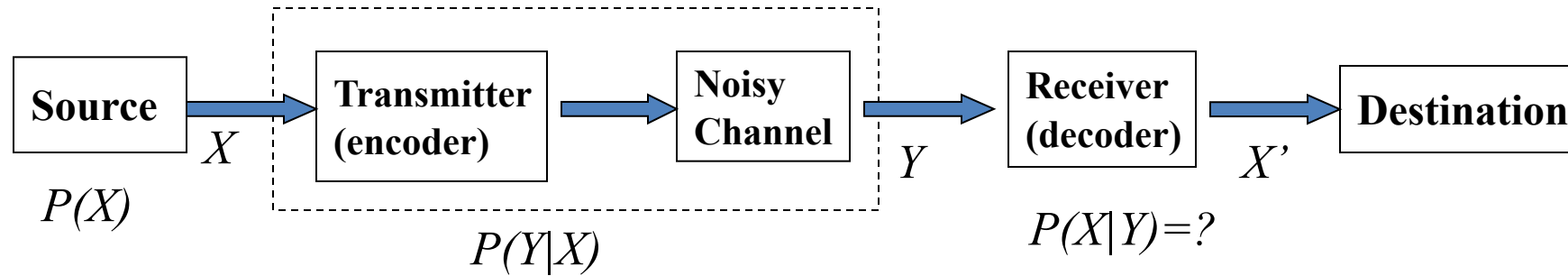


Application 1: As Prior in Bayesian Inference:



$$\hat{X} = \arg \max_X p(X | Y) = \arg \max_X p(Y | X) p(X) \quad (\text{Bayes Rule})$$

When X is text, $p(X)$ is a language model

Many Examples:

Speech recognition:

X =Word sequence

Y =Speech signal

Machine translation:

X =English sentence

Y =Chinese sentence

OCR Error Correction:

X =Correct word

Y = Erroneous word

Information Retrieval:

X =Document

Y =Query

Summarization:

X =Summary

Y =Document