Vanilla Beam Decoding

k: beam size, M: maximum length,
$\mathcal{V}$: Vocabulary, $\text{score}(\cdot)$: scoring function.

$B_0 \leftarrow \{\langle 0, \text{BOS} \rangle\}$
for $t \in \{1, \ldots, M-1\}$:
  for $\langle s, \mathbf{y} \rangle \in B_{t-1}$:
    if $\mathbf{y}.\text{last}() = \text{EOS}$:
      $H.\text{add}(\langle s, \mathbf{y} \rangle)$
      continue
    for $y \in \mathcal{V}$:
      $s \leftarrow \text{score}(\mathbf{y} \circ y)$, $H.\text{add}(\langle s, \mathbf{y} \circ y \rangle)$
  $B_t \leftarrow \varnothing$
  while $|B_t| < k$: # Find top k from H.
    $\langle s, \mathbf{y} \rangle \leftarrow H.\max()$, $B_t.\text{add}(\langle s, \mathbf{y} \rangle)$
    $H.\text{remove}(\langle s, \mathbf{y} \rangle)$
  if $\mathbf{y}.\text{last}() = \text{EOS}, \forall \mathbf{y} \in B_t$: # All finished.
    return $B_t.\max()$
return $B_t.\max()$