Table 1. Performance Summary

Techonology	1P8M 0.13µm CMOS	
Power Supply	3V	
Active Area	0.9mm×0.4mm	
Memory	Non-Volatile Floating Gate	
Memory SNR	46dB	
Training Algorithm	Unsupervised Online Clustering	
Input Referred Noise	$56.23pA_{rms}$	
System SNR	45dB	
I/O Type	Analog Current	
Operating Frequency	Training Mode	4.5kHz
	Recognition Mode	8.3kHz
Power Consumption	Training Mode	27µW
	Recognition Mode	11.4µW
Energy Efficiency	Training Mode	480GOPS/W
	Recognition Mode	1.04TOPS/W

Measurement Results

The ADE occupies an active area of $0.36\ mm^2$ in a 0.13 µm CMOS process. With a 3 V power supply, it consumes 27 µW in training mode, and 11.4 µW in recognition mode.

Noise Performance To measure input-referred noise, we construct a histogram of classifications near a decision boundary with adaptation disabled. Assuming additive Gaussian noise, the relative frequency of one of the two classification results approximates the cumulative density function (cdf) of a normal distribution with standard deviation equal to the RMS current noise. The measured input-referred current noise is $56.23\ pA_{rms}$, which with an input full scale range of 10 nA corresponds to an SNR of 45 dB, or 7.5 bit resolution.

Clustering Test To test the clustering performance of the node, an input dataset of 40,000 8-D vectors was drawn from 4 underlying Gaussian distributions with different mean and variance. During the test, the centroid means were read out every 0.5 s; the locations are overlaid on the data scatter in Figure 5. For easier visual interpretation, 2-D results are shown. The performance of the starvation trace is verified by presenting the *node* with an ill-posed clustering problem where one centroid is initialized far from the input data, and is therefore never updated without the ST enabled (Figure 5, left). With the starvation trace enabled, the starved centroid is slowly pulled toward the area populated by the data, achieving a correct clustering result, shown in Figure 5 (right).

Feature Extraction Test We demonstrate the full functionality of the chip by performing feature extraction for pattern recognition with the setup shown in Figure 6(a). The input patterns are 16×16 bitmaps corrupted by random pixel errors. An 8×4 moving window selects the ADE's 32 inputs. The ADE is first trained on unlabeled patterns. After training, adaptation can be disabled and the circuit operates in recognition mode. The 4 belief states

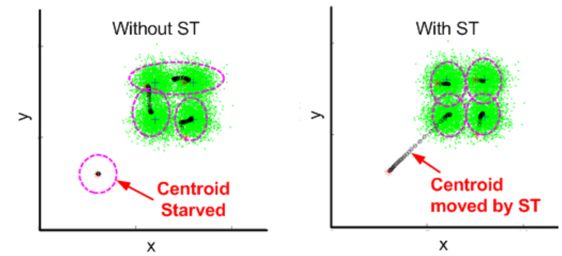

Figure 5. Clustering test results with bad initial condition without (left) and with the starvation trace enabled.

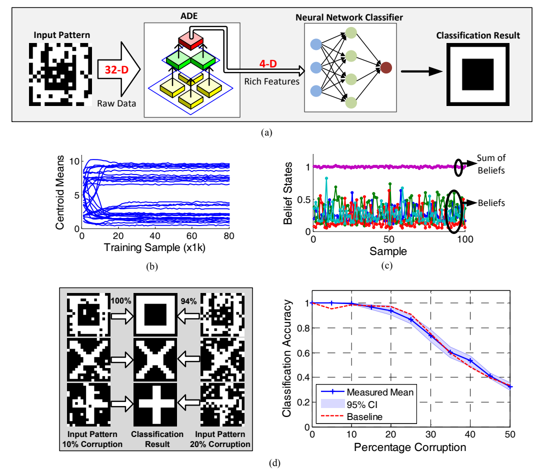

Figure 6. (a) Feature extraction test setup. (b) The convergence of centroids during training. (c) Rich features output from the top layer, showing the effectiveness of normalization. (d) Measured classification accuracy using the features extracted by the chip. The plot on the right shows the mean accuracy and 95% confidence interval (2□) from the three chips tested, compared to the software baseline.

from the top layer are used as rich features, achieving a dimension reduction from 32 to 4. A software neural network then classifies the reduced-dimension patterns. Three chips were tested and average recognition accuracies of 100% with pixel corruption level lower than 10% and 94% with 20% corruption are obtained, which is comparable to the floating-point software baseline, as shown in Figure 6(d), demonstrating robustness to the non-idealities of analog computation.

Efficiency Comparison The measured performance of the ADE is summarized in Table 1. It achieves an energy efficiency of 480 GOPS/W in training mode and 1.04 TOPS/W in recognition mode. Table 2 shows that, compared to other mixed-signal computational circuits, this work achieves very high energy efficiency in both modes.