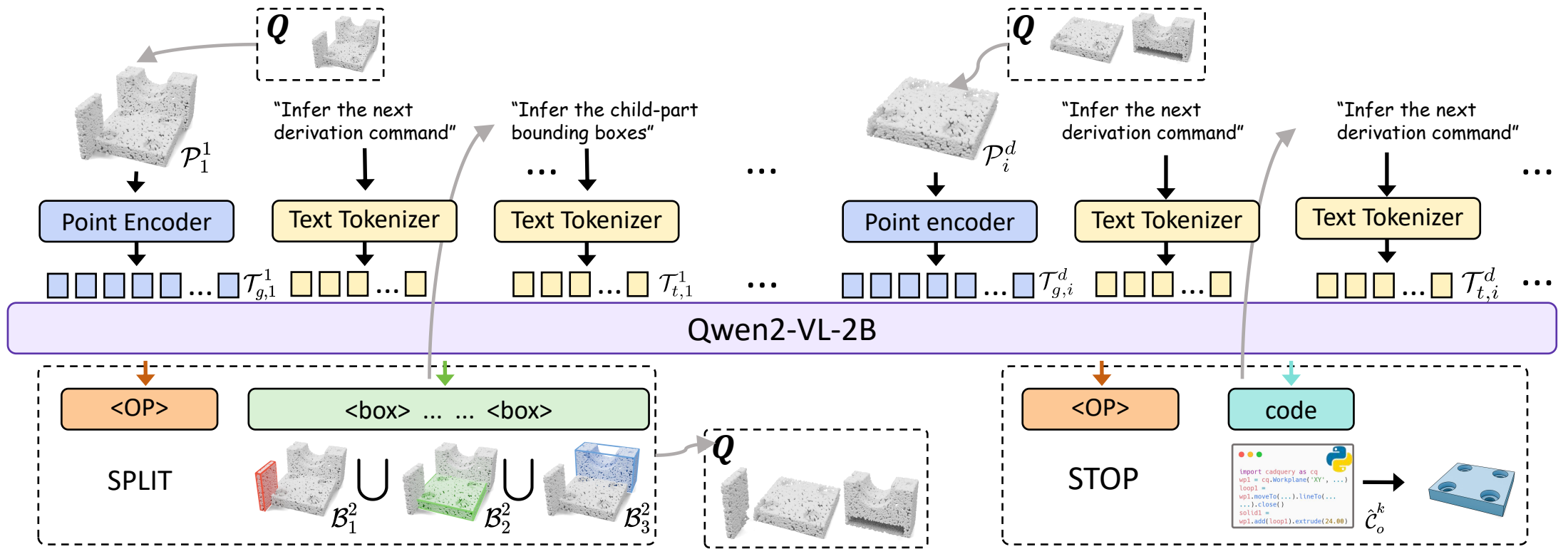
Q
$\mathcal{P}_1^1$
Point Encoder
"Infer the next derivation command"
Text Tokenizer
"Infer the child-part bounding boxes"
Text Tokenizer
$\mathcal{T}_{g,1}^1$
$\mathcal{T}_{t,1}^1$
Q
$\mathcal{P}_i^d$
Point encoder
"Infer the next derivation command"
Text Tokenizer
"Infer the next derivation command"
Text Tokenizer
$\mathcal{T}_{g,i}^d$
$\mathcal{T}_{t,i}^d$
Qwen2-VL-2B
<OP>
<box> ... ... <box>
SPLIT
$\mathcal{B}_1^2$
$\mathcal{B}_2^2$
$\mathcal{B}_3^2$
Q
<OP>
code
STOP
$\hat{\mathcal{C}}_o^k$