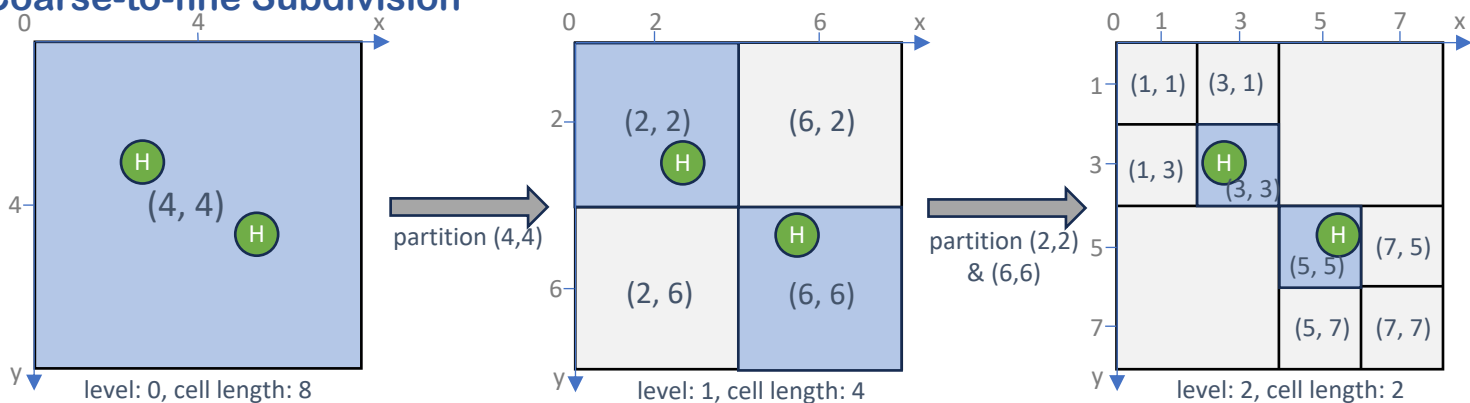
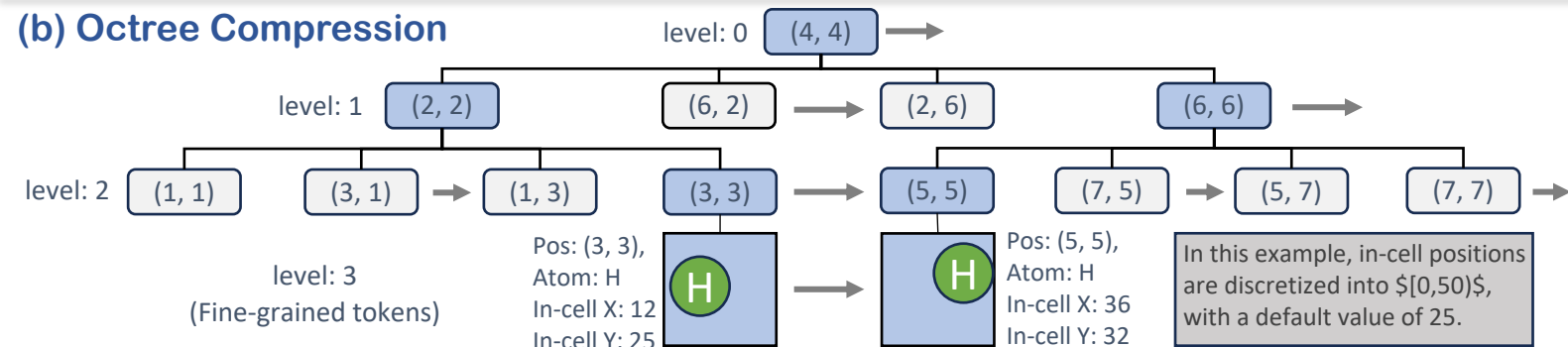


(a) Coarse-to-fine Subdivision



(b) Octree Compression



(c) Octree Spatial Tokens

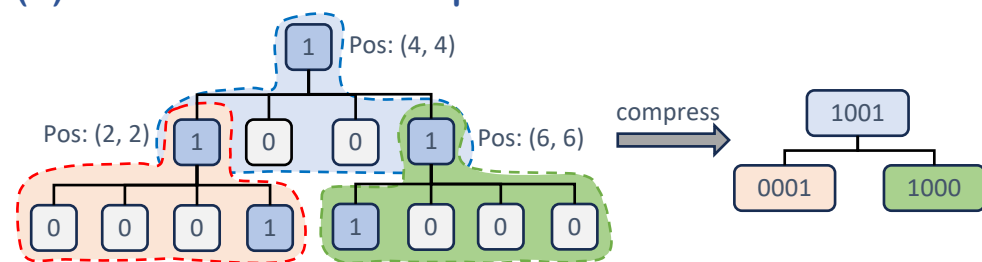
Pos	(4, 4)	(2, 2)	(6, 2)	(2, 6)	(6, 6)	(1, 1)	(3, 1)	(1, 3)	(3, 3)	(5, 5)	(7, 5)	(5, 7)	(7, 7)	(3, 3)	(5, 5)
Level	0	1	1	1	1	2	2	2	2	2	2	2	2	3	3
Type	1	1	0	0	1	0	0	0	1	1	0	0	0	H	H
In-cell X	25	25	25	25	25	25	25	25	25	25	25	25	25	12	36
In-cell Y	25	25	25	25	25	25	25	25	25	25	25	25	25	25	32

Sequence Order (Level-Wise)

Octree tokens

Fine-grained tokens

(d) 2-Level Subtree Compression



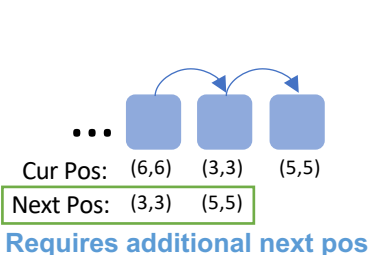
Compressed Tokens:

Pos	(4, 4)	(2, 2)	(6, 6)	(3, 3)	(5, 5)
Level	0	1	1	3	3
Type	9	8	1	H	H
In-cell X	25	25	25	12	36
In-cell Y	25	25	25	25	32

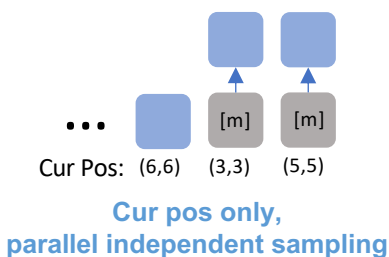
(e) Masked Next-Token Prediction

Challenge: dynamic token positions across samples

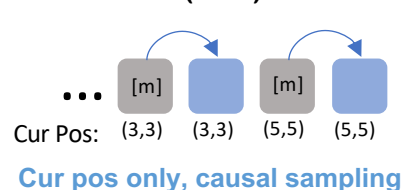
Next-Token Prediction



Masked Prediction



Masked Next-Token Prediction (Our)



(f) Unified Framework for 3D GU

