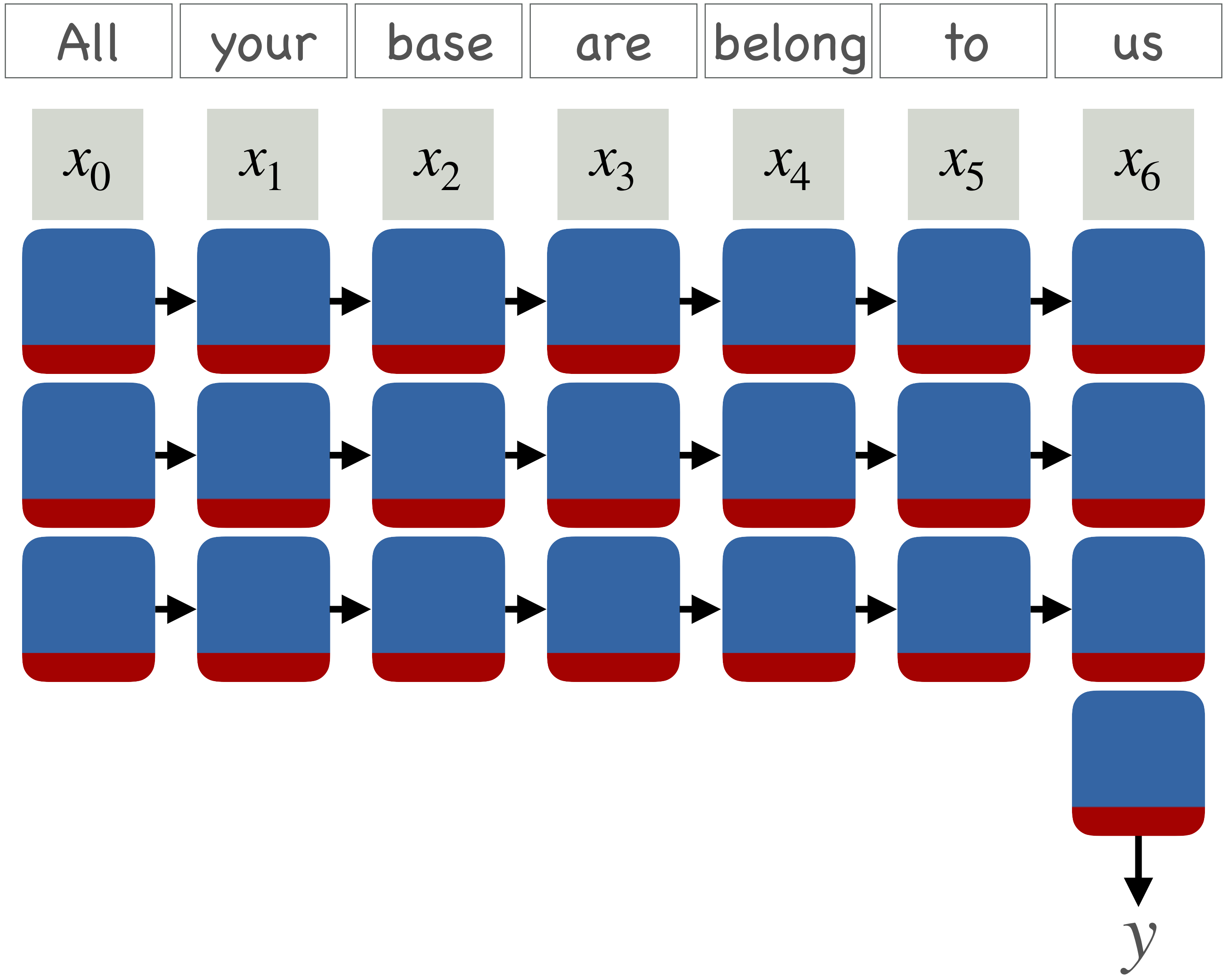
Temporal convolutions

Recurrent models

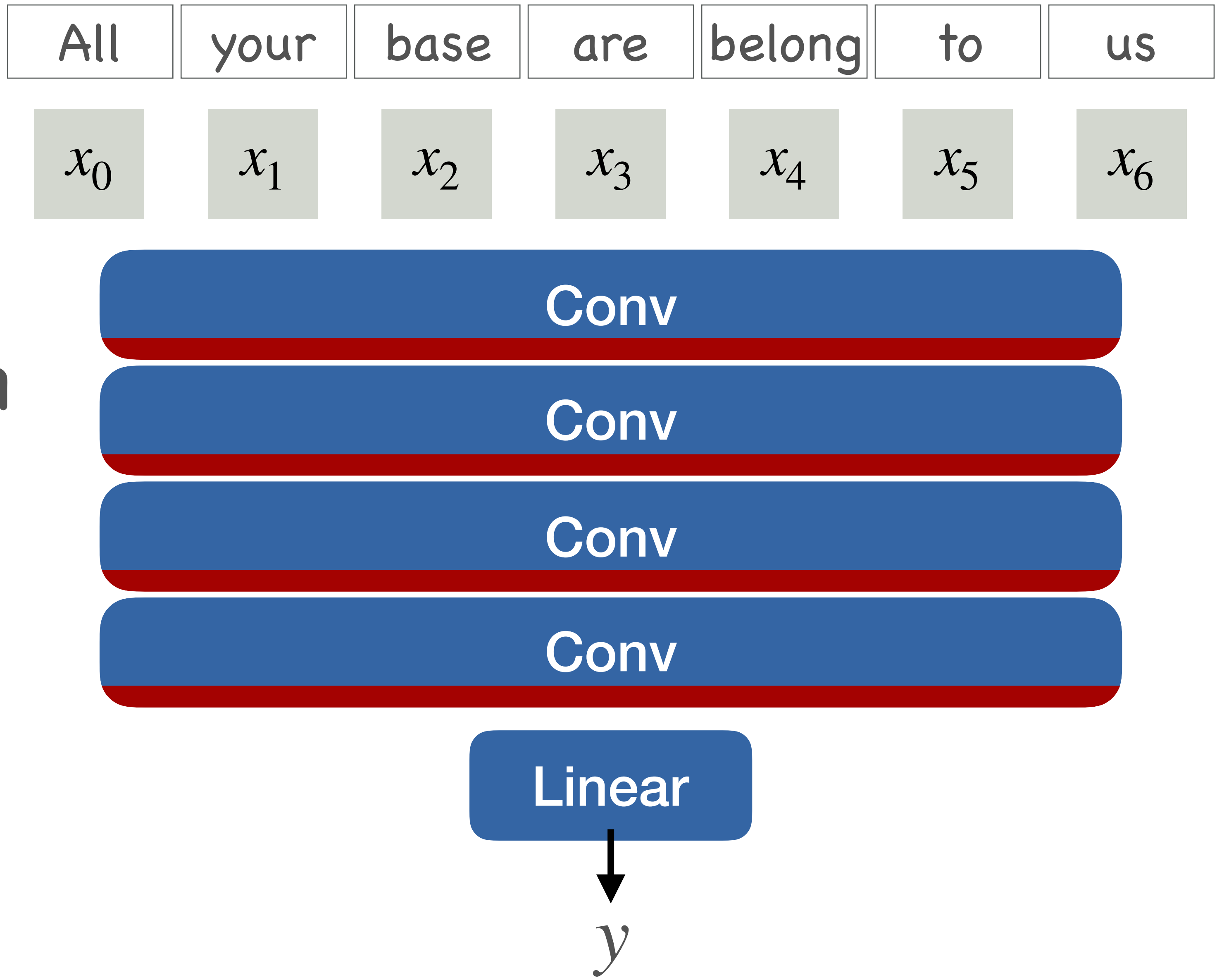
- Advantages
  - Variable input length
  - Variable output length
  - Structured output
- Disadvantage
  - Hard to train
  - Cannot learn dependencies longer than 100 steps



An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling, Bai etal., arXiv 2018

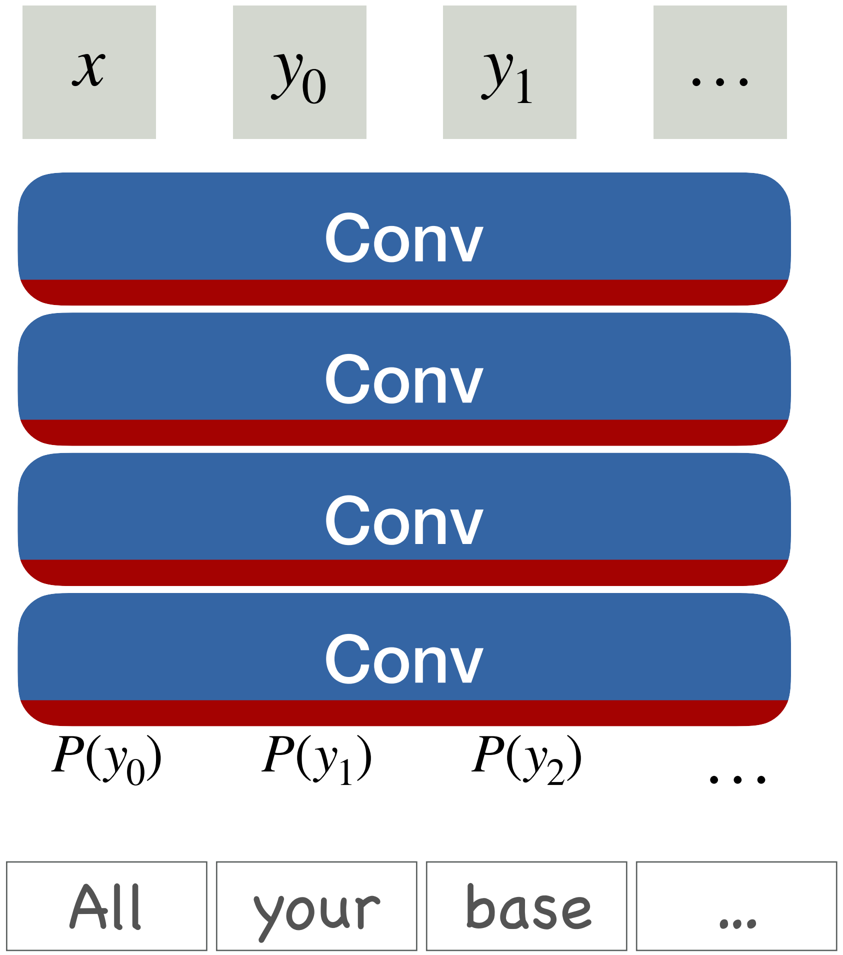
Temporal convolutional networks

- Dilated convolution
  - Exponential growth in receptive field
  - 5-10 layers, receptive field > 100 steps



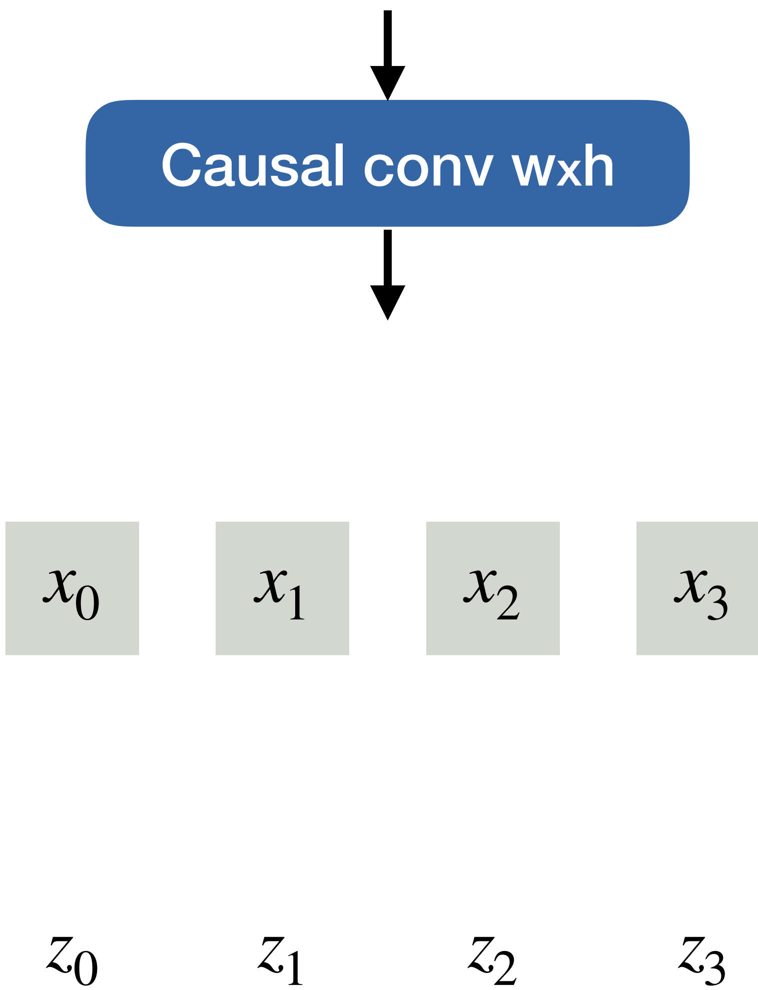
Sequence generation using convolutions

- Causal (masked) convolutions
  - Only look into past
- Auto-regressive model
  - $P(y_0 \mid x) \cdot P(y_1 \mid x, y_0) \cdot P(y_2 \mid x, y_0, y_1) \cdot \ldots$



- Conditional image generation with pixelcnn decoders, Van den Oord et al., *NIPS* 2016
- WaveNet: A generative model for raw audio, Van Den Oord et al., *arXiv* 2016

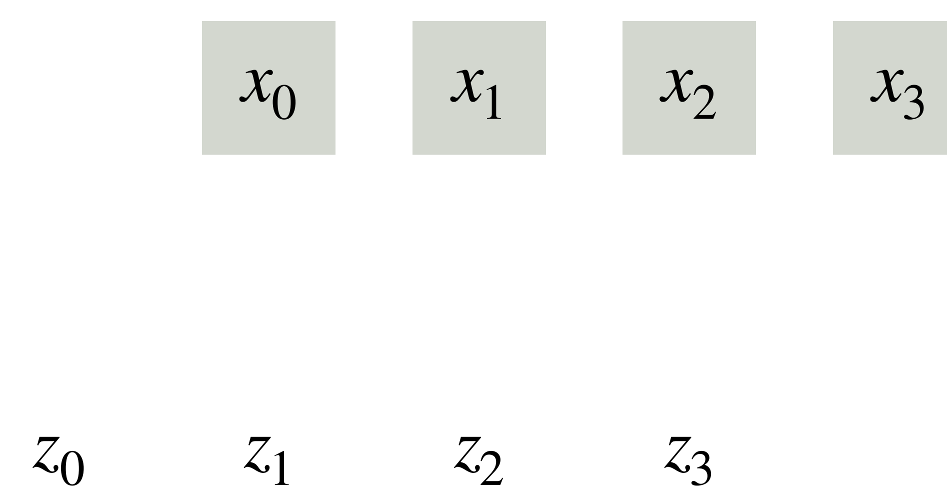
Causal convolution



- Input: $\mathbf{X} \in \mathbb{R}^{T \times C_1}$
- Kernel: $\mathbf{w} \in \mathbb{R}^{w \times C_1 \times C_2}$
- Bias: $\mathbf{b} \in \mathbb{R}^{C_2}$
- Output: $\mathbf{Z}_{t,b} = \mathbf{b}_c + \sum_{i=0}^{w-1} \sum_{j=0}^{C_1-1} \mathbf{X}_{t+i-w,b+j} \mathbf{w}_{i,j}$

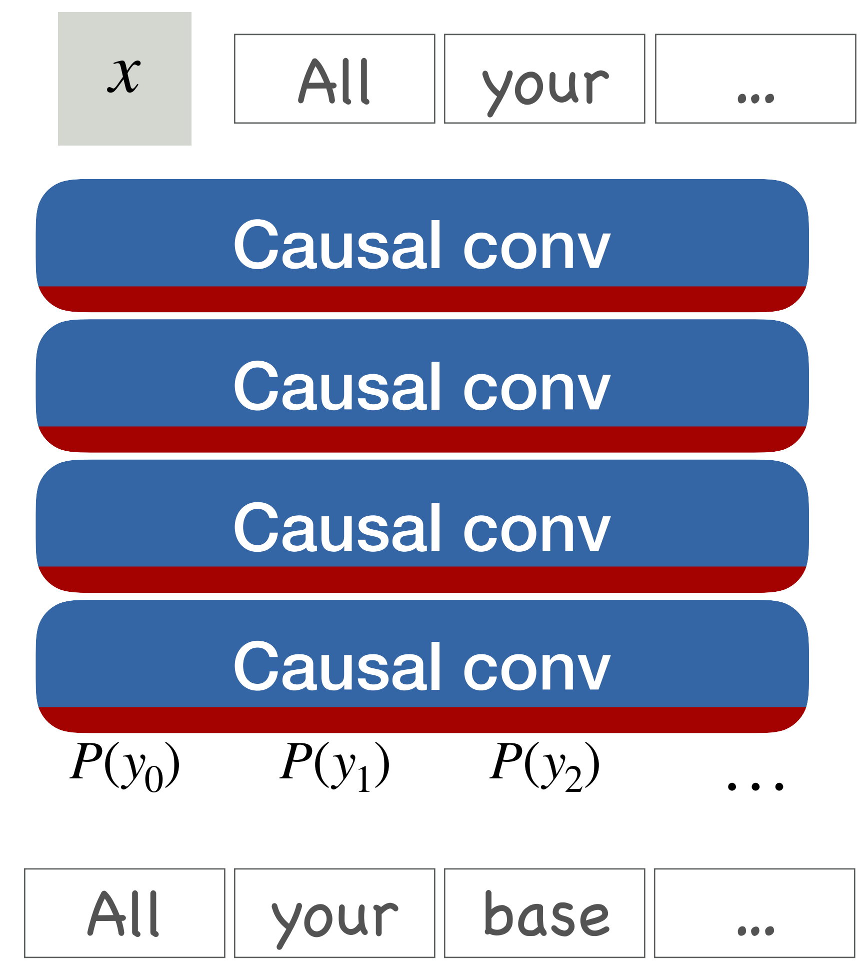
Causal convolution implementation

- Regular convolution
  - Shift output



Training with temporal convolutions

- Labels
    - input and output/loss
- Very efficient
    - fully convolutional



Inference with temporal convolutions

- Step by step
  - Harder to implement efficiently