Flatter Frequency Response from Feedback Delay Network Reverbs

Hans Anderson
Singapore University of Technology and Design
hans_anderson@mymail.sutd.edu.sg

Kin Wah Edward Lin
Singapore University of Technology and Design
edward_lin@mymail.sutd.edu.sg

Clifford So
Chinese University of Hong Kong
cliffso@cuhk.edu.hk

Simon Lui
Singapore University of Technology and Design
simon_lui@sutd.edu.sg

ABSTRACT

Increasing the size of a feedback delay network reverberator produces more uniform frequency response and higher density of echoes, thus improving the perceptual quality of the output. Feedback delay networks mix delay line outputs back to the inputs using a matrix-vector multiplication that requires $O(n^2)$ scalar multiplications in the general case or $O(n \log n)$ for some special matrices, where n is the number of delays in the network. Using the sparse matrix we present here, the mixing operaion can be done with $O(n)$ multiplications, which permits us to use larger size networks without increasing computational cost. We show that this produces an impulse response with a flatter average FFT spectrum and discuss its effect on echo density.

1. INTRODUCTION

Jot and Chaigne described a complete design of a feedback delay network (FDN) based reverberator with a late reverberation network composed of 16 delays [1]. Rochesso presented an implementation with 15 delays, choosing this size because his experiments showed that a minimum of 8 were necessary for acceptable sound quality [2]. Although the theory of feedback delay networks generalizes to networks of any size, we did not find any descriptions of implementations larger than size sixteen in the literature. In this work, we consider how best to make use of increasing computational resources as we scale the network up to larger sizes. We propose a network structure that circulates energy through a longer signal path, permitting us flexibility to control the amplitude and echo density independently for each output channel at specific times along the impulse response.

We evaluate our proposed design by measuring spectral uniformity and echo density, taking measurements on a lossless prototype version of the network that circulates energy indefinitely without damping. A well designed lossless prototype has maximally uniform spectral output [3, 4]. We construct a practical reverberator from the lossless prototype by adding damping filters. Although the reverberator with damping filters does not produce spectrally uniform output, it's important that the lossless prototype be spectrally neutral so that the output of the complete system will correspond only to the filters we designed into it and not be affected by unintentional colouration.

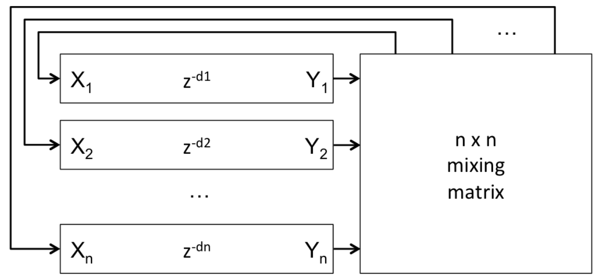

Figure 1. The basic structure of a feedback delay network. Audio input (not shown) is typically mixed in at the beginning of each delay line. Output taps may be inserted anywhere in the network.

The basic structure of a lossless feedback delay network prototype is shown in figure 1. If we let Y be a column vector that represents the output from each of the n delay lines and let X be the row vector representing the input to each of the delays then the mixing matrix M controls the mixing of feedback from the output of each delay to the input of all delays according to equation (1).

$$X = M \cdot Y \tag{1}$$

Rochesso [2] and Jot [5] recommend designing networks to produce the maximum possible density of echoes in the impulse response. Maximizing density is important for small networks because they naturally have sparser output. Examples of mixing matrices that maximise echo density include Hadamard matrices, Householder matrices [5] and circulant matrices [3]. Rochesso proposes a recursive algorithm, based on the efficient method of multiplying circulant matrices given in [3], that does maximally diffusive mixing in $O(n \log n)$ operations, where n is the network size. A fast Walsh-Hadamard transform equivalent to multiplication by a Hadamard matrix can also be done in $O(n \log n)$ time[6]. Our proposed method of mixing is intended for use in larger sized neworks; hence it does not maximize echo density.

2. THE BLOCK-CIRCULANT MIXING MATRIX

Figure 2 shows a 16×16 sparse block-circulant matrix whose only non-zero blocks are 4×4 Hadamard matrices located in the sub-diagonal positions and the upper-right corner. This structure generalises to any $n \times n$ matrix where n is a multiple of four.

The reason for using this structure is to avoid sending the output from any of the delay lines directly back to itself. In other words, we want to increase the time it takes for each individual echo to complete a circuit around the network. A sparse block-circulant matrix like this one increases the minimum round-trip time by a factor of $n/4$. This reduces the frequency of repetition for distinctly audible transients in the impulse response and retards the buildup of modal resonances.

The more slowly circulating nature of this structure also permits us to use strategic placement of delay taps to simulate time varying effects. For example, we can pan the outputs in some of the blocks right or left to produce an effect of reverberated sound rocking gently from side to side or reduce the number of output taps in some blocks to create pulsating density effects as the signal circulates from block to block. Griesinger recommends using effects like these to produce a more dramatic stereo image [7]. We may also use different amounts of damping on each block, combined with strategic positioning of output taps to produce various other non-uniform decay envelopes such as keeping the amplitude envelope at a constant level between 50 and 200 milliseconds, which Griesinger recommends as a strategy for increasing intelligibility of vocal sounds.

Another advantage of the sparse block-circulant mixing matrix is that the number of arithmetic operations required to compute the matrix-vector multiplication scales linearly proportional to n rather than n^2 or $n \log n$. This comes at the cost of reduced echo density, but as we will see in section 4, the cost may be offset by the ability to efficiently scale up the network.

The matrix in figure 2 can be generalized to use blocks of any size but we prefer size 4×4 because smaller blocks build echo density too slowly and larger sized blocks would limit our ability to lengthen the round-trip signal propagation time. As an interesting side note, the size four network proposed by Stautner and Puckette is isomorphic to the network using a 4×4 block-circulant mixing matrix composed of 2×2 blocks [8].

In our implementation, we do not directly multiply the delay line output vector by the matrix in figure 2. Instead, we use an equivalent network topology that is a series loop composed of several modules like the one shown in figure 3. With this modular design, we can easily scale the network size up or down at runtime to make use of available computing resources or modify the perceptual characteristics.

$$\begin{bmatrix} 0 & 0 & 0 & \boxed{H_{4x4}} \\ \boxed{H_{4x4}} & 0 & 0 & 0 \\ 0 & \boxed{H_{4x4}} & 0 & 0 \\ 0 & 0 & \boxed{H_{4x4}} & 0 \end{bmatrix}$$

Figure 2. A 16×16 sparse block-circulant matrix composed of 4×4 Hadamard matrices and zero blocks.

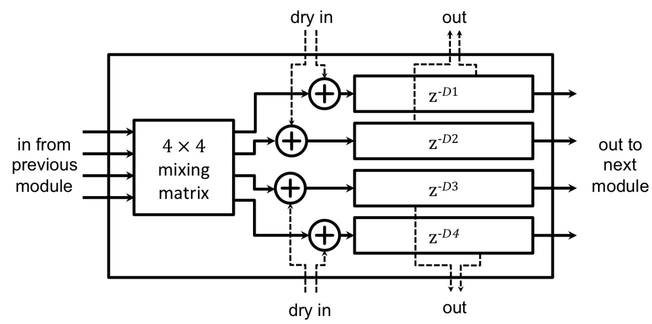

Figure 3. Connecting four of these modules together in series to form a loop produces a network structure equivalent to the one produced by a size sixteen network with a mixing matrix like the one in figure 2.

3. SPECTRAL FLATNESS AND ECHO DENSITY

To measure the spectral flatness, we use the following steps:

1. using a sliding window of 20 ms length on the first ten seconds of the impulse response, take the magnitude of the FFT of each window

2. calculate the mean energy for each FFT bin across all the windows

3. take the standard deviation of the result in step 2 divided by the mean of the result in step 2

Measurements of both spectral flatness and echo density are sensitive to changes in the average delay line length of the network. In order to make comparisons between networks of different sizes, we use the same minimum delay time of 7 ms, and maximum of 100 ms on all our reverb networks, with delay times between those limits randomised but uniformly distributed using the method proposed by Rubak [9].

3.1 Measuring echo density

Schroeder recommends that to achieve a flutter free sound, the reverb should produce a density of 1000 echoes per second [10]. According to Griesinger, the required density for late reverb is closer to 10,000 but in the first few milliseconds

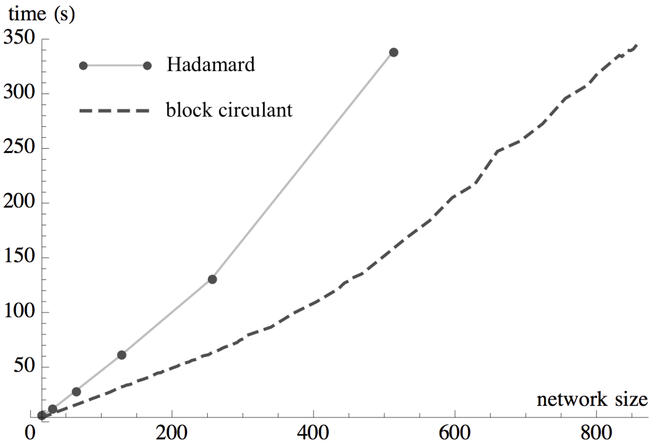

Figure 4. This graph shows the time required for various sizes of networks to do faster-than-real-time processing of 500 seconds of reverb. For Hadamard mixing networks, only networks of size $2^k, k \in \mathbb{Z}$ are shown because our fast Hadamard transform only works for powers of two. Sparse block circulant mixing works for all sizes that are multiples of four, hence the number of measurements is too large to show individual data points.

Fast Hadamard	16	32	64	128	256	512
Block circulant	20	52	116	248	456	848

Table 1. The top row shows network sizes for which we can use the fast Hadamard transform to do maximally diffusive mixing in $O(n \log n)$ time. Below, we list sparse block-circulant network sizes that run on our test computer with similar processing time to the Hadamard networks shown above.

it should be sparse. To measure the density of echoes, we use the following method taken from Griesinger [7]:

1. If the reverb decays exponentially, multiply the impulse response by an exponentially growing function to make the amplitude envelope constant. This is not necessary for lossless prototype networks.

2. Partition the impulse response using a sliding window, 20 ms in length.

3. Count the number of samples within 20db of the max amplitude sample.

4. RESULTS

One advantage of mixing with a sparse block-circulant matrix is that the number of arithmetic operations scales linearly as network size increases. For a given computational power limit, sparse mixing allows us to use a larger network size. Replacing a Hadamard-mixing network with a sparse block circulant network of equivalent speed lengthens the time for energy to circulate but it also retards the buildup of echo density. In order to make comparisons of the spectral flatness and echo density between the two types of networks, we identify pairs of networks that have similar computational power requirements. These pairs are shown in table 1.

As mentioned in section 1, there are several methods for performing the mixing operation on a maximally diffusive network. The fastest known algorithms scale in $O(n)$ time. For all of our measurements on maximally diffusive networks, we use the fast Walsh-Hadamard transform [6] to do the mixing.

To estimate the computational power requirements of each network, we measured the time to process a 500 second impulse response. All of the networks we tested, going up to a size of 848, completed the work faster than real-time, running on a Macintosh computer with a 2.7 GHz Intel Core i7 processor. Our implementation is written in c++ using only the standard libraries available in gcc and does not include any machine-specific commands.

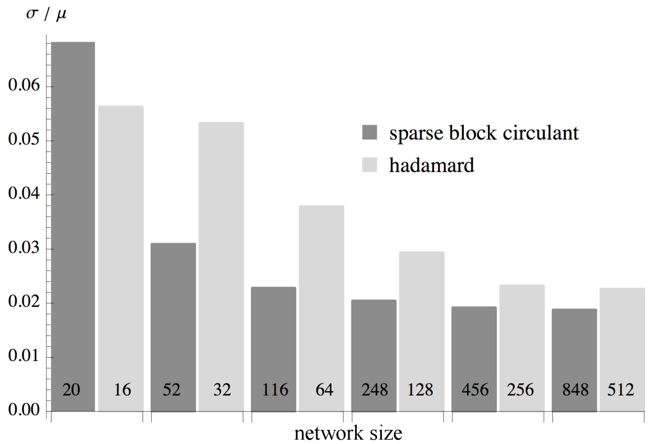

Figure 5. (Standard deviation / mean) of average FFT magnitude spectrum taken in 20 millisecond sliding windows on the impulse response. Lower values of σ/μ indicate flatter average magnitude spectrum. Adjacent bars on the graph are grouped into pairs, where each pair is a set of Hadamard and block-circulant mixed networks with unequal sizes but similar computational time requirements. The numbers at the bottom of each bar are the network sizes.

4.1 Spectral flatness

Figure 5 shows the (standard deviation / mean) of the average FFT magnitude spectrum, measured by the method described in section 3. Lower values of standard deviation indicate flatter frequency response.

When replacing a maximally diffusive network with a sparse block-circulant one of similar speed, the biggest gains in spectral uniformity occur when the original maximally diffusive network is size 32 or 64. This is partly due to the fact that the differences in performance between the two types of networks, shown in figure 4 are more pronounced for those sizes. This is probably a machine and platform-dependent effect related to the memory management and caching system.

4.2 Echo density

Figure 6 shows how echo density builds over time in several reverbs. Unlike spectral flatness, where we can generally say that flatter is better, we don't actually want to maximise echo density. Natural reverb starts with sparse early echoes and builds over time to a saturation point limited by the bandwidth of the recording [7]. Figure 6 shows that the echo density of a

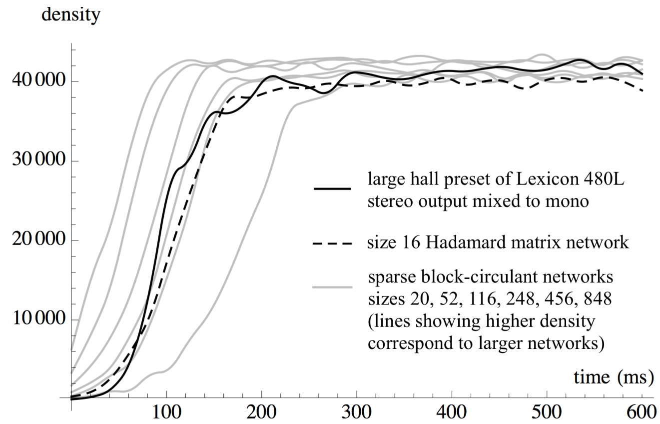

Figure 6. The echo density of sparsely mixed networks is lower than that of densely mixed networks of equal size. However, for sparse block circulant networks of sizes 52 and higher, echo density meets or exceeds both the size sixteen Hadamard network and the Large Hall preset of the Lexicon 480L. The sparse block-circulant network of size 52 has similar processor power requirements to the Hadamard network of size 32.

size sixteen maximally diffusive network is also very close to that of the Large Hall preset on the Lexicon 480L. Since size sixteen is also the largest size frequently mentioned in FDN literature, we use it as a benchmark representing a sufficient echo density level.

Since echo density has an upper limit, we can not maximise it beyond what the bandwidth of our audio system allows. Once a reverberator exceeds the desired echo density for its intended application, it is reasonable to sacrifice some of that density to acheive other design goals. Figure 6 shows that for designs where we would consider using a densely mixed network of size 32 or more, we may switch to a sparse block-circulant network of equal speed without being concerned about the reduced echo density.

Regarding the potential for building networks of very large size, figure 6 shows that even with the reduced echo density of block-circulant mixing we may run into problems with having too much density in the early part of the impulse response. In applications where we need to use such large networks, echo density buildup can be limited by increasing the length of the delays, by removing the output taps from some of the blocks or by not mixing input into all of the blocks.

5. FUTURE WORK

Most FDN reverbs in the literature are of size sixteen or less. Since computational power is more abundant now than when those papers were written, we recommend further study on the applications of very large networks. Block circulant mixing is especially interesting in this context because it parallelizes easily. Since graphics processing hardware is optimised for parallel computation and 4×4 matrix-vector multiplications, it may be useful to implement very large block-circulant networks on GPU hardware.

6. CONCLUSIONS

Sparse block-circulant mixing offers a trade-off of reduced echo density in exchange for flatter frequency response, longer circulation time, and more flexible design of amplitude envelope. When we consider echo density requirements, replacing a densely mixed network with a sparse one is an attractive option for densely mixed networks of size 32 or more.

7. ACKNOWLEDGMENTS

This work is supported by the SUTD-MIT International design center (IDC) Research Grant (IDG31200107 / IDD11200105 / IDD61200103).

8. REFERENCES

[1] Jot, Jean-Marc, and Antoine Chaigne. “Digital delay networks for designing artificial reverberators. *Audio Engineering Society Convention 90. Audio Engineering Society*, 1991.

[2] Rocchesso, Davide. “Maximally diffusive yet efficient feedback delay networks for artificial reverberation. *Signal Processing Letters, IEEE* 4.9 (1997): 252-255.

[3] Rocchesso, D., Smith, J.O., “Circulant and Elliptic Feedback Delay Networks for Artificial Reverberation, *IEEE Trans. Speech and Audio Proc*, 5(1):51-63, 1997.

[4] Dahl, Luke, and Jean-Marc Jot. “A reverberator based on absorbent all-pass filters. *DAFx-00: Proceedings*. 2000.

[5] Jot, J.M, “Study and Implementation of a Sound Spatializer Through Physical and Perceptual Models, *PhD Thesis, Telecom Paris*, 1992.

[6] Fino, Bernard J., and V. Ralph Algazi. “Unified matrix treatment of the fast Walsh-Hadamard transform. *IEEE Transactions on Computers* 25.11 (1976): 1142-1146.

[7] Griesinger, David. “Practical processors and programs for digital reverberation. *Audio Engineering Society Conference: 7th International Conference: Audio in Digital Times*. Audio Engineering Society, 1989.

[8] Stautner, John, and Miller Puckette. “Designing multichannel reverberators. *Computer Music Journal* (1982): 52-65.

[9] P. Rubak and L. G. Johansen, “Artificial reverberation based on a pseudo-random impulse response, *AES 104th Convention*, Preprint 4725, (Amsterdam), 1998.

[10] Schroeder, Manfred R. “Natural sounding artificial reverberation. *Journal of the Audio Engineering Society* 10.3 (1962): 219-223.