This presentation will explain a new efficient neural network design that can be applied to many different CNN models. This efficient design is called Butterfly Transform (BFT), a light-weight channel fusion method that reduces the computational complexity of the inference. This architecture is inspired by the famous Fast Fourier Transform (FFT). I also introduce a novel approach toward architecture learning via discovering neural wirings as the next step to architecture search for creating a sparse and accurate network design. In this model, we relax the typical notion of layers and instead enable channels to form connections independent of each other. This allows for a much larger space of possible networks.