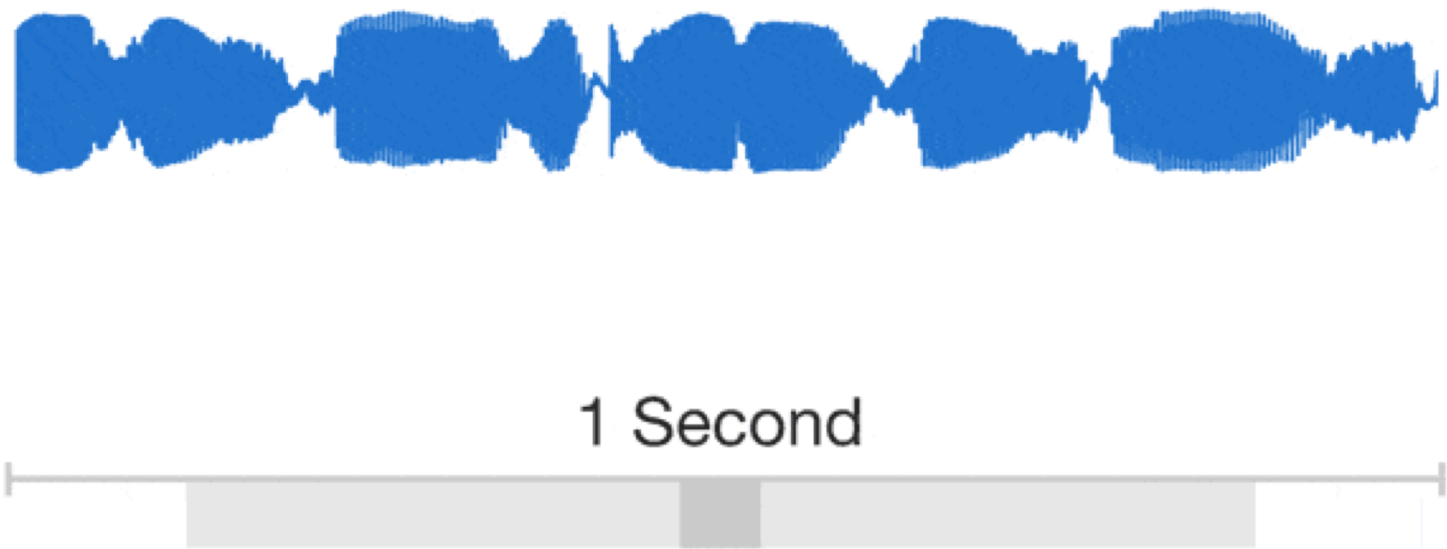
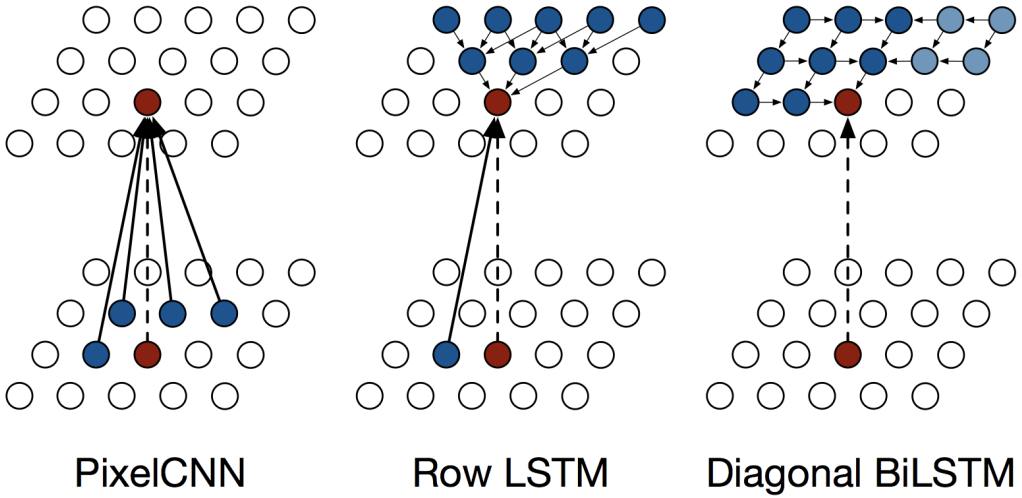
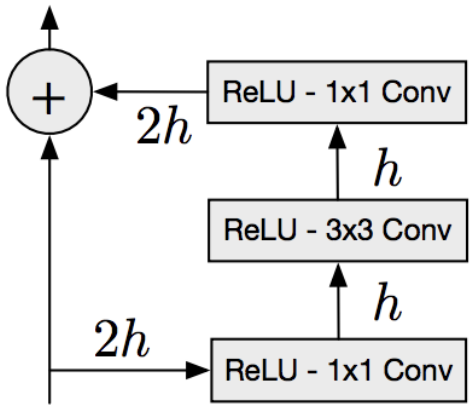
WaveNet: A Generative Model for Raw Audio

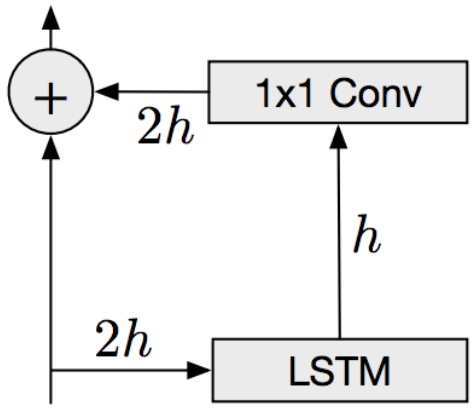
Chalange



PixelRNN

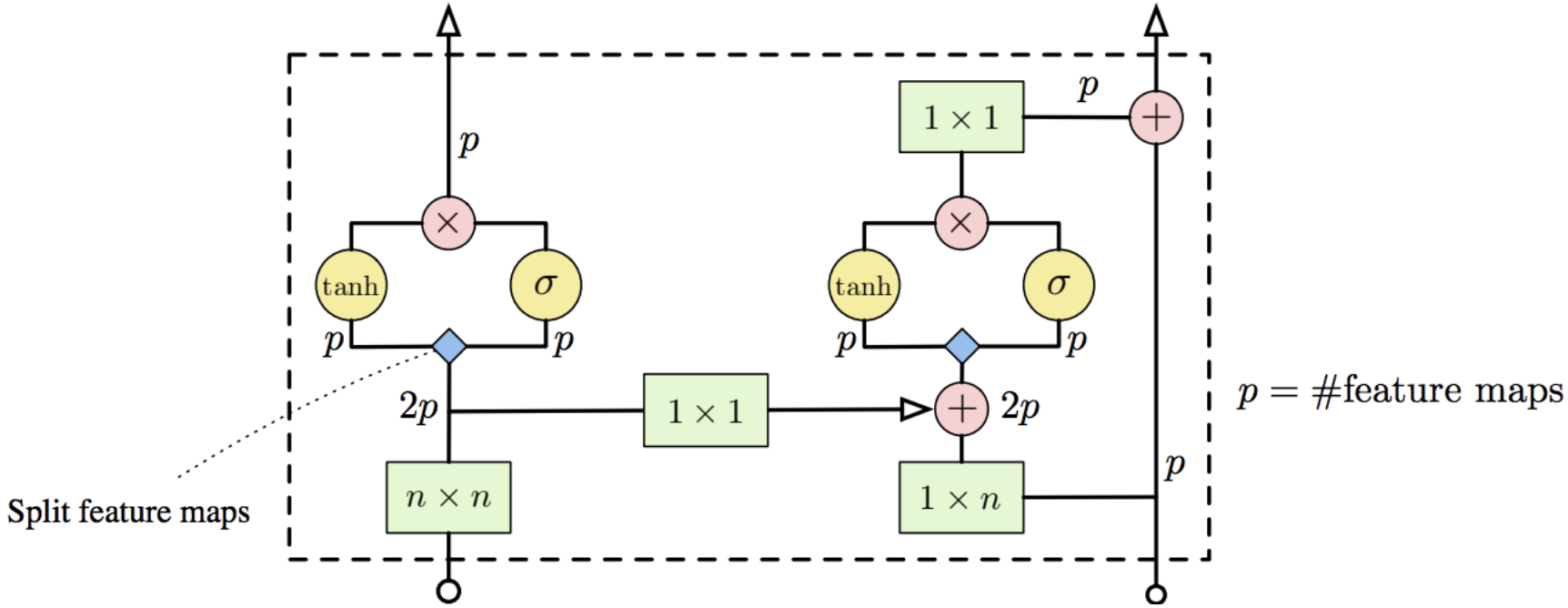

PixelCNN
Row LSTM
Diagonal BiLSTM


ReLU - 1x1 Conv
ReLU - 3x3 Conv
ReLU - 1x1 Conv
2h
h
h
2h


1x1 Conv
LSTM
2h
h
2h

PixelCNN

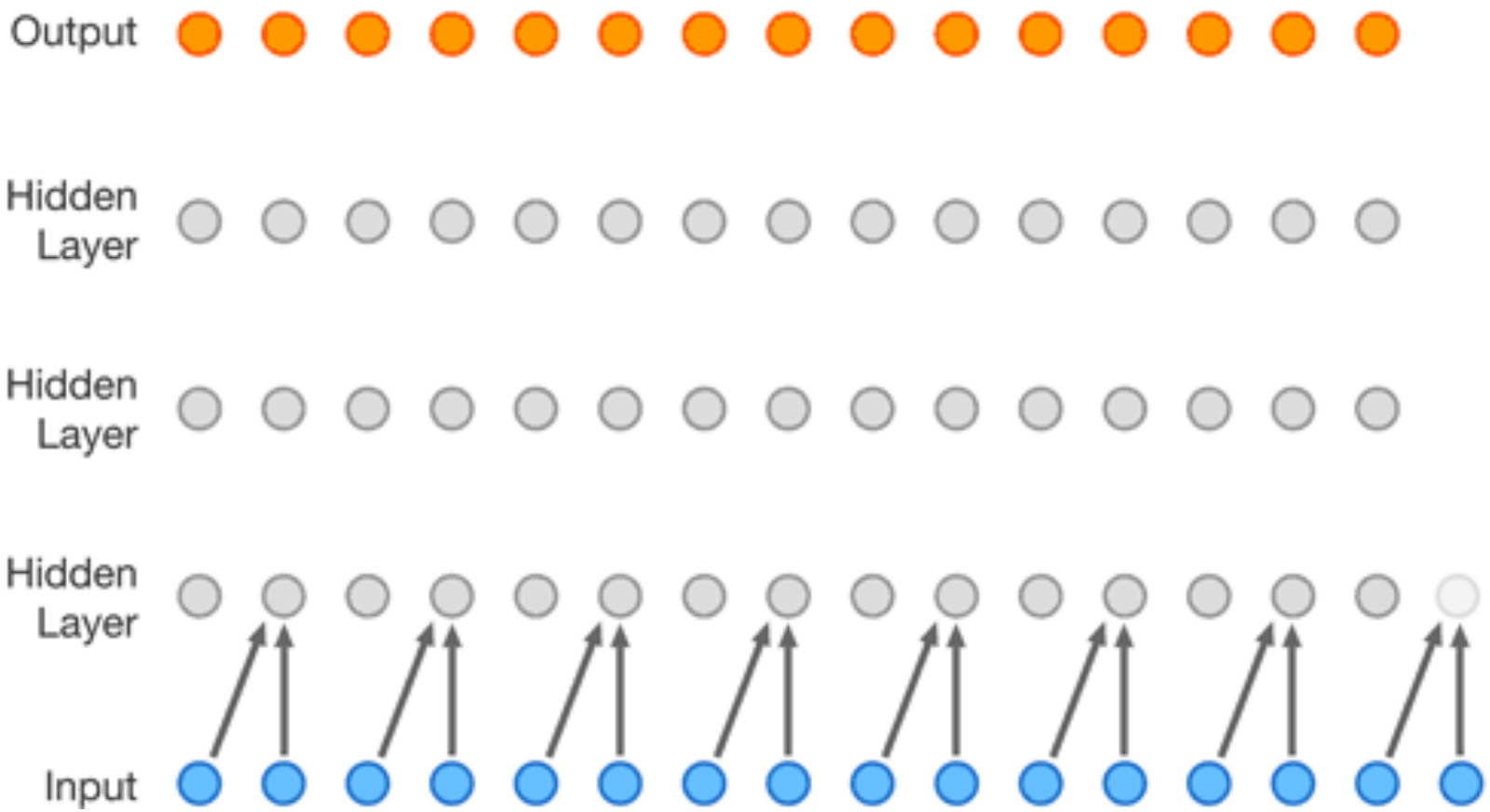
$$p(\mathbf{x}|\mathbf{h}) = \prod_{i=1}^{n^2} p(x_i|x_1, ..., x_{i-1}, \mathbf{h}).$$



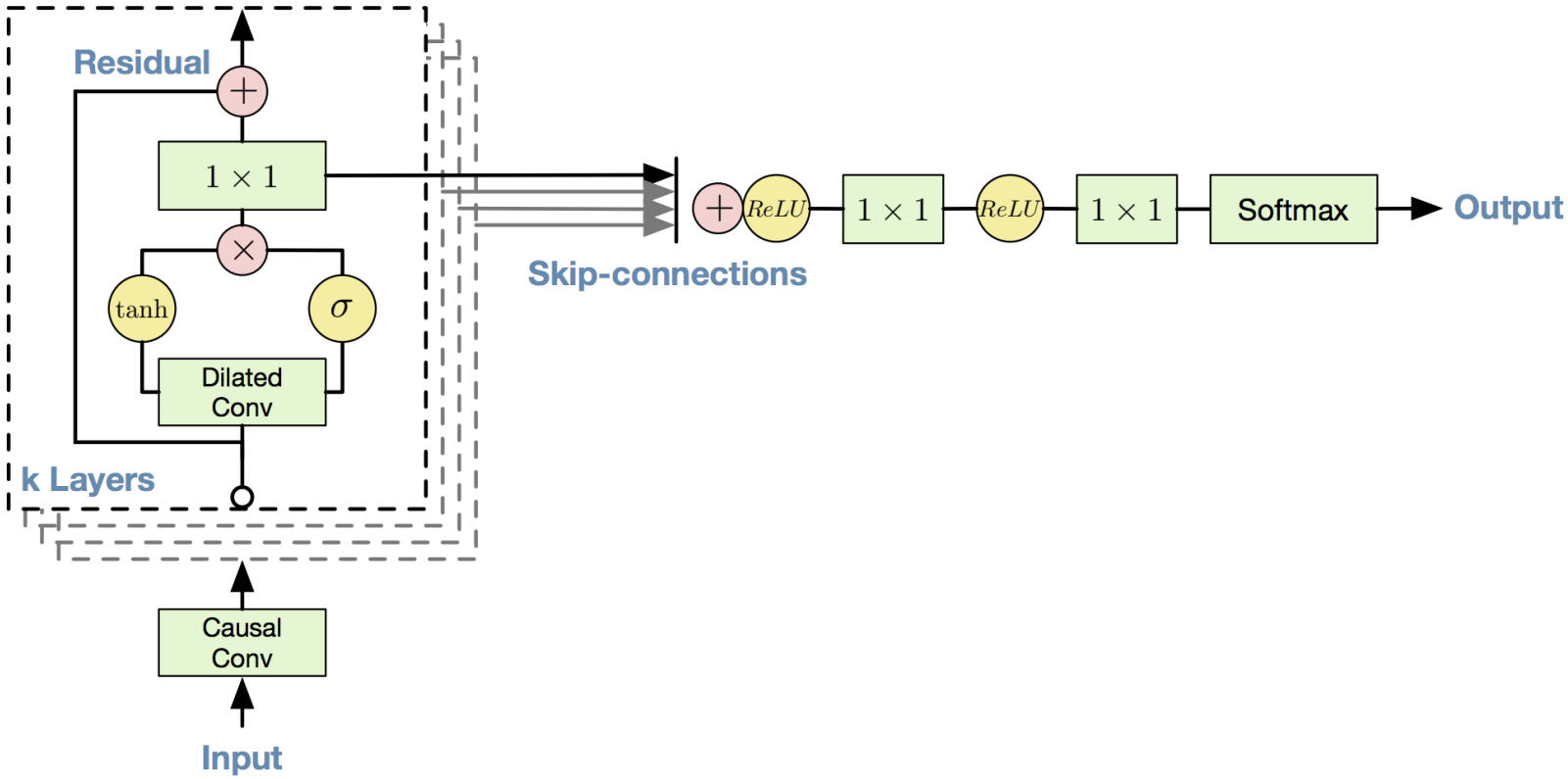
$$\mathbf{y} = \tanh(W_{k,f} * \mathbf{x} + V_{k,f}^T \mathbf{h}) \odot \sigma(W_{k,g} * \mathbf{x} + V_{k,g}^T \mathbf{h})$$

A single layer in the Gated PixelCNN architecture. Convolution operations are shown in green, element-wise multiplications and additions are shown in red. The convolutions with Wf and Wg from are combined into a single operation shown in blue, which splits the 2p features maps into two groups of p.

WaveNet is structured



Overview of the residual block and the entire architecture.

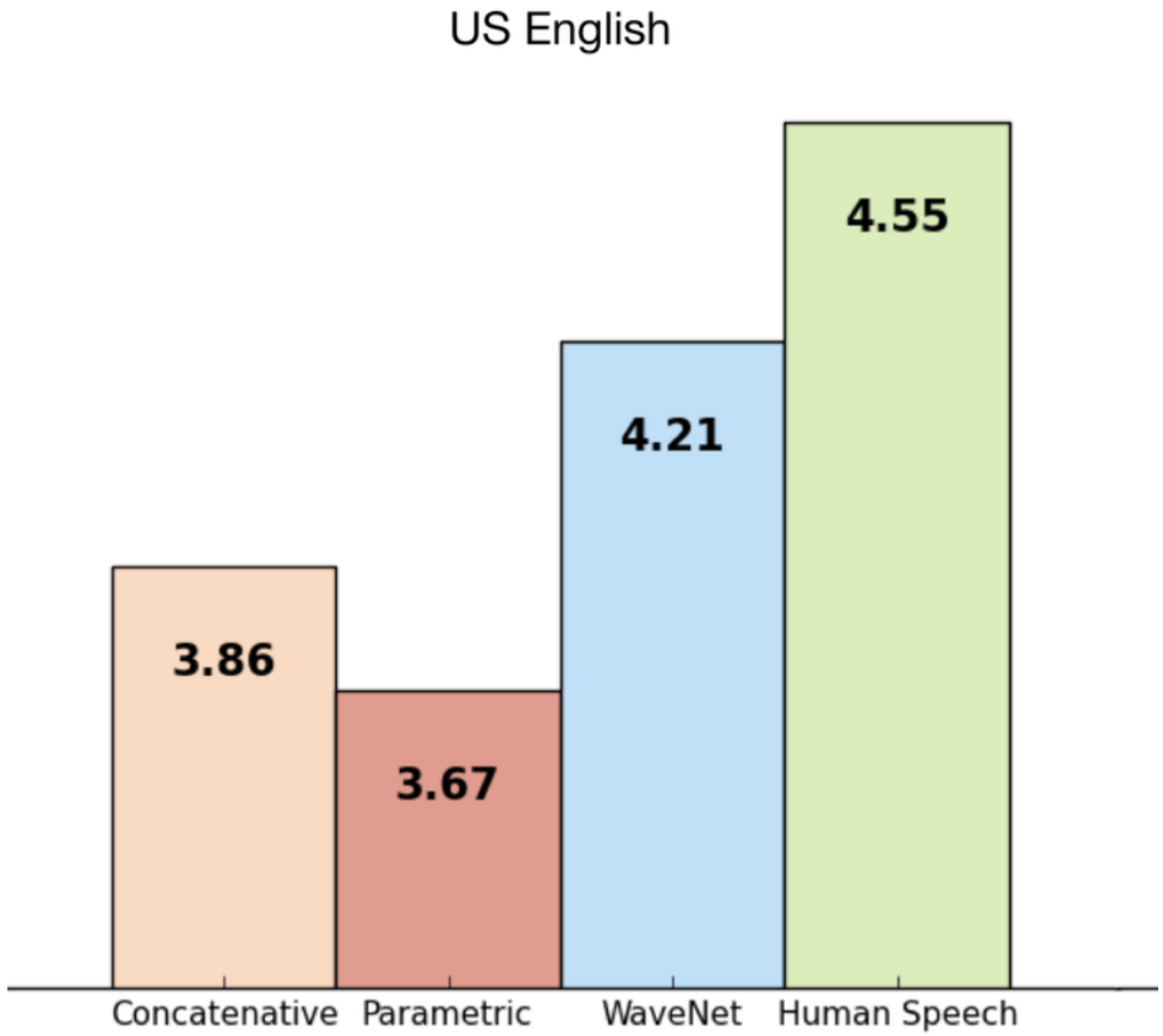


Comparison

Parametric

Concatenative

WaveNet



Knowing What to Say

Different identity

Generate music

Thank you for your attention!