3. Multirate Additive Synthesis

3.1 Overview

As concluded in section 1.4.3, an analysis of AS computation in the context of note-based music leads to the notion of computational optimisation by Multirate Additive Synthesis (MAS). To recap, the two chief areas of computational burden in MAS are

1. Interpolating S decimated sinusoids with independent sample rates up to f_s
2. The real-time scheduling of S sample rates (in all probability incommensurate).

The aim of this chapter is to develop the theoretical basis for a MAS algorithm by applying classical multirate concepts to the problems posed by (1) and (2). The conclusions of Chapter 1 are accepted *per se* (especially sections 1.2, 1.3.2 and 1.4.3) and form a starting point or 'requirements specification' to the task in hand. It transpires that an efficient solution for (1) leads to a great simplification in (2), which is finally solved in the proposed MAS Coprocessor (MASC) design of Chapter 8: discussion in this chapter is thus confined chiefly to topic (1).

Chapter 2 concluded that FFT^{-1} (Rodet and Depalle, 1992) is the contemporary 'state-of-the-art' in AS implementation because it did not compromise the goals of section 1.2 and is software-oriented. The well-spring for this chapter comes from the common knowledge that FFT / IFFT and multirate approaches are complementary methods for DSP optimisation and that, as yet, the latter has not been exploited systematically for AS optimisation, though there is latent interest (Haken, 1991). However, musical applications of multirate DSP outside the sphere of AS are discussed by Holm (1994) and Levine (1996a, 1996b). The fundamental premise for this chapter is that, in contrast to frequency-domain FFT / IFFT approaches, multirate techniques are time-domain and hence more compatible with a TOB architectural model; the optimal computational model for eqn. (1.1) as discussed in section 1.3.2.

3.1.1 Interpolation and Decimation in Multirate DSP

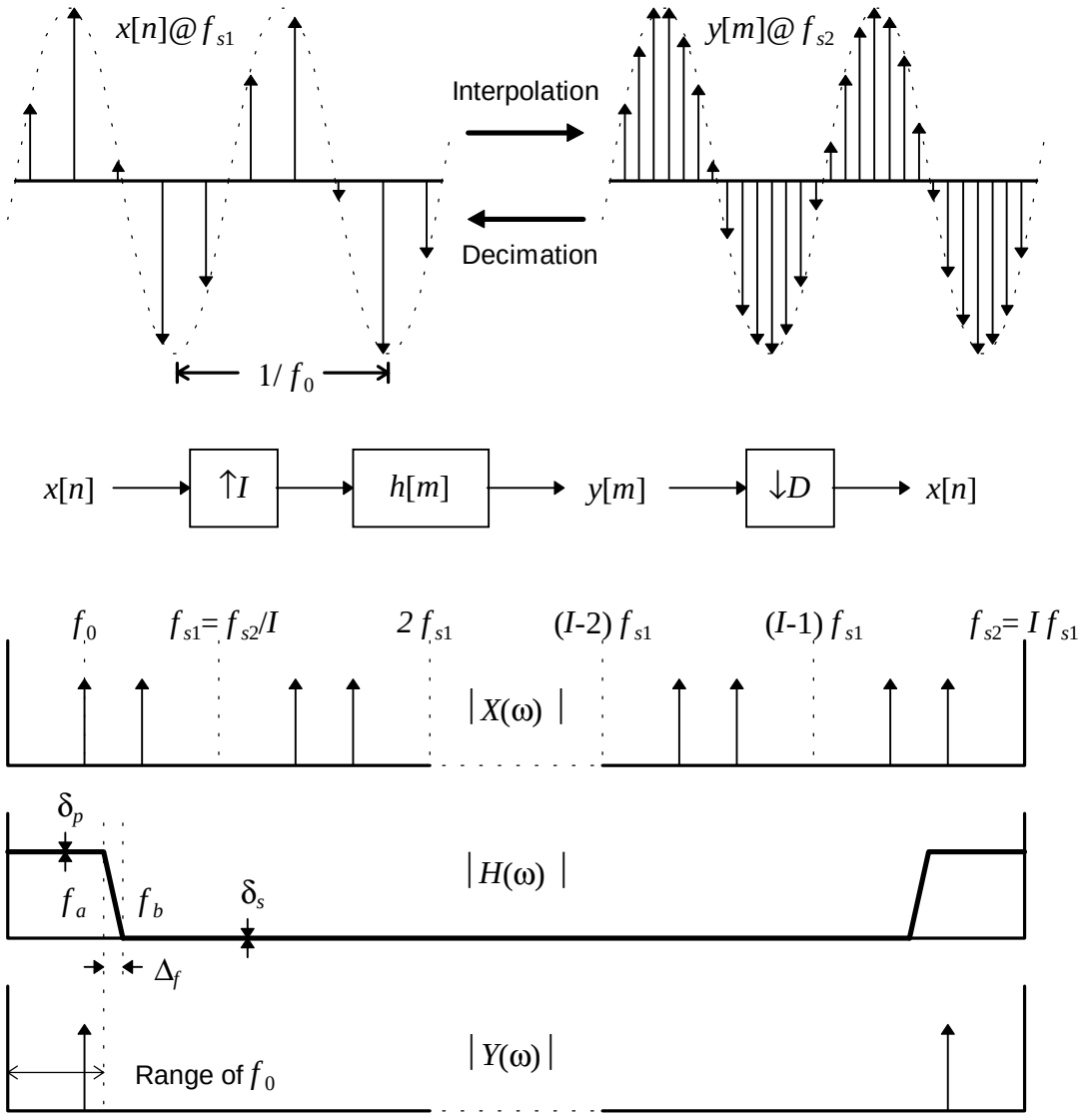

Figure 3.1 Interpolation and Decimation of Sinusoids

A definition of multirate DSP is the simultaneous processing of signals having sample rates (likely to be different) which are optimised to the signal bandwidths in question (Crochiere and Rabiner, 1983), (Vaidyanathan, 1993). A method enabling sample rate changes, without affecting signal content, is required. Two basic operations are therefore fundamental to multirate DSP: *interpolation* is an increase in sample rate by a factor *I* and *decimation* is a decrease by a factor *D*. Fig. 3.1 illustrates how these operations relate to a sinusoid in both time and frequency domain, since this is relevant to AS. A

stationary sinusoid $x[n]$ at frequency f_0 and sample rate of f_{s1} is interpolated by I (integer valued) to $y[m]$ at a new rate of $f_{s2}=I f_{s1}$, which is then decimated by $D=I$ to recover $x[n]$ showing that decimation is the inverse of interpolation.

Conceptually, the change from f_{s1} to f_{s2} is achieved by inserting I-1 zero-valued samples between each sample of $x[n]$. This has no effect on the signal spectrum in that images of the signal baseband (i.e. between DC and $f_{s1}/2$) occur about harmonics of f_{s1}. However, the frequency response images of a filter ($h[m]$) executed at f_{s2} occur about harmonics of f_{s2}. $H(\omega)$ is a low-pass filter which passes only the baseband component of $X(\omega)$ (self-evidently with those about harmonics of f_{s2}) and suppresses all intermediate components about harmonics of f_{s1} in order to generate $Y(\omega)$. These components constitute noise due to time quantisation. If f_{s2}>40kHz, then audible components up to $f_{s2}/2$ are removed leaving a pure audio frequency sinusoid in the baseband. Relevant to MAS is the observation that, in an AS oscillator bank, the decimated sinusoid $x[n]$ has a computational ratio of 1:I in comparison to its oversampled equivalent $y[m]$.

For $H(\omega)$ to have an economic implementation, a non-ideal design must be used which is parameterised by three tolerance variables determining performance; δ_s and δ_p set upper bounds on, respectively, passband and stopband ripple (gain relative to the ideal passband gain at 0dB) and f_a and f_b mark a finite transition region from stopband to passband where the frequency response gradually rolls off, often quoted as $\Delta_f=(f_b- f_a)/f_{s2}$. The relationship $f_a=f_{s1}-f_b$ places Δ_f symmetrically over the Nyquist frequency at f_{s1} of $f_{s1}/2$ such the stopband of $H(\omega)$ exactly coincides with range of f_0 harmonics requiring suppression. For high-quality interpolation with a flat frequency response, the range of f_0 is restricted to $0\leq f_0\leq f_a$ i.e. the passband of $H(\omega)$. It is undesirable for f_0 to enter Δ_f because (i) it is attenuated and (ii) a mirror-image harmonic in $f_{s1}/2$ is simultaneously boosted, with a risk of becoming perceptible. The range of f_0 in $x[n]$ is therefore oversampled by $f_{s1}/2f_b$ compared to an ideal H(ω).

I must be an integer for $h[m]$ to be a time-invariant filter. Conversion of non-integer ratios of f_{s2}/f_{s1} requires time-varying filtration with a resampling of the coefficients in $h[m]$ from a continuous time impulse response each sample period. If f_{s2}/f_{s1} is rational, in that they are synchronous and originate from the same clock, then a finite cycle of

coefficients of length $f_{s1}\, f_{s2}/c^2$ suffices, where c is the highest common factor of f_{s1} and f_{s2}. This can create a large cycle of coefficients with high storage demands making resampling from a high resolution look-up table more economic. An alternative is interpolation by I followed by decimation D such that $I/D = f_{s2}/f_s$: interpolation can use an efficient polyphase structure, outlined in the next section, with decimation implemented by commutating, modulo-D fashion, around the sub-filter outputs. An irrational ratio of f_{s2}/f_{s1} requires a complete recomputation of coefficients each sample period. This situation occurs when f_{s1} and f_{s2} are asynchronous and sourced from different clocks when linking equipment with incommensurate sample rates (e.g. CD at 44.1kHz and DAT at 48kHz), requiring a specialised convertor (Adams and Kwan, 1994).

3.1.2 Efficient Interpolator Design

There are two efficient alternatives for implementing interpolators with integer values of I that are pertinent to outline for future discussion, though the reader is referred to the literature for an in-depth analysis (Vaidyanathan, 1993). The first is the *multistage interpolator* which splits the interpolation process into its prime factors I_1, I_2.. I_N such that $I = I_1 \times I_2 .. \times I_N$. Each stage is unique, with passband and stopbands adjusted to correspond precisely to sidebands in the signals that require elimination. Moving up the cascade, relative transition widths increase as the signal becomes increasingly oversampled. Transition regions are of maximum width, and computation is minimised in that high-cost sharp roll-off filters are executed at a lower sampling frequency than low-cost gradual roll-off filters. The disadvantage of multistage interpolation is that it is limited to values of I that decompose into a convenient set of prime factors.

An alternative to the multistage interpolator is a single-stage 'polyphase' interpolator which functions for any integer value of I. The output $y[m]$ is taken by commutating between I 'subfilters' h_i ($0 \le i \le I-1$) each sample period of $x[n]$ where each h_i is, conceptually, an allpass filter introducing a pure phase delay of $2\pi i/I$. Only the subfilter connected to the commutator 'rotor' is computed each sample period of $y[m]$. The cost of computation reaches an asymptotic limit with increasing I because the complexity increase is accommodated by a proliferation of sub-filters with the order of each $H_i(z^I)$ remaining constant for a given (δ_p, δ_s). If the prototype interpolation filter $H(z)$ is FIR,

$H_i(z^I)$ ($0 \le i \le I$-1) corresponds to the modulo-I activation of coefficients caused by samples of $x[n]$ shifted into $H(z)$, separated by I-1 zero-valued samples

3.1.3 A General Model for MAS using Subband Decomposition

Consider the interpolation scheme of Fig. 3.1 modified such that we may exploit the computational savings implied by eqn. (1.4) by synthesising x in a baseband spanning DC to f_a where $f_{s1} \cong f_{opt}(x)$ (assuming an ideal H(ω)). x is then interpolated by $H(\omega)$ and heterodyned by $f_{min}(x)$ to the desired spectral location. For S independent oscillators, a unique instantiation of this scheme is required for each oscillator if the optimal value of $f_{opt}(x)$ were to be applied strictly leading to a costly and complex implementation. The reduction in oscillator update rate by using $f_{opt}(x)$ in place of f_s is outweighed by the cost of a high-quality interpolator executing at a rate of f_s. Furthermore, to restate the points of section 3.1 in the light of section 3.1.1, each optimum interpolation ratio $f_s/f_{opt}(x)$ is likely to be irrational leading to (i) structural complexity of individual interpolators and (ii) the problem of scheduling S incommensurate sample rates within a hard real-time context of T_{max}. S individual interpolator designs are also required. In contrast, consider the simplicity of the 'round-robin' schedule of the TOB in section 1.3.2. However, analogue time-varying (or 'tracking') interpolators for each voice are used successfully in samplers because the number of voices is low in comparison to AS (see section 2.2.1).

The solution is to use a finite set of K interpolators / frequency shifters which constitute a set of 'subbands' in frequency domain denoted s_k:$0 \le k \le K$-*1*, each defined by a unique set of bounds $\{f_{min}(s_k), f_{max}(s_k)\}$: such a set of interpolators summed into a single output at f_s is called a *synthesis filterbank.*. Each s_k is characterised conceptually by outputting $f_{min}(s_k)$ when an input oscillator is at DC, and $f_{max}(s_k)$ when the input is at the Nyquist rate of $f_s(s_k)/2$ and therefore includes the required frequency shift. Therefore the interpolation factor $I(s_k)=f_s/(f_{max}(s_k)-f_{min}(s_k))$ and $f_s(s_k)=2(f_{max}(s_k)-f_{min}(s_k))$, assuming ideal filters for convenience. The best subband for a partial x is that which has the lowest value of $f_s(s_k)$ and yet satisfies $f_{max}(s_k)>f_{max}(x)$ and $f_{min}(s_k)<f_{min}(x)$ ensuring that x will not alias during note lifecycle, if the *a priori* values for x are correct. However, the computation of $\Phi_i[n]$ and $F_i[n]$ assumed for eqn. (1.1) at f_s require linear scaling to $f_s(s_k)$ - the subband context - such that x appears at the output at f_s with correct phase and frequency. The $I(s_k)$'s may

now be preset to convenient integer-only values since they are independent of $\{f_{min}(x), f_{max}(x)\}$ leading to efficient filterbank implementations (as in section 3.1.2) and a set of easily scheduled oscillator sample rates.

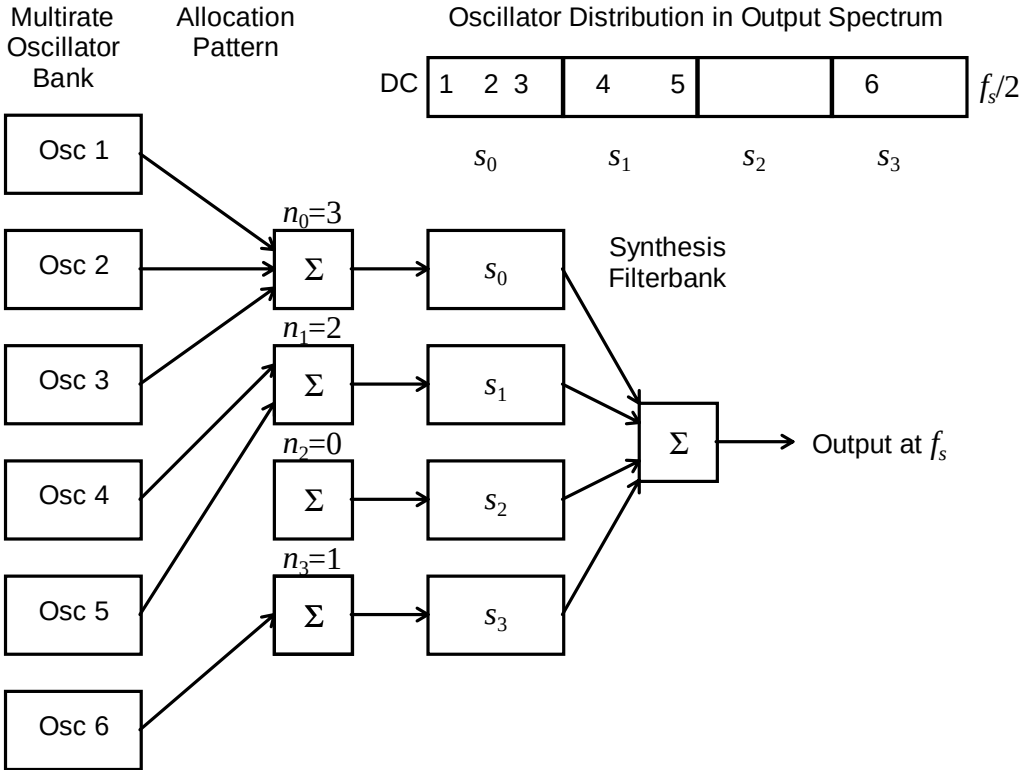

Figure 3.2 A General Model of Multirate Additive Synthesis

For illustration, a simplified example of MAS utilising a subband decomposition is shown in Fig. 3.2 with a scheme of four integer-spaced subbands (and hence four interpolators, each of $I=\uparrow 4$). The single summation of eqn. (1.1) is factorised in two by the introduction of K intermediate interpolators, utilising the superposition property of interpolation in that N parallel instantiations of a single prototype interpolator with post-summation is functionally equivalent to pre-summation followed by a single interpolator. Associated with each s_k is a set of n_k oscillators which execute and are accumulated at the input sample rate of subband s_k - f_s (s_k) - which interpolates up to f_s for final accumulation. When n_k is large, the interpolation overhead per constituent oscillator is small: a quantitative cost / benefit analysis is developed in section 3.1.7. Redundancy

exists in the suboptimal fit of x to s_k which can be minimised by a judicious choice of subband decomposition..

3.1.4 On Oscillator Q

$$Q(x) = \frac{f_{max}(x) + f_{min}(x)}{2(f_{max}(x) - f_{min}(x))} \tag{3.1}$$

For a partial x, a useful relationship between $f_{min}(x)$ and $f_{max}(x)$ is 'Q'; traditionally the 'quality' factor for bandpass filters (ratio of centre frequency to bandwidth). Broad-band signals are low-Q and narrow-band signals are high-Q. $Q(x)$ is calculated from eqn. (3.1) and is significant when sinusoids in the AS set are non-stationary, tending to force Heisenberg's inequality of eqn. (1.5) towards an equality. If AS tones are relatively stationary (e.g. resynthesis of a fixed-pitch resonant system with constant excitation like an organ pipe) then $Q(x)$ is not of interest. There is a duality in how $Q(x)$ relates to AS which requires some explanation. For instance, a high frequency attack transient may have a fixed centre frequency, but generate short-term side-band energy characteristic of a low-Q signal requiring a wide dilation of $\{f_{min}(x), f_{max}(x)\}$. For differing reasons, the same result is true of a meandering $F_i[n]$ envelope as $\{f_{min}(x), f_{max}(x)\}$ is normalised to envelope extremities, though short-term side-band energy is negligible and the sinusoid therefore has a high-Q characteristic. This duality originates in the necessity for a time-invariant $f_{opt}(x)$ via eqn. (1.4) but both scenarios tend to form an equality out of eqn. (1.5) in that $Q(x)=c$ where c is some frequency-independent constant ($c>2$): known popularly as the constant-Q principle (Brown and Puckette, 1992). Two examples provide an illustration:

1. For resonant systems excited by an initial transient, the partial decay rate and bandwidth is proportional (quantified by c) to centre frequency (c.f. the Karplus-Strong plucked string algorithm in section 2.2.2).
2. For the harmonically-spaced partials of a relatively stationary note undergoing pitch modulation, c is related to the modulation range.

3.1.5 Non-Overlapping Subband Decompositions

An initial scheme to minimise redundancy in the suboptimal fit of x to s_k is a fine-resolution integer series of subbands, illustrated in Fig. 3.3 for K=16. A key advantage is that a single oscillator sample rate suffices, removing the necessity to schedule multiple rates; analogous to the frame rate in IFFT synthesis. However, the constant-Q principle is circumscribed leading to two problems. Depending on K, there will be a threshold, denoted f_{eq}, where subband width equals the expected value of $Q(x)$: dependent on the type of note synthesised. If $f(x)$ denotes the 'operating point' of x then if $f(x) > f_{eq}$ subbands become too narrow to accommodate $Q(x)$ risking aliasing of x. For $f(x) < f_{eq}$, subbands have excessive width for $Q(x)$, leaving x oversampled implying computational redundancy. Interpolation resources are poorly distributed: the ear has a logarithmic perception of pitch and $K/2$ subbands - that is half the filterbank computation - cover the top octave of perception. In its defence, an integer series remains an intuitive mapping for the harmonically-spaced partials of a musical note, particularly if the higher partials are relatively stationary and narrow-band i.e. f_{eq} is towards the top end of the audio spectrum because $Q(x)$ is very high.

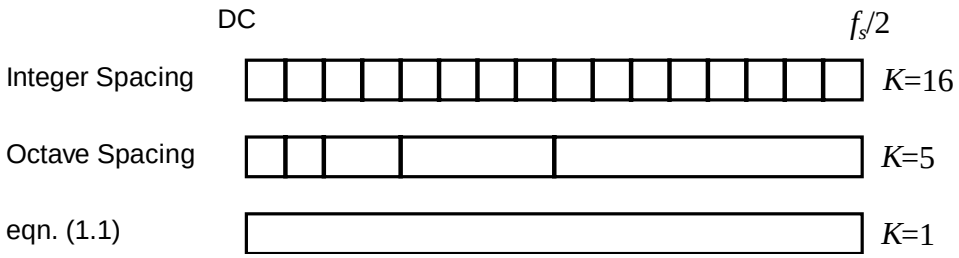

Figure 3.3 Classic Subband Decompositions

Octave subband spacing satisfies the constant-Q property and is illustrated in Fig. 3.3. An infinite series is required for true octave spacing starting from DC, and so the first two subbands are of equal width to pad the series. Two advantages of octave spacing, in comparison to integer spacing, are illustrated by the fact that it (i) uses but one interpolator (in place of $K/2$) to cover the top octave of perception and (ii) performs this at a high sample rate of $f_s/2$, sufficient for the control resolution and bandwidth required by low-Q high-frequency transients. With decreasing frequency, this difference lessens until it can be seen that the first two subbands in an octave-spaced series are integer-

spaced. To its disadvantage, octave spacing possesses progressively wider subbands implying increasing sample rates: high-Q high-frequency sinusoids are oversampled and are therefore computed more inefficiently than in an integer-spaced scheme. There is a proliferation to K-1 oscillator sample rates in place of the single rate for integer spacing indicating a potentially significant increase in oscillator scheduling overheads.

3.1.6 Overlapping Subbands

The common property of integer and octave spacing is that they are non-overlapping. The reason that they have been presented first is that they are the classical ways in which multirate DSP is used to analyse - and resynthesise - a fullband spectrum via subbands (Crochiere and Rabiner, 1983). The problem for MAS is that an oscillator x - paramaterised by $\{f_{min}(x), f_{max}(x)\}$- cannot be allocated across a boundary between adjacent subbands. This situation occurs when, for all subbands, $f_{max}(s_k)>f_{max}(x)$ and $f_{min}(s_k)<f_{min}(x)$ cannot both be satisfied meaning that x cannot be synthesised without aliasing out of a subband. However, any x that fails these conditions can be synthesised by including eqn. (1.1) in MAS which supports arbitrary $\{ f_{min}(x), f_{max}(x)\}$ if overlaid with the chosen non-overlapping subband series. However, fullband synthesis (i.e. classical AS), illustrated in Fig. 3.3 for comparison, is the most expensive of all.

An improvement is the fully-overlapping octave-spaced series illustrated in Fig. 3.4. The benefits of the octave-spaced series discussed in section 3.1.5 are retained. The only modification is that each subband is extended down to DC. By this means, $f_{min}(s_k)<f_{min}(x)$ is always satisfied and a good fit of s_k is identified by testing each s_k in the range $k=K$-1..0 until a subband that satisfies $f_{max}(s_k)>f_{max}(x)$ is found. A problem is that no account is taken of $f_{min}(x)$ which is desirable in order to arrive at the best approximation to the optimal sample rate $f_{opt}(x)$ via eqn. (1.4). For high values of $Q(x)$, this can result in excessive oversampling redundancy. Due to the fact that $f_{min}(s_k)$=DC rather than $f_{min}(s_k)= f_{max}(s_k)/2$, the fully-overlapping series has sample rates twice as high as the equivalent non-overlapping series in Fig. 3.3.

One option for reducing this overhead is to introduce a margin of overlap using a filterbank structure discussed in section 5.3. which permits arbitrary frequency displacements of subbands in the audio spectrum (Phillips et al, 1994). The amount of

overlap is related to an empirical assumption about the lowest value of $Q(x)$ permissible for most AS applications. Fig. 3.4 also illustrates a partially-overlapping subband series formed by setting $f_{min}(s_k)= (f_{min}(s_{k+1})+f_{max}(s_{k+1}))/2$, that is mid-band of s_{k+1}, with $f_{min}(s_K)$=DC and $f_{max}(s_0)=f_s/2$ while retaining an octave series in subband widths (and thus a set of $f_{opt}(x)$'s related by powers-of-2) for ease of scheduling. A significant handicap is that the constant-Q partial series of any note below a certain value of Q(x) cannot be accommodated without some partials aliasing.

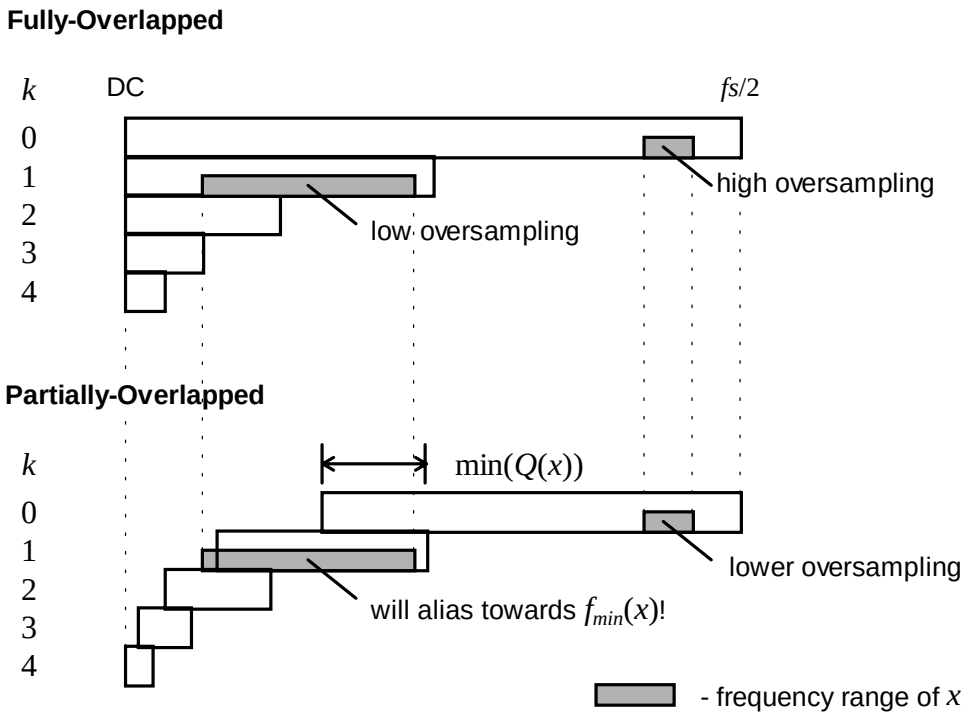

Figure 3.4 Fully and Partially Overlapping Octave-spaced Series for K=5

3.1.7 Towards an Objective Efficiency Comparison between AS and MAS

$$E = \frac{f_s u_{as} S}{v(K) + u_{mas} \sum_{k=1}^{K} f_s(s_k) n_k} \text{ where } S \equiv \sum_{k=1}^{K} n_k \qquad (3.2)$$

The general factors influencing the efficiency of MAS are summarised in eqn. (3.2) which represents the computational cost ratio E between AS via eqn. (1.1) and MAS: both for a single output stream $y[m]$. A successful MAS implementation should maximise E. The numerator is the net cost of a classical AS oscillator bank (e.g. the TOB architecture)

with oscillators updating at f_s at a unit cost of u_{as} per update Conversely, the denominator is the net cost of a MAS oscillator bank with costs $v(K)$ for filterbank computation (a function of K) and u_{mas} per oscillator update: n_k oscillators (S in total) are allocated to each s_k subband where $0 \le k \le K-1$. Note that u_{mas} is related to the distribution of the sample rates in s i.e. the set of $f_s(s_k)$'s. For instance, integer spacing in s has only a single sample rate implying low u_{mas}. In contrast, fully-overlapping octave spacing in s has K sample rates resulting in a higher multirate scheduling complexity and u_{mas}. The units for u_{as} and u_{mas} are implementation-dependent (e.g. CPU execution time or VLSI area) but have the same dimension.

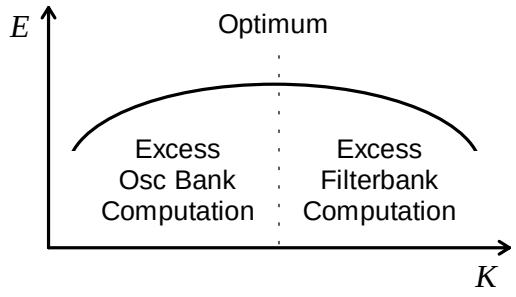

Figure 3.5 E versus v(K) given n

For a given scaleable filterbank scheme (e.g. fully-overlapping octave-spaced subbands), increasing K has the compound effect of increasing filterbank computation $v(K)$ and of providing a richer set of subbands in s. At a low value of K, filterbank computation is negligible, but a sparsely populated s results in the high probability of a poor fit of the selected 'best-fitting' subband to the typical $\{f_{min}(x), f_{max}(x)\}$ bounds of a oscillator x, and thus oversampling redundancy during synthesis. The extreme case is K=1 which is classical AS via eqn. (1.1). However, deploying excess filterbank resources with a high value of K yields an over-rich s that is poorly utilised as reflected by the resulting oscillator distribution n in eqn (3.2): the mean number of oscillators per subband will be low and, indeed, some may be empty. These trends are illustrated in Fig. 3.5. The maximum represents where K is optimised to the distribution of $\{f_{min}(x), f_{max}(x)\}$ bounds for the set of all S oscillators.

3.2 Proposal for a MAS Algorithm using QMF Filterbanks

Arising from the foregoing discussion, a subband decomposition exploiting the binary-tree structure of classical QMF (Quadrature Mirror Filter) filterbanks - with depth K - is identified which possesses the following desirable properties, facilitating a MAS algorithm which improves significantly upon the schemes discussed in sections 3.1.5. to 3.1.6:-

- No constraints are imposed upon $Q(x)$; even classical AS is included in the algorithm.
- Cost of note computation is optimised to $Q(x)$; fixed pitch notes are cheaper to compute (via integer-spaced subbands) than pitch-modulated notes (via octave-spaced subbands).
- Low computational cost (optimised through extensive research) and a simple scaleable filterbank topology using instantiations of a single protoype stage.
- A small set of K+1 subband sample rates (where all ratios are expressed in integer powers-of-two) reduces multirate oscillator bank and filterbank scheduling.
- A rich set of subbands (numbering 2^K-1) with which to guarantee "goodness-of-fit" of subband to the $\{f_{min}(x), f_{max}(x)\}$ bounds of an arbitrary oscillator x.

3.2.1 QMF Filterbanks

QMF Filterbanks are based upon the idealised prototype stage illustrated in Fig. 3.6 (Vaidyanathan, 1993). $H0(z)$ is an interpolating low-pass filter of $I=\uparrow 2$, known as a *half-band filter* because the cutoff frequency $z=\pi/2$ is at half the Nyquist frequency $z=\pi$. Observe that the stopband of $H0(z)$ suppresses the unwanted image of the baseband spectrum of $X0(z^2)$ outside $-\pi/2<z<\pi/2$ (use of z^2 indicates that z is with reference to the $y[m]$ sample rate). In conventional $I=\uparrow 2$ interpolation, this band contains no energy in the output spectrum $Y(z)$. However, the QMF stage accepts a second signal for interpolation, $X1(z^2)$, which is filtered by $H1(z)$, a high pass filter, that passes the image and suppresses the baseband. When added to the signal from $H0(z)$, this image fits in the "gap" outside $-\pi/2<z<\pi/2$. Note that the image is inverted in frequency. The term 'Quadrature Mirror' derives from the fact that $H1(z)$ is the mirror image of $H0(z)$ in

$z=\pi/2$, a property which economises on computation (as discussed in Chapter 4). An important conceptualisation of the QMF stage is as a node of a binary tree that divides its output band $y[m]$ into two equal-width, non-overlapping and contiguous subbands $x0[m]$ and $x1[n]$. The implications of non-ideal filters are dealt with in later chapters and, for the moment, it is preferable to discuss the ideal case.

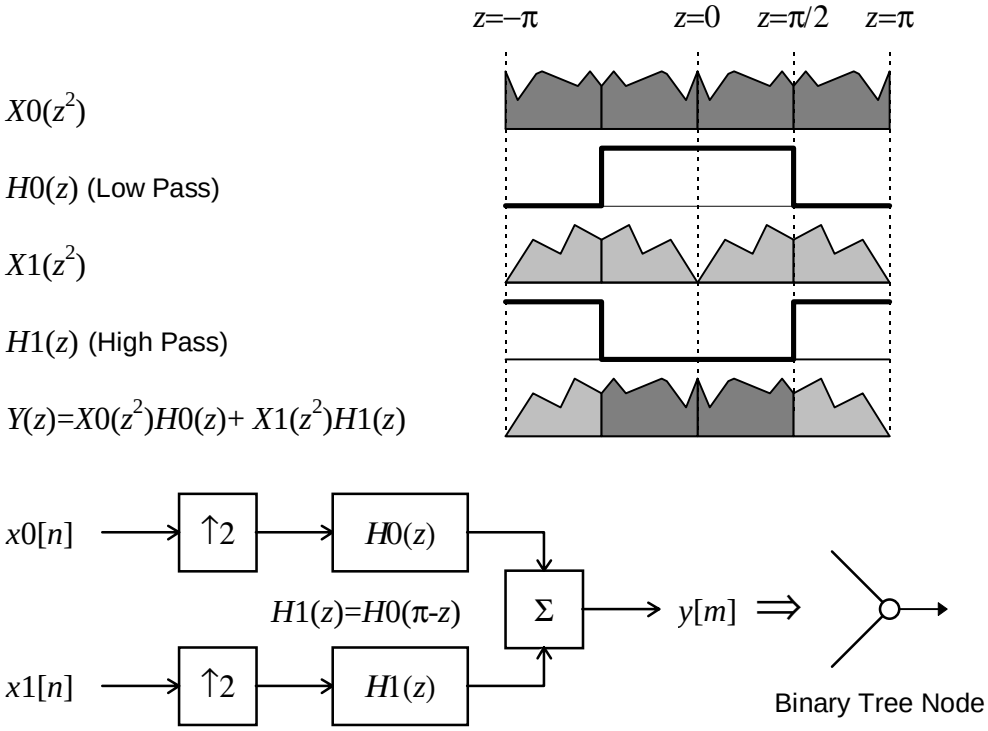

Figure 3.6 Operation of QMF Synthesis Filterbank Stage

3.2.2 QMF Subband Hierarchy

A QMF filterbank is created by assembling instantiations of the prototype stage of Fig. 3.6 into a binary tree of arbitrary completeness: 'complete' means that the number of stages on any path from the root fullband to any terminating subband is always K. An example is illustrated in Fig. 3.7 for K=3. A constraint for MAS is that worst-case latency from level K exponentiates with K: a low value is necessary for real-time synthesis within T_{max}. As justified in section 6.4.4, an optimal value appears to be K=3 which is the default value used in subsequent discussion. In spite of a ceiling on K, configurability of filterbank topology permits $v(K)$ in eqn. (3.2) to be optimised, given knowledge of the $\{f_{min}(x), f_{max}(x)\}$ distribution of the oscillator set, to maximise E. From

the perspective of MAS, a critical feature of a QMF filterbank is that it can be interpreted as a *subband hierarchy* when intermediate subbands between QMF stages are considered. These are usually ignored in conventional analysis / synthesis applications because QMF filterbanks are perceived exclusively as an efficient way to generate a series of 2^K integer-spaced subbands at level K: intermediate signals are of no interest (Crochiere and Rabiner, 1983). Fig. 3.7 illustrates both a QMF filterbank and its corresponding subband hierarchy.

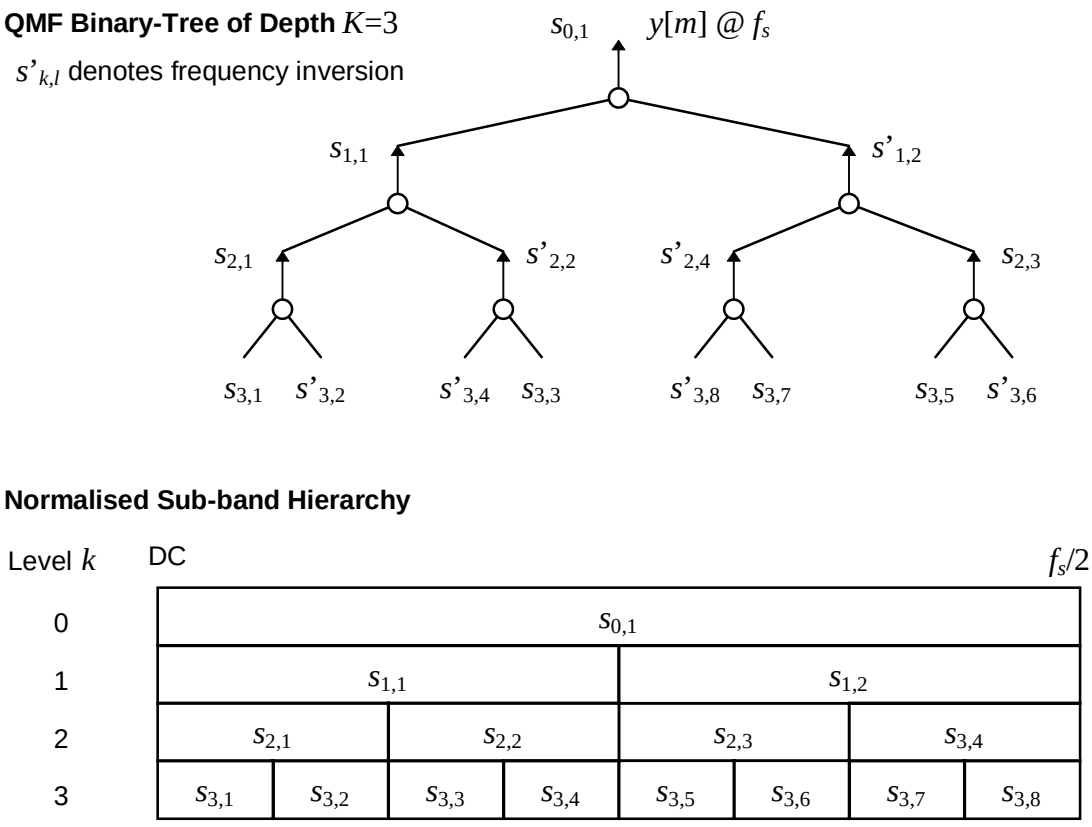

Figure 3.7 Example QMF Filterbank and Subband Hierarchy for K=3

Each subband in the hierarchy is denoted by $s_{k,l}:\{0\le k\le K,\ 1\le l\le 2^k\}$. l indexes the 2^k integer-spaced subbands at level k. The mapping from tree to hierarchy is indirect because, in an individual QMF stage, an inverted side-band image of the baseband of $x1[n]$ appears in the output $y[m]$, and therefore if $x1[n]$ subdivides into subbands itself, these are 'flipped' over in frequency domain. To generalise, if the filterbank path from a subband $s_{k,l}$ to the filterbank output includes an odd number of $H1(z)$ filters, $s_{k,l}$ is

inverted, and if the number is even, $s_{k,l}$ is non-inverted. Inverted subbands are denoted by $s'_{k,l}$ in Fig. 3.7. This peculiarity is easily compensated for in the proposed MAS algorithm by pre-inverting the $F_x[n]$ frequency envelope of an oscillator x allocated to an $s'_{k,l}$ (see section 4.2.5). However, it is convenient to consider a 'normalised' subband hierarchy, whilst acknowledging the indirect mapping to a non-normalised topology.

The internal structure of the subband hierarchy has many desirable properties. Firstly, there are K integer-spaced subband series from the terminal series at level K up to the fullband which is classical AS via eqn. (1.1). Secondly, the series constituted by $s_{3,1}$ and $s_{k,2}$:$\{1 \leq k \leq K\}$is a non-overlapping octave-spaced series whereas $s_{k,1}$:$\{0 \leq k \leq K\}$is the contrasting fully-overlapping equivalent. Therefore the subband decompositions of sections 3.1.5 and 3.1.6 are included in the hierarchy. Conceptually, if a note partial series does not fall into one of the allocation patterns presumed by one of these formal decompositions, others are available. The necessity for partially-overlapping subbands is avoided. Another advantage is that intermediate subbands are interpolated 'for free' in that the higher QMF stages are necessary structurally to recombine the terminal subbands. The computational cost of a complete tree is Kc (where c is the cost of the final stage executing at f_s) because at each extra level, though the number of stages doubles, the sample rate is halved, reflecting the principle of the multistage interpolator in section 3.1.2. Furthermore, a polyphase decomposition of the prototype QMF stage minimises computation as discussed in Chapter 4.

3.2.3 Allocation of $\{f_{min}(x), f_{max}(x)\}$ in a Subband Hierarchy

An allocation algorithm for determining the optimal subband $s_{k,l}$ from a subband hierarchy for an arbitrary oscillator x - bounded by $\{f_{min}(x), f_{max}(x)\}$ - must use the criteria of section 3.1.3 i.e. identify the minimum $f_s(s_{k,l})$ that satisfies $f_{max}(s_{k,l}) > f_{max}(x)$ and $f_{min}(s_{k,l}) < f_{min}(x)$. A conceptual refinement is to identify a level k such that subband width has the 'best-fit' - the minimum k that satisfies $f_s 2^{k-1} > (f_{max}(x) - f_{min}(x))$ - and test whether $\{f_{min}(x), f_{max}(x)\}$ is contained by a single subband. If so, then x can be allocated. If, however, $\{f_{min}(x), f_{max}(x)\}$ spans one of the vertical discontinuities in the hierarchy, and therefore risks aliasing, it is 'promoted' to the sub-band immediately above the termination of the discontinuity. In the extreme case of $f_{min}(x) < f_s/4 < f_{max}(x)$, promotion is

to the computationally most expensive top band $s_{0,1}$ which is classical AS. However, observe that in Fig. 3.7 discontinuities which promote to high subbands are few, and that those with smaller promotion distances are numerous. For the optimal level *k*, then, of its 2^k-1 discontinuities, the number promoting to the next level is 2^{k-1} i.e. over 50%.

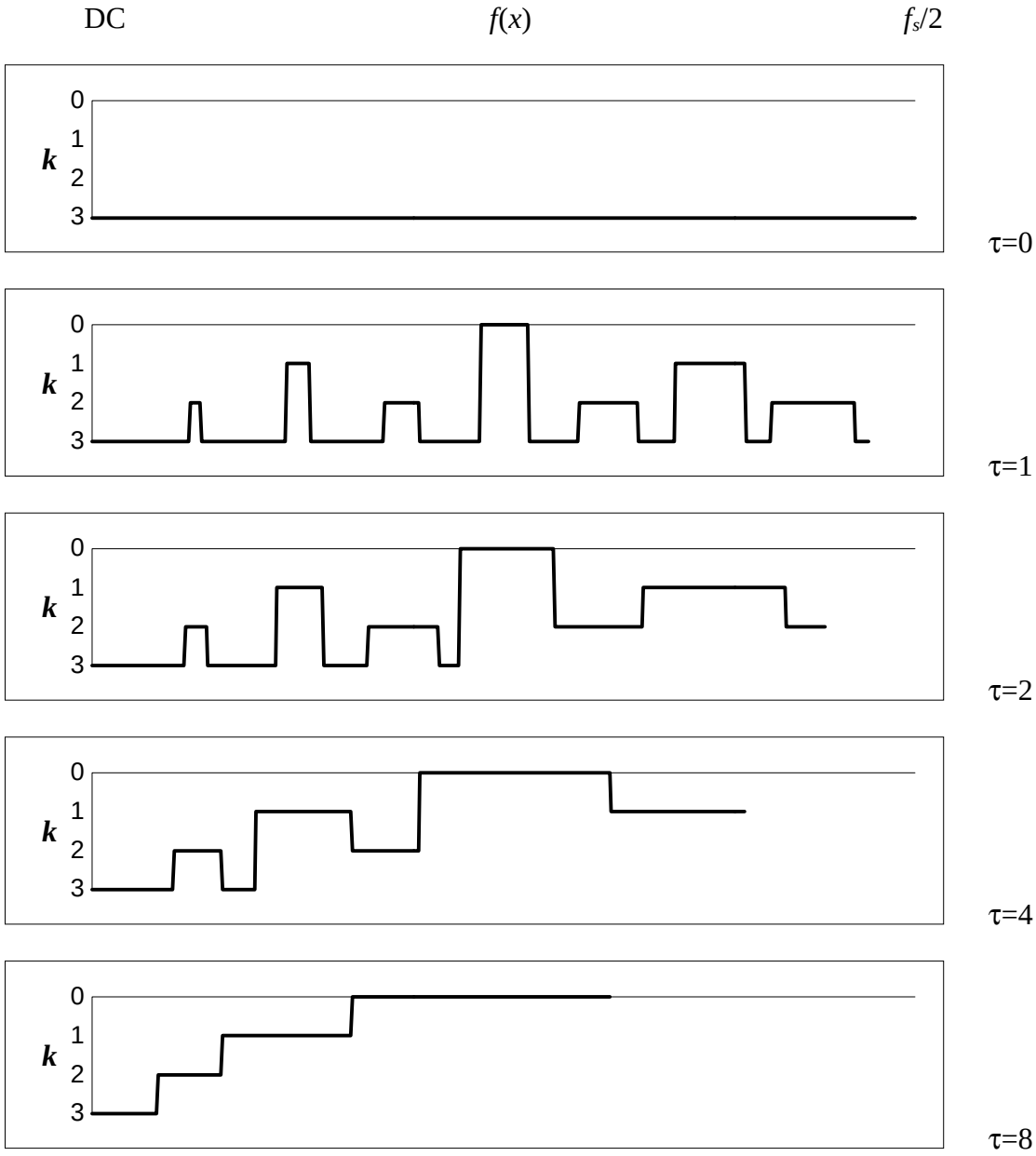

Figure 3.8 Evolution of Subband Hierarchy Allocation Pattern with Increasing τ

A hypothetical constant-Q scenario from section 3.1.4, demonstrating the advantage of a subband hierarchy, is to consider the allocation pattern of a note with partials in pitch-invariant harmony undergoing symmetrical modulation of ±τ semitones from its centre

pitch. If partial x has centre frequency $f(x)$ then $f_{min}(x)=f(x)/\delta_f$ and $f_{max}(x)=f(x)\delta_f$ where $\delta_f=2^{\tau/12}$. Fig. 3.8 illustrates the evolution of the pattern *versus* $\tau \in \{0,1,2,4,8\}$ where the abscissa is $f(x)$, and the ordinate is the allocation level of x in the hierarchy of Fig. 3.7. The resulting allocation patterns are independent of note pitch and a function of τ alone, serving as a template for the actual partial series. At $\tau=0$, partials have a very high-Q and are allocated to the deepest integer-spaced series $s_{3,1..8}$. As τ increases, the pattern gradually evolves into the fully-overlapping octave-spaced series $s_{0..3,1}$ that is suitable for low-Q partials. Evidence of computational efficiency is the mean subband sample rate expressed as a factor υ relative to classical AS at f_s in $s_{0,1}$ as a function of τ which is plotted in Fig. 3.9: theoretical υ is artificially high because it assumes a complete set of high treble partials up to the Nyquist limit $f_{max}(x) \leq f_s/2$. In practice, this band is likely to be sparsely occupied (Sandell, 1994) and thus practical results will produce a lower υ. Two properties of subband hierarchy are thus confirmed in that (i) cost of note computation is optimised to the expected value of τ and (ii), no constraints are placed on τ.

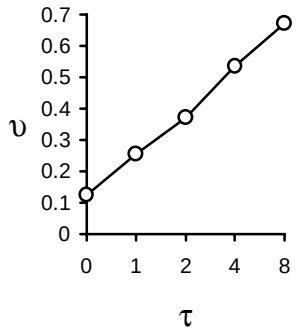

Figure 3.9 υ versus τ

3.3 A Comparison of AS via IFFT and Multirate Techniques

FFT^{-1} is a block transformation resulting in a coarse frame-rate control resolution, permissible because of the comparable coarse resolution of the human ear to envelope features. However, maintenance of the constant-Q principle is desirable, in order to reflect the properties of natural sounds. MAS supports constant-Q allocation in octave-

spaced subbands and facilitates a graduated tradeoff towards integer spacing, analogous to the single control rate of FFT^{-1}. Via a hierarchical subband decomposition, MAS quantises oscillator frequency range, but it is a logical reflection of the stationary nature of note pitch, and of a partial series bound to that pitch. FFT^{-1} does not require $\{f_{min}(x), f_{max}(x)\}$ of each oscillator x, but it is asserted that, for note-based music, such *a priori* information is available by default and therefore worthy of exploitation. An analogy is often drawn between IFFT / OLA synthesis with N frequency bins and an integer-spaced multirate filterbank with N subbands (Vaidyanathan, 1993). They represent complementary frequency-domain and time-domain approaches to optimising the tradeoff between Δ_t and Δ_f in Heisenberg's inequality of eqn. (1.5).

FFT^{-1} has an advantage over MAS for noise synthesis as the latter is dedicated to the optimal computation of sinusoids. However, closely spaced sinusoids randomly modulated in amplitude and frequency are capable of providing a satisfactory alternative to filtered noise in many circumstances (Jansen, 1991). Indeed, there is benefit to be gained from an integration of noise and sinusoids in the same model as, often, the distinction is artificial. For instance, the rapid decay of high frequencies in an attack transient is perceived as a short non-tonal noise. In contrast, SMS extracts a smoothed spectral envelope for noise and requires signal stationarity for partial classification (Masri and Bateman, 1996). Randomisation of envelopes is incompatible with PWL modelling which presupposes slow rates of control parameter change. A MAS oscillator bank may benefit from extensions to facilitate envelope randomisation using algorithms such as that proposed by Fitz and Haken (1995). 'Noisy' oscillators are likely to obey the constant-Q principle and therefore be compatible with allocation in a subband hierarchy.