Please watch the next lecture video. You may find some explanation.

In filter h, let's consider one dimension at a time. The cos term can be expressed as sum of two exponential terms and the sine-over-(pi\*n) is a sinc function. The DTFT of the exponential functions are shifted delta functions, and the DTFT of the sinc function is a centered rectangle function. When there are multiplied in the spatial domain, their DTFT are convoluted in frequency domain:

x[n]y[n] DTFT <-> 1 /2 π ∫ 2 π X ( θ ) Y ( Ω − θ ) d θ

This is continuous convolution. If you have not studied continuous time signal processing before, this maybe a little challenging. It's actually not that difficult. You can take X(theta)  for the two deltas, and Y(theta) for the centered rectangle function. Y(-theta) does not change anything. The deltas will make the integral disappear, and what's left will be two copies of the rectangle waveform, each centered at the location of one delta. This is in general the principle of frequency modulation.