Multirate Additive Synthesis

Desmond PHILLIPS *
d.k.phillips@lboro.ac.uk

Alan PURVIS
alan.purvis@dur.ac.uk

Simon JOHNSON
simon.johnson@dur.ac.uk

Durham Music Technology *

School of Engineering, University of Durham, Science Laboratories,
South Road, Durham City, County Durham, DH1 3LE, UK.

Abstract

The advantages of musical signal modelling in frequency domain via sinusoidal additive synthesis are well known, as is the prime disadvantage of high computational overhead. This paper proposes a reduction in overhead by the application of standard multirate DSP techniques to optimise the sample rate of oscillators, with a view to VLSI implementation. The optimisation works by adapting computation to oscillator frequency ranges, which are often known *a priori* for partials in note-based music, and places few restrictions on functionality. The cost of signal interpolation up to an industry-standard sample rate is minimised by grouping oscillators, spatially related in the final sound image, to common synthesis QMF filterbanks which have a hierarchical sub-band decomposition that balances interpolation cost with goodness-of-fit to expected oscillator frequency ranges. Finally, an overview is provided of the mapping to a hypothetical "single chip" coprocessor.

1.   Introduction

Additive Synthesis (AS) is a low-level transformation that underpins all spectral modelling paradigms [Serra & Smith, 1990]. The potential wide application and simplicity of AS make VLSI an attractive solution. Intriguingly, a problem is *over-generality* in the context of musical signals which are computed with substantial redundancy. However, optimising AS to assumed signal properties (e.g. harmonicity) restricts the application of AS to that particular signal class, compromising generality and making VLSI less attractive. To this end, a fresh solution is proposed - Multirate Additive Synthesis (MAS) - that has a graduated trade-off between generality and cost.

2.   Background

AS computes the Inverse Fourier Transform (IFT) of spectra comprising discrete lines, which is a property of tonal sounds such as musical timbres. In eqn. (1) $y[n]$ is the sum of $1 \leq i \leq N$ sinusoids, each modulated by individual amplitude and frequency envelopes $A_i[n]$ and $F_i[n]$; n is a time index at a sample rate of $f_s \geq 40\text{kHz}$. By this means, the spectral evolution of any tonal sound can be described over time.

$$y[n] = \sum_{i=1}^{N} A_i[n]\sin(\Phi_i[n])$$
$$\Phi_i[n] = \Phi_i[n-1] + 2\pi F_i[n]/f_s \qquad (1)$$

The prime disadvantage is that for N oscillators, $2N$ envelope data streams are required at f_s. For example, a 100-voice ensemble with 40 sinusoids per voice (to ensure synthesis quality) at f_s=44.1kHz requires a net oscillator update rate of 176.4MHz and an envelope data bandwidth of 352.8Mhz. Fortunately, envelope data may be short-time averaged with little perceptible loss of quality using a piecewise linear approximation described by a compact breakpoint set. This technique is reported as achieving data compression ratios of 100:1 and is central to AS and many spectral modelling paradigms [Serra & Smith, 1990].

However, for playback, uncompression is required in real-time for AS via an oscillator bank. Control data bandwidths of the magnitude discussed previously recur and swamp the throughput of CPUs optimised for the execution of high complexity instruction streams. AS is a fine-grain data-parallel algorithm and a traditional "form follows function" solution is to map eqn. (1) into a direct-form dedicated coprocessor exploiting deep pipelining, thus freeing CPU throughput for high-level tasks more suited to its design. A recently proposed alternative is to simulate an oscillator bank in software via the Inverse Fast Fourier Transform (IFFT) in an overlap-add scheme [Rodet & Depalle, 1992]. The algorithm is characterised by the data conditioning required to generate smooth, linear inter-frame changes in $A_i[n]$, $F_i[n]$ [Goodwin & Kogon, 1995].

---

3. Multirate Additive Synthesis

Direct-form oscillator banks have lost research momentum because they inherit the over-generality of AS, but a consideration of multirate DSP techniques permits their revision. In music synthesis, a chief requirement is the synthesis of notes which have properties of pitch, finite lifespan and timbres definable as an AS partial series. Assuming, for simplicity, that partial frequency ratios are pitch-invariant, then the upper and lower frequency bounds $\{f_{min}(x), f_{max}(x)\}$ of each partial x are related by the constant $\alpha = f_{max}(x)/f_{min}(x)$ which represents the frequency modulation range required by the note over its lifecycle for e.g. vibrato. If the note is subject to real-time modulation (e.g. pitch bend) α is chosen to provide sufficient headroom.

MAS exploits a reinterpretation of Nyquist's sampling theorem in that the critical time-invariant sample rate for alias-free synthesis of x is $f_{opt}=2(f_{max}(x)-f_{min}(x))$ rather than the uniform rate of f_s in eqn (1). However, x as synthesised by a digital sine oscillator executing at f_{opt} is a baseband signal that spans from DC to $(f_{max}(x)-f_{min}(x))$. Sidebands appear around harmonics of f_{opt} within the audio spectrum representing audible quantisation noise which must be eliminated by interpolation up to f_s. x is then heterodyned by $f_{min}(x)$ to its desired spectral location, bounded by $\{f_{min}(x), f_{max}(x)\}$. For a single oscillator, this procedure is extravagant, but since AS sums N sinusoids into a single stream, a single multirate synthesis filterbank can be applied to a sum of oscillators with identical $\{f_{min}(x), f_{max}(x)\}$. A problem for filterbank design is that $\{f_{min}(x), f_{max}(x)\}$ is unique to each partial x and therefore N incommensurate sample rates are generated, each requiring separate interpolation and creating a scheduling nightmare. At the other extreme, a uniform f_s in eqn. (1) is shown to have computational redundancy. A compromise is to employ a filterbank such as the QMF that has a finite sub-band set with a structure that matches expected values of $\{f_{min}(x), f_{max}(x)\}$.

MAS reduces the update rate of an oscillator and also the control bandwidth of piecewise linear envelope uncompression. Potential savings are indicated by the ratio f_s/f_{opt}. If a note is fixed pitch (e.g. from a keyboard instrument) then α is small and large savings are made. However, savings diminish as α increases passing through for example, from flute vibrato to trombone glissando. A factor that places a lower bound on f_{opt} is Heisenberg's time / frequency inequality, which in MAS terms, states that oscillator control bandwidth is related to operating frequency. High frequency attack transients require a greater peak control rate than lower frequency partials (AS via the IFFT imposes a frequency-invariant rate).
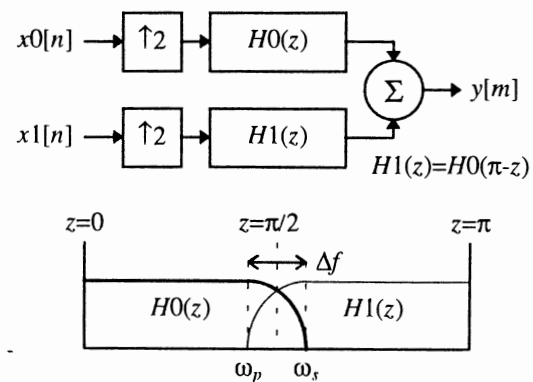
4. QMF Filterbanks

$$H1(z)=H0(\pi-z)$$

Figure 1. General Model of a QMF Stage

Synthesis Quadrature Mirror Filterbanks (QMF's) are constructed from a prototype stage illustrated in Figure 1. Output $y[m]$ comprises the sum of input signals $x0[n]$, $x1[n]$ applied, respectively, to half-band interpolation filters $H0(z)$ (low-pass) and $H1(z)$ (high-pass) which have mutual symmetry about $z=\pi/2$. This property permits efficient implementation via FIR, or allpass IIR, polyphase structures. Non-ideal filters have passband and stopband ripple (δ_p, δ_s), and most significantly for MAS, a non-infinitesimal transition region between passband and stopband $\Delta f=(\omega_s-\omega_p)/2\pi$. δ_s reduces quantisation noise in the audio band for high-fidelity synthesis e.g. δ_s=-80dB.
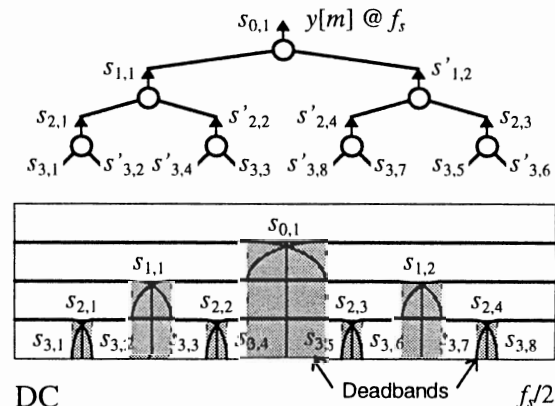

Figure 2. Sub-band Hierarchy for K=3

A QMF stage is a two-input node which divides the output band into two sub-bands. A binary tree filterbank of arbitrary completeness may be constructed from QMF stages to provide a recursive sub-band decomposition. For MAS, consider a complete tree of depth K, with levels $k=0..K$, and at each level k, an integer series of sub-bands denoted by $s_{k,l}$ where $l=1..2^k$. The result is a *sub-band hierarchy* illustrated in Figure 2. for the case of $K=3$. $H1$ causes a frequency inversion in each $x1$ sub-band such that the mapping of filterbank to hierarchy is indirect: $s'_{k,l}$ indicates that the sub-band requires pre-

inversion during normalisation of $F_x[n]$ to the associated sub-band such that x appears in $y[m]$ correctly modulated by $F_x[n]$.

The hierarchy has many desirable properties. Signals comprising summed oscillator sets allocated to each $s_{k,l}$ are introduced mid-filterbank as well as at the terminal sub-band series at level K (used for conventional analysis / synthesis applications) and thus intermediate sub-bands are interpolated 'for free'. For oscillator allocation, if x is not accommodated by level K - because sub-bands are too narrow or the range of x spans two adjacent sub-bands - then each level K-1..1 is tested in turn. Failing that, any $\{f_{min}(x), f_{max}(x)\}$ may be accommodated at $k=0$ which represents classical AS. The integer-spaced series $s_{K,1..2}{}^K$ facilitates efficient computation of fixed pitch notes with low α, whereas increasing α causes a partial series to be promoted ultimately into the octave-spaced series $s_{0..K,1}$. A constraint is that x may not be allocated into a transition region because it is attenuated and amplitude becomes, undesirably, a function of frequency. This results in 'dead-bands' in the hierarchy, which are inherited by sibling QMF stages because of cascaded filtration. Dead-bands are 'logically excluded' by a simple revision of the bounds of each $s_{k,l}$. If $\Delta f << 2^{-K}$, then dead-band influence on the hierarchy is slight.

MAS is efficient when net savings in oscillator computation exceed filterbank costs. The ratio in computation r between AS and MAS - interpretable as 'speedup' - is summarised in eqn. (2) where n_{fb} is the number of filterbanks, c_{qmf} is the cost of a single filterbank (due to multirate operation, $c_{qmf} \propto K$ for a complete QMF tree), c_{mas} and c_{as} are the costs of, respectively, MAS and AS oscillator updates and $n_{k,l}$ is the total number of oscillators allocated across sub-band $s_{k,l}$ of all n_{fb} filterbanks. The 2^{-k} term represents the sample rate reduction of allocation to level k in MAS relative to AS and also facilitates $c_{mas} \cong c_{as}$. n_{fb} filterbanks are necessary to provide sufficient streams for building up the subsequent sound-image via a DSP post-processor: for instance in locating orchestral sections within an acoustic volume.

$$r = \frac{c_{as} \sum_{k=0}^{K} \sum_{l=0}^{2^k} n_{k,l}}{n_{fb} c_{qmf} + c_{mas} \sum_{k=0}^{K} 2^{-k} \sum_{l=0}^{2^k} n_{k,l}} \quad (2)$$

Clearly, r is dependent upon the statistics of $n_{k,l}$. The largest proportion of energy in an audio signal is concentrated below about 5kHz, and that above is perceived as treble 'detail'. This infers that most oscillators in MAS will be allocated below this threshold. Low-Q sub-bands above (e.g. $s_{3,5..8}$) are likely to contain few oscillators and be uneconomic to

compute, particularly those badly affected by dead-bands. A solution is to prune filterbanks, in the light of $n_{k,l}$, so that the greatest concentration of sub-bands is in the lower audio frequencies where economisation potential is greatest.

5.   Filterbank Design Issues

If FIR QMF's are used, then the latency of each stage is a constant M samples, where $M \propto 1/\Delta f$. Latency from level k is $M(2^k-1)/f_s$, and to ensure that envelope features between partials allocated in different levels are time-aligned, delay-lines of length $(2^{K-k}-1)M$ are introduced to each $s_{k,l}$ oscillator set to normalise latency to the maximum from level K. The relationship $c_{qmf} \propto 1/\Delta f$ implies that reducing dead-bands (permitting greater optimality in oscillator allocation) increases filterbank costs. For a given $n_{k,l}$, r has a global maximum against Δf which determines optimum Δf, though the requirement for imperceptible real-time latency of <10ms (dominated by coprocessor delays) puts an upper bound on K and favours high Δf. In respect of the proposed coprocessor, $K=3$, $0.05 < \Delta f < 0.1$ seem optimal.

Polyphase all-pass IIR QMF's are characterised by narrow Δf's in comparison to FIR's of the same complexity, and are the best solution for minimising dead-bands [Hart et al, 1993]. A disadvantage is phase non-linearity which precludes latency normalisation with delay-lines. Indeed, a wider problem with the QMF structure is that $H1$'s pass only sidebands of x requiring non-trivial control of baseband phase to give x desired phase at $y[m]$. However, sub-bands $s_{0..K,1}$ comprise a pure cascade of $H0$'s which preserve baseband phase in an FIR implementation, enabling the maintainance of partial phase relationships in the lower audio frequencies.

The QMF MAS strategy relies upon logical exclusion of Δf. Another interesting solution is 'physical exclusion' with an oversampled complex-signalled filterbank. The first MAS proposal employed the technique, without hierarchy [Phillips et al, 1994], but the principle is easily extended to a perfect dead-band free hierarchy. Due to oversampling, and the overheads of complex signal processing, it is more expensive than the QMF approach, but as only $H0$-style filters are used, complete phase transparency from all sub-bands is facilitated, if at all required.

6.   A MAS Coprocessor

A VLSI MAS coprocessor therefore comprises two processing systems: (i) a multirate oscillator bank (MOB) and (ii) a filterbank processor (FBP). Figure 3. illustrates the architecture under consideration, which uses a memory hierarchy external to the host to contain AS control data proliferation. It is a multiprocessor system using a shared memory (SM)

model of interprocessor communication which is single-port permitting the MOB and FBP to share a multiplexed bus. Though the architecture is applicable to AS, MAS minimises the control bandwidth taken by the critical MOB-SM bottleneck. With eqn. (2) expressed in terms of SM bandwidth, tentative simulations of orchestral synthesis with real music, arrangements and timbres [Sandell, 1994] determine a baseline speedup of about $r\cong3$. The maximum number of oscillators envisaged is $N=rf_{mob}/f_s$ where f_{mob} is the MOB clock rate. For example, with f_{mob}=50MHz and f_s=44.1kHz, $N\cong3400$.
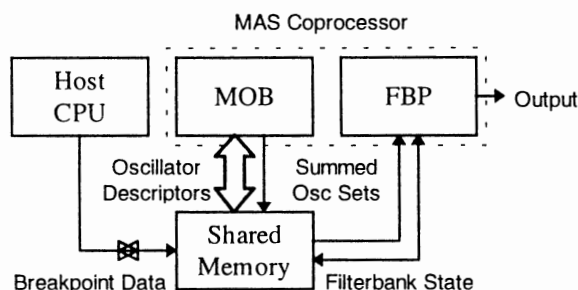
MAS Coprocessor

Figure 3. MAS Coprocessor

The SM is memory mapped into the host address space so that the host may write piecewise linear breakpoint data directly into oscillator descriptors (SM size is $2^{18}..2^{20}$ words for sufficient descriptor size and number). Each descriptor comprises oscillator state variables and breakpoint caches for $A_i[n]$ and $F_i[n]$. The MOB sequentially fetches, updates and writes descriptors. The SM bandwidth required is very high, but by quantising breakpoints to a multiple of oscillator sample rate (c.f. IFFT AS synthesis), oscillators are updated in bursts (determined as b=12 cycles for a 32-bit SM) and interleaved with SM fetch / write-back such that the MOB operates at maximum throughput. $\sin(\Phi_i[n])$ is computed by linear-interpolation: a pipelinable algorithm requiring a small ROM and multiplier.

Descriptors are stored in a linked list to facilitate resource allocation, and enable descriptors in a set allocated to a particular sub-band to be contiguous. This permits a concurrent burst accumulation of the set which is buffered on completion via SM to the FBP. Sub-band sets associated with a particular filterbank are also required to be contiguous to facilitate scheduling, which is driven by the FBP executing filterbanks 'round-robin' at a fixed frame rate with a period of $b/(f_s 2^{-K})$. QMF topology is host-configurable and is executed depth-first by the FBP, with caching of intermediate results before generation of the final output to a DSP post-processor. QMF state is saved via SM to minimise on-chip RAM.

The MOB is loosely coupled to the FBP, but is synchronised at frame level using a 'busy-waiting' algorithm to ensure mutually exclusive access to the SM buffers. Multirate operation is implemented by extending MOB bursts to $b2^{K-k}$ cycles to generate the higher sample rates for levels $k=K-1..0$. The MOB has lowest priority access to the SM as it may operate with a degree of asynchronicity to the FBP. It can be blocked by the host and FBP which require much a lower SM bandwidth. Predicted MOB throughput is not seriously degraded as a result. The FBP has a hard real-time schedule and therefore has priority over the host as bus conflicts are infrequent. Mutually exclusive host / MOB descriptor access is ensured via an MOB-driven interrupt to the host making worst-case system latency from e.g. host MIDI event to FBP output as 3 frames + filterbank latency.

7.   Summary

The proposed MAS algorithm using a QMF sub-band hierarchy allied to an optimised coprocessor offers a substantial performance improvement over a traditional oscillator bank, without compromising the fundamental generality of AS. However, further research is required in two areas: (i) a methodical investigation of the statistics of eqn. (2) to refine aspects of the MAS algorithm, and (ii) a coprocessor simulation to verify its design and identify areas for further optimisation. It is hoped that MAS is a contribution of interest to the ongoing debate about AS and spectral modelling.

References

[Goodwin & Kogon, 1995], M. Goodwin & A. Kogon, "Overlap-Add Synthesis of Nonstationary Sinusoids", *Proc. International Comp. Music Conf.*, pp. 355-356, 1995.

[Hart *et al*, 1993], J. Hart, P. Naylor & O. Tanrikulu, "Polyphase Allpass Structures for Sub-band Acoustic Echo Cancellation", EUROSPEECH'93, Berlin, Germany, Vol. 3, pp. 1813-1816.

[Phillips *et al*, 1994] D. Phillips, A. Purvis & S. Johnson, "A Multirate Optimisation for Real-Time Additive Synthesis", *Proc. International Comp. Music Conf.*, pp. 364-367, 1994.

[Rodet & Depalle, 1992], X. Rodet and P. Depalle, "A new additive synthesis method using inverse Fourier transform and spectral envelopes", *Proc. International Comp. Music Conf.*, pp. 410-411, 1992.

[Sandell, 1994], G. J. Sandell, "SHARC Timbre Database", release 0.90, Sussex University, UK, November 1994.

[Serra & Smith, 1990], X. Serra & J. Smith, "Spectral Modelling Synthesis: A Sound Analysis System Based on a Deterministic plus Stochastic Decomposition", *Comp. Music Jnl.*, vol. 14, no. 4, pp. 12-24, 1990.