

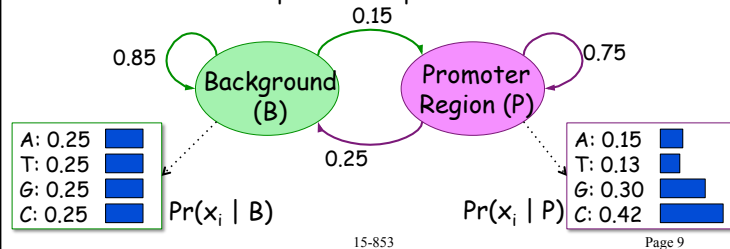




## Hidden Markov Model (HMM)

HMM consists of:

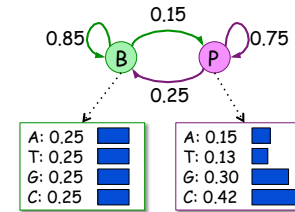
- A set of states and some initial distribution
- Transition probabilities among states — given the current state, what's the probability of the next?
- Emission probabilities — given the current state, a distribution over possible output characters



15-853

Page 9

## Hidden Markov Models



**Markov Property:** Transition and emission probabilities depend on current state only, not the past

**Output:** A sequence of emissions (e.g., GCAAATGC)

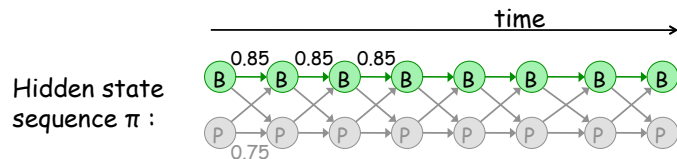
Main goal of today: What is the sequence of hidden states that best explains the output sequence?

(i.e., which subsequences are promoters?)

15-853

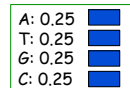
Page 10

## Scoring a path



Emissions X : G C A A A T G C

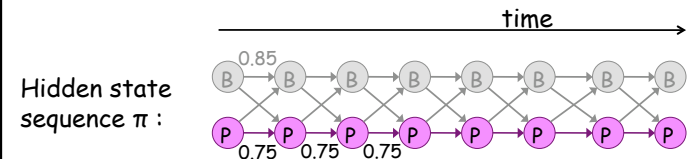
$$\begin{aligned} \Pr(X, \pi) &= \Pr(\text{B initial}) \times \Pr(G|B) \times \Pr(B \rightarrow B) \times \Pr(C|B) \times \\ &\Pr(B \rightarrow B) \times \Pr(A|B) \dots \\ &= \Pr(\text{B initial}) \times 0.85^7 \times 0.25^8 \\ &= \Pr(B) \times 4.9 \times 10^{-6} \end{aligned}$$



15-853

Page 11

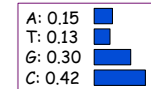
## Scoring a path



Emissions X :

G C A A A T G C  
0.30 0.42 0.15 0.15 0.15 0.13 0.30 0.42

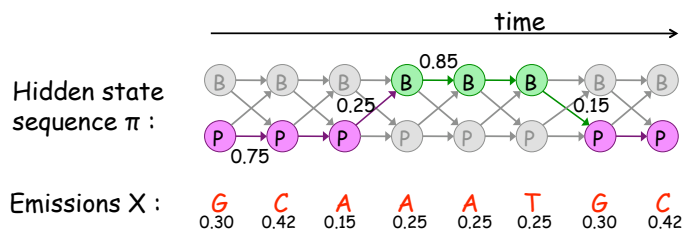
$$\begin{aligned} \Pr(X, \pi) &= \Pr(\text{P initial}) \times \Pr(G|P) \times \Pr(P \rightarrow P) \times \Pr(C|P) \times \\ &\Pr(P \rightarrow P) \times \Pr(A|P) \dots \\ &= \Pr(\text{P initial}) \times 0.75^7 \times 0.3^2 \times 0.42^2 \times 0.15^3 \times 0.13 \\ &= \Pr(P) \times 9.30 \times 10^{-7} \end{aligned}$$



15-853

Page 12

## Scoring a path

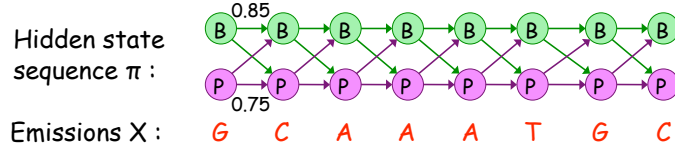


$$\begin{aligned} \Pr(X, \pi) &= \Pr(\text{P initial}) \times \Pr(G|P) \times \Pr(P \rightarrow P) \times \Pr(C|P) \times \\ &\quad \Pr(P \rightarrow P) \times \Pr(A|P) \dots \\ &= \Pr(\text{P initial}) \times 0.75^3 \times 0.85^2 \times 0.25 \times 0.15 \\ &\quad \times 0.25^3 \times 0.30^2 \times 0.42^2 \times 0.15 \\ &= \Pr(P) \times 4.25 \times 10^{-7} \end{aligned}$$

15-853

Page 13

## Viterbi Decoding



Find the highest-scoring path  $\pi^* = \operatorname{argmax}_{\pi} \Pr(x, \pi)$

Naïve: score all paths and choose the best one

$$\pi_1 = B \rightarrow B \rightarrow B \rightarrow B \rightarrow B \rightarrow B \rightarrow B \rightarrow B$$

$$\Pr(X, \pi_1) = 4.9 \times 10^{-6}$$

$$\pi_2 = P \rightarrow P \rightarrow P \rightarrow P \rightarrow P \rightarrow P \rightarrow P \rightarrow P \quad \Pr(X, \pi_2) = 9.3 \times 10^{-7}$$

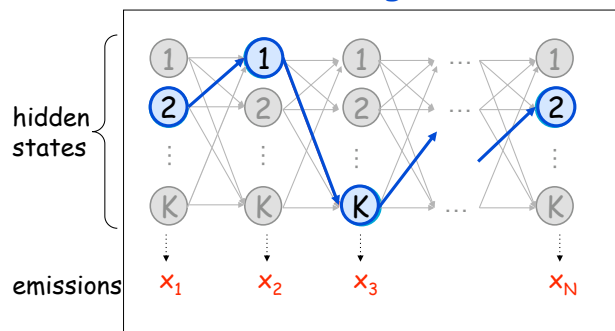
$$\pi_3 = P \rightarrow P \rightarrow P \rightarrow B \rightarrow B \rightarrow P \rightarrow P \rightarrow P \quad \Pr(X, \pi_3) = 4.2 \times 10^{-7}$$

...

Better: Viterbi algorithm

Page 14

## Viterbi Algorithm



Find path  $\pi^*$  that maximizes total joint probability

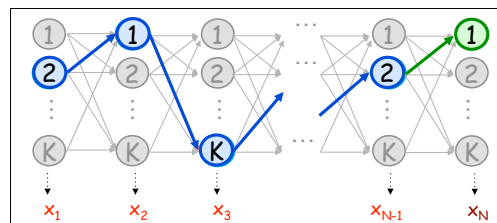
$$\Pr(X, \pi) = \Pr(\pi_0) \times \prod_{\pi_i} (\Pr(x_i | \pi_i) \times \Pr(\pi_i \rightarrow \pi_{i+1}))$$

start      emission      transition

15-853

Page 15

## Recursive substructure



Suppose best path for  $x_1 x_2 \dots x_{N-1} x_N$  is  $\pi_1 \pi_2 \dots \pi_{N-1} \pi_N$

$$\text{i.e., } \pi_1 \pi_2 \dots \pi_{N-1} \pi_N = \operatorname{argmax}_{\pi} (\Pr(x_1 \dots x_{N-1} x_N | \pi) \times \Pr(\pi))$$

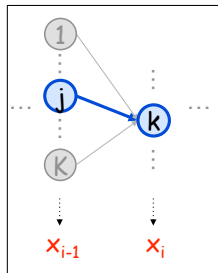
Then  $\pi_1 \pi_2 \dots \pi_{N-1} = \operatorname{argmax}_{\pi \text{ ending at } \pi_{N-1}} (\Pr(x_1 \dots x_{N-1} | \pi) \times \Pr(\pi))$

15-853

Page 16

## Viterbi Algorithm

Let  $V_k(i)$  be the probability of the most likely path that ends at state  $\pi_i=k$  for the emissions  $x_1x_2\dots x_i$ , i.e.,  $V_k(i) = \max_{\pi_1\dots\pi_{i-1}} \Pr(x_1x_2\dots x_i, \pi_1\pi_2\dots\pi_{i-1}, \pi_i=k)$   
Calculate  $V_k(i)$  recursively



$$V_k(i) = \Pr(x_i|k) \times \max_j (V_j(i-1) \times \Pr(j \rightarrow k))$$

↑ emission at state k
↑ transition from j to k

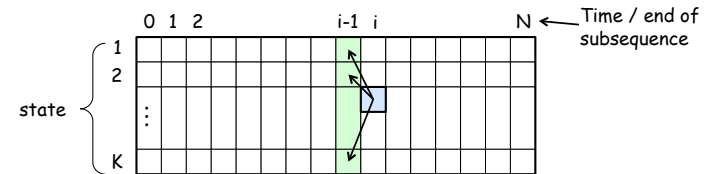
↑ best path to j :

$$\max_{\pi_1\dots\pi_{i-2}} \Pr(x_1\dots x_{i-1}, \pi_1\dots\pi_{i-2}, \pi_{i-1}=j)$$

15-853

Page 17

## Viterbi Algorithm: Dynamic Programming



Given emissions  $X = x_1x_2\dots x_N$

Traceback:

Follow max pointer back

**Initialization:**

$V_k(0) = \text{initial probability for } k$

**Calculations in practice:**

Use log probabilities — adding instead of multiplying

**Iteration:**

$$V_k(i) = \Pr(x_i|k) \times \max_j (V_j(i-1) \times \Pr(j \rightarrow k))$$

**Running time and space:**

Time:  $O(K^2N)$

Space:  $O(KN)$

**Termination:**

$$\Pr(X, \pi^*) = \max_k V_k(N)$$

15-853

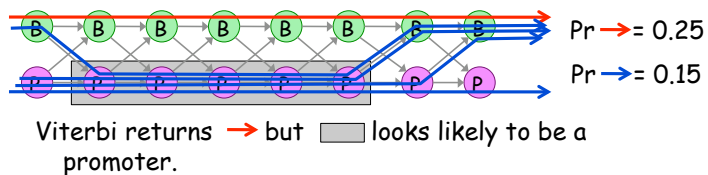
Page 18

## To Viterbi or not to Viterbi?

Viterbi returns the most likely path given the emissions.

How good is this answer?

- Can we tell how much more likely is this path than other paths?
- What if many paths are nearly equally likely?



15-853

Page 19

## Scoring over all paths

Subcomponent of another decoding approach, called posterior decoding

We scored paths before with  $\Pr(X, \pi) = \Pr(X|\pi)\Pr(\pi)$  — given the emissions and a particular path, give the probability.

What is the probability of the emission over *all* paths?

- Calculate  $\Pr(X) = \sum_{\pi} \Pr(X, \pi)$
- Modify decoding to keep total probability passing through each state.

15-853

Page 20

## Scoring over all paths

Associate with each path a probability

- Some paths passing through a state at step  $i$  are likely, other are unlikely: sum them up
- Return the total probability that emissions are observed, summed over all paths
- Equivalent question : what is the probability that the model generates the emission sequence? (can be used to learn model)

Naïve algorithm:

- Enumerate all paths. Score each of them as before. Take the sum.

15-853

Page 21

## Forward algorithm derivation

Define the "forward" probability of an emission sequence  $x_1 \dots x_i$  ending at a state  $k$

$$\begin{aligned}
 f_k(i) &= \Pr(x_1 \dots x_{i-1} x_i, \pi_i = k) \\
 &= \sum_{\pi_1 \dots \pi_{i-1}} \Pr(x_1 \dots x_{i-1}, \pi_1 \dots \pi_{i-2} \pi_{i-1}, \pi_i = k) \times \Pr(x_i | \pi_i = k) \\
 &= \sum_{\pi_1 \dots \pi_{i-1}} \Pr(x_1 \dots x_{i-1}, \pi_1 \dots \pi_{i-2} \pi_{i-1}) \times \Pr(\pi_{i-1} \rightarrow \pi_i) \times \Pr(x_i | \pi_i = k) \\
 &= \sum_j \left[ \sum_{\pi_1 \dots \pi_{i-2}} \Pr(x_1 \dots x_{i-1}, \pi_1 \dots \pi_{i-2}, \pi_{i-1} = j) \right] \times \Pr(j \rightarrow k) \\
 &\quad \times \Pr(x_i | \pi_i = k) \\
 &= \sum_j \boxed{f_j(i-1)} \times \Pr(j \rightarrow k) \times \Pr(x_i | \pi_i = k) \\
 &= \Pr(x_i | \pi_i = k) \times \sum_j f_j(i-1) \times \Pr(j \rightarrow k)
 \end{aligned}$$

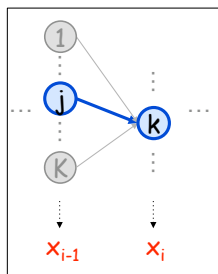
15-853

Page 22

## Forward Algorithm

Let  $f_k(i)$  be the "forward" probability of an emission sequence  $x_1 x_2 \dots x_i$  ending at state  $k$

Calculate  $f_k(i)$  recursively



$$f_k(i) = \Pr(x_i | k) \times \sum_j (f_j(i-1) \times \Pr(j \rightarrow k))$$

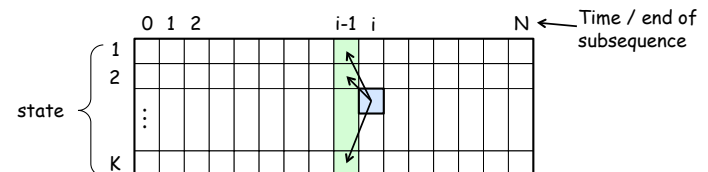
emission at state  $k$                       transition from  $j$  to  $k$

forward probability to  $j$  :  
 $\sum_{\pi_1 \dots \pi_{i-2}} \Pr(x_1 \dots x_{i-1}, \pi_1 \dots \pi_{i-2}, \pi_{i-1} = j)$

15-853

Page 23

## Forward algorithm: dynamic programming



Given emissions  $X = x_1 x_2 \dots x_N$

Calculations in practice:

Initialization:

$f_k(0) = \text{initial probability for } k$

- Using log scores is tricky here, because of the summation.
- Values are rounded and scaled

Iteration:

$$f_k(i) = \Pr(x_i | k) \times \sum_j (f_j(i-1) \times \Pr(j \rightarrow k))$$

Running time and space:

Time:  $O(K^2 N)$   
 Space:  $O(KN)$

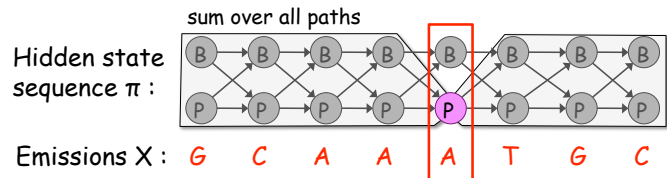
Termination:

$$\Pr(X) = \sum_k f_k(N)$$

15-853

Page 24

## Posterior decoding: probability of a particular state



Calculate the most probable state at each  $\pi_i$  position  $i$   
 $\Pr(\pi_5 = P | X)$   
 Sum over all paths that emit  $x_i = A$  at position  $i = 5$

15-853

Page 25

## Posterior decoding formula

We want  $\Pr(\pi_i = k | X)$  – the probability that the  $i$ th state is  $k$  given the entire emission sequence  $X$

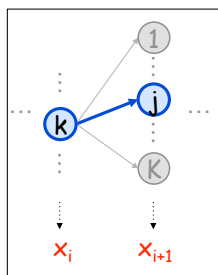
$$\begin{aligned} \Pr(\pi_i = k | X) &= \Pr(\pi_i = k, X) / \Pr(X) \\ &= \Pr(x_1 \dots x_i, x_{i+1} \dots x_N, \pi_i = k) / \Pr(X) \\ &= \Pr(x_1 \dots x_i, \pi_i = k) \times \Pr(x_{i+1} \dots x_N | \pi_i = k, x_1 \dots x_i) / \Pr(X) \\ &= \underbrace{\Pr(x_1 \dots x_i, \pi_i = k)}_{\text{forward score}} \times \underbrace{\Pr(x_{i+1} \dots x_N | \pi_i = k)}_{\text{backward score}} / \Pr(X) \\ &= \text{forward score} \times \text{backward score} / \Pr(X) \\ &= f_k(i) \times b_k(i) / \sum_k f_k(N) \end{aligned}$$

15-853

Page 26

## Backward Algorithm

Let  $b_k(i)$  be the "backward" probability of an emission sequence  $x_{i+1}x_i \dots x_N$  beginning from state  $\pi_i = k$   
 Calculate  $b_k(i)$  recursively



$$b_k(i) = \sum_j (\Pr(x_{i+1} | j) \times b_j(i+1) \times \Pr(k \rightarrow j))$$

emission at state  $k$

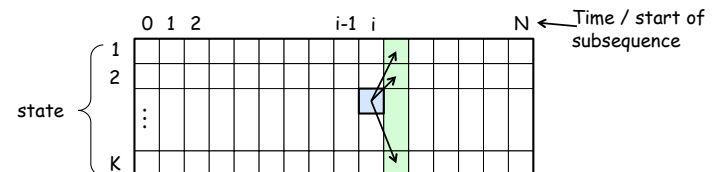
transition from  $k$  to  $j$

backward probability from  $j$  :  
 $\sum_{\pi_{i+2} \dots \pi_N} \Pr(x_{i+2} \dots x_N | \pi_{i+1} = j)$

15-853

Page 27

## Backward algorithm



Similar to the forward algorithm (and Viterbi), except minor details and table fills from right to left.

Given emissions  $X = x_1 x_2 \dots x_N$  Termination:  
 $\Pr(X) = \sum_k f_k(N)$

Initialization:  
 $f_k(N) = 1$

Iteration:  
 $b_k(i) = \sum_j (\Pr(x_{i+1} | j) \times b_j(i+1) \times \Pr(k \rightarrow j))$

15-853

Page 28

## Putting posterior decoding together

$D_k(i)$  = probability that  $i$ th state is  $k$  given the entire emission sequence  $X$

We had  $D_k(i) = \Pr(\pi_i=k | X) = f_k(i) \times b_k(i) / (\sum_k f_k(N))$

1. Run forward and backwards algorithms
2. Compute  $D_k(i)$  for each  $i, k$  (in any order)
3. Most likely state at position  $i$  is  $\hat{\pi}_i = \operatorname{argmax}_k D_k(i)$

For a probability distribution over states for the  $i$ th position, normalize by the sum of the probabilities

- Can normalize columns while doing forwards/backwards — may ameliorate calculations

Note: posterior "path"  $\hat{\pi}_1 \hat{\pi}_2 \dots \hat{\pi}_N$  may be infeasible

15-853

Page 29

## Summary so far

Score likelihood of a sequence and emissions

- Single path:  $\Pr(X, \pi)$
- All paths:  $\Pr(X)$  — forwards algorithm

Decoding (determining hidden states / classifying subsequences)

- Single path:  $\operatorname{argmax}_\pi \Pr(X, \pi)$  — Viterbi algorithm
- All paths:  $\Pr(\pi_i=k | X)$  — posterior decoding

Learning the model (not covered)

Next:

- Adding "memory" to the HMM
- Generalized HMMs

15-853

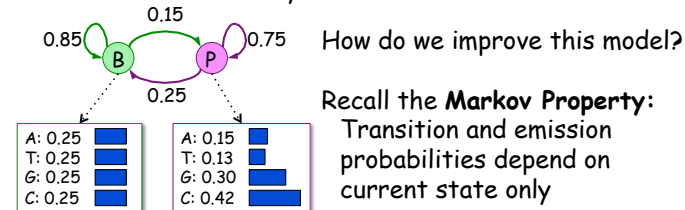
Page 30

## Refining the promoter model

Promoters are likely to be "CpG islands"

A **CpG island** is a stretch with a high concentration (maybe 50%) of CG dinucleotides

Old model counts only frequency of G's and C's, but we want to see if they're next to each other

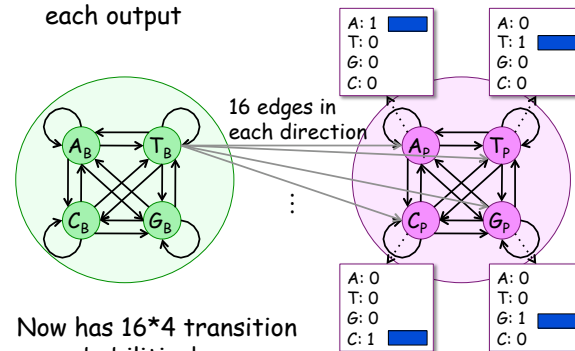


15-853

Page 31

## Adapting HMMs to dinucleotides

Increase the number of states: unique state for each output

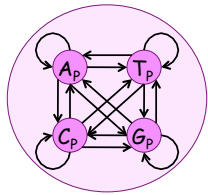


15-853

Page 32



## Adapting HMMs to dinucleotides



Combine two Markov chain models:

- P model: a CpG island
- B model: background

Transition probabilities in each model

- Encode differences in dinucleotide frequencies

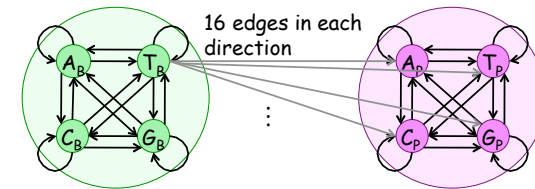
P	A	C	G	T
A	.180	.274	.426	.120
C	.171	.368	.274	.188
G	.161	.339	.375	.125
T	.079	.355	.384	.182

15-853

B	A	C	G	T
A	.300	.205	.285	.210
C	.322	.298	.078	.302
G	.248	.246	.298	.208
T	.177	.239	.292	.292

Page 33

## Adapting HMMs to dinucleotides



Can assign probabilities to all 32 edges in between  
Or use  $\Pr(B \rightarrow B)$ ,  $\Pr(B \rightarrow P)$ ,  $\Pr(P \rightarrow B)$ ,  $\Pr(P \rightarrow P)$  and multiply by probabilities in previous slide.

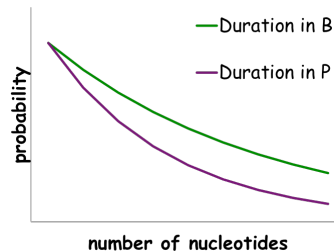
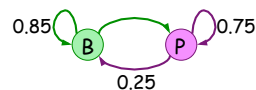
Can extend approach to codons (trinucleotides) or more, but usually try to group states to reduce transition table size

15-853

Page 34

## Duration in each HMM state

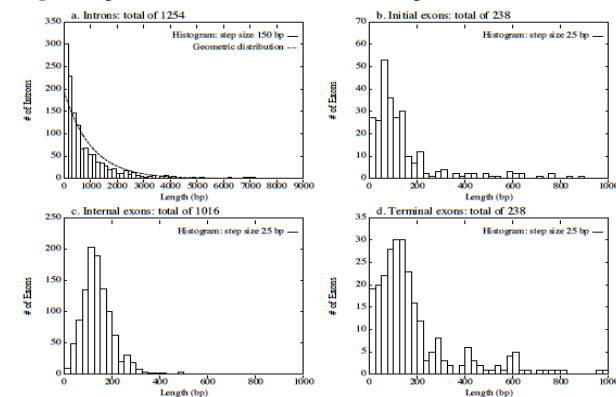
HMMs give a geometric distribution over the duration in each state



15-853

Page 35

## Human exon lengths



from Burge, MIT PhD thesis

15-853

Page 36

## Generalized HMMs (GHMMs)

Model duration at each state

- Emit multiple characters at each state, according to duration
- Can model arbitrary distribution

Disadvantages

- Decoding is much more expensive — analogous to arbitrary distance penalty in sequence alignment

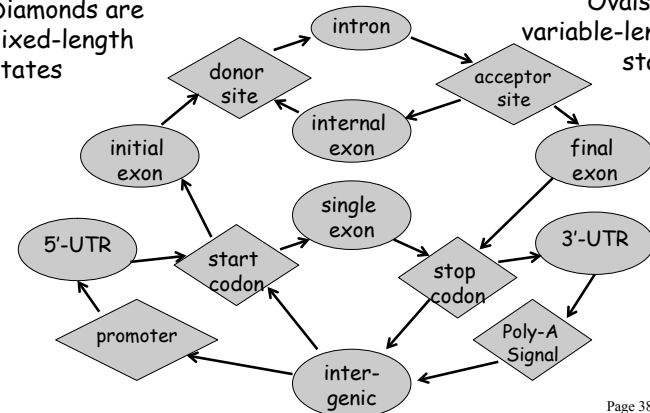
15-853

Page 37

## A "simple" GHMM for Gene Finding

Diamonds are  
fixed-length  
states

Ovals are  
variable-length  
states



<http://geneprediction.org/book/classroom.html>

Page 38