PhD Dissertation Defense

Performance Optimization of Transient Pressure Acoustic Simulation by Coarse-grain Diffusion Model

Ian Flood
Construction Management
University of Florida

Hassan Azad
Architecture
University of Florida

Chaofeng Wang
Construction Management
University of Florida

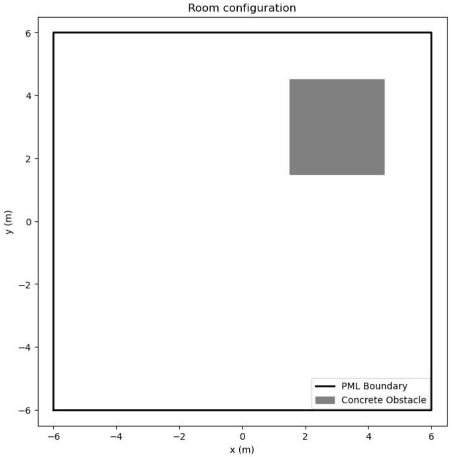
Kejun Huang
Computer Science
University of Florida

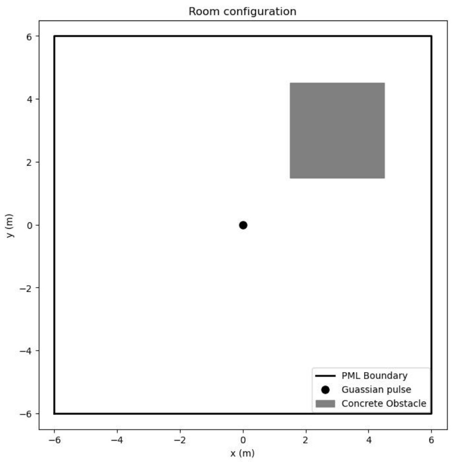
Qian Shi, Ph.D. Candidate

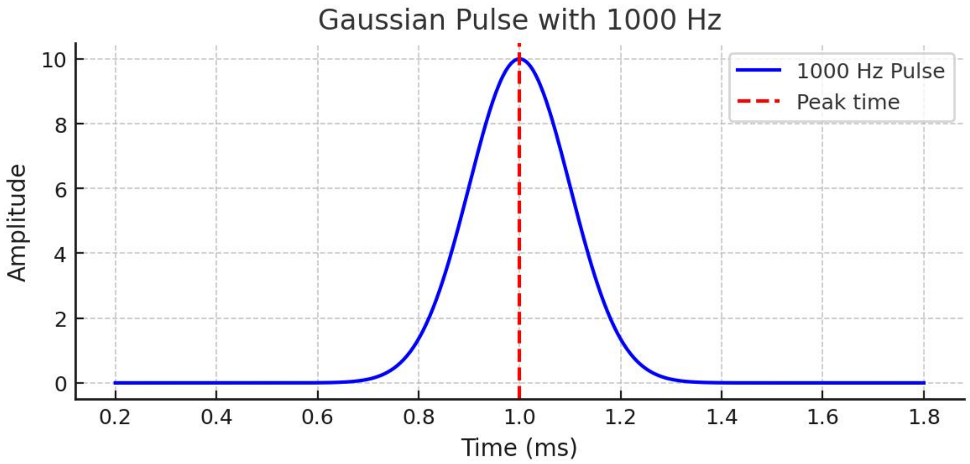
University of Florida

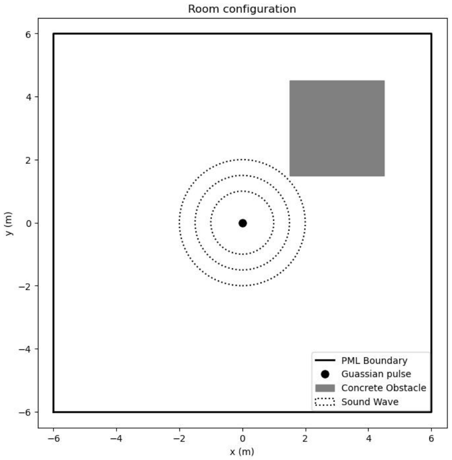
UF UNIVERSITY of FLORIDA

What is the transient acoustic pressure simulation ?

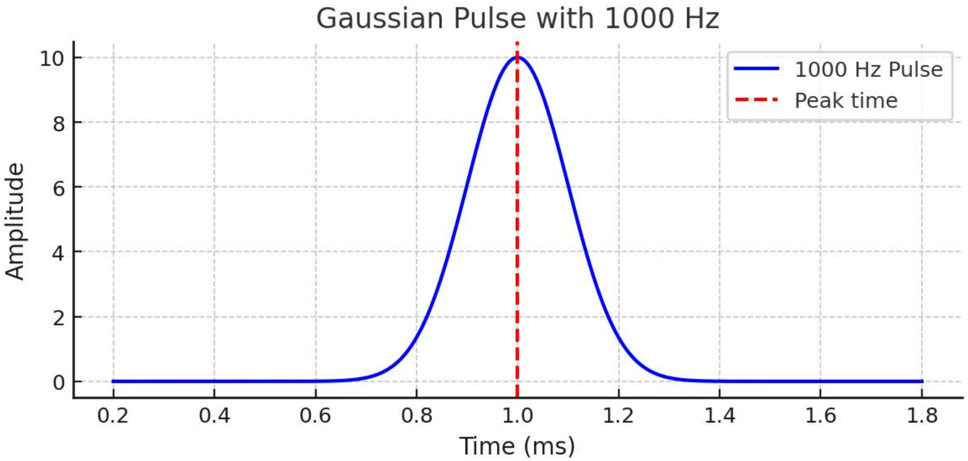

Room configuration
PML Boundary
Concrete Obstacle
y (m)
x (m)
6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6


Room configuration
y (m)
x (m)
PML Boundary
Guassian pulse
Concrete Obstacle

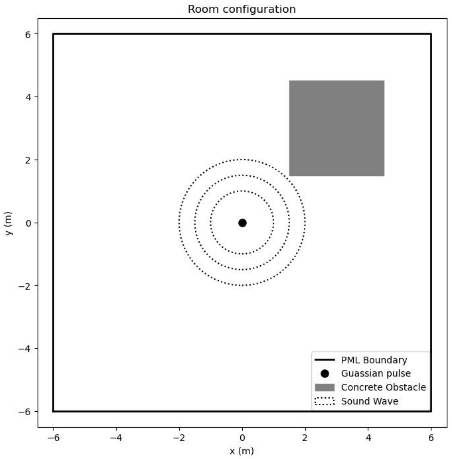

Gaussian Pulse with 1000 Hz
Amplitude
Time (ms)
1000 Hz Pulse
Peak time

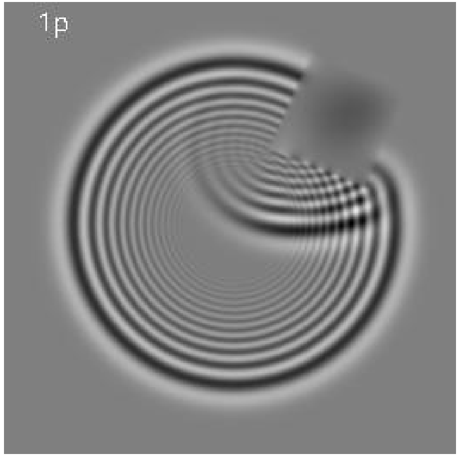


Room configuration
PML Boundary
Guassian pulse
Concrete Obstacle
Sound Wave
y (m)
x (m)



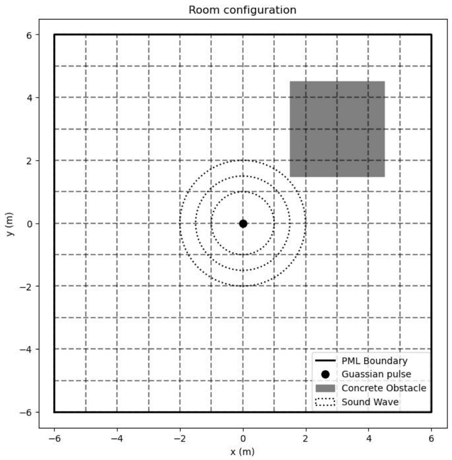
Gaussian Pulse with 1000 Hz
1000 Hz Pulse
Peak time
Amplitude
Time (ms)





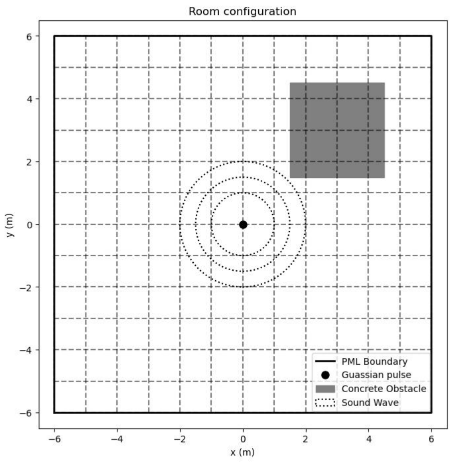
- Reflection
- Diffraction
- Refraction
- Absorption
- Transmission

Spatial-temporal data



Finite element method

Comsol multiphysics



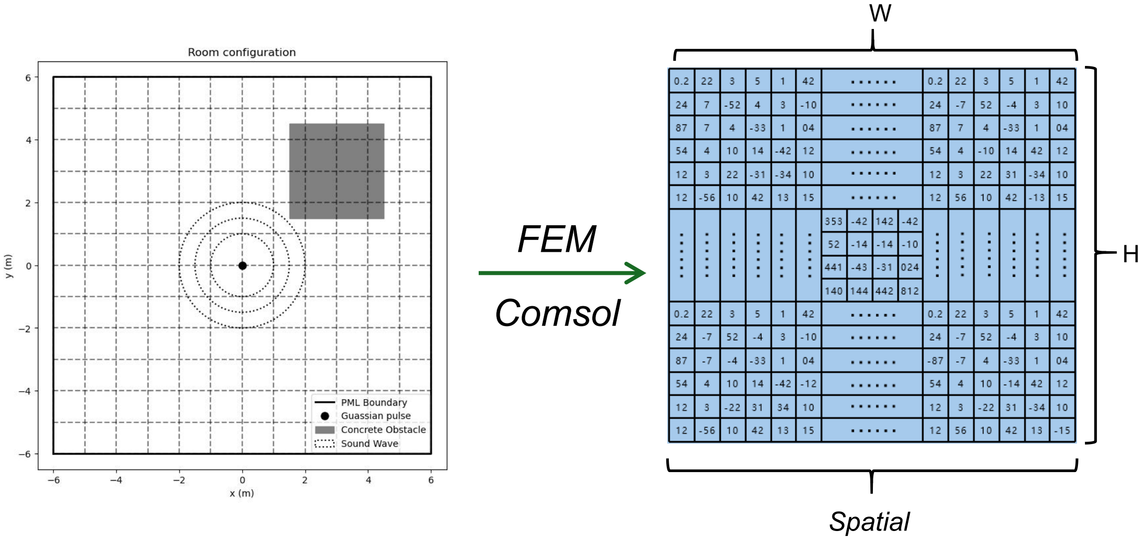
Finite element method

Comsol multiphysics

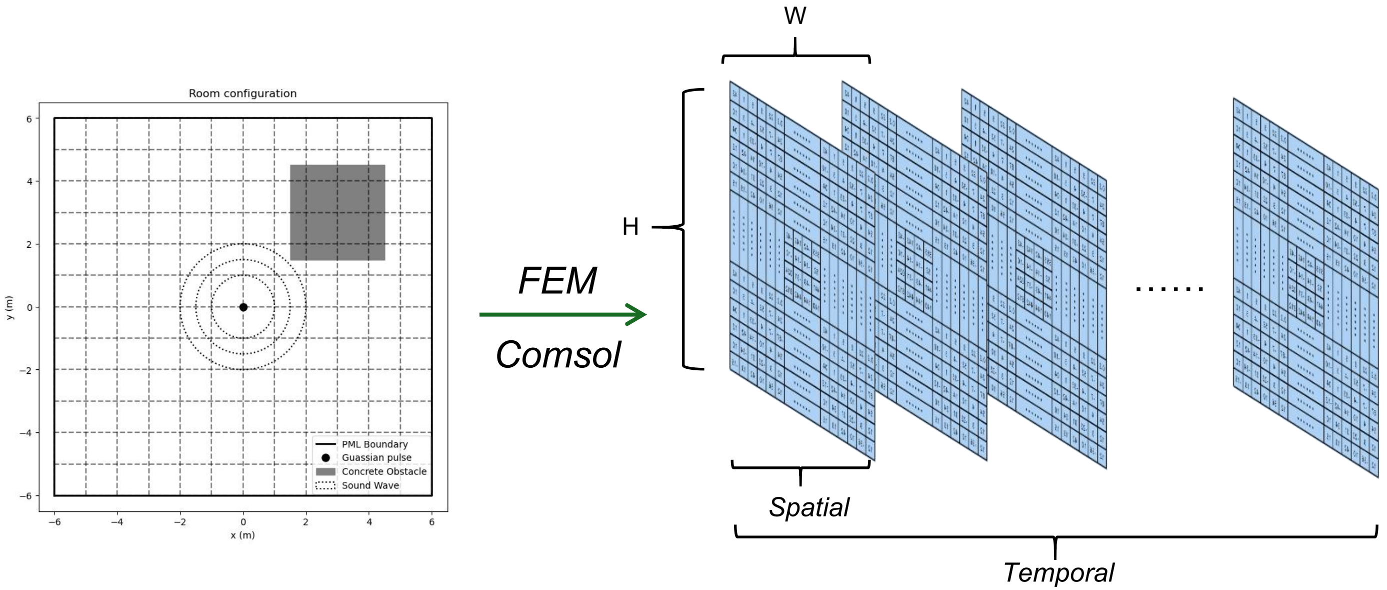
$$\frac{\partial^2 p}{\partial t^2} + \sigma(x,y)\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left(\frac{1}{\rho(x,y)}\frac{\partial p}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{1}{\rho(x,y)}\frac{\partial p}{\partial y}\right)$$

- *Wave equation (PDE) governing Time-dependent propagation, reflection, and interaction of sound waves.*

(Morse & Ingard, 1986)



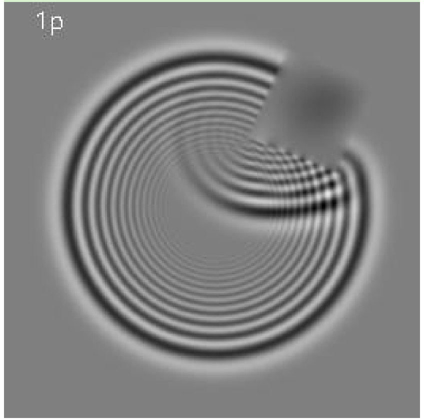
Room configuration
PML Boundary
Guassian pulse
Concrete Obstacle
Sound Wave
x (m)
y (m)
FEM
Comsol
W
H
Spatial



Spatial-temporal data

Other Spatial-temporal data ?

Transient pressure field



$$\frac{\partial^2 p}{\partial t^2} = c^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \right)$$

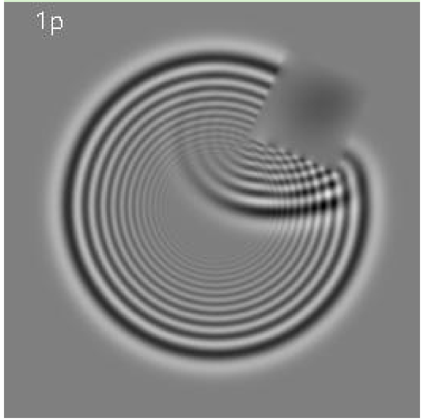
p(x,y,t)

KTH dataset (CV)

$$I(x, y, t+1) \approx \mathcal{F}(I(x, y, t), I(x, y, t-1), \ldots)$$

No explicit physical equation

Transient pressure field



Vortex Evolution in a Planar Jet Flow (fluid)

$$\frac{\partial^2 p}{\partial t^2} = c^2 \left(\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} \right)$$

p(x,y,t)

$$\frac{\partial \mathbf{u}}{\partial t} = -\mathbf{u} \cdot \nabla \mathbf{u} - \nabla p + \nu \nabla^2 \mathbf{u}$$

u(x,y,t) *p(x,y,t)*

why is transient acoustic pressure so important ?

① Optimizing Room Acoustics for Speech and Music

- Simulates acoustics to optimize audio clarity in concert halls, classrooms, and offices.

(Gade, 2014)

② Understanding and Controlling Noise & Vibration

- Enhances sound insulation by reducing noise transmission between rooms.

(Vér & Beranek, 2005)

③ Enhancing HVAC System Design

- Reduces noise from HVAC systems and machinery to improve indoor comfort.

(Yao, 2016)

Current Challenge !

① Physically-based Simulation Are Still Slow

- Obtain High-quality acoustic Results taking a lot of time Courant-Friedrichs-Lewy （CFL）

(Courant, R., Friedrichs, K., & Lewy, H. 1928)

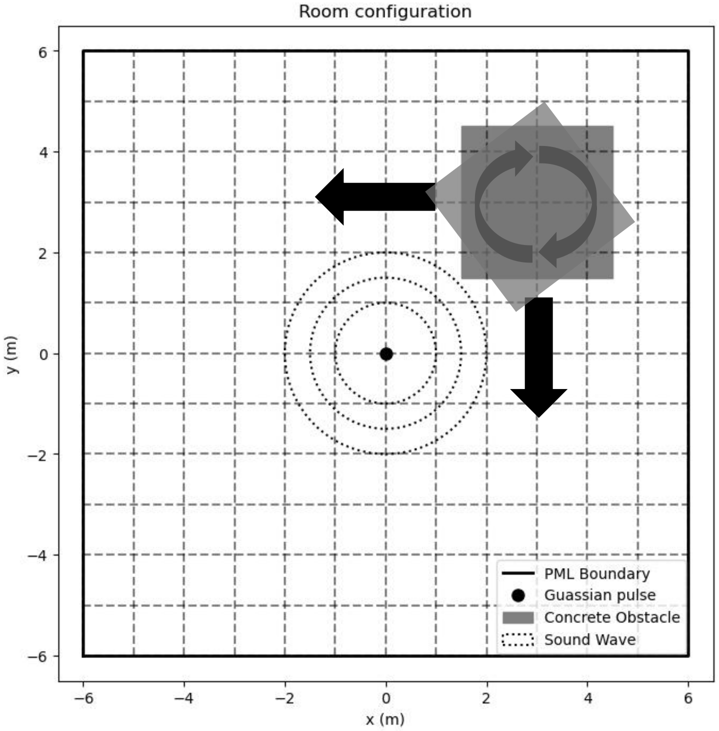
Frequency F	*1000 HZ*	*2500 HZ*	*5000 HZ*	*10000 HZ*
Wave length λ (m)	*0.343*	*0.1372*	*0.0686*	*0.0343*
Wave period T (s)	*0.001*	*0.0004*	*0.0002*	*0.0001*
mesh size Δx (m) ($\Delta x \leq \lambda/6$)	*0.0570*	*0.0229*	*0.0114*	*0.0057*
Number of element (12 x 12 m)	*210 x 210*	*524 x 524*	*1048 x 1048*	*2096 x 2096*
Step size Δt (s) ($\Delta t \leq T/20$)	*0.00005*	*0.00002*	*0.00001*	*0.000005*
Number of time instances (0.01s)	*200*	*500*	*1000*	*2000*

Model	***CPU Mem***	***CPU core***	***CPU Hours***
Comsol FEM	*8 G*	*AMD EPYC 9654 96-Core*	*6.23 h*

② Limited Support For Follow-up Operation

- Changing An Existing Simulation Entails Error

(Kim & Delaney, 2013)



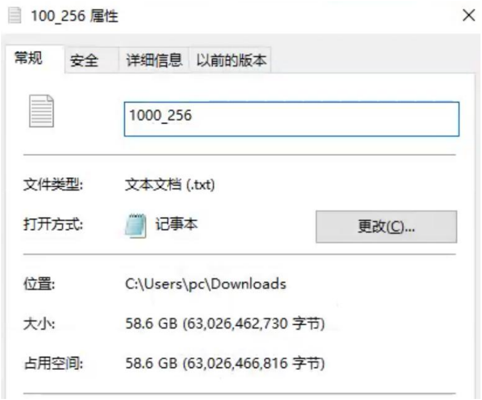
I. Orientation
II. Location

③ Producing Data Consumes Large Amount Of Storage Space

- Growing Need To Reuse Stored Simulation Data

(Kim & Delaney, 2013)

Frequency F	*1000 HZ*
Wave length λ (m)	*0.343*
Wave period T (s)	*0.001*
mesh size Δx (m) ($\Delta x \le \lambda/6$)	*0.0570*
Number of element (12 x 12 m)	*210 x 210*
Step size Δt (s) ($\Delta t \le T/20$)	*0.00005*
Number of time instances (0.02s)	*1000*



Main challeges need to be solved:

① Slow simulation speed

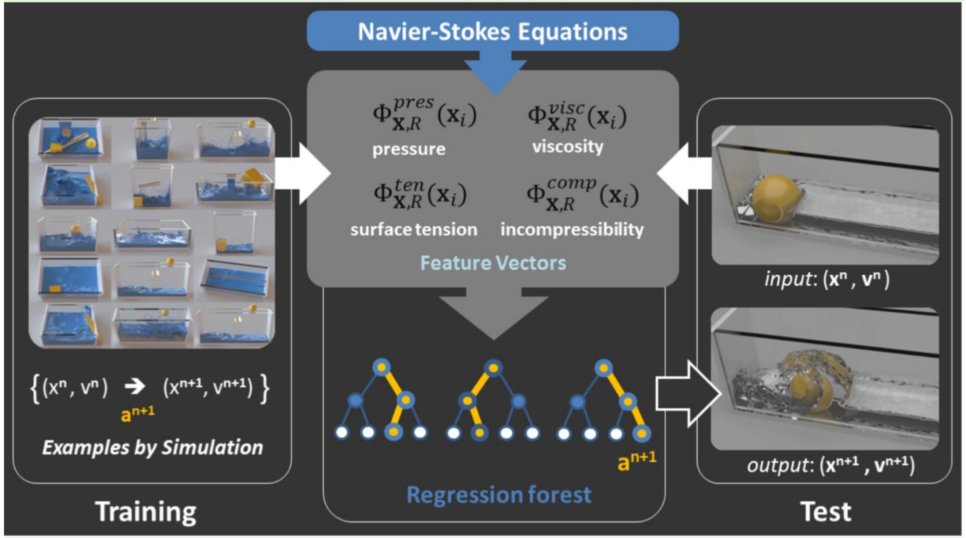
② Lack of interpolation ability

③ large storage space

How to solve them ? Machine learning !

Previous studies proposed some models :

Data-driven Fluid Simulations using Regression Forests



First ML model for Spatial temporal data (fluid) by (Ladický et al., 2015)

◆ First ML model for fluid

◆ Data-driven approximation of physics

☒ High Memory for Training

☒ Limited generalization

Speed up simulation	Interpolation ability	Save storage
✓	✓	✗

CNN?

☑ Performs well on unseen convex shapes

☑ Effectively captures diffraction, occlusion, and spatial coherence.

☒ No time-line modeling

☒ Non-autoregressive modeling,

End to End

Speed up simulation	Interpolation ability	Save storage
✓	✓	✗

Fast acoustic scattering unsing CNN

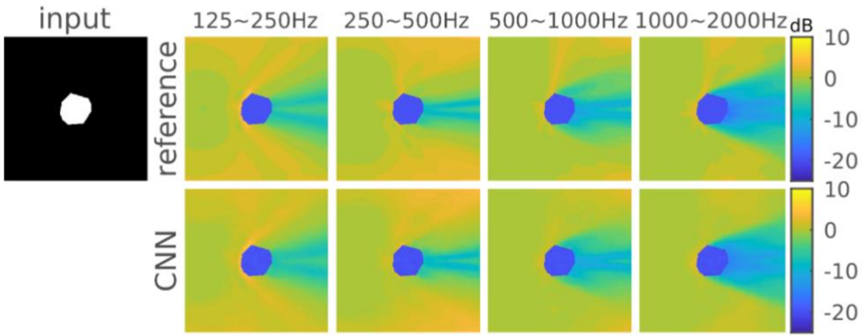


Fig. 1: Acoustic scattering formulated as 2D image-to-image regression. Input object shape is specified as a binary image (left). A point source, not shown, is placed to the left of the object. Numerical wave simulation is used to produce reference scattered loudness fields in frequency bands (top row). Our CNN produces a close approximation at over 100× speedup (bottom row).

CNN mapping for acoustic frequency loudness field by (Fan et al., 2020)

Latent Space Physics: Towards Learning the Temporal Evolution of Fluid Flow

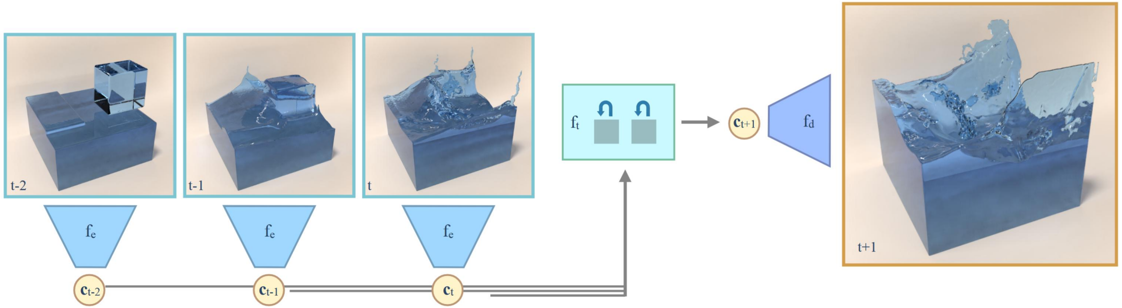


Figure 1: *Our method encodes multiple steps of a simulation field, typically pressure, into a reduced latent representation with a convolutional neural network. A second neural network with LSTM units then predicts the latent space code for one or more future time steps, yielding large reductions in runtime compared to regular solvers.*

Speed up simulation	✓
Interpolation ability	✓
Save storage	✓

(Wiewel et al., 2019)

RNN ?

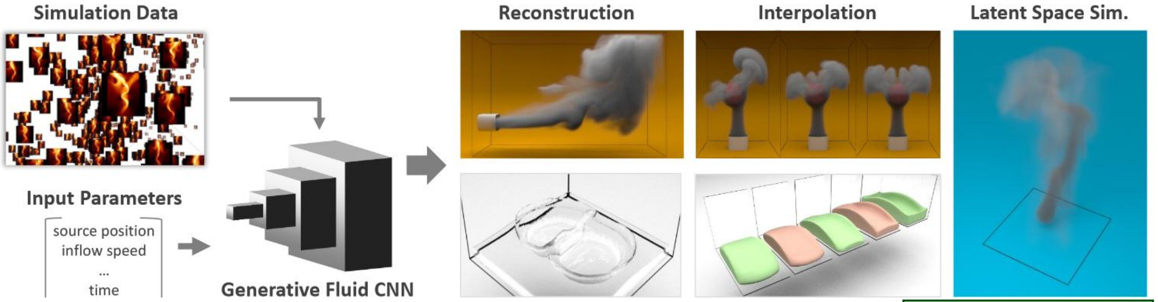
☑ Improved Spatiotemporal Modeling

☑ Generalizes to new scenarios with realistic results

☒ Increased Computational Complexity

☒ Training Instability

Deep Fluids: A Generative Network for Parameterized Fluid Simulations



(Kim et al., 2019)

Speed up simulation	✓
Interpolation ability	✓
Save storage	✗

GAN ?

☑ High-Resolution Generation
☑ Interpolation and Extrapolation
☑ Add reference point to prevent data draft

☒ Training Instability
☒ High Computational Cost

Modeling Heat Transfer in Fluid Systems Using Vision Transformers and U-Net

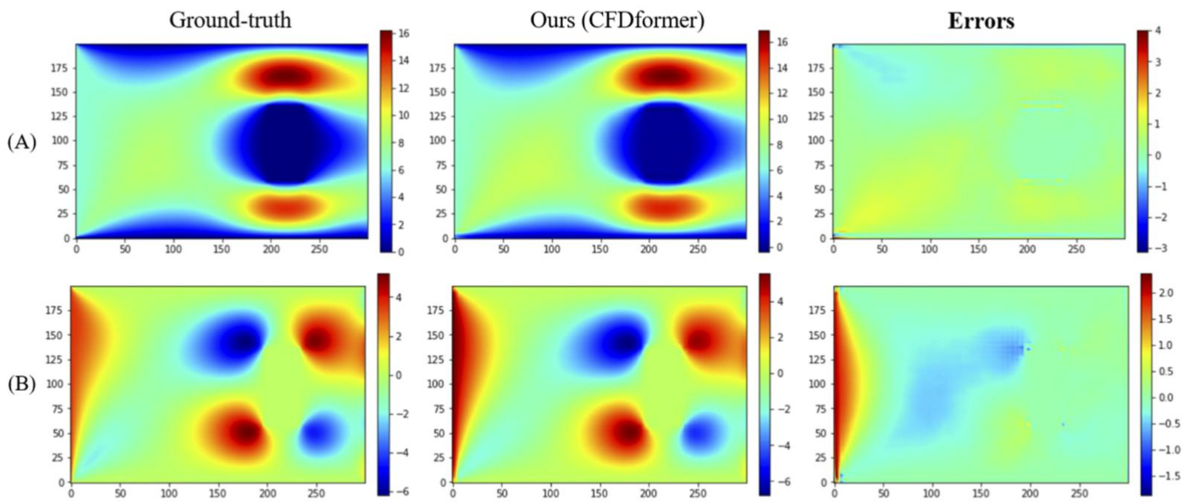


FIG. 7. The interpolation results for two velocity components at $u_{x_0} = 0.23$ (m/s) and $u_{y_0} = 0.17$ (m/s) (a) $u_{x(m/s)}$ (b) $u_{y\,(m/s)}$.

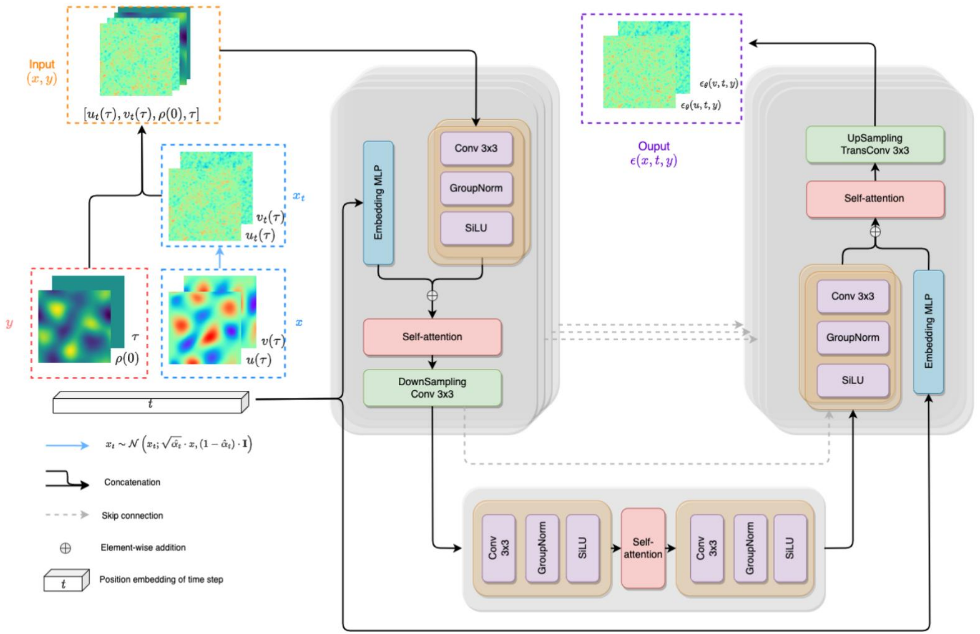
Speed up simulation	✗
Interpolation ability	✓
Save storage	✓

(Kang et al., 2023)

Transform?
☑ Handles both local and global features (via CNN + ViT)
☑ Stable Training

☒ Slower Generation Speed Compared to GANs

A Denoising Diffusion Model for Fluid Field Prediction



(Yang & Sommer, 2023)

Speed up simulation	✓
Interpolation ability	✓
Save storage	✗

Diffusion Models Beat GANs
on Image Synthesis
(Dhariwal & Nichol, 2021)

☑ High resolution generation
☑ More Stable Training

☒ Non-autoregressive, directly inference

Research goal:

An auto-regressive model is developed to efficiently reconstruct spatio-temporal sequences of transient acoustic pressure, with built-in interpolation capability. To further evaluate the model's approximation ability, two additional experiments on complex scene are conducted.

Speed up simulation	Interpolation ability	Save storage
✓	✓	✓

Research goal:

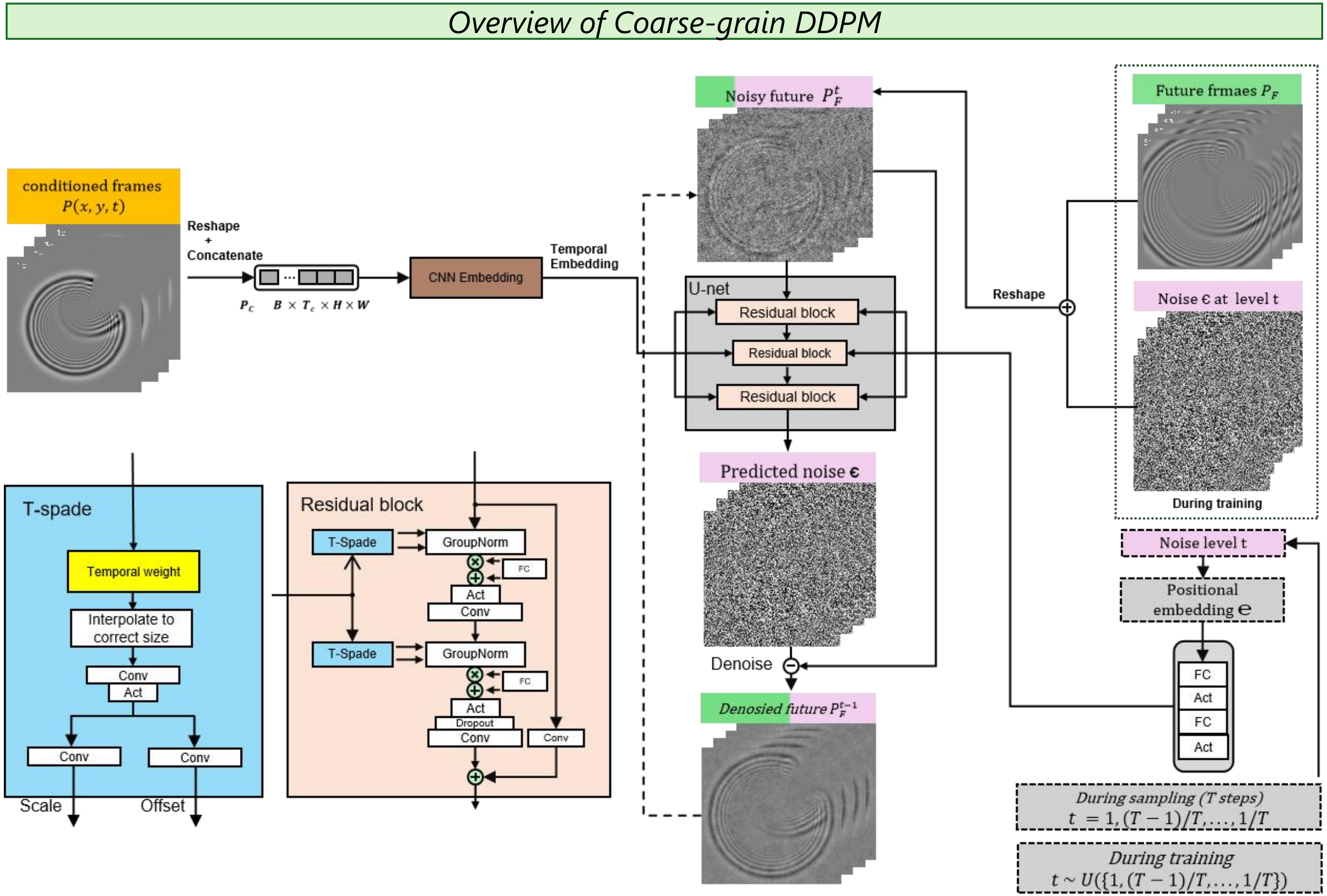
An auto-regressive model is developed to efficiently reconstruct spatio-temporal sequences of transient acoustic pressure, with built-in interpolation capability. To further evaluate the model's approximation ability, two additional complex experiments are conducted.

- Objective 1: Build a T-spade enhanced DDPM to auto-regressively predict acoustic pressure in a 2D room.
- Objective 2: A coarse-grained integrated method is used to accelerate simulation while maintaining acceptable accuracy.
- Objective 3: An extended, more complex scene is used to further evaluate the model's approximation and interpolation ability

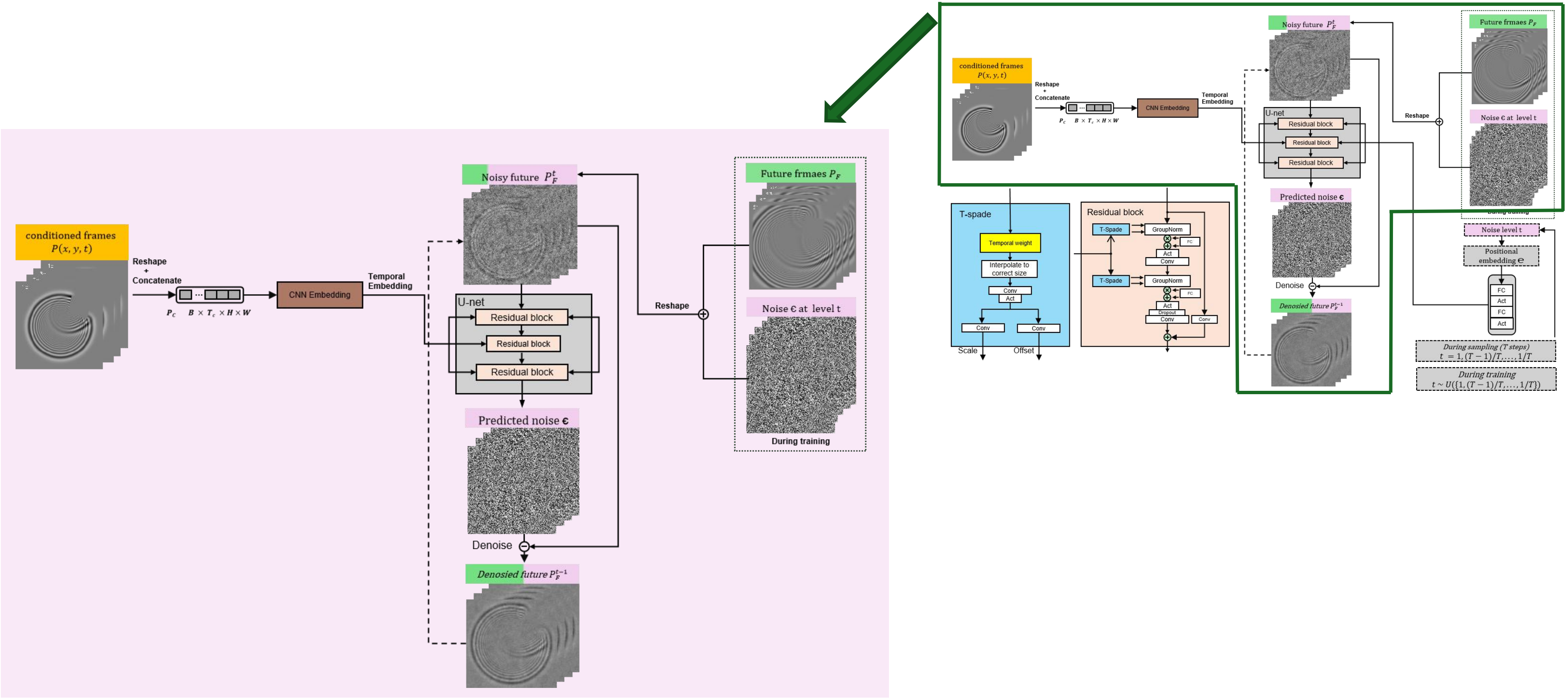
Methodology

Objective 1:

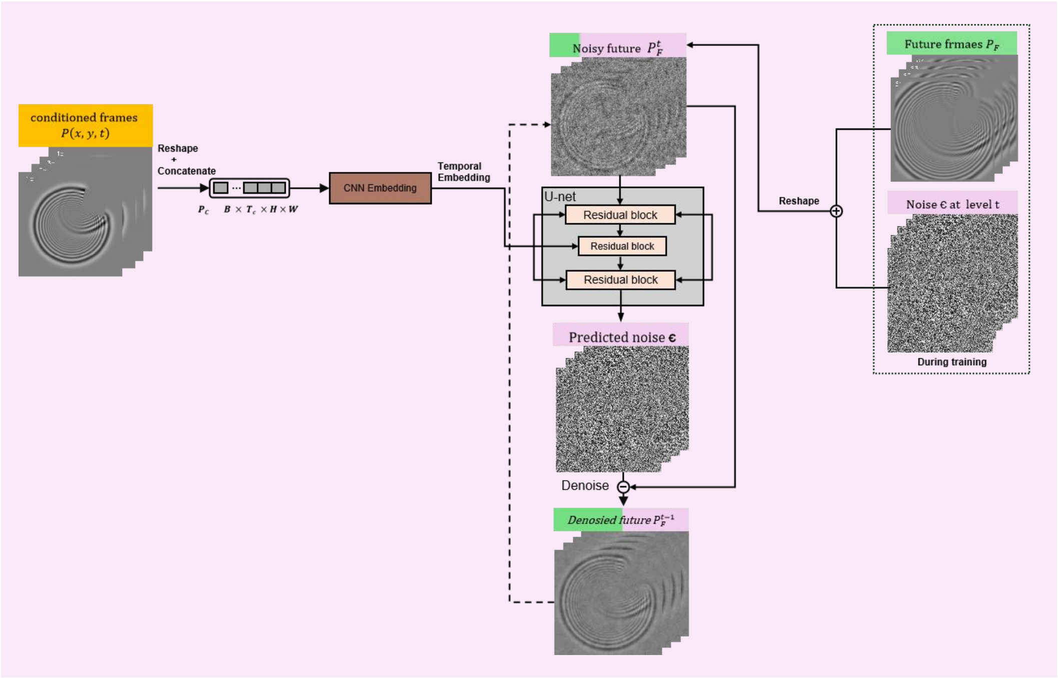
Build a T-spade enhanced DDPM to auto-regressively predict sound propagation in a 2D room and and support interpolation beyond observed regions.

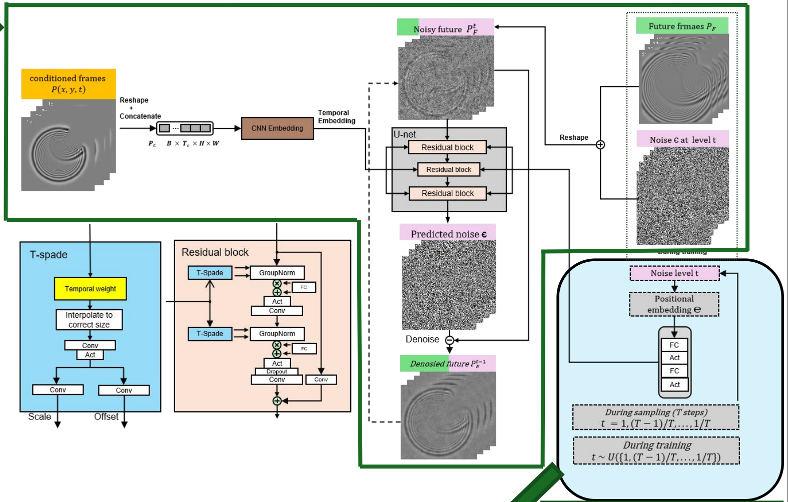


Overview of Coarse-grain DDPM
conditioned frames
$P(x, y, t)$
Reshape
+
Concatenate
P_C
$B \times T_c \times H \times W$
CNN Embedding
Temporal
Embedding
Noisy future P_F^t
U-net
Residual block
Residual block
Residual block
Predicted noise ϵ
Denoise
Denoised future P_F^{t-1}
Reshape
Future frmaes P_F
Noise ϵ at level t
During training
Noise level t
Positional
embedding e
FC
Act
FC
Act
During sampling (T steps)
$t = 1, (T-1)/T, ..., 1/T$
During training
$t \sim U(\{1, (T-1)/T, ..., 1/T\})$
T-spade
Temporal weight
Interpolate to
correct size
Conv
Act
Conv
Conv
Scale
Offset
Residual block
T-Spade
GroupNorm
FC
Act
Conv
T-Spade
GroupNorm
FC
Act
Dropout
Conv
Conv

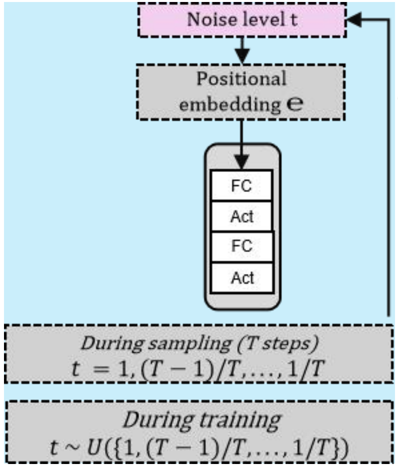


conditioned frames $P(x, y, t)$
Reshape + Concatenate
P_c
$B \times T_c \times H \times W$
CNN Embedding
Temporal Embedding
Noisy future P_F^t
U-net
Residual block
Residual block
Residual block
Predicted noise ϵ
Denoise
Denosied future P_F^{t-1}
Future frmaes P_F
Reshape
Noise ϵ at level t
During training
T-spade
Temporal weight
Interpolate to correct size
Conv
Act
Conv
Conv
Scale
Offset
Residual block
T-Spade
GroupNorm
FC
Act
Conv
T-Spade
GroupNorm
FC
Act
Dropout
Conv
Conv
Noise level t
Positional embedding ϵ
FC
Act
FC
Act
During sampling (T steps)
$t = 1, (T-1)/T, \ldots, 1/T$
During training
$t \sim U(\{1, (T-1)/T, \ldots, 1/T\})$

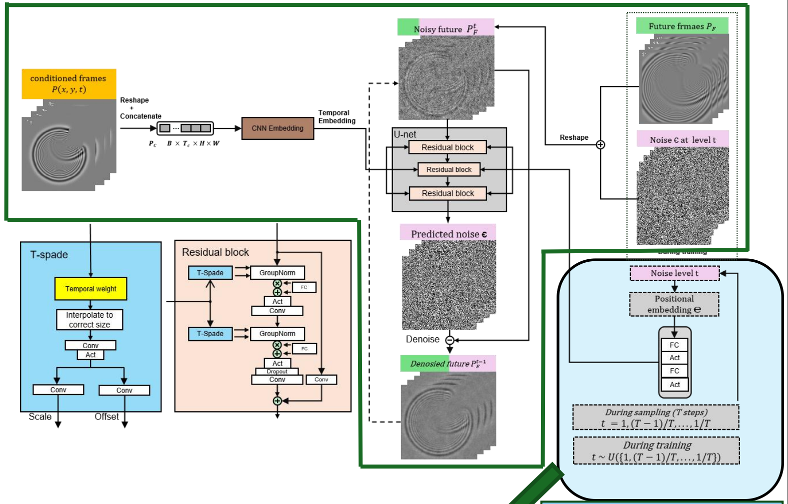




(Voleti et al., 2022)



Encode time information
(which diffusion step the model is at)



(Voleti et al., 2022)

Denoising Diffusion Implicit Models

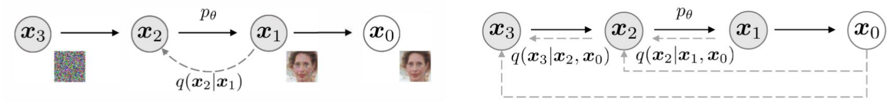


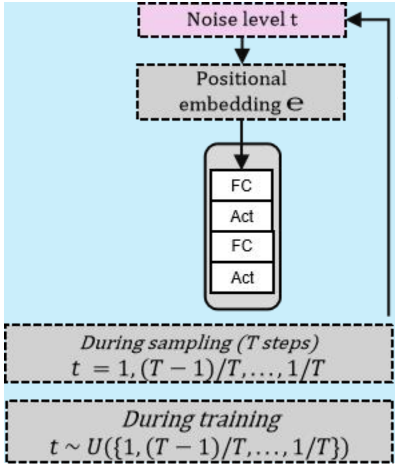
Figure 1: Graphical models for diffusion (left) and non-Markovian (right) inference models.

① Deterministic sampling: no randomness at each step.

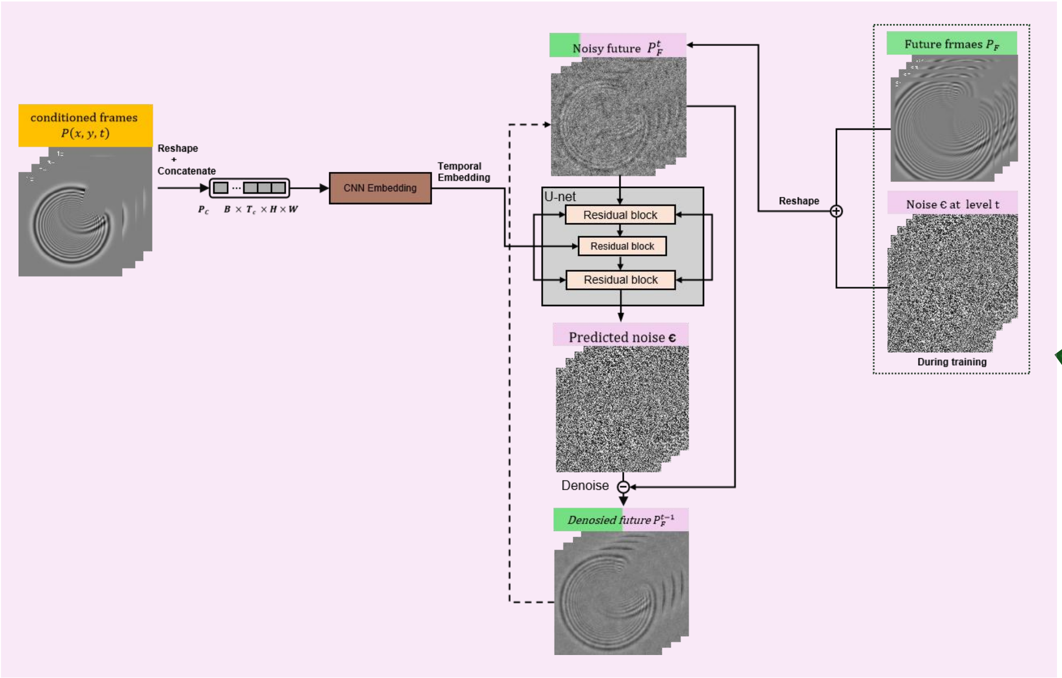
② Enables large step sizes and fewer total steps.

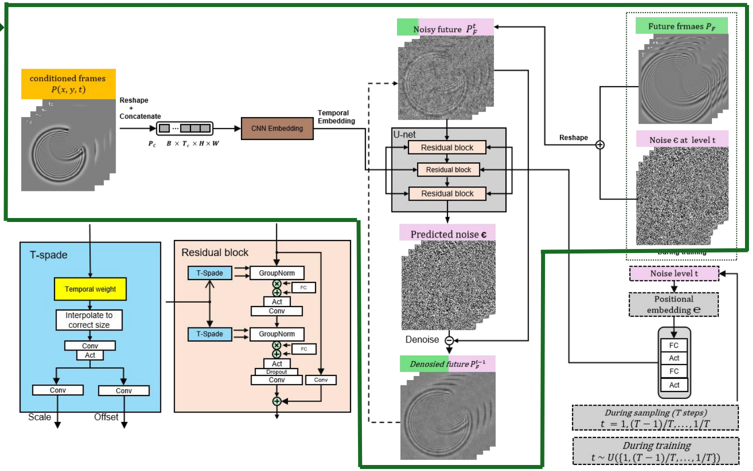
③ Can generate high-quality samples in ~50 steps vs. 1000 in DDPM.

(Song et al., 2022)

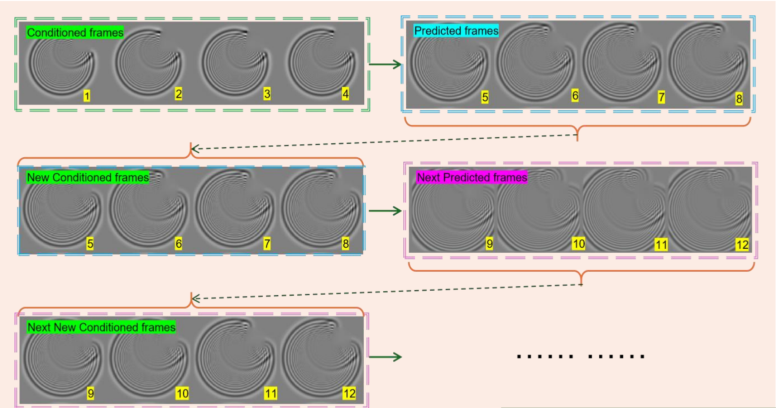


Encode time information
(which diffusion step the model is at)

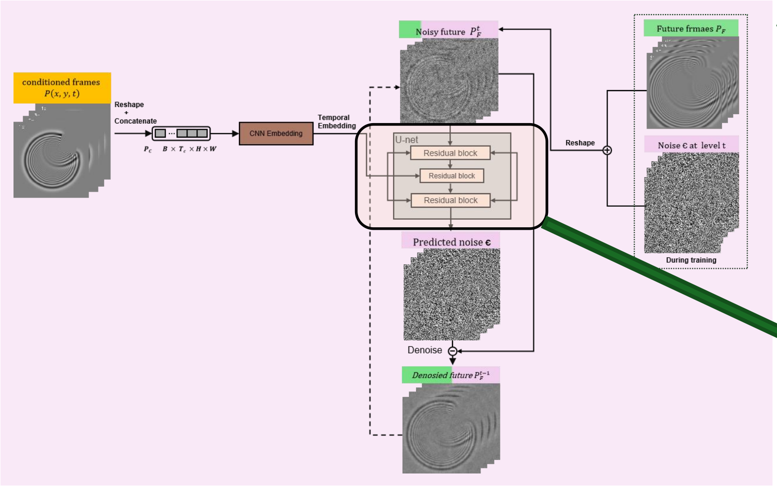


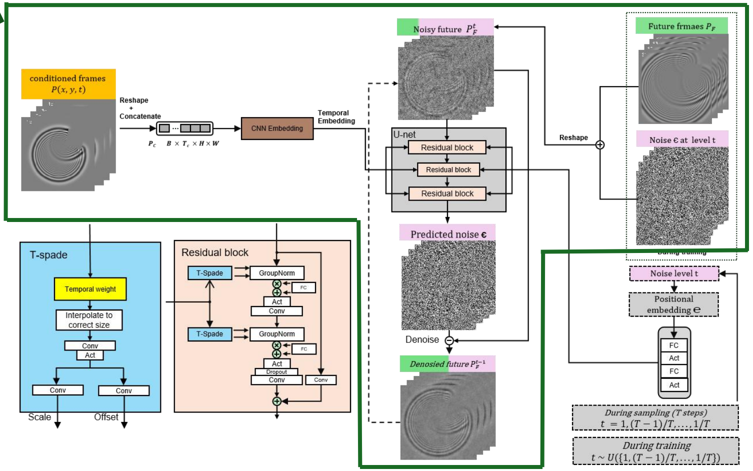


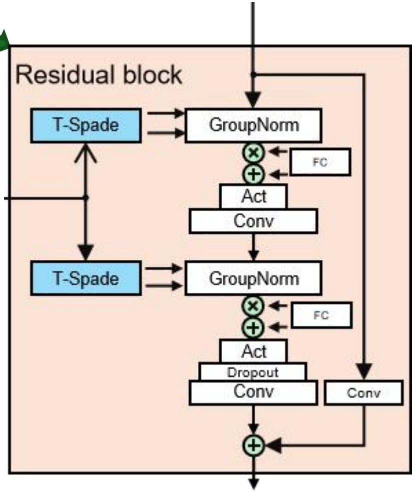
(Voleti et al., 2022)



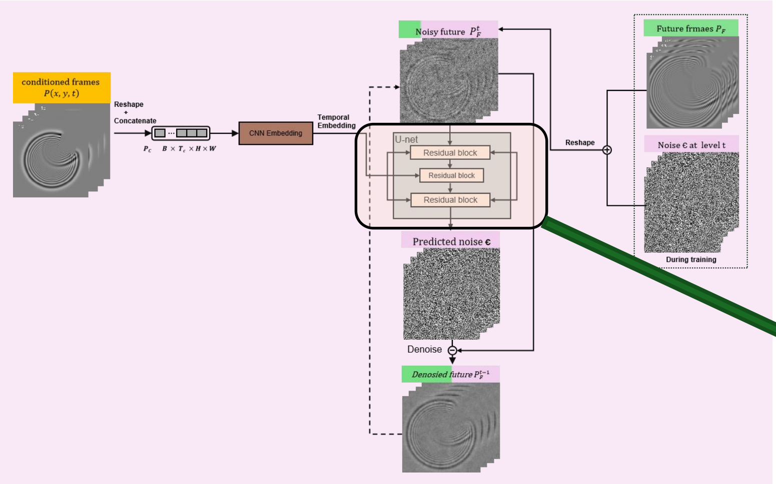
Auto-regressive flow

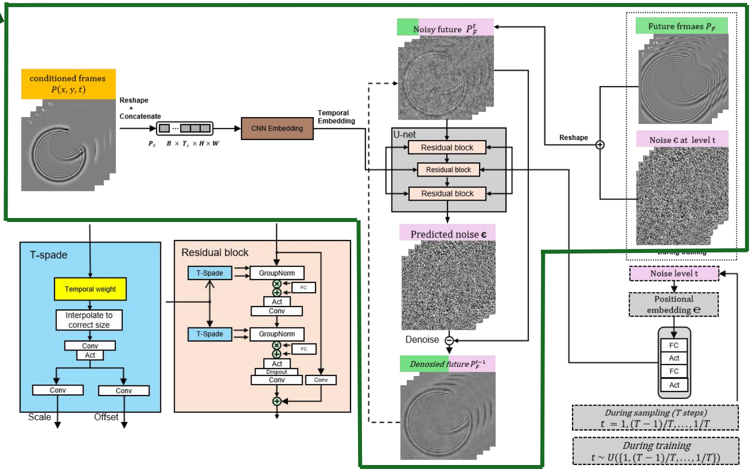






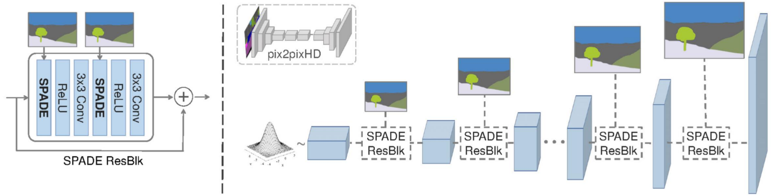
(Voleti et al., 2022)



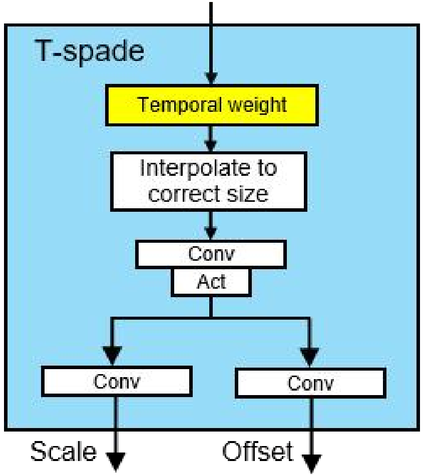


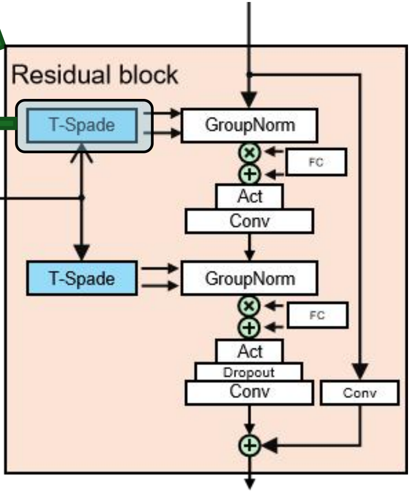
(Voleti et al., 2022)

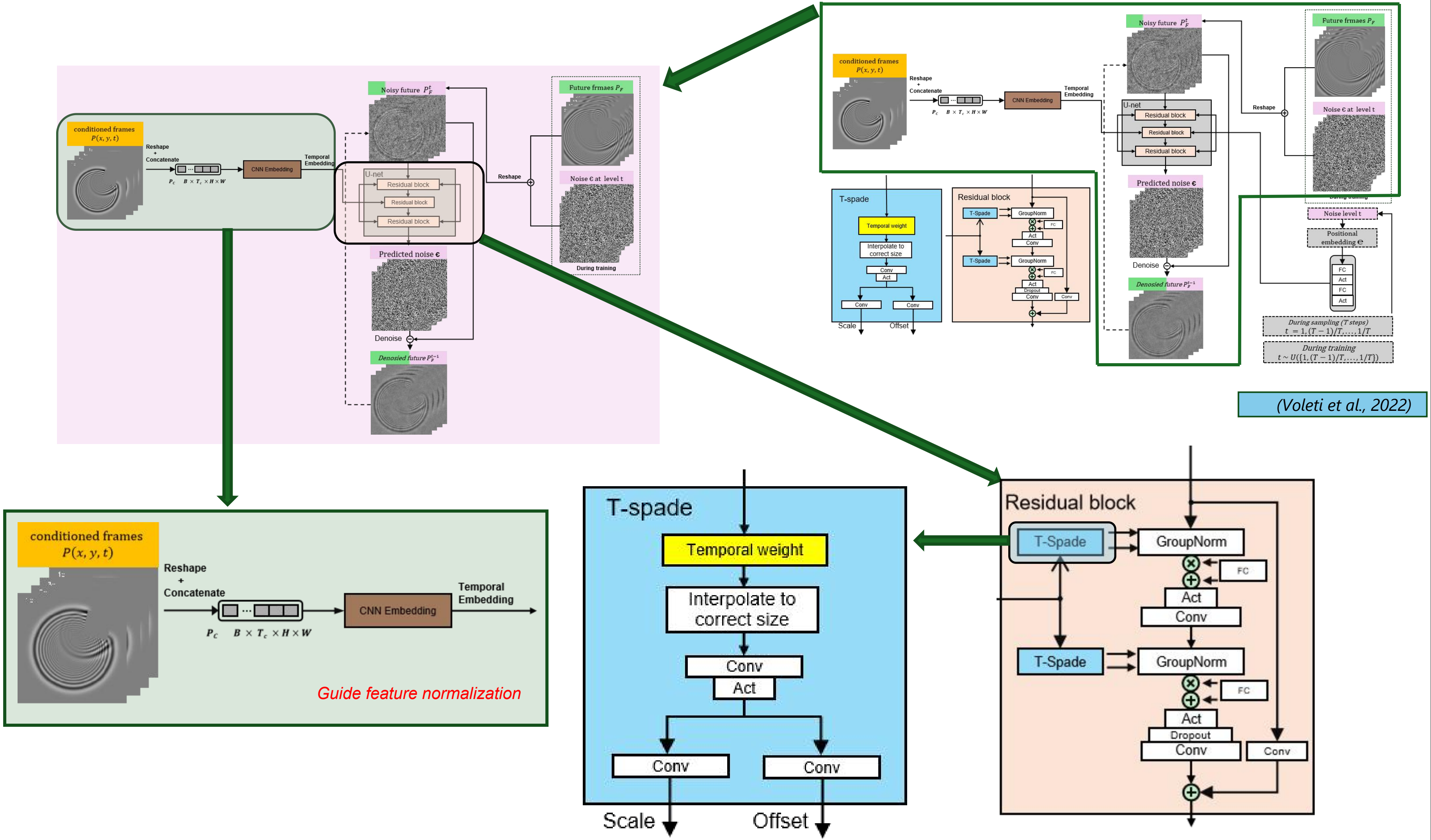
Semantic Image Synthesis With Spatially-Adaptive Normalization



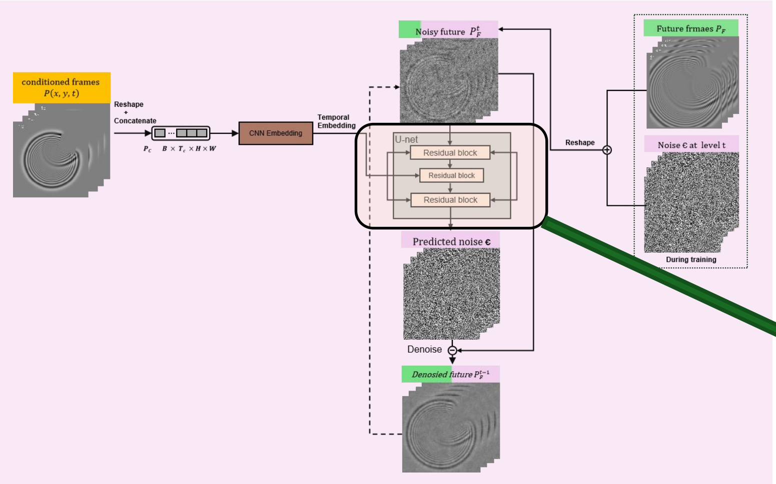
(Park et al., 2019)

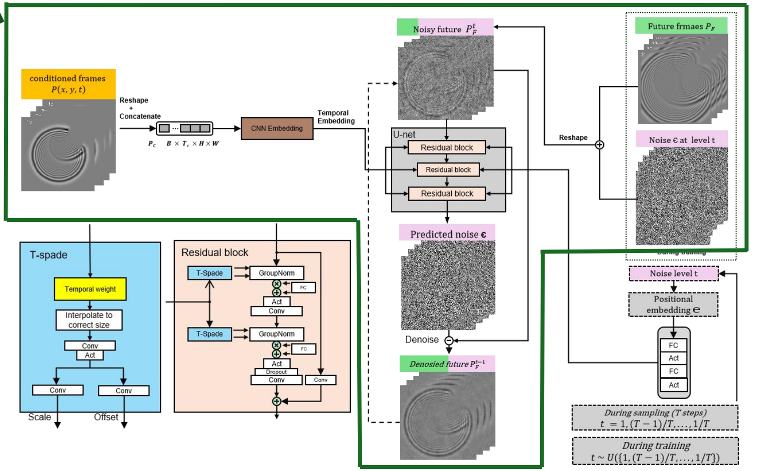






conditioned frames
P(x, y, t)
Reshape
+
Concatenate
P_c
B × T_c × H × W
CNN Embedding
Temporal Embedding
Noisy future P_F^t
U-net
Residual block
Residual block
Residual block
Predicted noise ϵ
Denoise
Denosied future P_F^{t-1}
Future frmaes P_F
Noise ϵ at level t
Reshape
During training
Noise level t
Positional embedding e
FC
Act
FC
Act
During sampling (T steps)
t = 1, (T − 1)/T, ..., 1/T
During training
t ~ U({1, (T − 1)/T, ..., 1/T})
(Voleti et al., 2022)
Guide feature normalization
T-spade
Temporal weight
Interpolate to correct size
Conv
Act
Conv
Conv
Scale
Offset
Residual block
T-Spade
GroupNorm
FC
Act
Conv
T-Spade
GroupNorm
FC
Act
Dropout
Conv
Conv

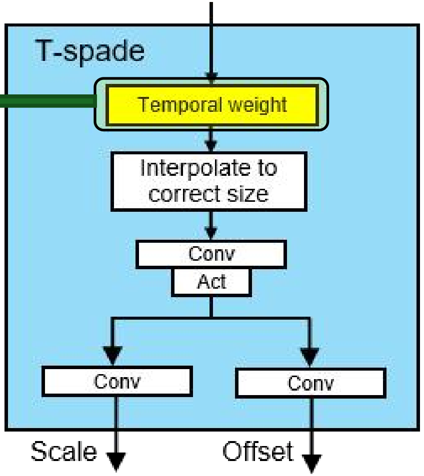


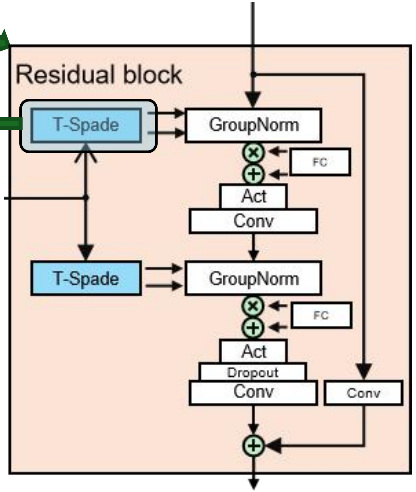


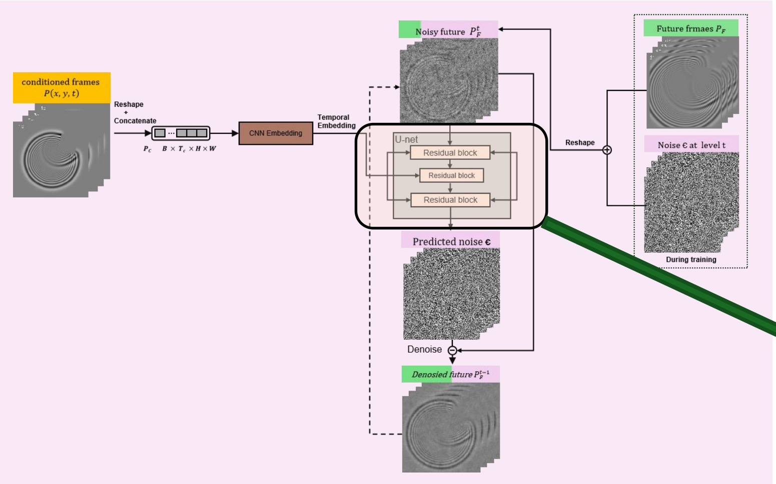
(Voleti et al., 2022)

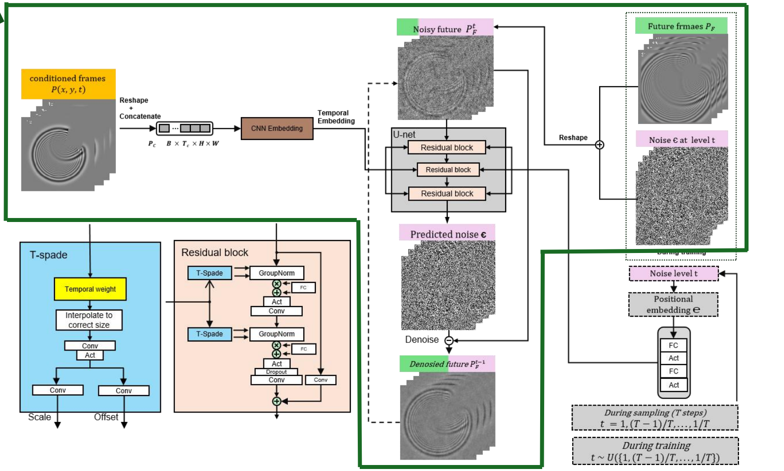
$$w_t = \frac{e^{-\lambda(t_c - t)}}{\sum_{t' \le t_c} e^{-\lambda(t_c - t')}}$$

$$\gamma(p) = \sum_{t<t_c} w_t \cdot \gamma_t, \quad \beta(p) = \sum_{t<t_c} w_t \cdot \beta_t$$

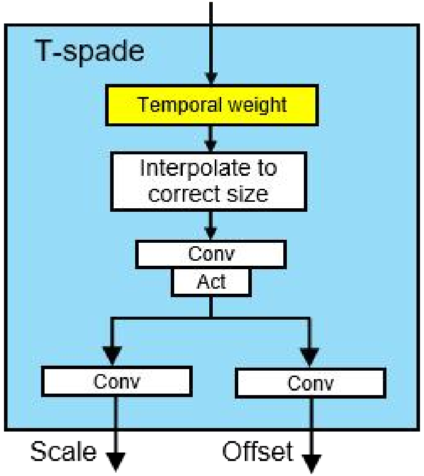


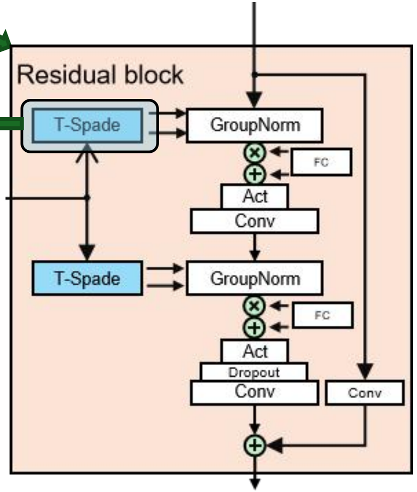






(Voleti et al., 2022)





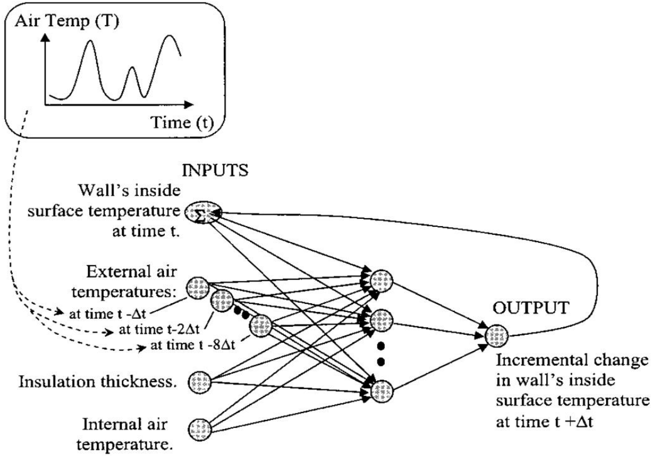
- Adaptive normalization enhances both spatial and temporal consistency during generation.
- Autoregressive sampling enables flexible and controllable sequence generation over time, and supports interpolation beyond the observed scene.
- Diffusion models produce high-fidelity results with fine-grained detail, making them well-suited for complex, dynamic data.

Objective 2:

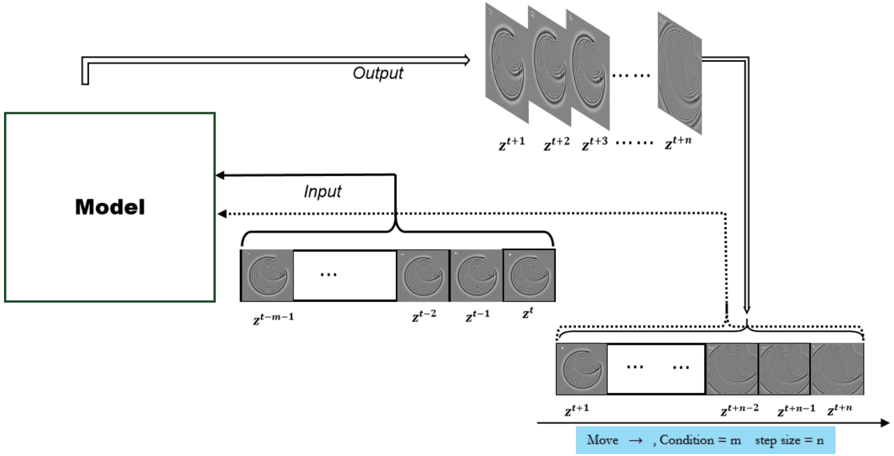
A coarse-grained integrated method is used to accelerate simulation while maintaining acceptable accuracy

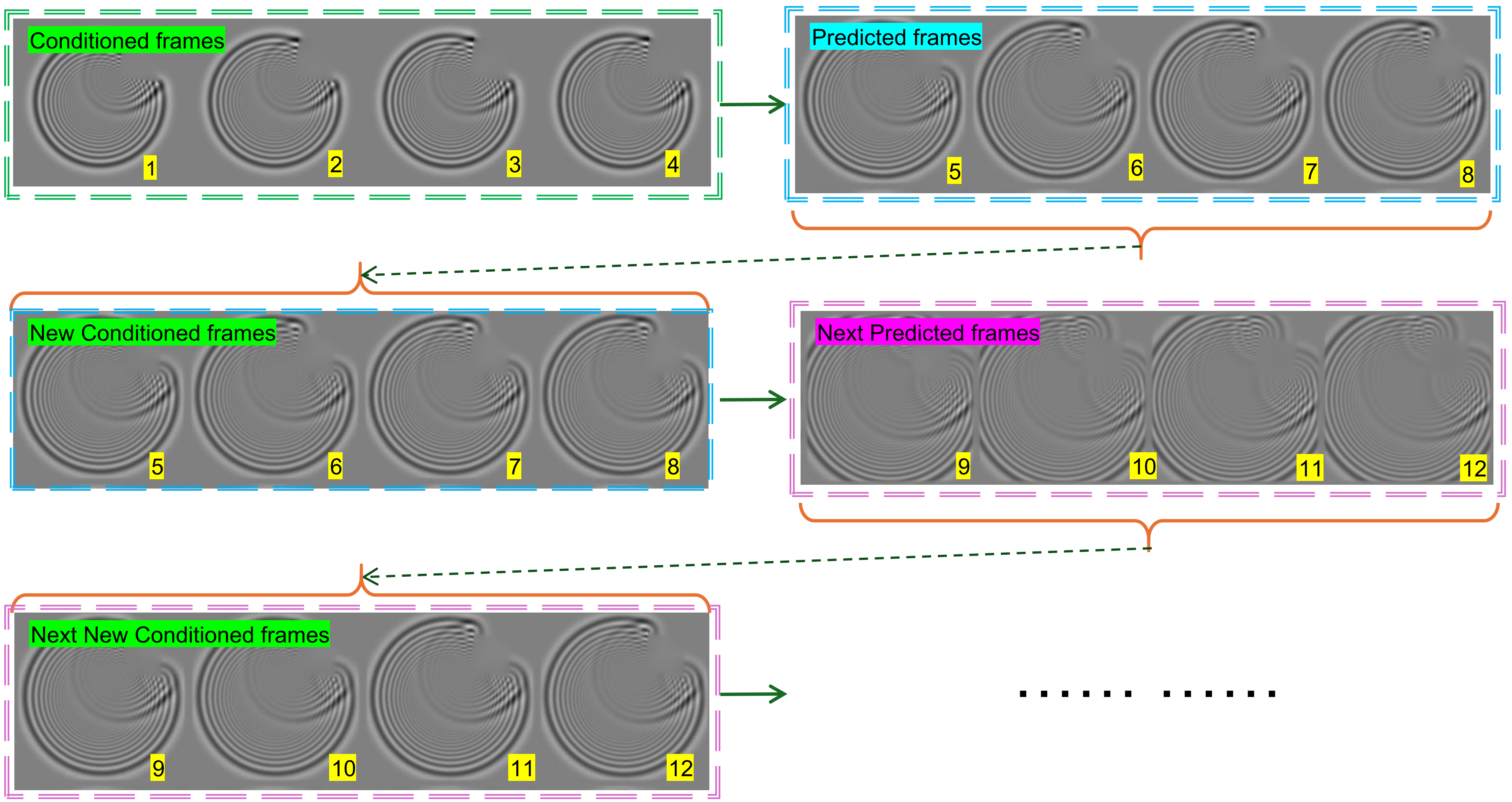
Temporal coarseness

Corse-grain model for thermal behaviour of Buildings



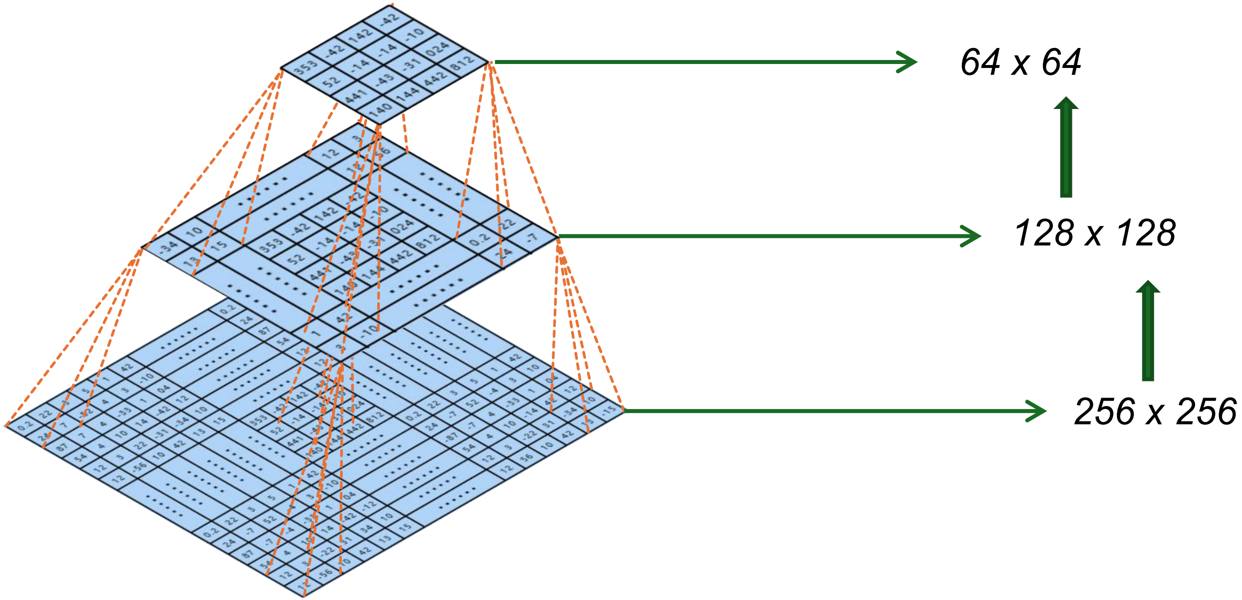
(Flood et al., 2008)





Conditioned frames
1
2
3
4
Predicted frames
5
6
7
8
New Conditioned frames
5
6
7
8
Next Predicted frames
9
10
11
12
Next New Conditioned frames
9
10
11
12

Spatial coarseness



64 x 64
128 x 128
256 x 256

① Average Pooling (LeCun et al., 1998)

$$y_{i,j} = \frac{1}{k^2} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x_{si+m,\, sj+n}$$

②Max Pooling (LeCun et al., 1998)

$$y_{i,j} = \max_{0 \le m,n < k} \left(x_{si+m,\, sj+n} \right)$$

③Strided Convolution (Dumoulin and Visin, 2016)

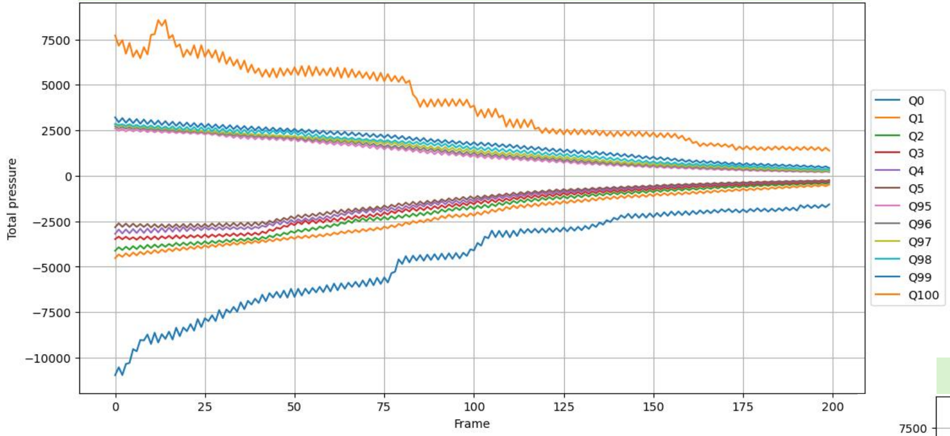
$$y_{i,j} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} w_{m,n} \cdot x_{si+m,\, sj+n} + b$$

❌ Blurring ⟶ Blurs wavefront details and smooths reflection edges, making it unsuitable for modeling shock waves and other high-gradient acoustic fields.

❌ Distortion ⟶ Tends to lose background energy, overly emphasizes local extrema, and may distort the overall sound pressure distribution.

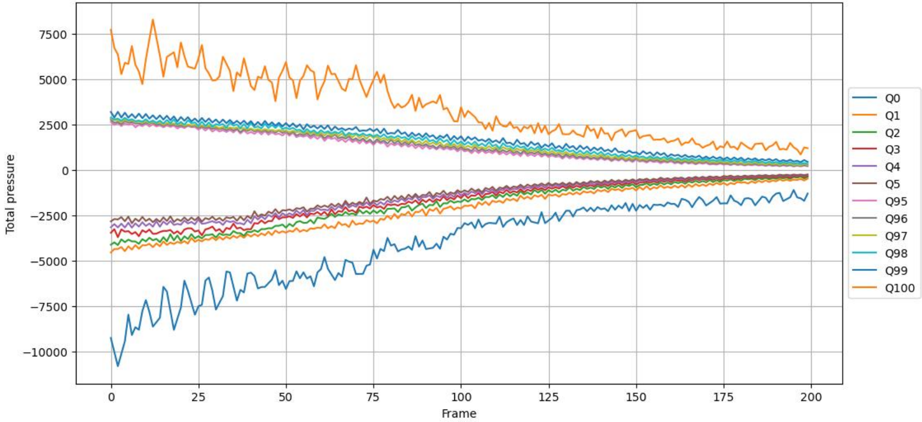
✓ Awareness ⟶ Strided convolution effectively downsamples spatial data while preserving critical features such as wavefronts and boundary reflection

256 x 256 qutantile evolution plot

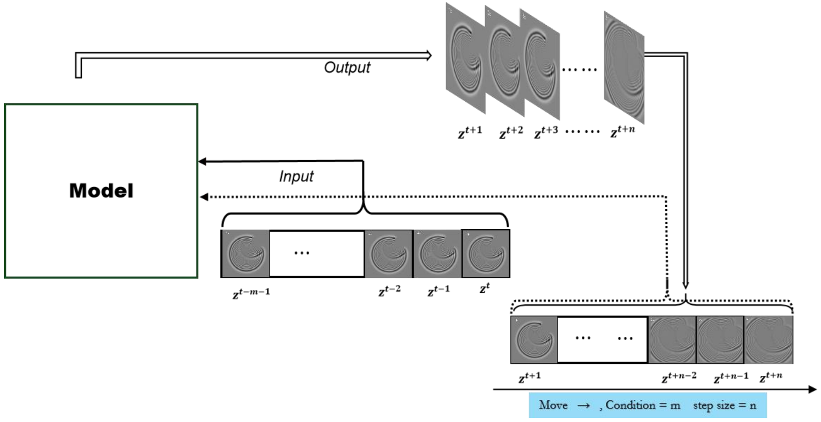


Strided Convolution downsampling

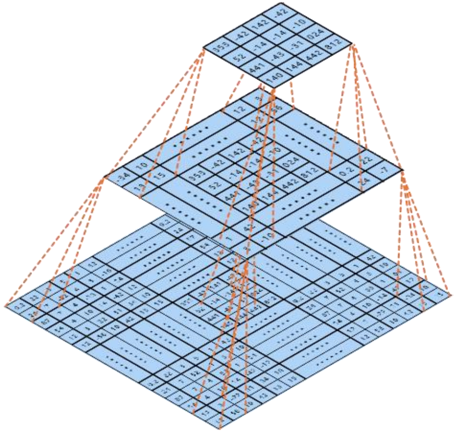
64 x 64 qutantile evolution plot



Temporal-spatial coarseness



temporal



spatial

Temporal-spatial coarseness

Evaluation of the model's inference speed and accuracy sensitivity to temporal and spatial coarseness

Acoustic Evaluation metric

4	-33	1	04	....				87	7	4	-33
10	14	-42	12	....				54	4	-10	14
22	-31	-34	10	....				12	3	22	31
10	42	13	15	....				12	56	10	42
⋮	⋮	⋮	⋮	353	-42	142	-42	⋮	⋮	⋮	⋮
				52	-14	-14	-10				
				441	-43	-31	024				
				140	144	442	812				
3	5	1	42	....				0.2	22	3	5
52	-4	3	-10	....				24	-7	52	-4
-4	-33	1	04	....				-87	-7	4	-33
10	14	-42	-12	....				54	4	10	-14

$$L_P(\mathbf{x}) = 20\log_{10}\left(\frac{|P(\mathbf{x})|}{p_{\mathrm{ref}}}\right), \quad L_{\hat{P}}(\mathbf{x}) = 20\log_{10}\left(\frac{|\hat{P}(\mathbf{x})|}{p_{\mathrm{ref}}}\right)$$

$$MAE(P,\hat{P}) = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} L_P(\mathbf{x}) - L_{\hat{P}}(\mathbf{x})$$

where: $P(\mathbf{x})$ is the actual sound pressure at location $\mathbf{x}$ (Pa).

$\hat{P}(\mathbf{x})$ is the predicted sound pressure (Pa).

$p_{\mathrm{ref}} = 20\ \mu\mathrm{Pa}$ (air), $\quad 1\ \mu\mathrm{Pa}$ (water)

(Craik, 1990)

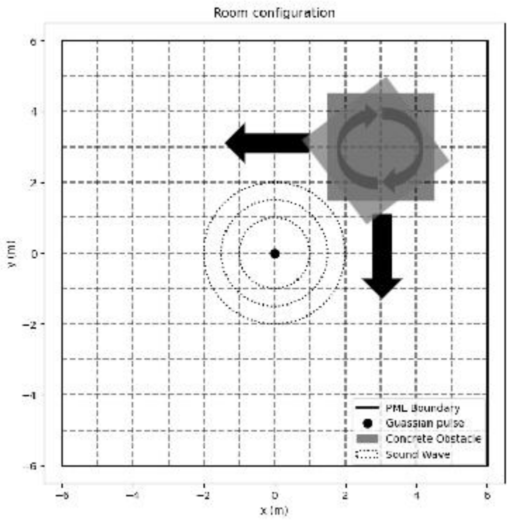
Metric	*≤ 1−2 dB*	*2−5 dB*	*>5 dB*
SPL error	*high fidelity*	*acceptable*	*No*

Introduce the MAE sound pressure level error to better measure coarse-grain model's ability to balance accuracy and speed, to make evaluation more human perceptual

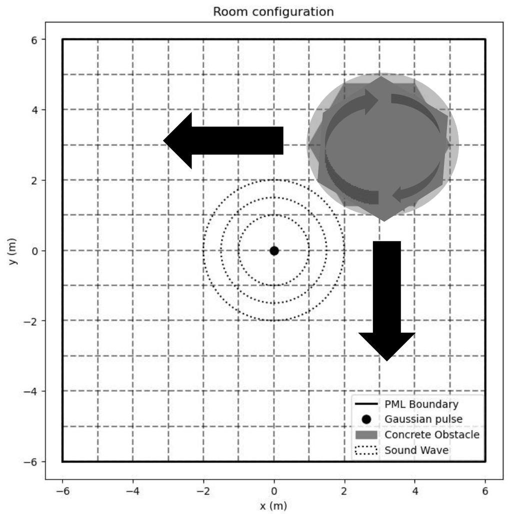
Objective 3:

An extended, more complex scene is used to further evaluate the model's approximation ability and interpolation ability

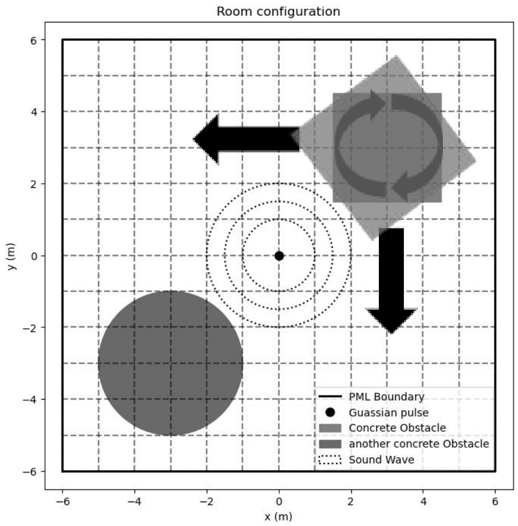
Basic scene (1 square)



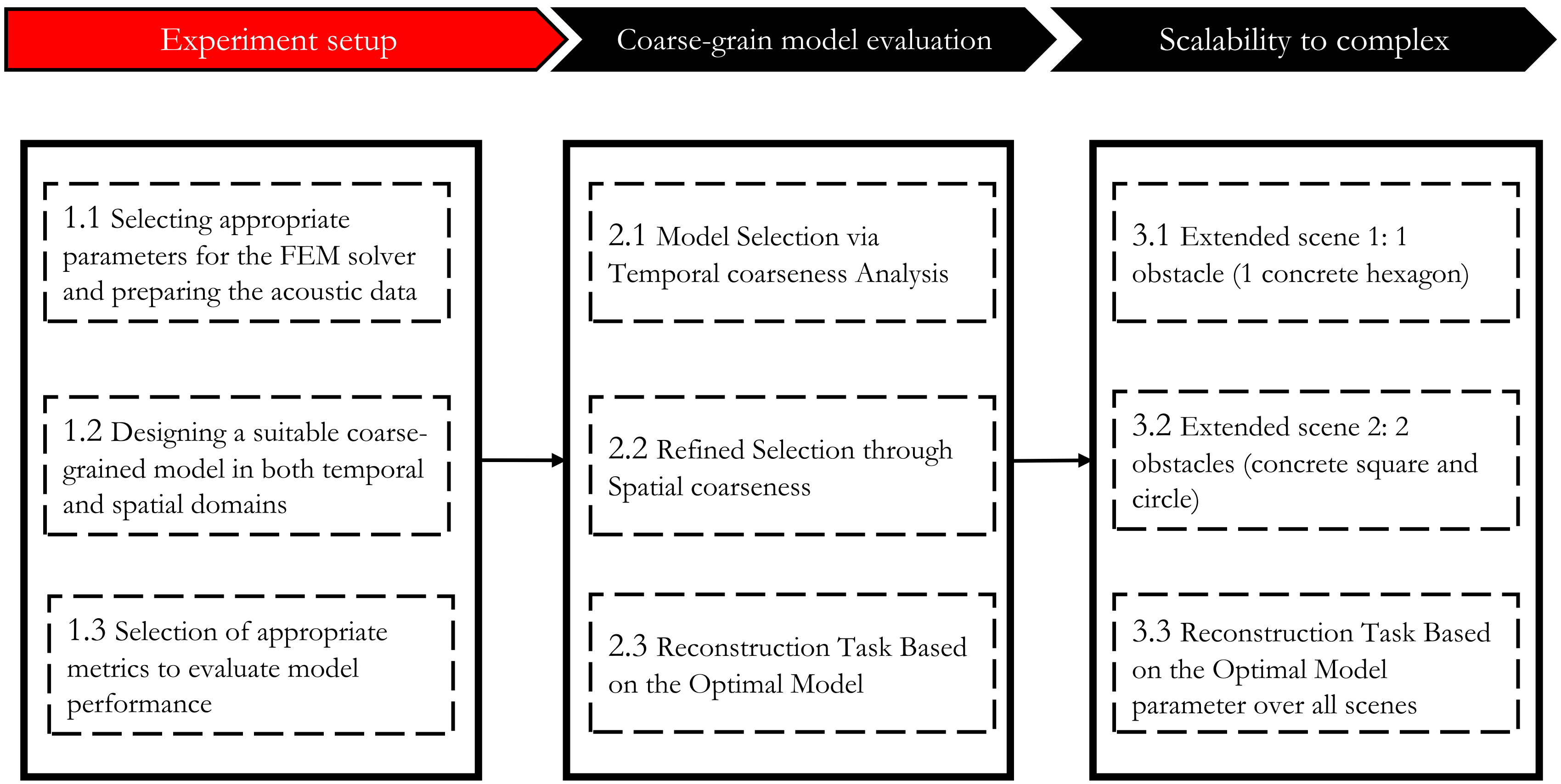
Extended 1 (1 hexagon)



Extended 2 (1 square and 1 circle)



Experiment

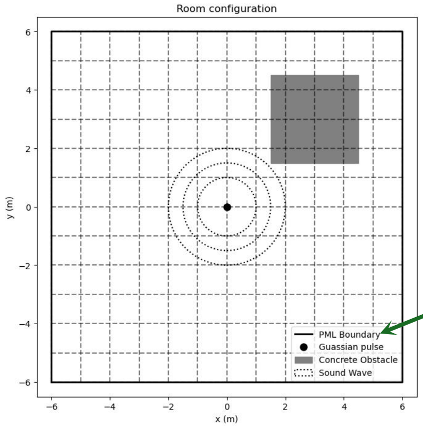

Experiment setup
Coarse-grain model evaluation
Scalability to complex
1.1 Selecting appropriate parameters for the FEM solver and preparing the acoustic data
1.2 Designing a suitable coarse-grained model in both temporal and spatial domains
1.3 Selection of appropriate metrics to evaluate model performance
2.1 Model Selection via Temporal coarseness Analysis
2.2 Refined Selection through Spatial coarseness
2.3 Reconstruction Task Based on the Optimal Model
3.1 Extended scene 1: 1 obstacle (1 concrete hexagon)
3.2 Extended scene 2: 2 obstacles (concrete square and circle)
3.3 Reconstruction Task Based on the Optimal Model parameter over all scenes

Experiment setup	Coarse-grain model evaluation	Scalability to complex

1.1 Selecting appropriate parameters for the FEM solver and preparing the acoustic data

1.2 Designing a suitable coarse-grained model in both temporal and spatial domains

1.3 Selection of appropriate metrics to evaluate model performance



Frequency F	1000 HZ
Wave length λ (m)	0.343
Wave period T (s)	0.001
mesh size Δx (m) ($\Delta x \leq \lambda/6$)	0.0570
Number of element (12 x 12 m)	210 x 210
Step size Δt (s) ($\Delta t \leq T/20$)	0.00005
Number of time instances (0.02s)	1000

FEM in Comsol

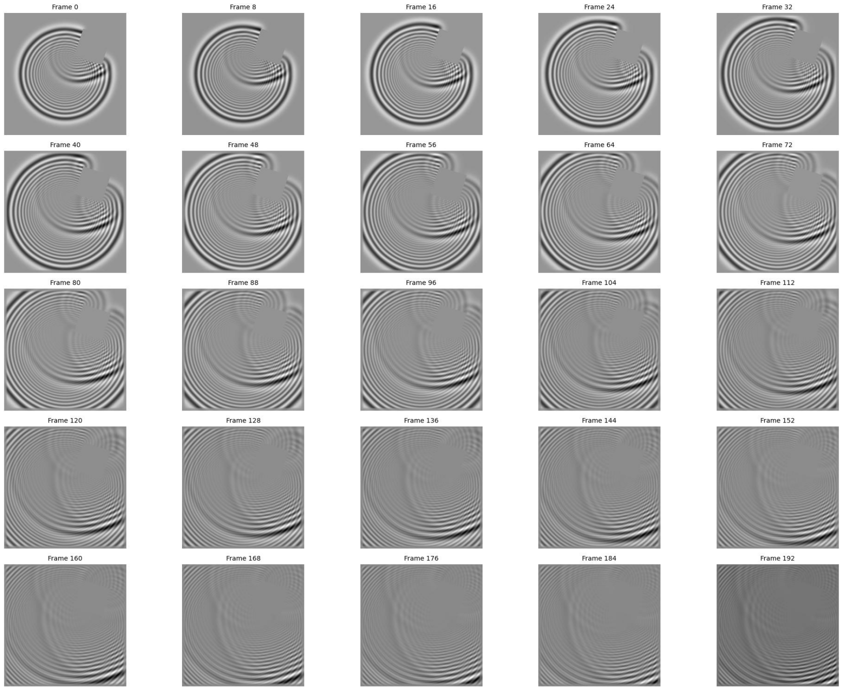
Parameters	Global	Room	Room dimension	12 * 12 m
			Boundary Conditions	PML
			Medium	air
		Obstacle	Material	Concrete
			Densitity	3000 kg/m³
		Sound source	Type	Gaussian pulse
			Frequency	1000 HZ
			Peak time	0.001 s
			Amplitude	10 m²/s
		Solver	Type	FEM
			Mesh size (Δx)	0.0469 m
			Mesh resolution	256 x 256
			Time step (Δt)	T/10 = 0.0001 s
			Time range	0.0160 - 0.0260 s
			Total frames	100
	Parametric sweep	Obstacle	Location of obstacle	(x1,y1)
			Oriatation of obstacle	(a1)
Dependent Variable			Sound pressure map (P)	Pa

Experiment setup | Coarse-grain model evaluation | Scalability to complex

1.1 Selecting appropriate parameters for the FEM solver and preparing the acoustic data

1.2 Designing a suitable coarse-grained model in both temporal and spatial domains

1.3 Selection of appropriate metrics to evaluate model performance



- Experiment setup
- Coarse-grain model evaluation
- Scalability to complex

1.1 Selecting appropriate parameters for the FEM solver and preparing the acoustic data

1.2 Designing a suitable coarse-grained model in both temporal and spatial domains

1.3 Selection of appropriate metrics to evaluate model performance

Coarse-grain Diffusion model			
Basic scene (1 square)	256 x 256 100 frames 128 samples 75% Training 25% Testing	Temporal coarseness exp C + P = 24 C= S (256 x 256)	C1S1P23
			C2S2P22
			C4S4P20
			C8S8P16
			C12S12P12
		Spatial coarseness exp (C8S8P16)	256 x 256
			128 x 128
			64 x 64
Extended scene	200 samples 75% Training 25% Testing	Best Model	1 Hexagon
	128 samples 75% Training 25% Testing	Best Model	2 obstacles

Accuracy sensitivity to T/S coarseness

C - number of conditioned frames
S - number of frames generated at a time
P - number of supervisor frames

C+P = 24
T = 24 * 0.0001 s =2.4 ms

Experiment setup | Coarse-grain model evaluation | Scalability to complex

1.1 Selecting appropriate parameters for the FEM solver and preparing the acoustic data

1.2 Designing a suitable coarse-grained model in both temporal and spatial domains

1.3 Selection of appropriate metrics to evaluate model performance

$$L_P(\mathbf{x}) = 20\log_{10}\left(\frac{|P(\mathbf{x})|}{p_{\mathrm{ref}}}\right), \quad L_{\hat{P}}(\mathbf{x}) = 20\log_{10}\left(\frac{|\hat{P}(\mathbf{x})|}{p_{\mathrm{ref}}}\right)$$

$$MAE(P,\hat{P}) = \frac{1}{|\Omega|}\sum_{\mathbf{x}\in\Omega} L_P(\mathbf{x}) - L_{\hat{P}}(\mathbf{x})$$

where: $P(\mathbf{x})$ is the actual sound pressure at location $\mathbf{x}$ (Pa).

$\hat{P}(\mathbf{x})$ is the predicted sound pressure (Pa).

$p_{\mathrm{ref}} = 20\ \mu\mathrm{Pa}$ (air), $\quad 1\ \mu\mathrm{Pa}$ (water)

(Craik, 1990)

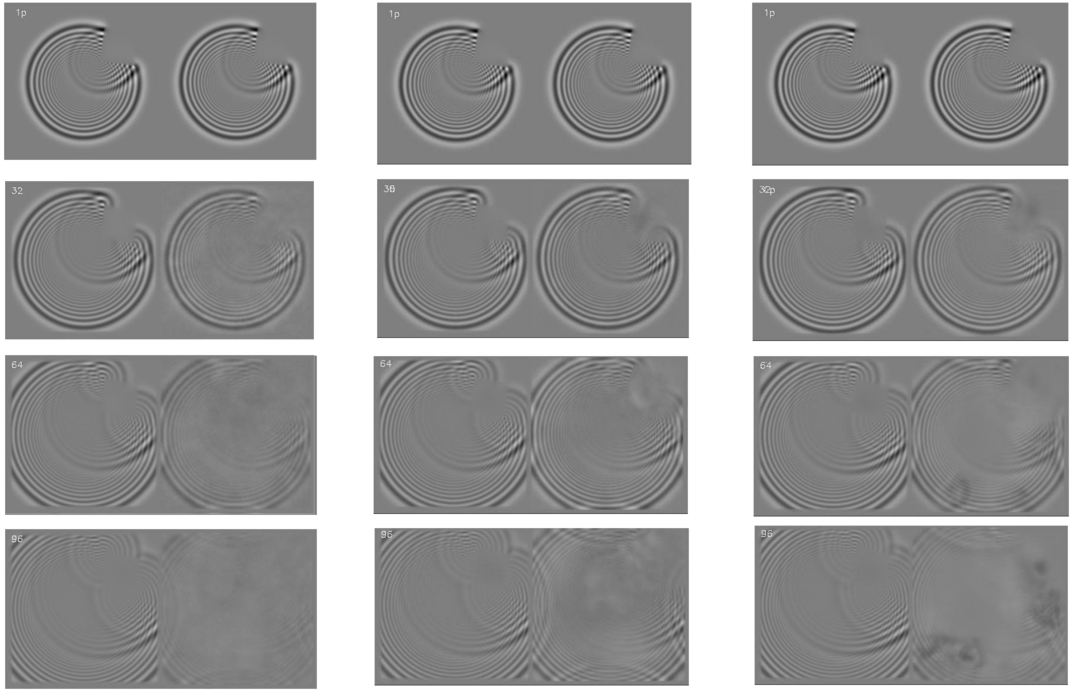
Metric	*≤ 1−2 dB*	*2−5 dB*	*>5 dB*
SPL error	*high fidelity*	*acceptable*	*No*

2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Temporal coarseness exp | C + P = 24 | C= S | (256 x 256) | Reconstruct 100 frames task				
C1S1P23	*C2S2P22*	*C4S4P20*	*C8S8P16*	*C12S12P12*
Small temporal coarseness		*middle*	*large*	



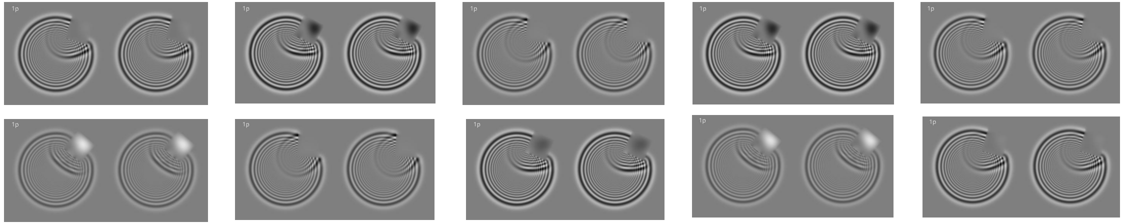
Experiment setup | **Coarse-grain model evaluation** | Scalability to complex

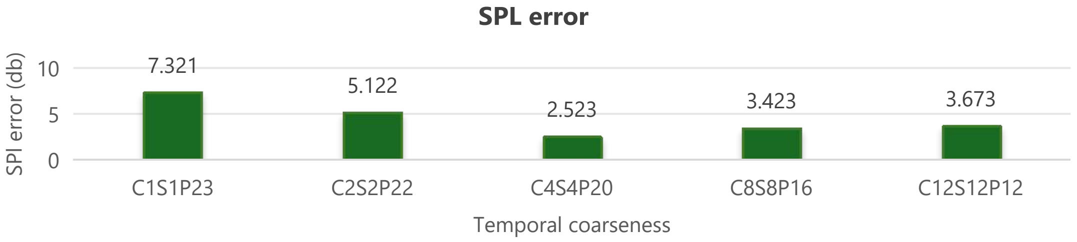
2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Temporal coarseness exp | C + P = 24 | C= S | (256 x 256) | Reconstruct 100 frames task				
C1S1P23	*C2S2P22*	*C4S4P20*	*C8S8P16*	*C12S12P12*





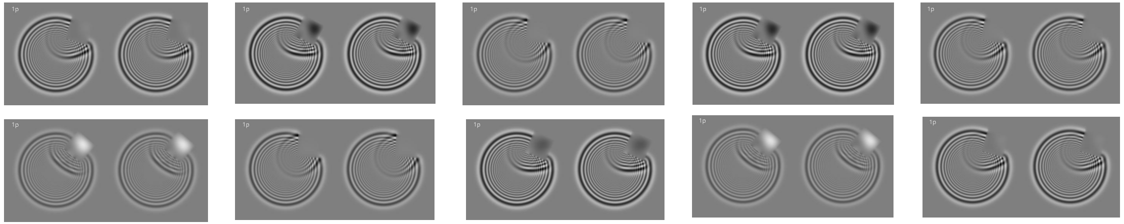
Experiment setup | **Coarse-grain model evaluation** | Scalability to complex

2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Temporal coarseness exp | C + P = 24 | C= S | (256 x 256) | Reconstruct 100 frames task				
C1S1P23	*C2S2P22*	*C4S4P20*	*C8S8P16*	*C12S12P12*



Model	SPL error	Reconstruct time (h)	Best model
C1S1P23 | 256	7.321	7.2	
C2S2P22 | 256	5.122	3.6	
C4S4P20 | 256	2.523	1.8	✓
C8S8P16 | 256	3.423	1.0	
C12S12P12 | 256	3.673	0.6	

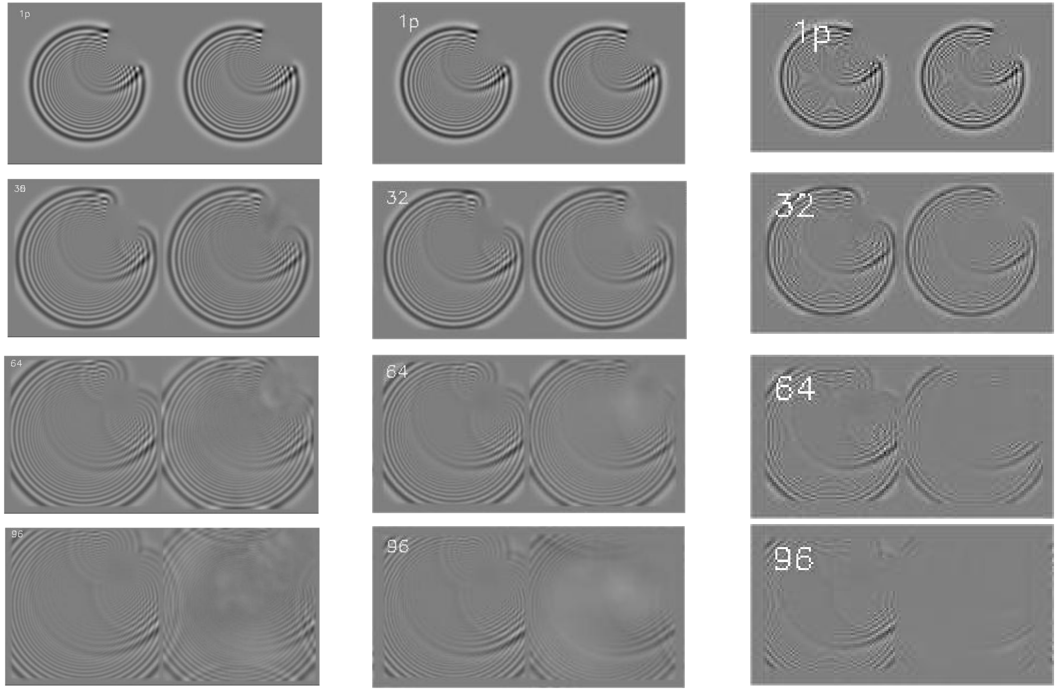
Experiment setup | **Coarse-grain model evaluation** | Scalability to complex

2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Spatial coarseness exp | (C4S4P20) | Reconstruct 100 frames task		
256 x 256	*128 x 128*	*64 x 64*



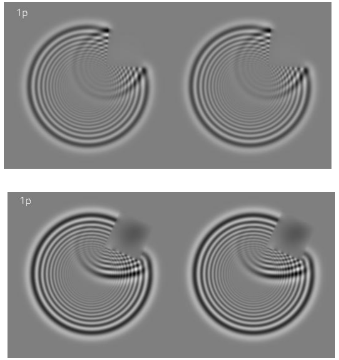
Experiment setup | Coarse-grain model evaluation | Scalability to complex

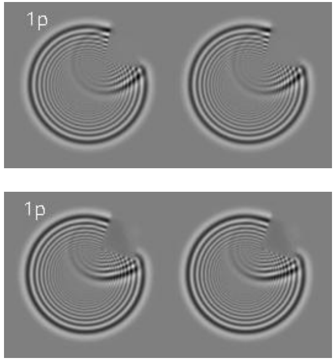
2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

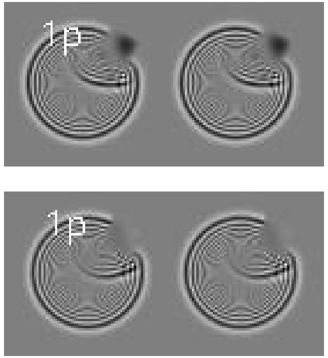
2.2 Refined Selection through Spatial coarseness

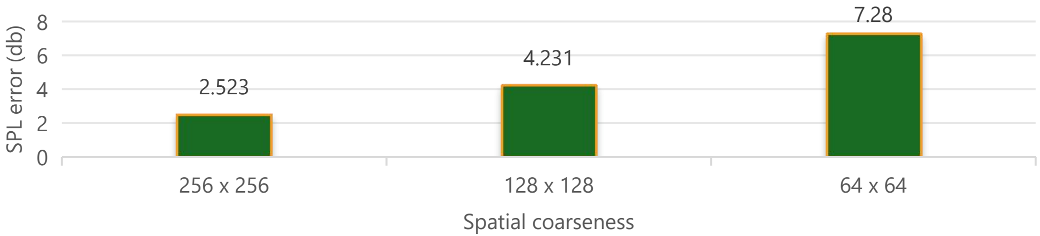
2.3 Reconstruction Task Based on the Optimal Model

Spatial coarseness exp | (C4S4P20) | Reconstruct 100 frames task		
256 x 256	128 x 128	64 x 64









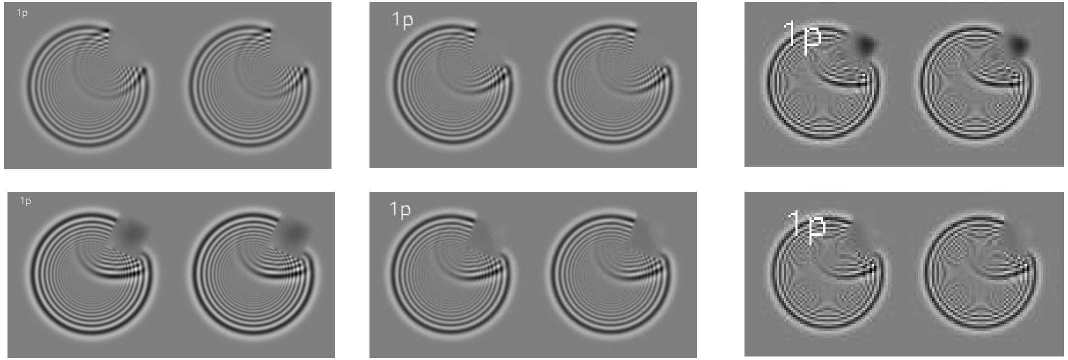
Experiment setup | Coarse-grain model evaluation | Scalability to complex

2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Spatial coarseness exp | (C4S4P20) | Reconstruct 100 frames task		
256 x 256	*128 x 128*	*64 x 64*



Model	SPL error	Reconstruct time (h)	Best model
C4S4P20 | 256	2.523	1.80	
C4S4P20 | 128	4.231	0.40	✓
C4S4P20 | 64	7.280	0.16	

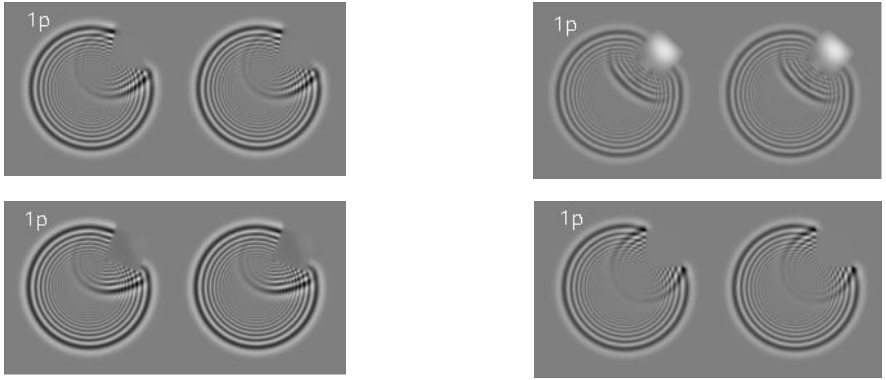
Experiment setup | **Coarse-grain model evaluation** | Scalability to complex

2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Best coarse-grain model | 128 | (C4S4P20) | Reconstruct 100 frames task



Model	SPL error	Reconstruct time (h)	Speed up ratio
FEM solver	N/A	6.23	N/A
C4S4P20 | 128	4.231	0.40	**15.575**

Experiment setup | Coarse-grain model evaluation | Scalability to complex

2.1 Model Selection via Temporal coarseness Analysis with whole spatial resolution

2.2 Refined Selection through Spatial coarseness

2.3 Reconstruction Task Based on the Optimal Model

Computing overhead | All models

Numerical Model	CPU Mem	CPU core	CPU Hours
Comsol FEM	8 G	AMD EPYC 9654 96-Core	6.23 h

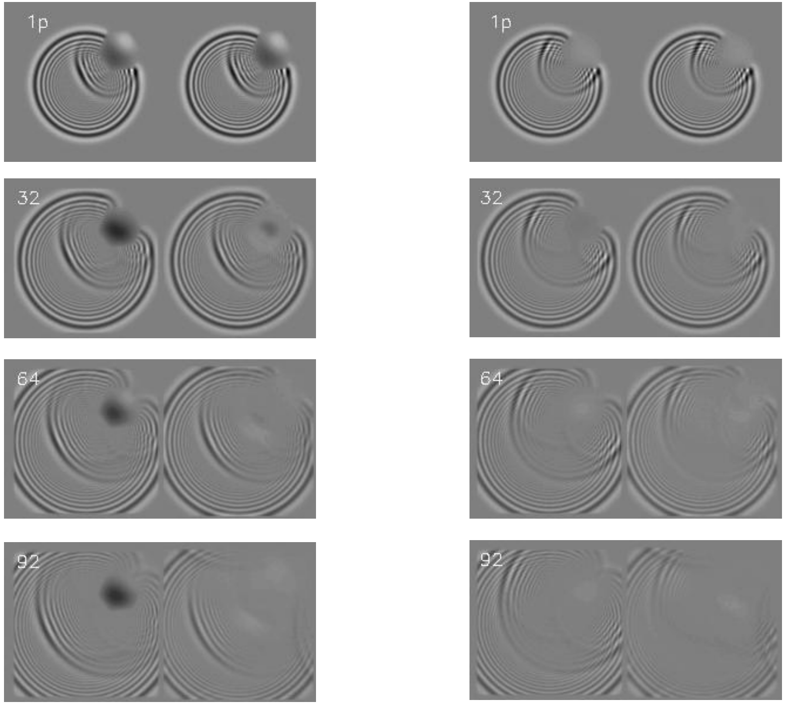
Diffusion model (Training)	Params	CPU Mem	batch size	GPU	GPU Mem	Best steps	GPU hours (training)
C1S1P23 | 256	58.0 M	2.352 G	8	RTX 4090	16.32 G	5000	6.12 hours
C2S2P22 | 256	58.0 M	2.37 G	8	RTX 4090	16.32 G	5000	6.23 hours
C4S4P20 | 256	58.0 M	2.30 G	8	RTX 4090	16.36 G	7000	7.4 hours
C8S8P16 | 256	58.0 M	2.70 G	8	RTX 4090	16.40 G	5000	5.92 hours
C12S12P12 | 256	58.0 M	2.79 G	8	RTX 4090	16.31 G	5000	5.60 hours
C4S4P16 | 128	57.4 M	2.81 G	16	RTX 4090	19.42 G	4000	5.02 hours
C4S4P20 | 64	41.1 M	1.81 G	16	RTX 4090	7.33 G	4000	3.33 hours

3.1 Extended scene 1: 1 obstacle (1 concrete hexagon)

3.2 Extended scene 2: 2 obstacles (concrete square and circle)

3.3 Reconstruction Task Based on the Optimal Model parameter over all scenes

1 Obstacle | Hexagon | C4S4P20 | 128 | 100 frames generation task



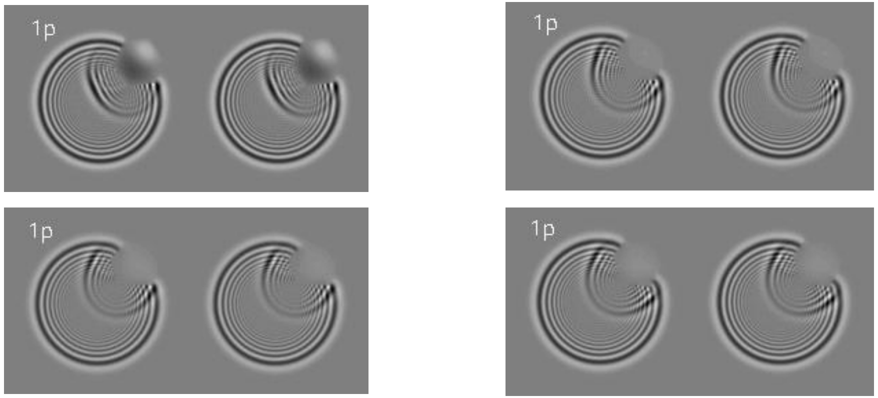
Experiment setup | Coarse-grain model evaluation | Scalability to complex

3.1 Extended scene 1: 1 obstacle (1 concrete hexagon)

3.2 Extended scene 2: 2 obstacles (concrete square and circle)

3.3 Reconstruction Task Based on the Optimal Model parameter over all scenes

1 Obstacle | Hexagon | C4S4P20 | 128 | 100 frames generation task



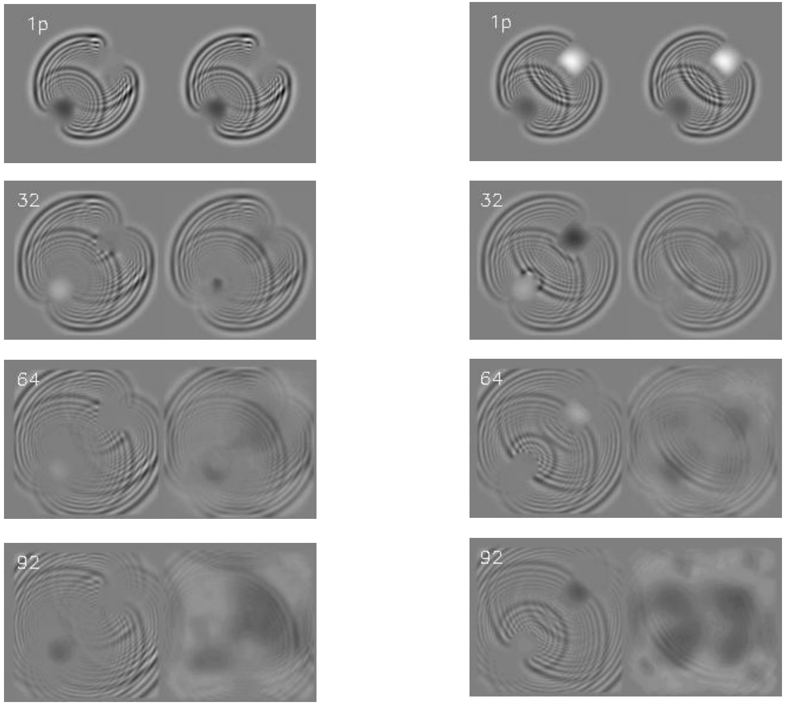
Model	SPL error	Reconstruct time (h)	Speed up ratio
FEM solver	N/A	6.56	N/A
C4S4P20 | 128	4.768	0.60	**10.94**

3.1 Extended scene 1: 1 obstacle (1 concrete hexagon)

3.2 Extended scene 2: 2 obstacles (concrete square and circle)

3.3 Reconstruction Task Based on the Optimal Model parameter over all scenes

2 Obstacles | Square and circle | C4S4P20 | 128 | 100 frames generation task



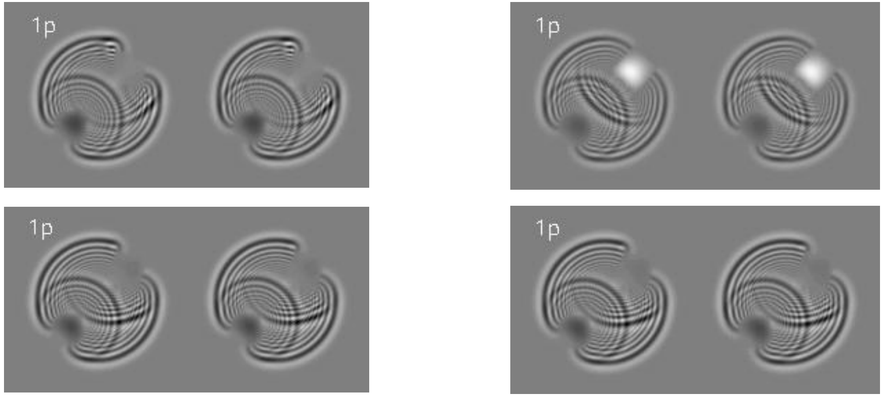
Experiment setup | Coarse-grain model evaluation | **Scalability to complex**

3.1 Extended scene 1: 1 obstacle (1 concrete hexagon)

3.2 Extended scene 2: 2 obstacles (concrete square and circle)

3.3 Reconstruction Task Based on the Optimal Model parameter over all scenes

2 Obstacle | Square and circle | C4S4P20 | 128 | 100 frames generation task



Model	SPL error	Reconstruct time (h)	Speed up ratio
FEM solver	N/A	6.32	N/A
C4S4P20 | 128	8.293 ×	0.40	**15.80**

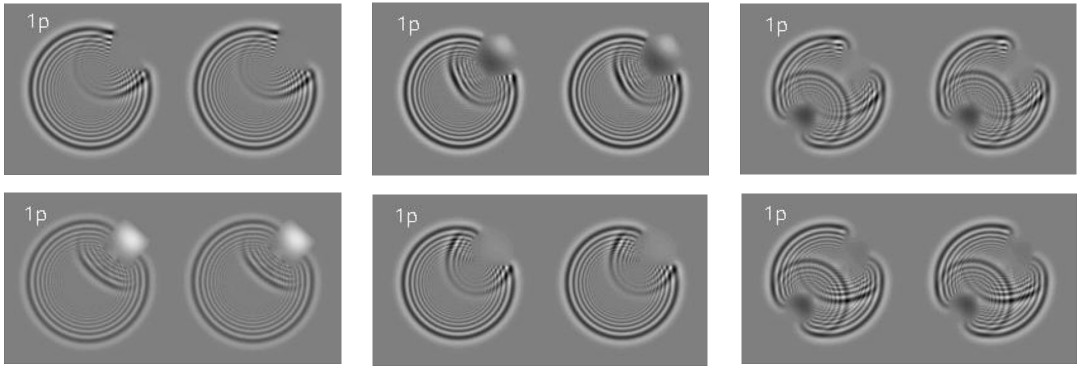
Experiment setup | Coarse-grain model evaluation | Scalability to complex

3.1 Extended scene 1: 1 obstacle (1 concrete hexagon)

3.2 Extended scene 2: 2 obstacles (concrete square and circle)

3.3 Reconstruction Task Based on the Optimal Model parameter over all scenes

Optimal model over all scenes | *C4S4P20* | *128* | *Reconstruct 100 frames task*		
1 Square	*1 Hexagon*	*1 square and 1 circle*



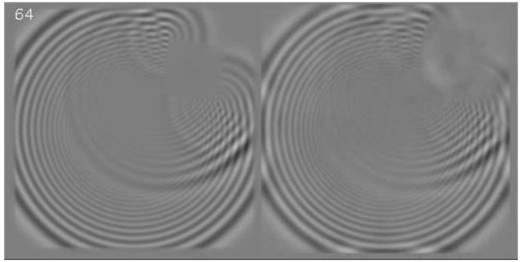
Model	SPL error	Reconstruct time (h)	Speed up ratio	
1 Square | C4S4P20 | 128	4.231	0.40	15.575	✓
1 Hexagon | C4S4P20 | 128	4.768	0.60	10.94	✓
1 square and 1 circle | C4S4P20 | 128	8.293	0.40	15.80	×

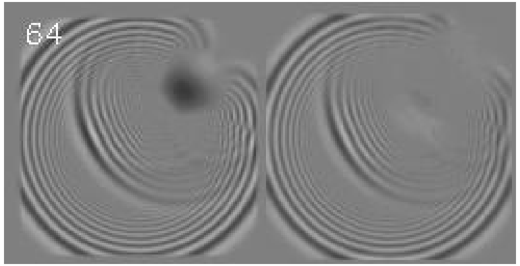
Conclusion

- ☑ Developed a T-Spade enhanced DDPM for autoregressive acoustic pressure prediction in 2D environments.
- ☑ Introduced a coarse-grained integration strategy that achieved a 15.575× speed-up while maintaining a low SPL error of 4.231.
- ☑ Validated model generalization on extended acoustic scenes: in Scene 1, a 10.94× speed-up was achieved with SPL error kept below 5, whereas in Scene 2, despite a 15.58× speed-up, the SPL error reached 8.293, indicating a significant loss in accuracy.

Limitations

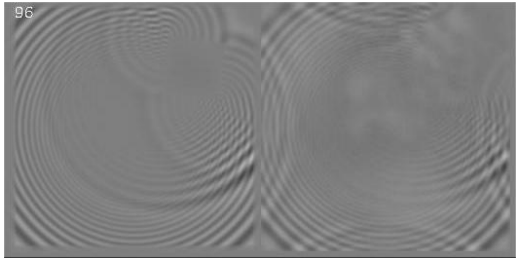
- ☒ As the number of frames increases, the model shows data drift and distortion in areas inside and near concrete obstacles

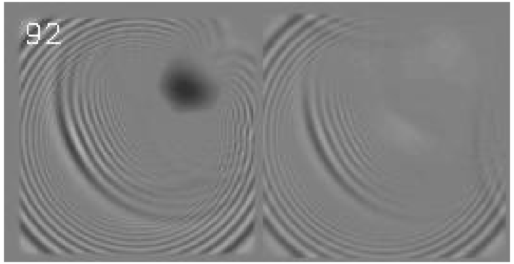




Limitations

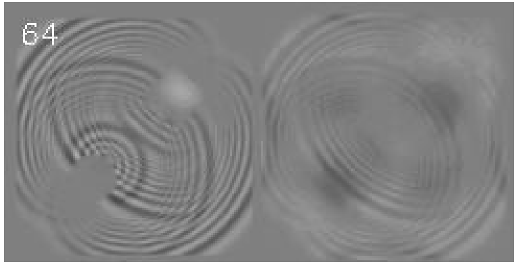
☑ For frames beyond 100, the absence of boundary conditions information causes residual wave patterns that should have been absorbed by the PML.

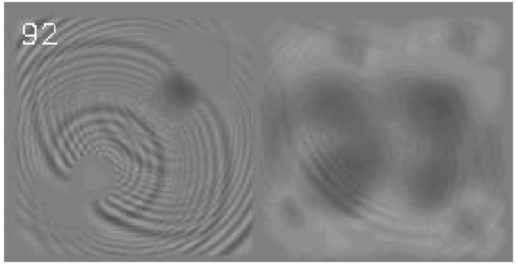




Limitations

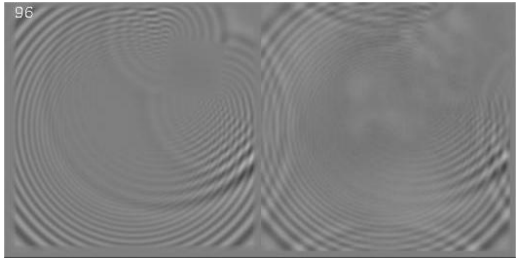
☑ In Extended Scene 2, the model fails to effectively handle persistent reflection between two closely positioned obstacles.





Future work

↑ Physics-Constrained Boundary Correction to solve incorrect boundary reflections

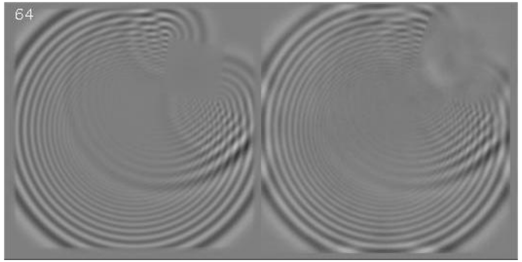


$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_{\text{phys}} \cdot \mathcal{L}_{\text{phys}} + \lambda_{\text{bdy}} \cdot \mathcal{L}_{\text{boundary}}$$

Symbol	**Description**
$\mathcal{L}_{\text{total}}$	Total loss function used for training
$\mathcal{L}_{\text{data}}$	Data loss, measuring prediction error
$\mathcal{L}_{\text{phys}}$	Physics-based loss, enforcing wave equation
$\mathcal{L}_{\text{boundary}}$	Boundary loss, reducing boundary reflections
$\lambda_{\text{phys}}, \lambda_{\text{bdy}}$	Weights for balancing each term

Future work

↑ Physics-Constrained stable loss to solve distortion around obstacle



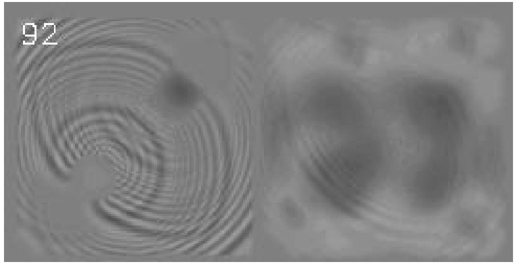
$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_1 \mathcal{L}_{\text{reg}} + \lambda_2 \mathcal{L}_{\text{stable}}$$

$$\mathcal{L}_{\text{stable}} = \quad f_{t+1}^{\text{struct}} - \mathcal{F}\left(f_t^{\text{struct}}, f_{t-1}^{\text{struct}}\right) \quad {}^{2}$$

Symbol	Description
$\mathcal{L}_{\text{total}}$	Total loss function
$\mathcal{L}_{\text{data}}$	Data loss
$\mathcal{L}_{\text{reg}}$	Regularization loss
$\mathcal{L}_{\text{stable}}$	Stability loss near obstacles
f_t^{struct}	Prediction in structure region
$\mathcal{F}(\cdot)$	Temporal predictor
$\lambda_1,\ \lambda_2$	Loss weights

Future work

↑ Structure-Aware Reflection Guidance to solve distortion around obstacle



$$\mathcal{L}_{\text{reflect}} = \left\| \mathcal{G}(f_{t+1}) - \sigma\left(\|\nabla\rho(\mathbf{x})\|\right) \cdot \mathcal{G}(f_t) \right\|_{\Omega}^{2}$$

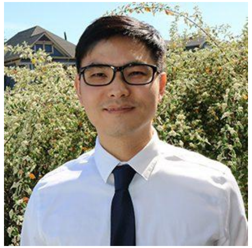
Symbol	Description
$\mathcal{L}_{\text{reflect}}$	Reflection loss guided by object structure
$f_t,\ f_{t+1}$	Predictions at time t and t+1
$\mathcal{G}(\cdot)$	Feature extractor for reflection behavior
$\rho(\mathbf{x})$	Density field indicating object distribution
$\nabla\rho(\mathbf{x})$	Gradient of density (high at object boundaries)
$\sigma(\cdot)$	Activation function modulating reflection strength
Ω	Region where reflection is likely to occur

Acknowledgement

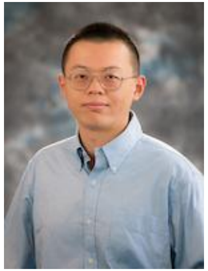
Committee members:

Ian Flood
Construction Management
University of Florida

Hassan Azad
Architecture
University of Florida



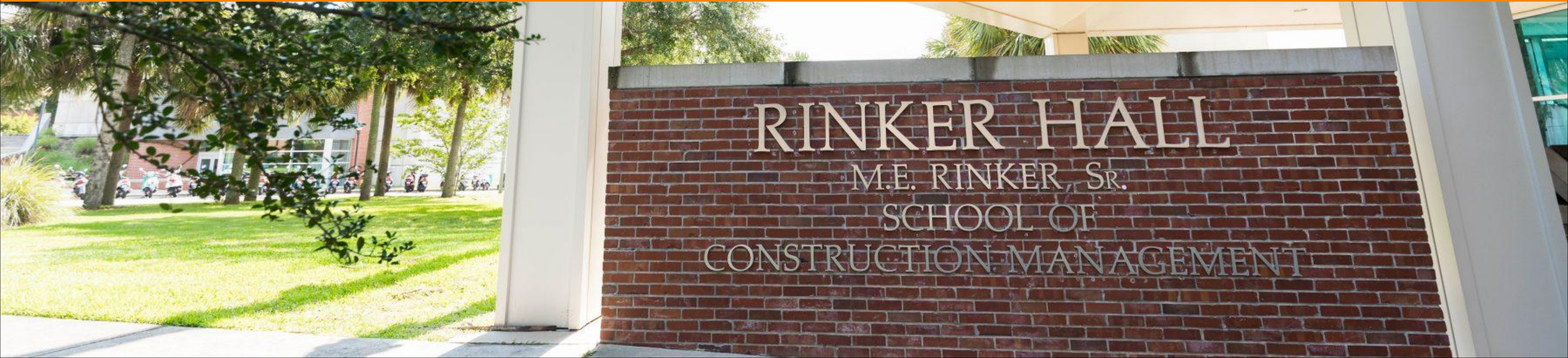
Chaofeng Wang
Construction Management
University of Florida



Kejun Huang
Computer Science
University of Florida

Thanks to everyone for supporting my PhD studies

Thank you !


RINKER HALL
M.E. RINKER, SR.
SCHOOL OF
CONSTRUCTION MANAGEMENT