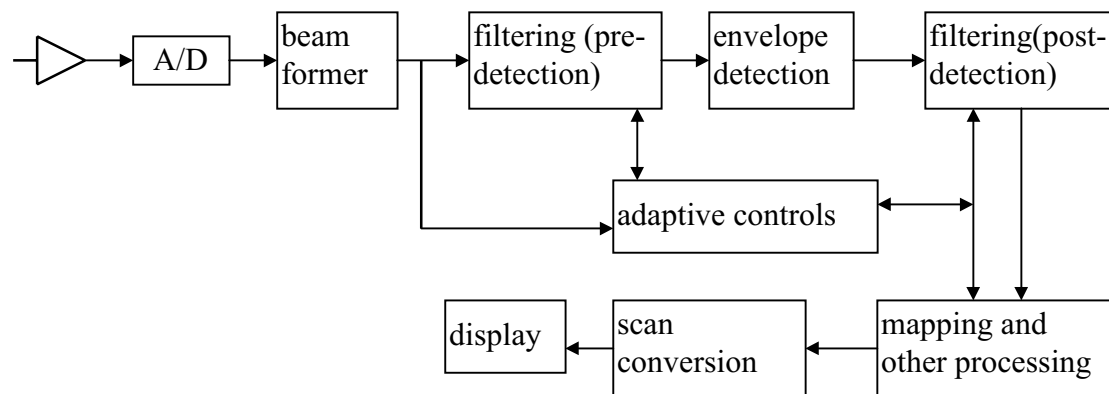


Chapter 6 : Real-Time Image Formation

I. Generic Digital Imaging System

- An ideal transmitter is capable of sending out a pulse with an arbitrary waveform (i.e., arbitrary pulse shape, amplitude and time delay). The pulse shape determines the frequency contents of the transmitted signal. The amplitude of the pulse varies across the array in order to provide aperture weighting (a.k.a. apodization). It can also be varied in order to adjust the acoustic power delivered to the body. Finally, the time delay for each individual channel is related to focusing quality.
- Since a fixed transmit focus is usually used for each firing, the transmitter hardware is relatively easy. In addition, some systems are only using an impulse for the pulse excitation such that system cost can be further reduced.
- A generic digital receiver is illustrated in the following figure



- Signals received by the transducer array are first amplified in order to compensate for the depth dependent signal loss during propagation. This is also known as time gain compensation (TGC). The amplified signals are then digitized and sent to the receive beam former for dynamic receive focusing.
- The pre-detection filter following the receive beam former serves the purpose of pulse shaping (axial filtering), beam shaping (lateral filtering) and temporal filtering (filtering between image frames). In other words, it is used to alter the frequency contents (both temporal and spatial) of the received signals.

Additionally, it can be used for data re-sampling (interpolation or decimation).

- Typically, only the envelope of the received signal is displayed. The envelope detector is used to filter out the carrier frequency component and to keep only the magnitude of the remaining signal. The envelope detected signal is also called the video signal. Two common methods for envelope detection are shown in the following. Assume that the output of the pre-detection filter is

$$S(t) = A(t) \cos 2\pi f_0 t = \text{Re}\{A(t) e^{j2\pi f_0 t}\},$$

where f_0 is the carrier frequency. The envelope $A(t)$ can be extracted by demodulation followed by low pass filtering, i.e.,

$$A(t) = \text{LPF}\{S(t) \cos 2\pi f_0 t\},$$

where *LPF* stands for low pass filtering. The detected signal ($D(t)$) is simply the magnitude of $A(t)$, i.e.,

$$D(t) = \text{abs}(A(t)).$$

Alternatively, the envelope can be detected by converting the real signal $S(t)$ to a complex signal via Hilbert transform. The magnitude of the resultant complex signal is the detected signal. In other words,

$$\begin{aligned} S(t) + j \cdot H.T.\{S(t)\} &= 2A(t) e^{j2\pi f_0 t} \\ D(t) &= \text{abs}(S(t) + j \cdot H.T.\{S(t)\}) / 2 \end{aligned}$$

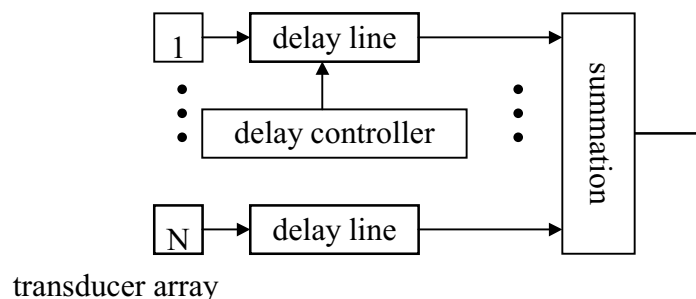
where *H.T.* stands for Hilbert transform. The above methods are also applicable to more complicated signals.

- The post-detection filter is applied to the detected signal for the purpose of smoothing, edge enhancement or a combination of both. It is typically used to enhance contrast or spatial resolution such that particular features in the image can be better perceived. Additional data re-sampling can also be performed at this stage. Other processing, such as re-mapping (gray scale or color) and digital gain adjustment, can be applied following the post-detection filter.

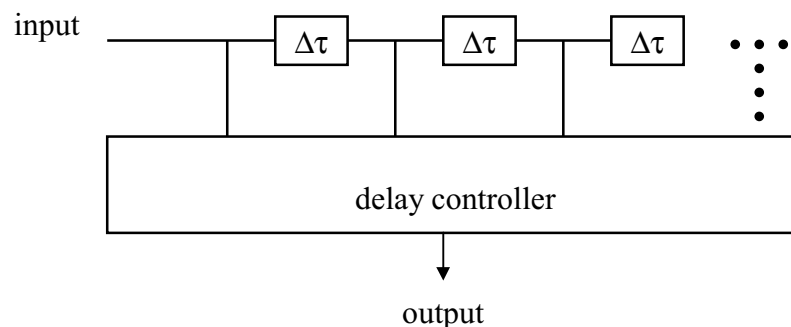
- Scan conversion is the process of converting the acquired ultrasound data (e.g., polar coordinate data for a sector scan format) into the Cartesian raster data used by standard display monitors. In order not to introduce artificial artifacts, such as a Moire pattern, two-dimensional interpolation is usually involved for most pixels in the image and hence it is computationally demanding. Scan conversion will be discussed in detail later in this chapter.
- Ideally, the imaging parameters in all the aforementioned functional blocks should be controlled and adjusted adaptively. In other words, the system should be optimized automatically based on clinical needs, patient characteristics and users' personal preference. However, due to implementation difficulties, most systems are not fully adaptive and the optimization is usually done manually.

II. Receive Beam Former Architecture

- Compared to the hardware for transmission, the receive beam formation hardware is relatively complicated due to the need for dynamic focusing. As mentioned previously, the receive beam formation is indeed a “delay-and-sum” process and can be represented by the following simple model



- Each array element has its own delay value and the value is time (depth) varying. In an analog system, the delay is typically produced with a tapped delay line shown as below



In a digital system, receive signals need to be sampled at an interval of $\Delta\tau$ and the data are stored in memory. In either case, the ideal continuous delay value is quantized and the taps are continuously changed as a function of time (or range) for dynamic focusing. Therefore, the quantized delay value for the i -th array element becomes

$$\tau^R(x_i, R, \theta) = -\frac{x_i \sin \theta}{c} + \frac{x_i^2 \cos^2 \theta}{2Rc} \approx n\Delta\tau.$$

Note that the round trip propagation delay at the center of the array is not included for simplicity since we are only interested in relative propagation delays. Since $2R = ct$, it is straightforward to see that the tap number n is a function of time t . In other words,

$$n(t) \approx -\frac{x_i \sin \theta}{c\Delta\tau} + \frac{x_i^2 \cos^2 \theta}{c^2 t \Delta\tau}.$$

At a fixed steering angle θ (which is always true for a single receive beam former), a tap is changed when the delay between two different ranges (or times) is sufficiently large, i.e.,

$$n(t_1) - n(t_2) = 1 = \frac{x_i^2 \cos^2 \theta}{c^2 \Delta\tau} \left(\frac{1}{t_1} - \frac{1}{t_2} \right).$$

For a particular array element and a steering angle, the above function is only dependent on t_1 and t_2 . The purpose of the delay controller is to perform this calculation in real-time (or by looking up a pre-computed table) and to determine when to switch the delay line or to choose the next sample.

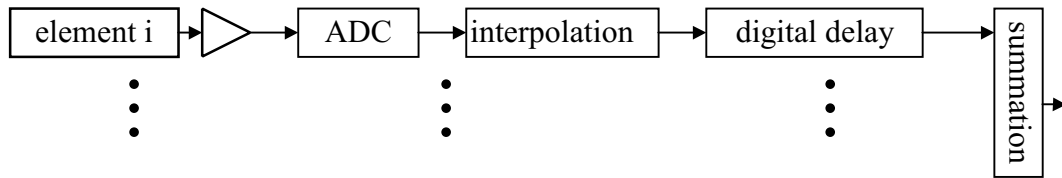
- To choose $\Delta\tau$ such that high quality beam formation can be achieved, $\Delta\tau$ has to be sufficiently small in order to minimize quantization errors. At a center frequency of f_0 , $\Delta\tau$ is equivalent to a phase $\Delta\theta$ where $\Delta\theta = 2\pi f_0 \Delta\tau$. To reduce the quantization errors for high focusing quality, $\Delta\theta$ needs to be below $2\pi/32 \approx 11.25^\circ$ (as a rule of thumb). Therefore

$$\Delta\tau = \frac{\Delta\theta}{2\pi f_0} \leq \frac{1}{32f_0}.$$

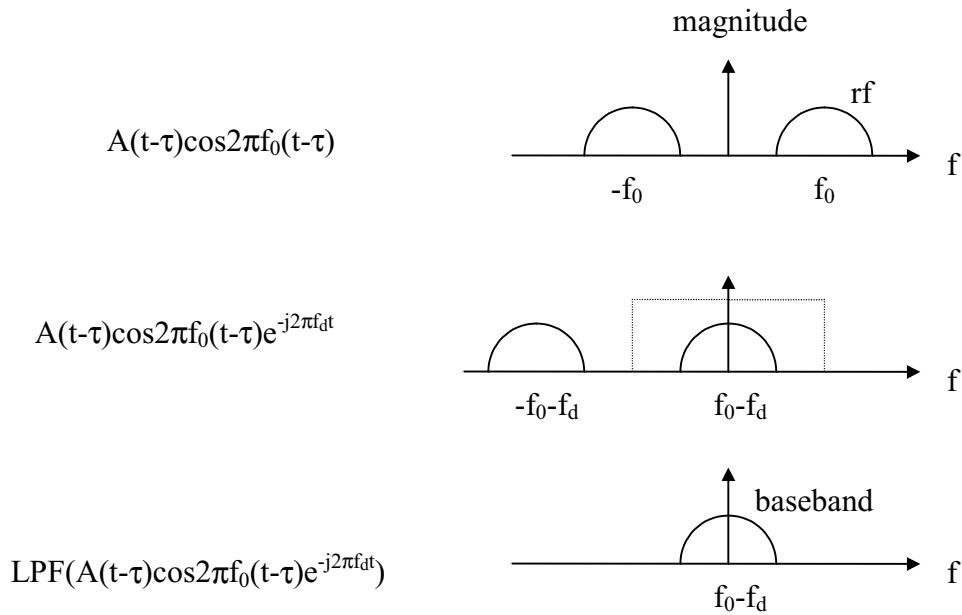
- Most commercial ultrasonic imaging systems have frequencies ranging from

1.5MHz to 15MHz. At 10MHz, for example, the above equation indicates that $\Delta\tau$ needs to be less than $\frac{1}{320} \mu\text{sec} \approx 3.1 \text{ nsec}$. In a digital system, this means that the clock frequency needs to be as high as 320MHz. Even if the frequency of the A/D converter only needs to satisfy Nyquist criterion, the beam former still requires a much higher sampling frequency for high focusing quality (i.e., the A/D converter output needs to be up-sampled to a higher rate). At the current technological stage, this is very difficult (i.e., expensive) to achieve. Even in an analog system, this type of accuracy is very hard to maintain.

- A simple model for digital beam forming is shown below



- An alternative way to achieve high focusing accuracy at relatively low clock frequencies in a digital system is to perform beam formation in baseband. In other words, the original radio frequency (rf) signals are demodulated down to baseband before delays are applied. In the frequency domain, demodulation can be illustrated as the following



Assume that the original rf signal is $A(t - \tau) \cos 2\pi f_0 (t - \tau)$, where τ is associated with the propagation delay. By multiplying the above signal with $e^{-j2\pi f_d t}$ (f_d is the demodulation frequency and is close to f_0) followed by a low pass filter, the spectrum is simply shifted (i.e., demodulated) down to baseband with the center frequency at around DC. In general, the demodulated signal is a complex signal and the real part of the signal is referred to as the in-phase, or I, component. The imaginary part of the signal is referred to as the quadrature phase, or Q, component. In other words, we have

$$\begin{aligned}
 I &= LPF\{A(t - \tau) \cos 2\pi f_0 (t - \tau) \cos 2\pi f_d t\} \\
 &= LPF\left\{\frac{A(t - \tau)}{2} (\cos 2\pi ((f_0 - f_d)(t - \tau) - f_d \tau) + \cos 2\pi ((f_0 + f_d)(t - \tau) + f_d \tau))\right\} \\
 &= \frac{A(t - \tau)}{2} \cos 2\pi ((f_0 - f_d)(t - \tau) - f_d \tau) \\
 Q &= LPF\{-A(t - \tau) \cos 2\pi f_0 (t - \tau) \sin 2\pi f_d t\} \\
 &= LPF\left\{\frac{A(t - \tau)}{2} (\sin 2\pi ((f_0 - f_d)(t - \tau) - f_d \tau) - \sin 2\pi ((f_0 + f_d)(t - \tau) + f_d \tau))\right\} \\
 &= \frac{A(t - \tau)}{2} \sin 2\pi ((f_0 - f_d)(t - \tau) - f_d \tau)
 \end{aligned}$$

Therefore, the complex baseband signal ($BB(t)$) after demodulation is

$$\begin{aligned}
 BB(t) &= I + jQ \\
 &= \frac{A(t - \tau)}{2} e^{j2\pi \Delta f (t - \tau)} e^{-j2\pi f_d \tau},
 \end{aligned}$$

where $\Delta f \equiv f_0 - f_d$.

- Note that $A(t - \tau) e^{j2\pi \Delta f (t - \tau)}$ is simply a low frequency version of the original signal. Using the previous criterion, the delay quantization interval $\Delta \tau$ only needs to satisfy the following equation

$$\Delta \tau = \frac{\Delta \theta}{2\pi \Delta f} \leq \frac{1}{32\Delta f}.$$

Since f_d is chosen to be close to f_0 , Δf can be made much smaller than f_0 and $\Delta \tau$ can be much coarser (i.e., cheaper) than that in a rf beam former.

- Although the delays are much cheaper, there is still a phase term $e^{-j2\pi f_d \tau}$ that

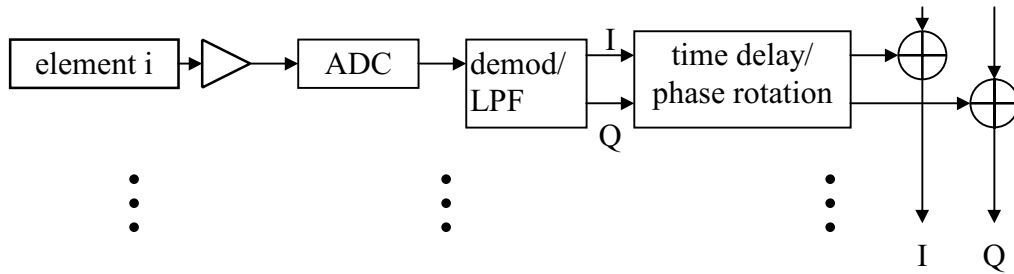
needs to be taken care of. Generally, this is done very accurately by simply rotating the phase of the baseband signal using processors such as CORDIC.

- After applying both time delays (τ'_i) and phase rotations (θ_i) for each individual channel, the summed signal becomes

$$O(t) = \sum_{i=1}^N \frac{A(t - \tau_i + \tau'_i)}{2} e^{j2\pi\Delta f(t - \tau_i + \tau'_i)} e^{-j2\pi f_d(\tau_i - \theta_i)}.$$

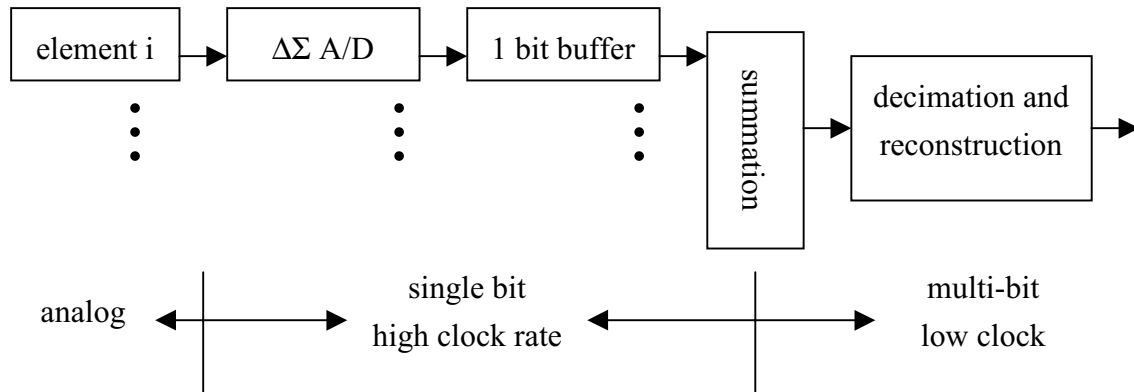
Note that τ'_i is a quantized value and θ_i can be made arbitrarily close to τ_i .

- A simple block diagram of a baseband beam former is shown below



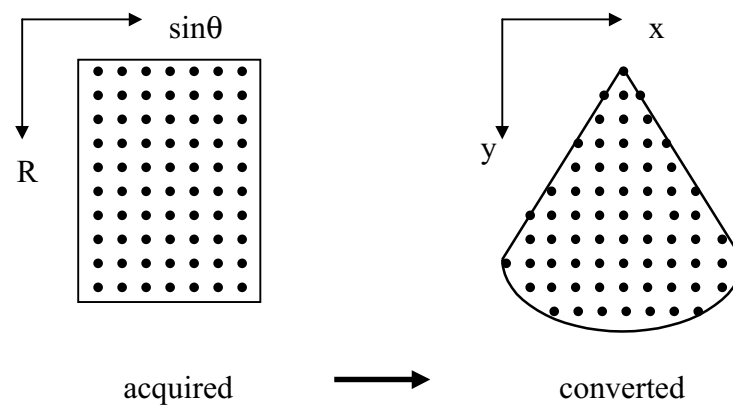
- Coarse time delays as well as accurate phase rotations on baseband signals are indeed an alternative way of interpolation. Digital baseband beam formers provide high quality focusing at a lower cost by trading accuracy in time delays for complex signal processing.
- With the rapid advancement of electronic technologies, alternative beam formation architectures have become feasible. One example is the Delta-Sigma beam former which takes advantage of high clock rate. With the further development of advanced electronic technologies, new digital beam formers should be continued to explore.
- Conventional digital beamformers use 8-10 bit A/D converters at clock frequencies between 20-40 MHz. Delta-Sigma converters, on the other hand, typically use 1 bit conversion at a very fast clock rate. Note that if the clock rate is 32 to 64 times the carrier frequency of the sound waves, delay architecture can be significantly simplified. In addition, 1 bit conversion may also dramatically reduce the size and cost of a digital beamformer.

- A block diagram of the Delta-Sigma based beamformer is shown in the following figure.

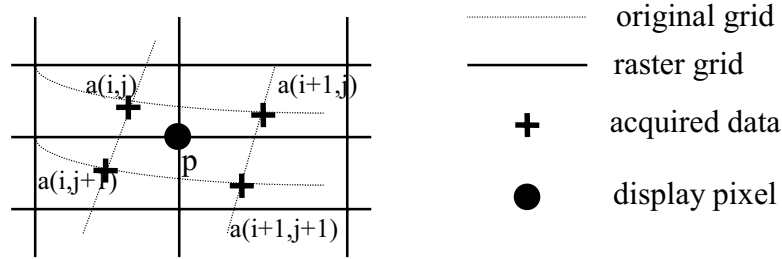


III. Scan Conversion

- Scan conversion is the process of converting the acquired ultrasound data format into the Cartesian raster data format used by standard display monitors. For example, an image with a sector scan format is acquired on a polar grid (R, θ) and needs to be converted to a Cartesian raster grid (x, y) for display.



- Generally the exact location of an acquired data point does not correspond to an allowable display pixel. Hence, the data must be converted before it can be written to the display buffer. This conversion is typically done by bi-linearly interpolating the nearest acquired data points shown below.

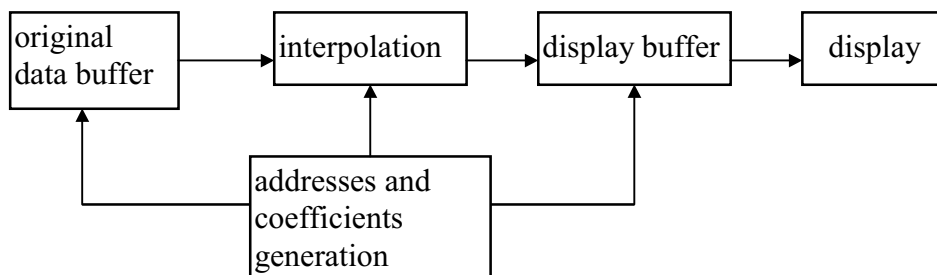


Let $a(i, j)$ be the original acquired sample at the i -th line and the j -th range point. If the output pixel $p(m, n)$ (the (m, n) point on the raster grid) falls in between $a(i, j), a(i+1, j), a(i, j+1)$ and $a(i+1, j+1)$, then

$$p(m, n) = c_{m,n,i,j}a(i, j) + c_{m,n,i+1,j}a(i+1, j) + c_{m,n,i,j+1}a(i, j+1) + c_{m,n,i+1,j+1}a(i+1, j+1),$$

where $c_{m,n,i,j}$'s are pre-determined interpolation coefficients and are a function of the distance between $p(m, n)$ and four neighboring points.

- With less accurate interpolation schemes (e.g., nearest neighbor mapping), image artifacts, such as the Moire pattern, may be present. The artifacts are not only aesthetically unpleasant, they may also interfere with the true diagnostic information in the image.
- A generic scan converter is shown in the following diagram



- In modern ultrasonic systems, scan conversion is done in digital forms which usually utilize special purpose hardware in order to handle its large amount of complex computations. Recently, the scan conversion function has also been implemented on a general purpose microprocessor (TMS320C80. [1]). Therefore, additional capabilities and/or functional changes can be realized in software (flexible), instead of hardware (not flexible).

- Reference

1. C. Basoglu, Y. Kim, and V. Chalana, “A real-time scan conversion algorithm on commercially available microprocessors”, *Ultrasonic Imaging*, Vol. 18, pp. 241-260, 1996.

IV. Frame Rate

- Fundamentally, frame rate (i.e., number of image frames per second) is determined by the propagation time of sound waves. For a sound velocity of $1.54\text{mm}/\mu\text{sec}$ in soft tissue, the round trip propagation time required for a depth of 20cm is simply $t_{line} = 2 \cdot 200 / 1.54 \approx 259.74\mu\text{sec}$. On the other hand, a typical image consists of around 100 to 150 ultrasound lines (determined by the spatial sampling criterion that we discussed previously). Therefore, the frame time required for an image consisting of 120 lines is $t_{frame} = 120 \cdot t_{line} \approx 31.17\text{ms}$ and the frame rate is $1 / 31.17\text{ms} \approx 32\text{frames}/\text{sec}$.
- Generally speaking, images are displayed in “real-time” if the frame rate is around or above 30 (30 for NTSC standards and 24 for motion pictures with 2:1 interlace). For a typical B-mode image, such as the above example, the frame rate is high enough such that the perception of flickering is at its minimum. However, there are many situations in which the frame rate is significantly lower than 30Hz (e.g, 10Hz) due to dense line sampling and/or repetitive firing for blood flow estimation. Under these situations, the motion of tissue may not appear very smoothly. However, the frame rate is still high enough such that users can easily interact with patients and adjust the image in a timely fashion. This type of direct interaction between the users and the patients is indeed the essence of “real-time” diagnostic imaging.
- It is also likely that the frame rate is much higher than 30Hz (e.g., 100Hz). Although information may be lost during live scanning for systems which can only refresh the screen at 30Hz, the information can still be used by displaying still images or by performing slow motion video playback. This is why high frame rates (higher than 30Hz) are still diagnostically important for fast moving structures (such as heart valves).
- Although a 30Hz frame rate with 2:1 interlace is sufficient to avoid the perception

of flickering, it does not guarantee that the temporal sampling rate satisfies the Nyquist criterion. This is also why higher frame rates may be clinically useful and an important feature of a system. There are many methods that can increase the frame rate (e.g., by parallel beam formation or by increasing the beam sampling interval). The details, however, are beyond the scope of this course.