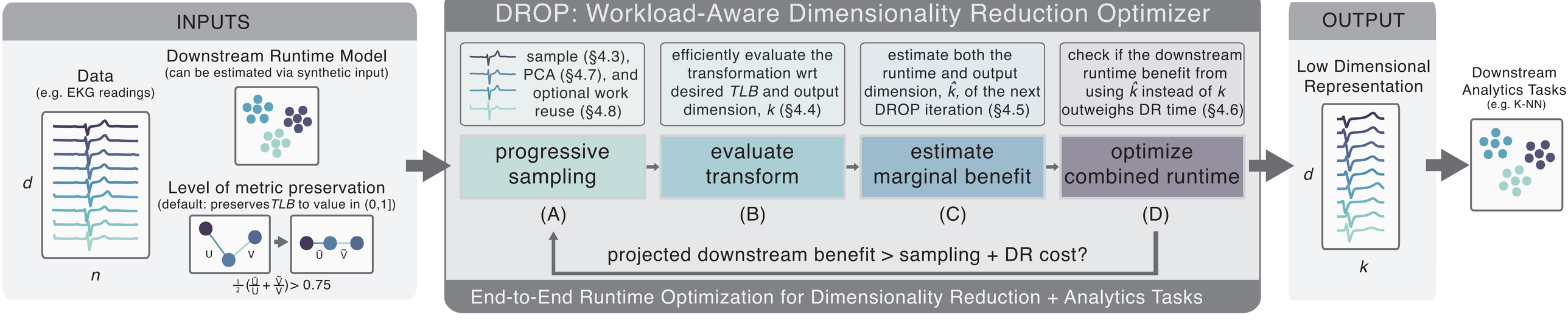
INPUTS
Data
(e.g. EKG readings)
d
n
Downstream Runtime Model
(can be estimated via synthetic input)
Level of metric preservation
(default: preserves TLB to value in (0,1])
U
V
Ũ
Ṽ
$\frac{1}{2}(\frac{\tilde{U}}{U}+\frac{\tilde{V}}{V})>0.75$
DROP: Workload-Aware Dimensionality Reduction Optimizer
sample (§4.3), PCA (§4.7), and optional work reuse (§4.8)
efficiently evaluate the transformation wrt desired TLB and output dimension, k (§4.4)
estimate both the runtime and output dimension, $\hat{k}$, of the next DROP iteration (§4.5)
check if the downstream runtime benefit from using $\hat{k}$ instead of k outweighs DR time (§4.6)
progressive sampling
evaluate transform
estimate marginal benefit
optimize combined runtime
(A)
(B)
(C)
(D)
projected downstream benefit > sampling + DR cost?
End-to-End Runtime Optimization for Dimensionality Reduction + Analytics Tasks
OUTPUT
Low Dimensional Representation
d
k
Downstream Analytics Tasks
(e.g. K-NN)