THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 52, No. 7, September, 1973
Printed in U.S.A.

Adaptive Quantization in Differential PCM Coding of Speech

By P. CUMMISKEY, N. S. JAYANT, and J. L. FLANAGAN

(Manuscript received March 12, 1973)

We describe an adaptive differential PCM (ADPCM) coder which makes instantaneous exponential changes of quantizer step-size. The coder includes a simple first-order predictor and a time-invariant, minimally complex adaptation strategy. Step-size multipliers depend only on the most recent quantizer output, and input signals of unknown variance can be accommodated. We derive appropriate multiplier values from computer simulations with speech signals and with Gauss-Markov inputs. We compare performance of the ADPCM coder with conventional log-PCM, using both objective and subjective criteria. Finally, we describe an economical integrated hardware implementation of the ADPCM coder. We believe that at bit rates of 24 to 32 kb/s, ADPCM provides a robust and efficient technique for speech communication and for digital storage of speech.

I. INTRODUCTION

The advantages of coding speech digitally are well known.[1] Expected benefits include low costs per line, ease of maintenance, and high-quality signal regeneration at repeaters. Furthermore, digital coding is well matched to current technology in terms of readily available integrated circuit hardware. Results from speech-coding research are now beginning to specify techniques that are nearly optimal for a given bit rate, a given channel quality, and a given degree of coder complexity. Finally, the subject of direct digital conversion between alternative code formats is being widely studied, and simple techniques have already been proposed for some specific conversions.

The coder discussed in this paper is believed to be efficient and robust for speech coding at bit rates of 24 to 32 kilobits/second (kb/s). Other refinements of differential PCM (DPCM)[2] are based, at least in part, on adaptive prediction.[3–6] These techniques offer considerable potential for bandwidth compression,[3] but are typically hard to implement. Therefore, for the type of bit rates mentioned earlier, it seems much more reasonable to tap the advantages of a more simply implemented adaptive quantizer.

Our adaptive DPCM (ADPCM) coder, therefore, operates on the basis of a fixed, first-order predictor in the DPCM loop, and a time-invariant, adaptation strategy for instantaneous changes of quantizer step-size. The technique has obvious advantages over conventional PCM (due to redundancy removal) and over conventional DPCM (due to increased dynamic range). Further, the quality of speech reproduction in the 24- to 32-kb/s range is believed to be perceptually better than that provided by adaptive delta modulation (ADM) which, however, has the advantage of even greater simplicity.[7,8]

Besides digital telephone applications, appropriate utilizations of ADPCM coding are seen in computer storage of digital speech (for voice answer-back, "voice-wiring," and similar functions), in mobile radio telephony, and in special applications such as deep-space communication and digital encryption.

II. DEFINITION OF THE ADPCM CODER: ADAPTIVE QUANTIZATION WITH A ONE-WORD MEMORY

A schematic block diagram of the coder appears in Fig. 1. It follows the conventional differential PCM structure with a first-order, fixed

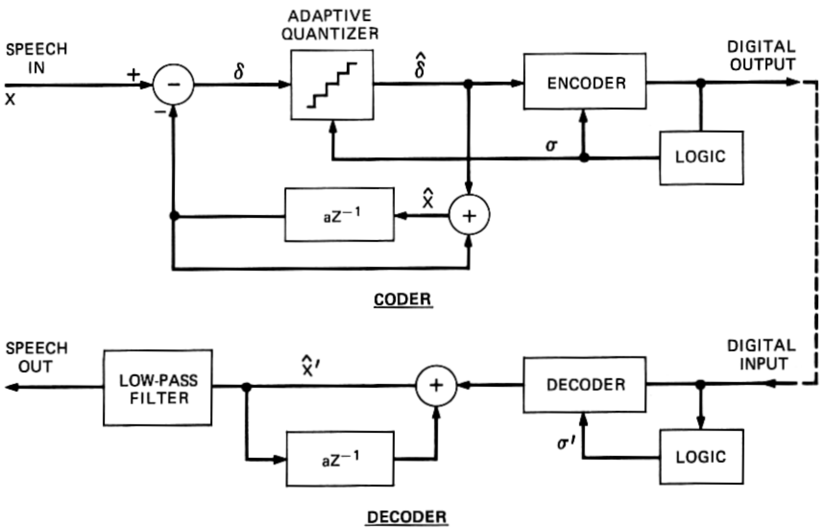

Fig. 1—Block diagram of ADPCM coder.

predictor in the feedback loop.[2] It has, however, the additional box labeled LOGIC, which provides adaptation of quantizer step-size on the basis of the most recent quantizer output. In the absence of channel errors, the step-size controls σ and σ' are identical, and so are the signal estimates $\hat{x}$ and $\hat{x}'$.

Step-size adaptations are motivated by the assumption that the variance of the quantizer input δ is unknown. The empirical adaptation rule is that, for every new input sample, the step-size is changed by a factor depending only on the knowledge of which quantizer slot was occupied by the previous signal sample.

Formally, if the outputs of a uniform B-bit quantizer are of the form

$$Y_u = P_u \frac{\Delta_u}{2}; \qquad \pm P_u = 1, 3, \cdots 2^B - 1; \qquad \Delta_u > 0,$$

the step-size Δ_r is given by the previous step-size multiplied by a time-invariant function of the code-word magnitude $|P_{r-1}|$;

$$\Delta_r = \Delta_{r-1} \cdot M(|P_{r-1}|),$$

subject, of course, to maximum and minimum limits on Δ_r, as specified in specific implementations. Step-size multipliers for a 3-bit uniform quantizer are illustrated in Fig. 2. Note that there are only four distinct

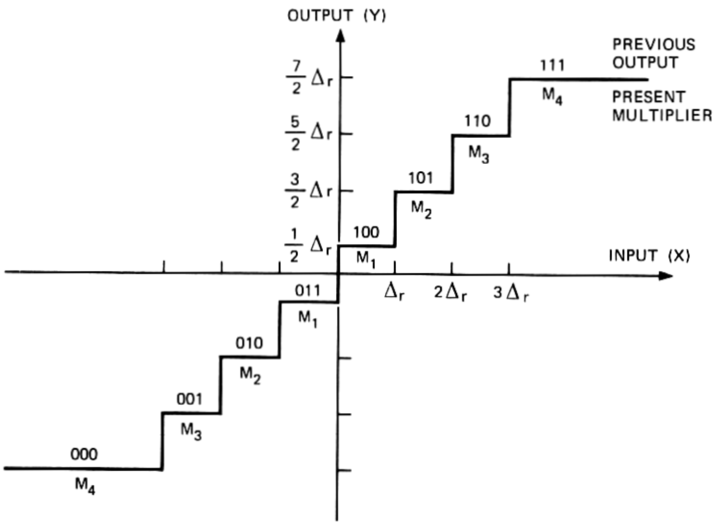

Fig. 2—Adaptation multipliers associated with quantizer levels for 3-bit ADPCM coder.

TABLE I—ADAPTATION MULTIPLIERS FOR SPEECH AND FOR GAUSS-MARKOV INPUTS (3-BIT CODER)

Previous Output Word	Multiplier	Speech Input	Gauss-Markov Input (Correlation Between Adjacent Samples = 0.5)
111 or 000	M_4	2	1.75
110 or 001	M_3	5/4	1.25
101 or 010	M_2	7/8	0.90
100 or 011	M_1	7/8	0.90

multipliers because the polarity of quantizer output is not utilized in the adaptation logic. Furthermore, meaningful adaptation requires that the step-size be increased on the detection of quantizer overload ($M_4 > 1$) and decreased during underload ($M_1 < 1$), and that $M_1 \leqq M_2 \leqq M_3 \leqq M_4$. Derivation of specific multiplier values is outlined in the next section.

III. DESIGN OF STEP-SIZE MULTIPLIERS

Two conflicting requirements are encountered in designing step-size multipliers. The first is the need to respond quickly to abrupt changes of input variance (suggesting the use of $M_4 \gg 1$, $M_1 \ll 1$ for the 3-bit example). The second requirement is the prevention of excessive step-size alterations in a stationary or steady-state situation (suggesting the use of $M_4 = 1 + \epsilon_4$; $M_1 = 1 - \epsilon_1$; $\epsilon_4, \epsilon_1 \rightarrow 0$). Compromise values of multipliers are therefore suggested for an input signal, or for a class of input signals.

Extensive computer simulations were carried out to determine the most desirable multiplier values for an illustrative speech sample. The sample was a male utterance of "This circuit operates on the same principle as N. S. Jayant's simulation." The speech was bandpass-filtered (200–3200 Hz), and was sampled at 8 kHz. Multiplier values were sought that maximized the signal-to-quantization-error (power) ratio (SNR), as averaged over the entire duration of the above utterance. Rounded values of these multipliers are shown in Table I for a 3-bit quantizer. Also shown are the values found to be desirable for the quantization of a Gauss-Markov input with an input signal correlation similar to that expected for Nyquist-sampled speech.[2] The similarity is interesting, particularly because the speech quantizer had a Max nonuniformity[9] (to take into account the observed Gaussian tendencies of the quantizer input), while the Gauss-Markov simulation utilized a uniform quantizer. The latter simulation also showed that desirable multiplier values are only slightly dependent on signal correlation,

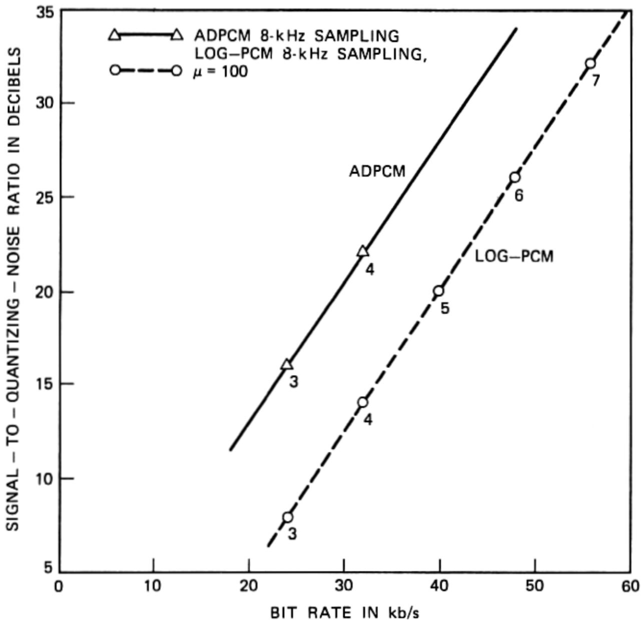

Fig. 3—Signal-to-quantizing-noise ratios for speech signals.

suggesting a robust adaptation strategy. Finally, the similarity of the multiplier values found for speech and for Gauss-Markov inputs suggests that the coder has a versatility that might extend to facsimile and video signals.[10]

The general problem of determining most desirable multiplier values is discussed at length in a companion paper.[10] That paper also points out the possibility of near-optimal adaptation strategies that have nontrivial ($\neq 1$) values only for the end multipliers (M_1 and M_4 in the 3-bit example), and compares our adaptation logic with that of Stroh.[4]

IV. PERFORMANCE COMPARISONS OF THE ADPCM CODER WITH CONVENTIONAL PCM

4.1 *SNR Data*

Computer simulations using speech input showed the signal-to-error-power ratio to be 16 dB for a 3-bit (24 kb/s) ADPCM coder which has the multipliers of Table I and a maximum step-size which was $D = 128$ times the minimum. The SNR of 16 dB represents an 8-dB gain over 3-bit logarithmic PCM with $\mu = 100$.[11]* It turns out that

* This value was chosen on the basis of past experience with speech coders. Present trends are toward a higher value for μ.

Table II—Comparison of Objective and Subjective Performance of ADPCM and Log-PCM

Objective Rating (SNR)	Subjective Rating (Preference)	
7-bit PCM	7-bit PCM	(High)
6-bit PCM	4-bit ADPCM	
4-bit ADPCM	6-bit PCM	
5-bit PCM	3-bit ADPCM	
3-bit ADPCM	5-bit PCM	
4-bit PCM	4-bit PCM	(Low)

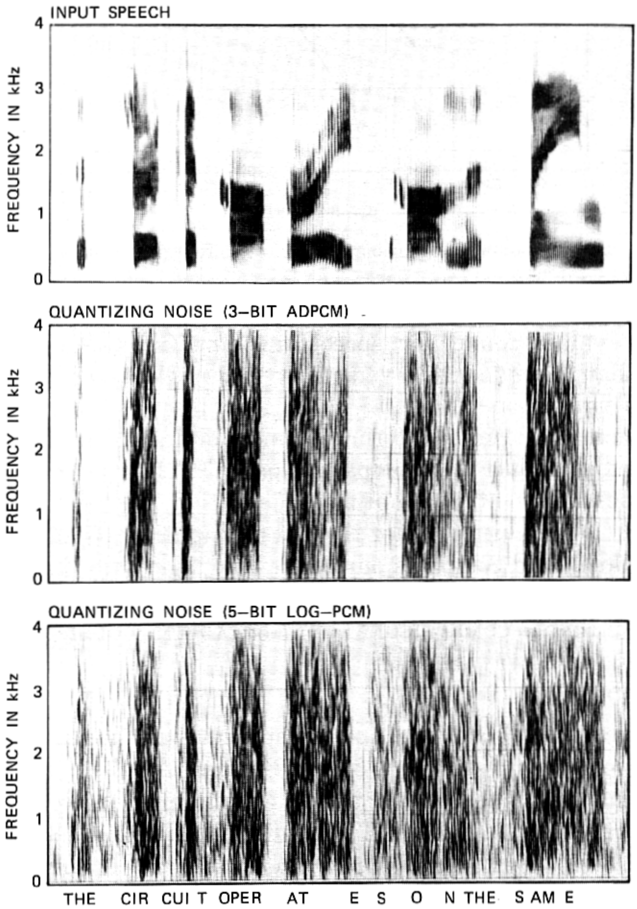

Fig. 4—Spectograms of speech and quantization error.

this improvement includes a 4-dB gain due to differential encoding and a 4-dB gain due to the quantizer adaptation. Figure 3 gives a more complete comparison of speech signal SNR's measured for the ADPCM and for log-PCM.

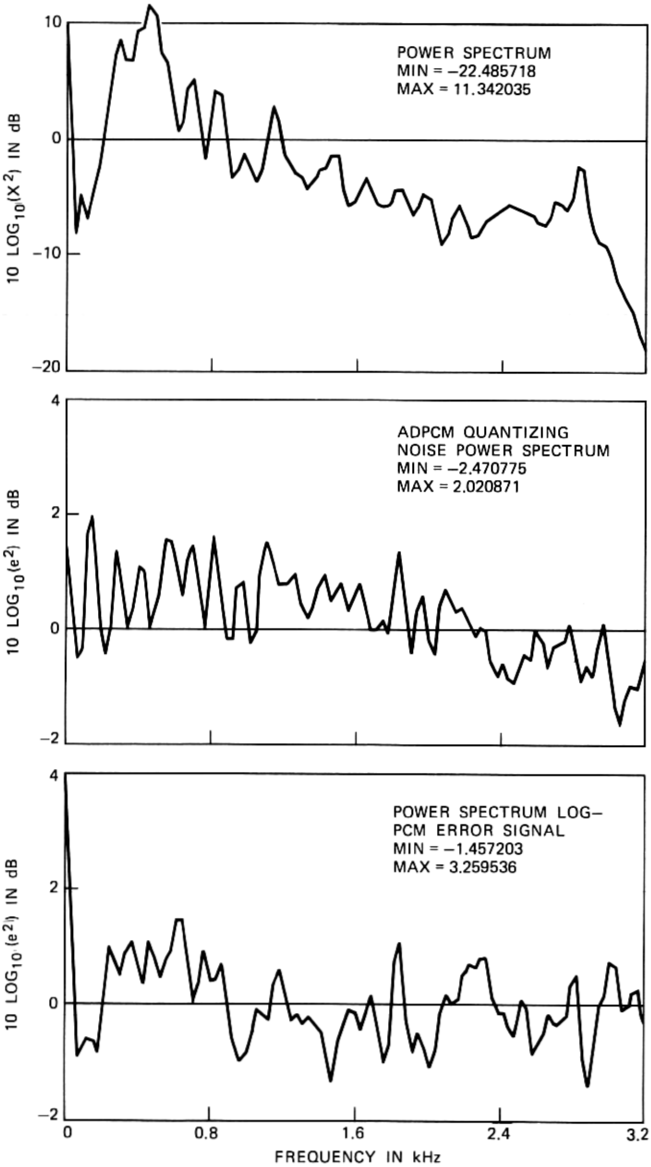

Fig. 5—Long-term error spectra.

4.2 *Error Spectrograms and Long-Term Spectra*

Figure 4 displays spectrograms of a section of the input speech and the associated quantizing noise spectrograms with 3-bit ADPCM and 5-bit log-PCM. Note that ADPCM provides considerably less noise during the silent intervals in speech, although the total noise power is greater in this coder (Fig. 3). This suggests that adaptive quantization in ADPCM provides greater dynamic range than the logarithmic compandor used in PCM. (The observation results, no doubt, from the specific numerical values $\mu = 100$ and $D = 128$ in Section 4.1. However, these values are believed to be very representative.)

Figure 5 illustrates another interesting difference between the quantizing noise in ADPCM and that in PCM. Note the high-frequency rolloff in ADPCM noise and the relative whiteness of the noise spectrum in PCM.

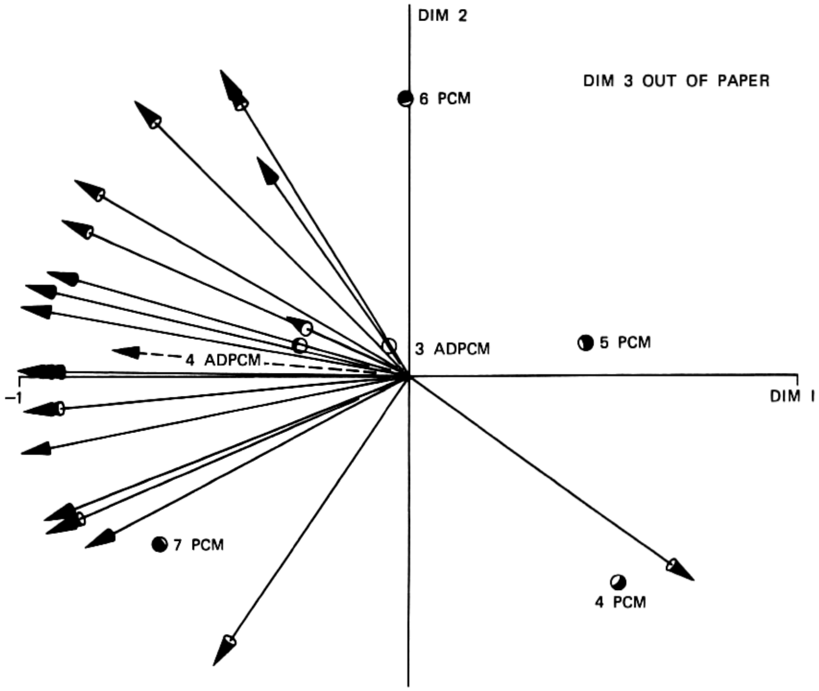

Fig. 6—Subjective preference judgments of various ADPCM and log-PCM codings. Dimensions 1 and 2 account for most of intersubject variance. Increasing preference is in the $-x$ direction. Individual subject vectors are plotted, and projection of coding conditions onto a subject vector indicates how that individual rank-ordered the coding systems.

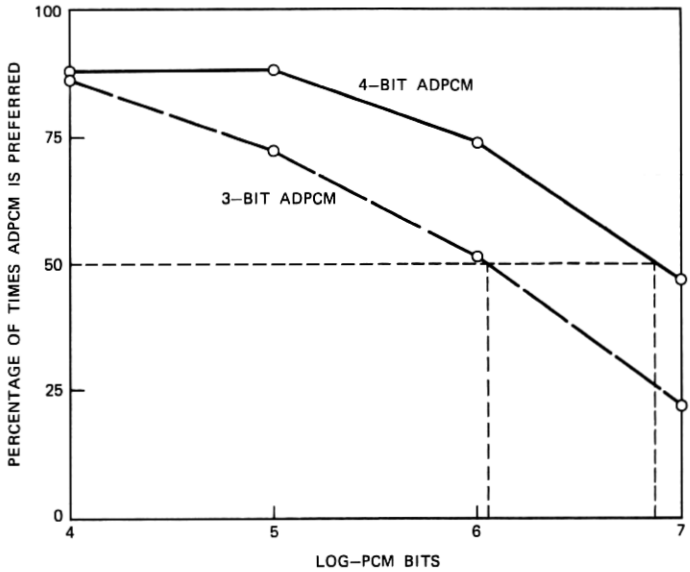

Fig. 7—Alternative interpretation of (ADPCM-versus-PCM) preference scores.

4.3 *Subjective Tests*

Apart from the measured differences mentioned in Section 4.2, informal listening tests indicated that ADPCM noise had a perceptually more palatable character* than PCM noise of equal variance; in other words, the quality of speech reproduction from the ADPCM (relative to PCM) was much better than was suggested by the SNR comparisons in Fig. 3. This observation was borne out in the following perceptual experiment.

The experiment involved 3- and 4-bit ADPCM stimuli and 4-, 5-, 6-, and 7-bit log-PCM stimuli. The total number of cross comparisons possible was 16 (2 stimuli $\times$ 4 stimuli $\times$ 2 orders of presentation). Twenty-two listeners participated in the tests and made preference judgments of signal quality for each of the 16 A-B comparisons. The preference judgments were submitted to a multidimensional scaling program,[12] and the results were plotted in terms of two subjective dimensions which accounted for most of the perceived differences. Dimension 1, in particular, accounted for 75 percent of the variance in the preference data.

The results are shown in Fig. 6. Individual subject vectors are

* Related, perhaps, to a lesser proportion of the noise getting into the idle circuit; and, also, due to some correlations of ADPCM noise with pitch information.

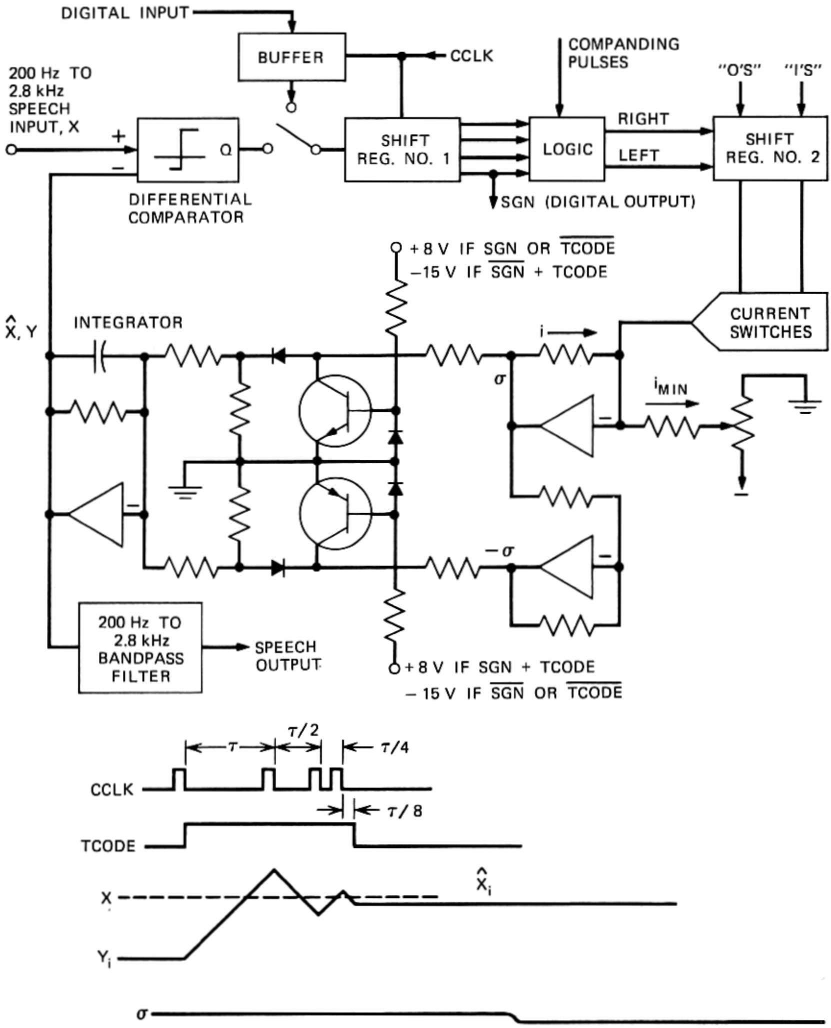

Fig. 8—Circuit block diagram for the hardware ADPCM coder.

displayed (solid lines), and projection of the coding conditions into a subject's vector reveals how that individual rank-ordered the signal qualities.* A resultant of the subject vectors is also shown (dashed), and projections onto this resultant indicate subject consensus in rank-

* One subject (vector in quadrant IV) apparently misunderstood the test instructions and gave essentially complementary preference judgments.

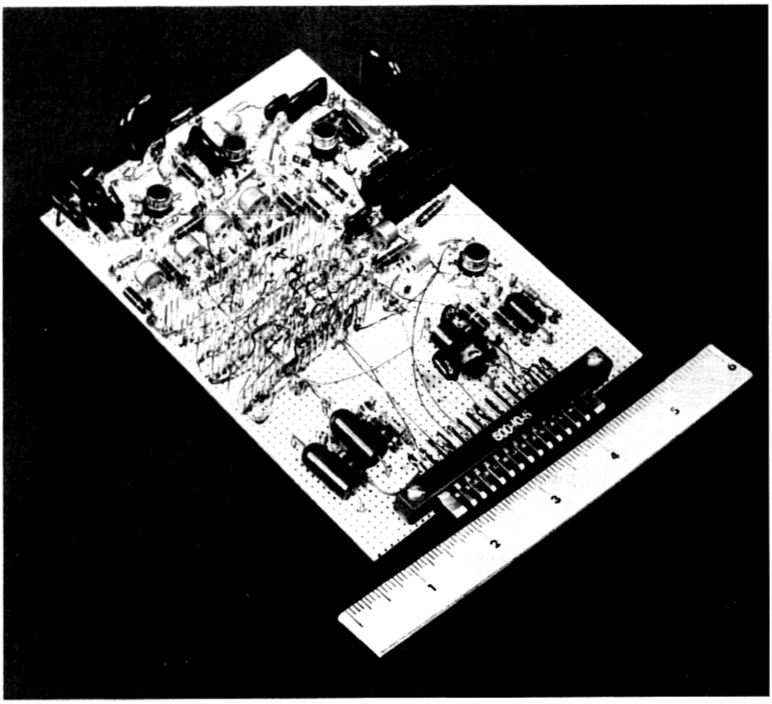

Fig. 9—ADPCM coder.

ing the qualities. This averaged subjective ranking can be compared with the objective SNR performance for the same codings (using the data of Fig. 3).

A comparison of the objective (SNR) performance and the subjective (preference) ranking of the codings tested is shown in Table II. It is clear from these data that subjectively the ADPCM does an even better job than the objective SNR's indicate. Arrows indicate two subjective "promotions" of the ADPCM. Table II (and, of course, Fig. 6) show, too, that 4-bit ADPCM is perceptually better than 6-bit log-PCM.

Figure 7 provides yet another means of comparing ADPCM and PCM using the original preference scores from the perceptual test. Ordinates in Fig. 7 are overall percentages (including all listeners) of A-B judgments where an ADPCM stimulus was preferred to a certain PCM stimulus as shown on the x-axis of Fig. 7. The 50-percent probability-of-preference line intersects the curves at points whose ab-

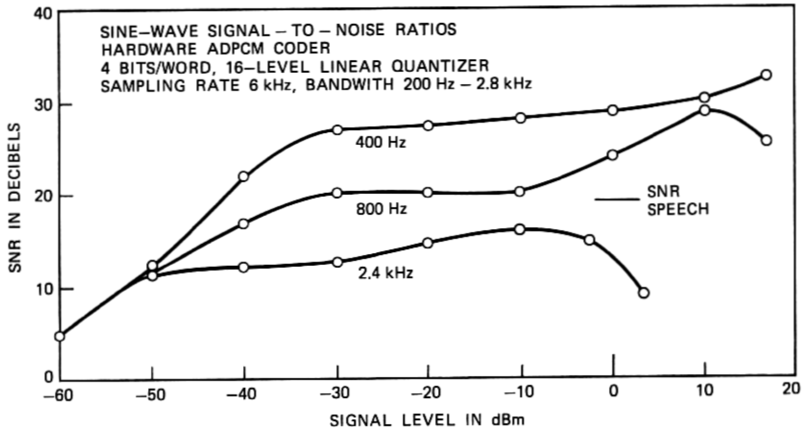

Fig. 10—Signal-to-quantizing-noise ratios for sine-wave input to the hardware ADPCM coder.

scissas represent quantitative log-PCM equivalences for 3- and 4-bit ADPCM.

V. HARDWARE DESIGN OF ADPCM CODER

The computer simulations described above established the design criteria for the ADPCM. To assess hardward viability of the technique, we constructed a 4-bit ADPCM coder in integrated circuit hardware. A circuit block diagram of the hardware coder is given in Fig. 8, and a photograph of the circuit card is shown in Fig. 9. State-of-the-art circuit technology is used and all circuit components are "off the shelf."

The circuit incorporates a uniform quantizer realized by using a serial logic (shown at the bottom of Fig. 8) to code the difference between the input X and the signal stored on the integrator Y. The logic provides four consecutive increments of integrator voltage within a duration much smaller than the sampling period. Each of these increments follows the latest sign of $(X - Y)$ and has a magnitude that is one-half that of the previous increment in the cycle.

Step-size adaptations are controlled by the current switches which provide a dictionary of 21 step-sizes. These are spaced with a ratio of $2^{\frac{1}{3}}$ between adjacent steps, and therefore provide an overall step-size range of 128:1. The step-size multipliers for speech (Table I) are approximated in the circuit as positive and negative exponents of $2^{\frac{1}{3}}$.

Measurements on the hardware realization confirm the computer observations on SNR for speech signals. Also, SNR's for sine-wave inputs are conventionally used to assess digital coders (although sine-wave performance can be very deceptive in terms of perceptual acceptability). Figure 10 shows SNR measured on the hardware coder for sine-wave input. The 800-Hz behavior is reasonably consistent with the SNR measured for speech input.

VI. DIRECT DIGITAL CONVERSION BETWEEN ADPCM AND OTHER SIGNAL FORMATS

An important issue in the compatibility of digital systems is the provision of graceful and virtually transparent conversions among different code formats.[13] Digital techniques for directly converting between ADPCM and the conventional formats of DPCM and PCM have been proposed and are being studied.[14] One of the indications of these studies is that direct conversion between ADPCM and DPCM is quite feasible, especially when the conversion incorporates an intermediate stage of PCM.

VII. CONCLUSION

Results of this study indicate that the ADPCM technique leads to an economic, efficient digital coding of speech for the bit-rate range 24 to 32 kb/s. This range constitutes a channel capacity saving of over 2:1 compared to conventional PCM and produces a signal coding of comparable quality. Hardware implementation is relatively straightforward and noncritical.

Studies presently in progress are examining ADPCM coding for operation at 18 kb/s. Preliminary indications are that signal quality attractive for mobile radio application can be achieved at this low bit rate. This low rate also makes digital encryption for privacy attractive in mobile telephone.

Although not specifically discussed in this exposition, ADPCM proves reasonably robust in the presence of errors in the transmit channel. The computer simulations described above incorporated preliminary studies of error vulnerability which show the coder to perform well for channels with error probability $\leqq 10^{-4}$. Typical error rates in "clean" PCM channels are routinely maintained lower than this.

Further study of ADPCM is anticipated in objective analysis of its quantizing characteristics. This should be coupled with more complete perceptual tests to better understand the "perceptual palatability" of ADPCM coding. Further, a close competitor is adaptive delta modulation,[7,8] and subjective comparisons are planned that will include this coding technique.

One present utilization of the hardware ADPCM coder is in a computer voice response system for generating computer-spoken wiring instructions.[15] Speech coding at 24 kb/s provides economy of digital storage and simple A/D–D/A communication with the computer.

REFERENCES

1. Flanagan, J. L., "Focal Points in Speech Communication Research," IEEE Trans. Commun. Technology, December 1971 (Part I), pp. 1006–1015.
2. McDonald, R. A., "Signal-to-Noise and Idle Channel Performance of Differential Pulse Code Modulation Systems—Particular Applications to Voice Signals," B.S.T.J., *45*, No. 7 (September 1966), pp. 1123–1151.
3. Atal, B. S., and Schroeder, M. R., "Adaptive Predictive Coding of Speech Signals," B.S.T.J., *49*, No. 8 (October 1970), pp. 1973–1986.
4. Stroh, R. W., "Optimum and Adaptive Differential Pulse Code Modulation," Ph.D. Thesis, Polytechnic Institute of Brooklyn, 1970.
5. Noll, P. W., "Non-Adaptive and Adaptive DPCM of Speech Signals," Overdruk int Polytechnisch Tijdschrift, editie elektrotechnick-Elektronica, Nr. 19, 1972.
6. Cummiskey, P., "Adaptive Differential Pulse-Code Modulation for Speech Processing," Ph.D. Thesis, Newark College of Engineering, 1973.
7. Jayant, N. S., "Adaptive Delta Modulation with a One-Bit Memory," B.S.T.J., *49*, No. 3 (March 1970), pp. 321–342.
8. Jayant, N. S., Cummiskey, P., and Flanagan, J. L., "Design and Implementation of an Adaptive Delta Modulator," Proc. IEEE Int. Conf. Speech Commun. and Processing, Boston, Massachusetts, April 1972.
9. Max, J., "Quantization for Minimum Distortion," IRE Trans. Information Theory, *IT-6*, March 1960, pp. 7–12.
10. Jayant, N. S., "Adaptive Quantization With a One-Word Memory," B.S.T.J., this issue, pp. 1119–1144.
11. Smith, B., "Instantaneous Companding of Quantized Signals," B.S.T.J., *36*, No. 3 (May 1957), pp. 653–709.
12. Carroll, J. D., "Individual Differences and Multidimensional Scaling," in R. N. Shepard, A. K. Romney, and S. Nerlove (eds.), *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, vol. I, New York: Seminar Press, 1972, pp. 105–155.
13. Goodman, D. J., and Flanagan, J. L., "Direct Digital Conversion between Linear and Adaptive Delta Modulation Formats," Proc. Int. Conf. Commun., Montreal, June 1971, pp. 1–33 to 1–36.
14. Flanagan, J. L., U. S. patent application.
15. Shipley, K., Rabiner, L. R., Schafer, R. W., and Flanagan, J. L., "Computer-Spoken Instructions for ESS-1 Central Office Wiring," unpublished work.