demonstrates, the comparative energy of a single sample processed at the highest quantized rate is

approximately equivalent to 17 samples processed at $r = 0.25$, 5 samples at $r = 0.5$, and 2 samples at $r = 0.75$.

The *AQRS* algorithm uses the average buffered workload to estimate the required rate, and then selects the smallest quantized rate value that is sufficient to process the average buffered workload and at the same time empty one or more buffer locations. Because of this policy, the algorithm minimizes the rate values at the highest voltage quantization. Similar to the *RSA* algorithm, the *AQRS* algorithm prevents any buffer over-flow or under-flow conditions because the rate values are always selected from the actual total buffered workload.

The *AQRS* algorithm calculates the rate values in two steps: the first step calculates the average buffered workload (Equations 5 and 6), and the second step selects the smallest quantized rate value (Equation 7). The rate selection policy identifies the quantized rate r that is realistic ($\leq w_t$), clears the first buffer location ($\geq w[n]$), and satisfies the estimated average rate ($\geq r_a$). Equation 7 shows the details of the rate selection policy.

$$w_t = \sum_{k=0}^{B-1} w[n-k] \quad (5)$$

$$r_a = \frac{w_t}{B} \quad (6)$$

$$r = \begin{cases} r_0 & \text{if } w_t \geq r_0 \ \& \ w[n] \leq r_0 \ \& \ r_a \leq r_0 \\ r_1 & \text{if } w_t \geq r_1 \ \& \ w[n] \leq r_1 \ \& \ r_a \leq r_1 \\ \dots & \\ r_{p-2} & \text{if } w_t \geq r_{p-2} \ \& \ w[n] \leq r_{p-2} \ \& \ r_a \leq r_{p-2} \\ r_{p-1} & \text{if } w_t \geq r_{p-1} \\ w_t & \text{otherwise} \end{cases} \quad (7)$$

where, w_t is the total buffered workload, r_a is the integer portion of the average buffered workload, r is the selected rate value, B is the buffer size, $p(\geq 2)$ is the number of quantized rates (voltage levels), and r_0, \dots, r_{p-1} are quantized rate values ($0 < r_0 < r_1 < \dots < r_{p-2} < r_{p-1}$). If the rate values are normalized $r_{p-1} = 1$.

3.2 Sample Calculations

Figure 3 demonstrates how the *AQRS* algorithm calculates the rate r for a hypothetical data sequence. For this example a buffer size of 3 and a 4-level voltage quantization is used. The four normalized quantized rates are 0.25, 0.5, 0.75, and 1.0, corresponding to 16, 32, 48, and 64 non-zero coefficients, respectively, for the FMIDCT computation [4]. The Equations 5, 6, and 7 modified to suit this example are given by Equations 8, 9, and 10, respectively.

In the zeroth sample period, the first workload (15) is buffered into buffer location $w[n-2]$ and Equation 8 produces 15 as the total buffered workload w_t for this period. Next, the evaluation of Equation 9 produces 5 as the average rate r_a for this sample period. Since w_t is less than the smallest quantized rate (16), all the *if* conditions in the Equation 10 are evaluated to *FALSE* and the *OTHERWISE* part of the *if* condition sets the rate r to w_t ($r = 15$). In the first sample period, the w_t is 16, r_a is 5, and $w[n]$ is 0, so Equation 10 evaluates to *TRUE* for the first *if* condition and hence produces 16 as the rate. Similarly in the second, third, fourth and fifth sample periods, the algorithm produces rates of 32, 32, 48, and 48, respectively.

Workloads: 15, 16, 63, 62, 62, 64,

Buffer Contents: Unprocessed Work

Data Shifting Direction: ←

Period	w[n]	w[n-1]	w[n-2]	Wt	ra	w[n]	r
0	0	0	15	15	5	0	15
1	0	0	16	16	5	0	16
2	0	0	63	63	21	0	32
3	0	31	62	93	31	0	32
4	0	61	62	123	41	0	48
5	13	62	64	139	46	13	48

Figure 3: AQRS Algorithm Sample Calculations

$$w_t = \sum_{k=0}^2 w[n-k] \quad (8)$$

$$r_a = \frac{w_t}{3} \quad (9)$$

$$r = \begin{cases} 16 & \text{if } w_t \geq 16 \ \& \ w[n] \leq 16 \ \& \ r_a \leq 16 \\ 32 & \text{if } w_t \geq 32 \ \& \ w[n] \leq 32 \ \& \ r_a \leq 32 \\ 48 & \text{if } w_t \geq 48 \ \& \ w[n] \leq 48 \ \& \ r_a \leq 48 \\ 64 & \text{if } w_t \geq 64 \\ w_t & \text{otherwise} \end{cases} \quad (10)$$

3.3 Complexity

The *AQRS* algorithm can be implemented by a constant number ($B-1$) of *add* operations, a single *divide* operation, and a data-dependent number of *compare* (n_c) and *and* (n_a) operations. The constant B is the buffer size, and n_c and n_a are defined by Equations 11 and 12, respectively.

$$2 \leq n_c \leq 3p-2 \quad (11)$$

$$2 \leq n_a \leq 2(p-1) \quad (12)$$

where, p is the number of quantized voltage levels. A comparison of the variation of operational complexity for *AQRS* algorithm, *RSA* algorithm, and the workload averaging technique as a function of buffer size is shown in Figure 4. Since the number of operations for *AQRS* algorithm and *RSA* algorithm are workload dependent, an upper and a lower bound for the number of operations have been plotted.

4. EXPERIMENTAL SETUP

To demonstrate the energy savings achievable with the *AQRS* algorithm we compare its energy dissipation results with the results of the conventional workload averaging technique [2]. For the energy analysis we use the Forward Mapped Inverse Discrete Cosine Transform (FMIDCT) algorithm [4] as the data-dependent computation and nine MPEG-2 compressed video sequences as test data. The video sequences

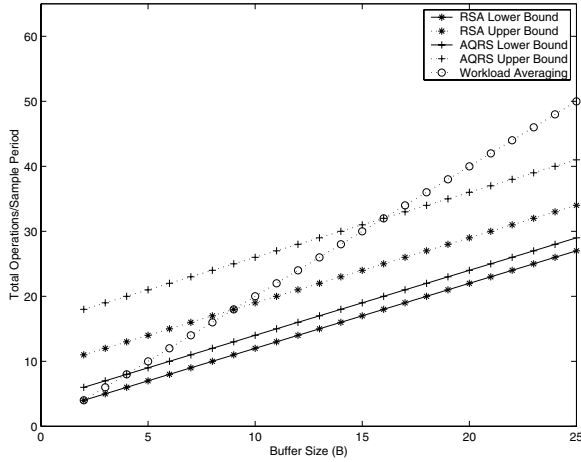


Figure 4: Operational Complexity vs. Buffer Size

Parameter	Value
Frames in a GOP	15
I/P Frame Distance	3
Horizontal Picture Size (Pixels)	176
Vertical Picture Size (Pixels)	144
Frame Rate	30 frames/sec
Bit Rate	4 Mbits/sec
Profile	Main
Level	Low
Chroma Format	4:2:0

Table 1: MPEG-2 Compression Parameters

comprise of 300-frame color video sequences in QCIF format commonly used in the research community. More specifically, the names of the video sequences are Akiyo, Carphone, Coastguard, Container, Foreman, Hall, Mother, News, and Silent [7]. The MPEG-2 codec software used for video compression and decompression was acquired from the MPEG Software Simulation Group [5].

The MPEG-2 video compression parameters used in our experimental work are given in Table 1.

To study the impact of buffer size on the performance of the algorithm, we varied the buffer size from 2 to 25. For energy calculations, we assumed the maximum supply voltage and the threshold voltage (V_t) to be 3.3V and 0.7V, respectively. Moreover, we assumed that energy loss in the DC-DC converter to be negligible.

The simulation tool used for our energy analysis was MATLAB (Version 5) from MathWorks Inc.

5. RESULTS

Figures 5 and 6 show the energy savings achieved by the *AQRS* algorithm compared to the conventional workload averaging technique. For *all* nine test video sequences the algorithm produced better energy savings than the workload averaging technique, and the energy savings ranged from 1% to 23% for the 300-frame video sequences as the buffer size was varied from 2 to 25.

The *AQRS* algorithm achieves these energy savings by

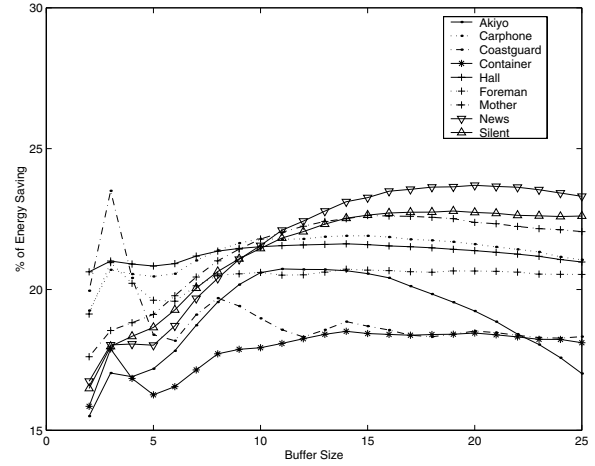


Figure 5: Energy Savings Comparison with Averaging (Undithered)

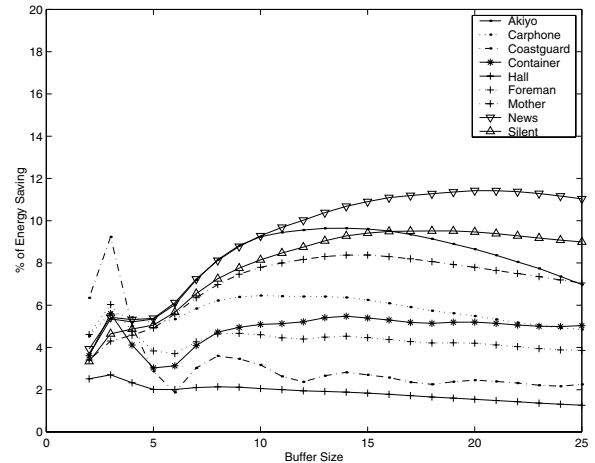


Figure 6: Energy Savings Comparison with Averaging (Dithered)

minimizing the higher quantized rates during rate selection. Table 2 demonstrates the changes in workload distribution compared to the corresponding distribution using the *RSA* algorithm. As this table shows, there is nearly a 100% reduction in workload distribution at $r = 1.0$ and a 75% or higher reduction for a majority of test cases at $r = 0.75$. Moreover, there is a small reduction in distribution at $r = 0.25$ and the majority of workload distributions are concentrated at $r = 0.5$. Finally Figure 7 shows the comparative workload distributions for all nine test video sequences using the workload averaging technique, the Rate Selection Algorithm [1], and the *AQRS* algorithm. For this comparison, a buffer size of 10 is used. This figure further highlights the efficiency of *AQRS* algorithm in reducing the higher quantized rates, particularly in the $0.5 > r \geq 1.0$ range as demonstrated in Table 2.

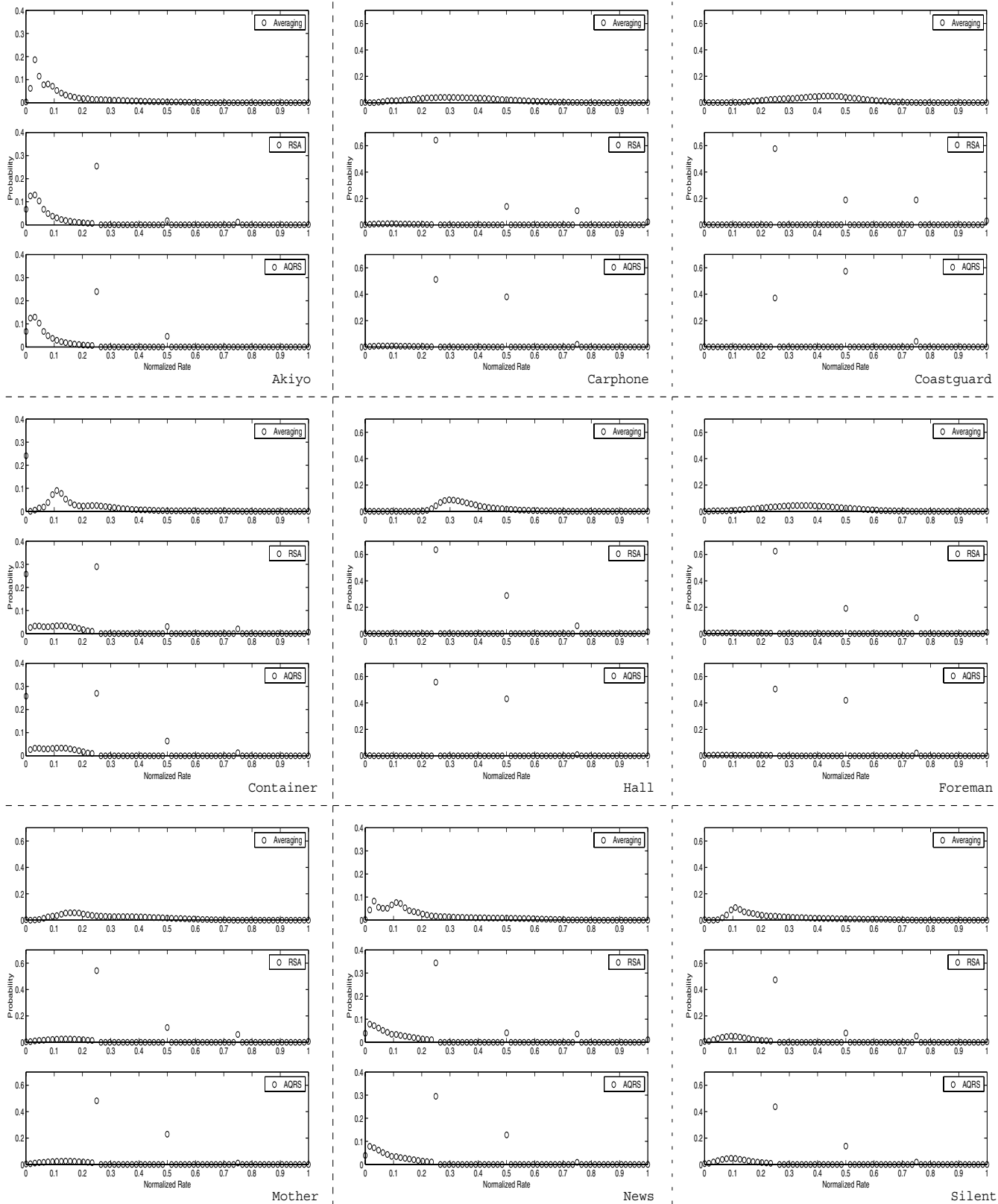


Figure 7: Rate Distribution Comparison for All Test Video Sequences

Sequence	r=0.25	r=0.5	r=0.75	r=1.0
Akiyo	-5.84%	155.78%	-96.76%	-100.0%
Carphone	-20.47%	172.18%	-80.69%	-99.48%
Coastguard	-35.91%	203.99%	-77.83%	-99.93%
Container	-6.82%	110.66%	-36.76%	-99.11%
Hall	-12.24%	49.07%	-85.61%	-99.96%
Foreman	-19.08%	119.55%	-82.73%	-99.82%
Mother	-11.36%	104.65%	-82.67%	-99.03%
News	-14.12%	209.67%	-75.21%	-99.26%
Silent	-7.59%	96.56%	-62.01%	-99.85%

Table 2: Workload Distribution Comparison - AQRS vs. RSA

6. CONCLUSION

In this paper we have presented a highly energy efficient alternative algorithm to the conventional workload averaging technique for voltage quantized dynamic voltage scaling. The algorithm named *Average-based Quantized Rate Selection* is based on the workload averaging technique [2] and our Rate Selection Algorithm [1]. The algorithm monitors the total buffered workload value, uses the average buffered workload to estimate the rate, and where possible *selects* rates that are equal to quantized rates. At quantized rates, the energy dissipation of voltage quantized DVS approaches the energy dissipation of continuous voltage level DVS. Moreover, the estimation of rate using the average buffered workload reduces higher quantized rates, there by reducing the overall energy dissipation. Our experimental work using the Forward Mapped Inverse Discrete Cosine Transform as the data-dependent computation for DVS, nine MPEG-2 test video sequences as test data, and a four-level voltage quantization shows that our algorithm produces better energy savings than the workload averaging technique for all test cases, and the energy savings achieved ranges from 1% to 23%.

Apart from the additional energy savings, our algorithm also has two key advantages over the averaging technique. First, the algorithm only requires one buffer or half the buffer size requirement of the averaging technique, making it very effective for applications with reduced memory and area requirements. Second, the algorithm eliminates the need for voltage dithering, there by improving the overall stability of the system by minimizing the number of voltage transitions.

Finally, as Figure 4 shows, the *AQRS* algorithm has a higher operational complexity compared to the *RSA* algorithm. However, comparison of the *AQRS* and the workload averaging technique shows that for buffer size $B > 15$, the overhead of the former (upper and lower bound) is smaller than the latter, making the *AQRS* algorithm a very energy efficient, low complexity replacement algorithm for the workload averaging technique in voltage quantized dynamic voltage scaling.

7. REFERENCES

- [1] L. H. Chandrasena and M. J. Liebelt. A Rate Selection Algorithm for Quantized Undithered Dynamic Supply Voltage Scaling. In *International Symposium on Low Power Electronics and Design*, July 2000.
- [2] V. Gutnik and A. P. Chandrakasan. Embedded Power Supply for Low-Power DSP. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 5(4):425–435, December 1997.
- [3] T. Ishihara and H. Yasuura. Voltage Scheduling Problem for Dynamically Variable Voltage Processors. In *International Symposium on Low Power Electronics and Design*, August 1998.
- [4] L. McMillan and L. Westover. A Forward-Mapping Realization of the Inverse Discrete Cosine Transform. In *Data Compression Conference*, pages 219–228, March 1992.
- [5] <http://www.bok.net/mpeg/MSSG/>.
- [6] L. S. Nielsen, C. Niessen, J. Sparsø, and K. van Berkel. Low-Power Operation Using Self-Timed Circuits and Adaptive Scaling of Supply Voltage. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2(4):391–397, December 1994.
- [7] [Ftp://dspftp.ece.ubc.ca/pub/tmn/qcif_source/](ftp://dspftp.ece.ubc.ca/pub/tmn/qcif_source/).