ICML 40 Years International Conference On Machine Learning

WISCONSIN UNIVERSITY OF WISCONSIN–MADISON

UIC UNIVERSITY OF ILLINOIS CHICAGO

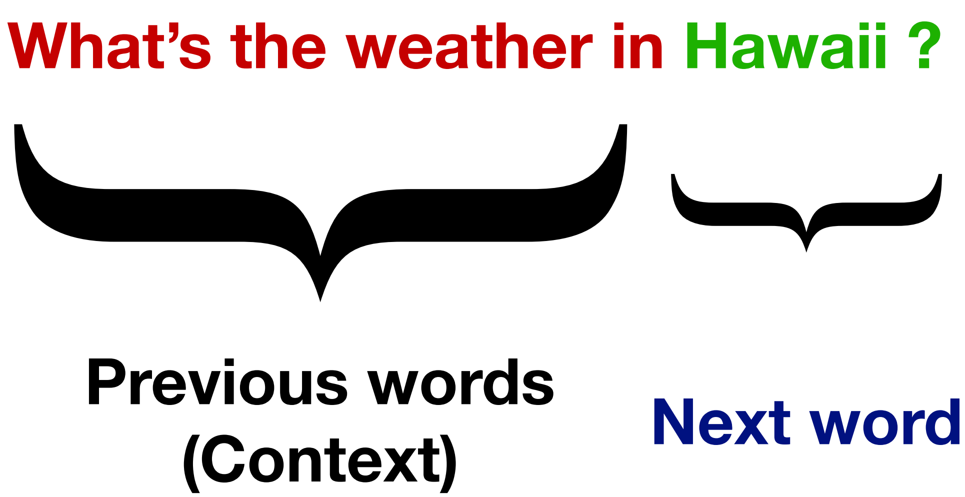
Controlled Differential Equations on Long Sequences via Non-standard Wavelets

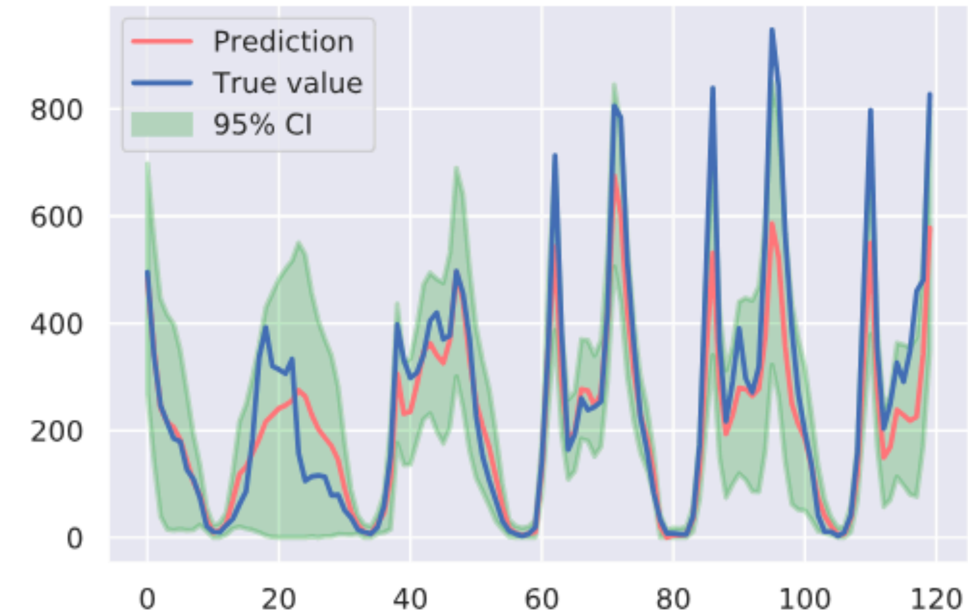
Sourav Pal

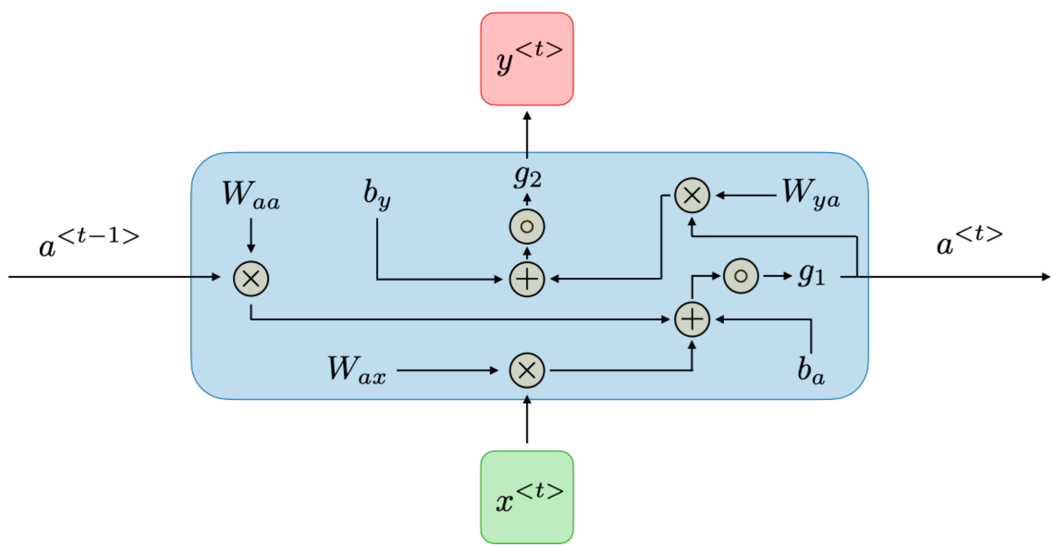
Zhanpeng Zeng

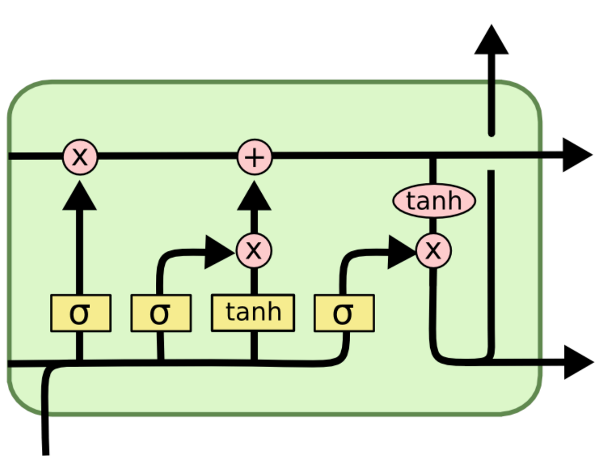
Sathya N. Ravi

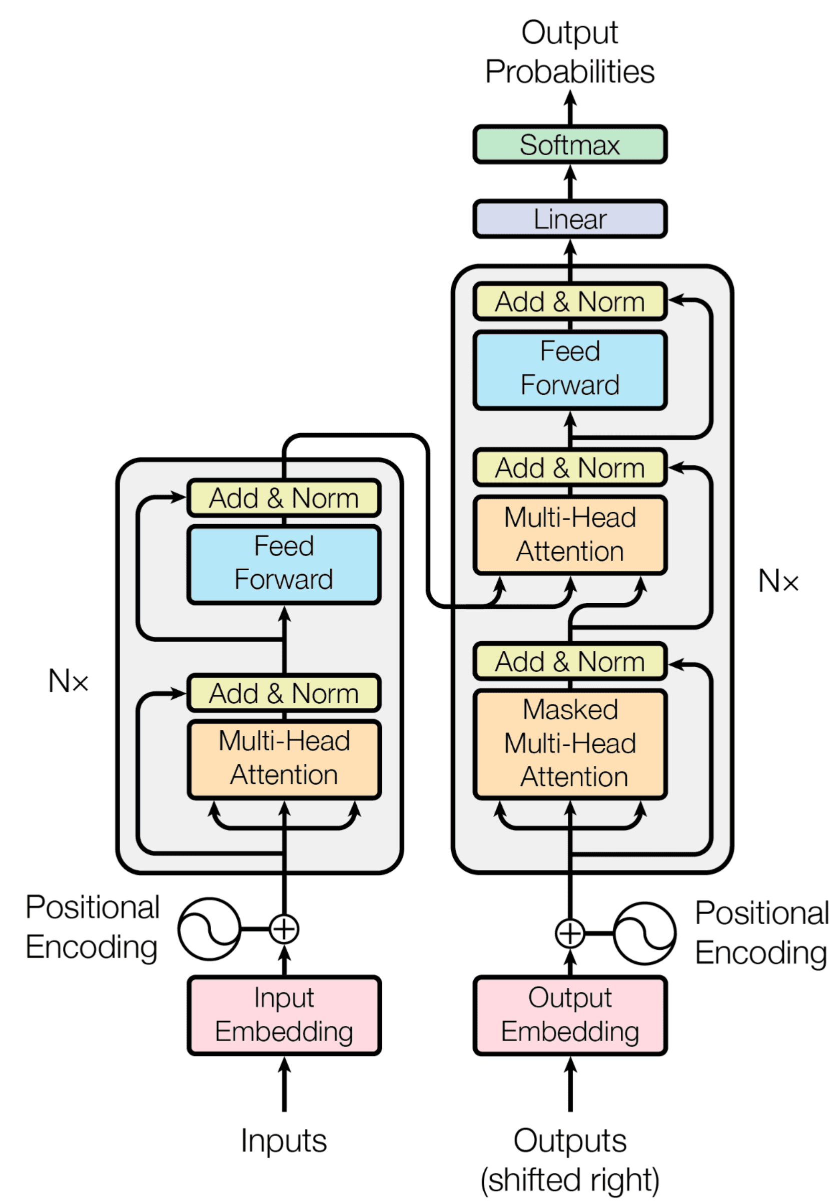
Vikas Singh


What's the weather in Hawaii ?
Previous words
(Context)
Next word


Prediction
True value
95% CI
800
600
400
200
0
0
20
40
60
80
100
120


$y^{<t>}$
W_{aa}
b_y
g_2
W_{ya}
$a^{<t-1>}$
$a^{<t>}$
g_1
W_{ax}
b_a
$x^{<t>}$


x
+
tanh
x
x
σ
σ
tanh
σ


Output
Probabilities
Softmax
Linear
Add & Norm
Feed
Forward
Add & Norm
Multi-Head
Attention
Add & Norm
Masked
Multi-Head
Attention
Nx
Add & Norm
Feed
Forward
Add & Norm
Multi-Head
Attention
Nx
Positional
Encoding
Positional
Encoding
Input
Embedding
Output
Embedding
Inputs
Outputs
(shifted right)

Neural Ordinary Differential Equations

Ri

Neural Laplace: Learning diverse classes of differential equations in the Laplace domain

Samuel Holt [1] Zhaozhi Qian [1] Mihaela van der Schaar [1]

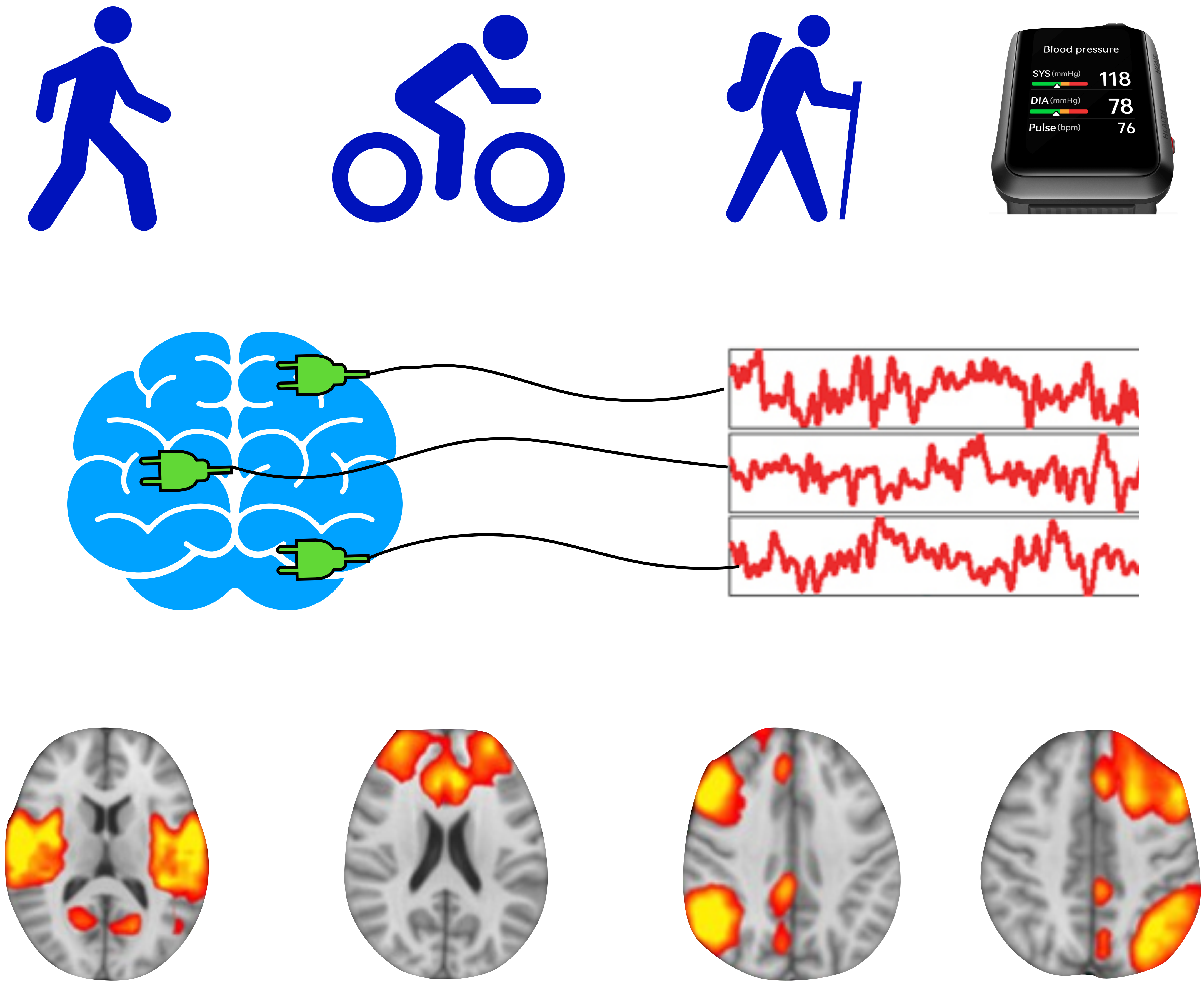
Neural Ordinary Differential Equation (NODE)

Neural Controlled Differential Equations for Irregular Time Series

Neural Rough Differential Equations for Long Time Series

James Morrill [1 2] Cristopher Salvi [1 2] Patrick Kidger [1 2] James Foster [1 2] Terry Lyons [1 2]

Neural Controlled Differential Equation (NCDE)


Blood pressure
SYS (mmHg)
118
DIA (mmHg)
78
Pulse (bpm)
76

NODE:

$$z_t = z_0 + \underbrace{\int_0^t f_\theta(z_s)}_{\text{Parameterized function}} \mathrm{d}s$$

NCDE:

$$z_t = z_0 + \int_0^t f_\theta(z_s) \underbrace{\mathrm{d}\mathbf{X}_s}_{\text{Control Path}}$$

$$= z_0 + \int_0^t f_\theta\left(z_s\right) \mathbf{X}'(s) \mathrm{d}s$$

Integral transform: $$\mathcal{A}(f)(t) = \int \underbrace{a(t, s)}_{\text{Kernel}} f(s)\mathrm{d}s$$

NCDE as integral transform: $$\underbrace{u_t}_{z_t - z_0} = \int_0^t a_\theta(t, s)\underbrace{v(s)}_{\underbrace{(h_\theta(z_s))^T}_{\text{Parameterized transform}} \times \underbrace{X'(s)}_{\text{Derivative of the control path}}}\mathrm{d}s$$

Parameterized transform

Derivative of the control path

$$u_t = \int_0^t a_\theta(t, s) v(s) \mathrm{d}s$$

Discretize

$$\underbrace{\mathbf{u}}_{T \times 1} = \underbrace{\mathbf{A}}_{T \times T} \underbrace{\mathbf{v}}_{T \times 1}$$

Sequence Length: T

Complexity: $O(T^2)$

Prohibitive for long sequences!

Multi-resolution Analysis

Scaling functions (Father wavelet): $\phi_k^l(\tau) = 2^{l/2}\phi(2^l\tau - k)$

Scaling: l

Translation: k

Wavelet functions (Mother wavelet): $\psi_k^l(\tau) = 2^{l/2}\psi(2^l\tau - k)$

Kernel: $a(t,s)$

$$\alpha_{km}^{l} = \iint \psi_k^l(t)\, a(t,s)\, \psi_m^l(s)\, \mathrm{d}t\mathrm{d}s$$

$$\beta_{km}^{l} = \iint \psi_k^l(t)\, a(t,s)\, \phi_m^l(s)\, \mathrm{d}t\mathrm{d}s$$

$$\gamma_{km}^{l} = \iint \phi_k^l(t)\, a(t,s)\, \psi_m^l(s)\, \mathrm{d}t\mathrm{d}s$$

$$A_{km}^{l} = \iint \phi_k^l(t)\, a(t,s)\, \phi_m^l(s)\, \mathrm{d}t\mathrm{d}s$$

α^l	β^l
γ^l	A^l

Forward Wavelet Transform: $\mathscr{W}^T$

Inverse Wavelet Transform: $\mathscr{W}$

$$\mathscr{W}^T A^l \mathscr{W} = \begin{bmatrix} \alpha^{l+1} & \beta^{l+1} \\ \gamma^{l+1} & A^{l+1} \end{bmatrix}$$

$$A^l = \mathscr{W} \begin{bmatrix} \alpha^{l+1} & \beta^{l+1} \\ \gamma^{l+1} & A^{l+1} \end{bmatrix} \mathscr{W}^T$$

Calderon-Zygmund Operator

FAST WAVELET TRANSFORMS AND NUMERICAL ALGORITHMS I.

G. Beylkin[1,2], R. Coifman[2], V. Rokhlin[3]

Yale University
New Haven, Connecticut 06520

Kernel: $a(t, s)$

$$|a(t,s)| \leq \frac{1}{|t-s|}$$

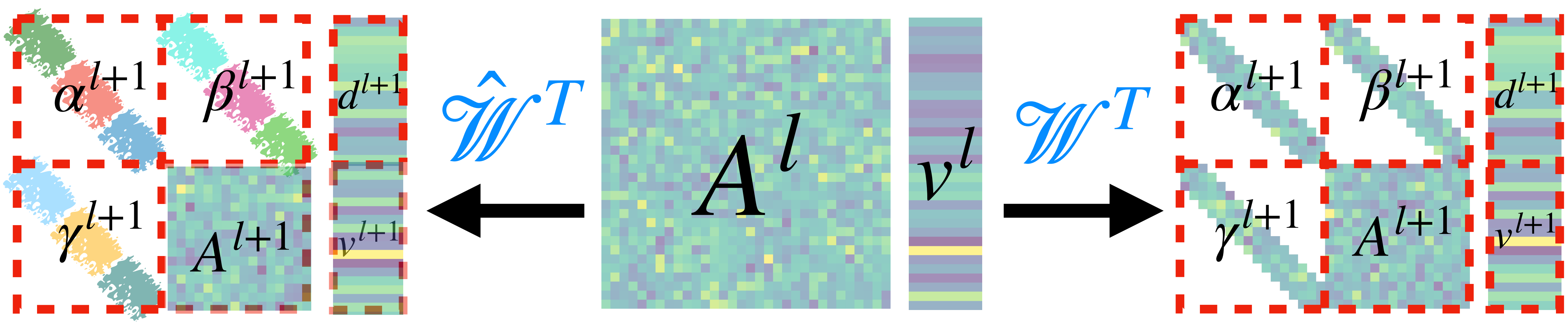
$$|\partial_t^M a(t,s)| + |\partial_s^M a(t,s)| \leq \frac{C_0}{|t-s|^{1+M}}$$

$$|\alpha_{km}^l| + |\beta_{km}^l| + |\gamma_{km}^l| \leq \frac{C_M}{1+|k-m|^{1+M}}$$

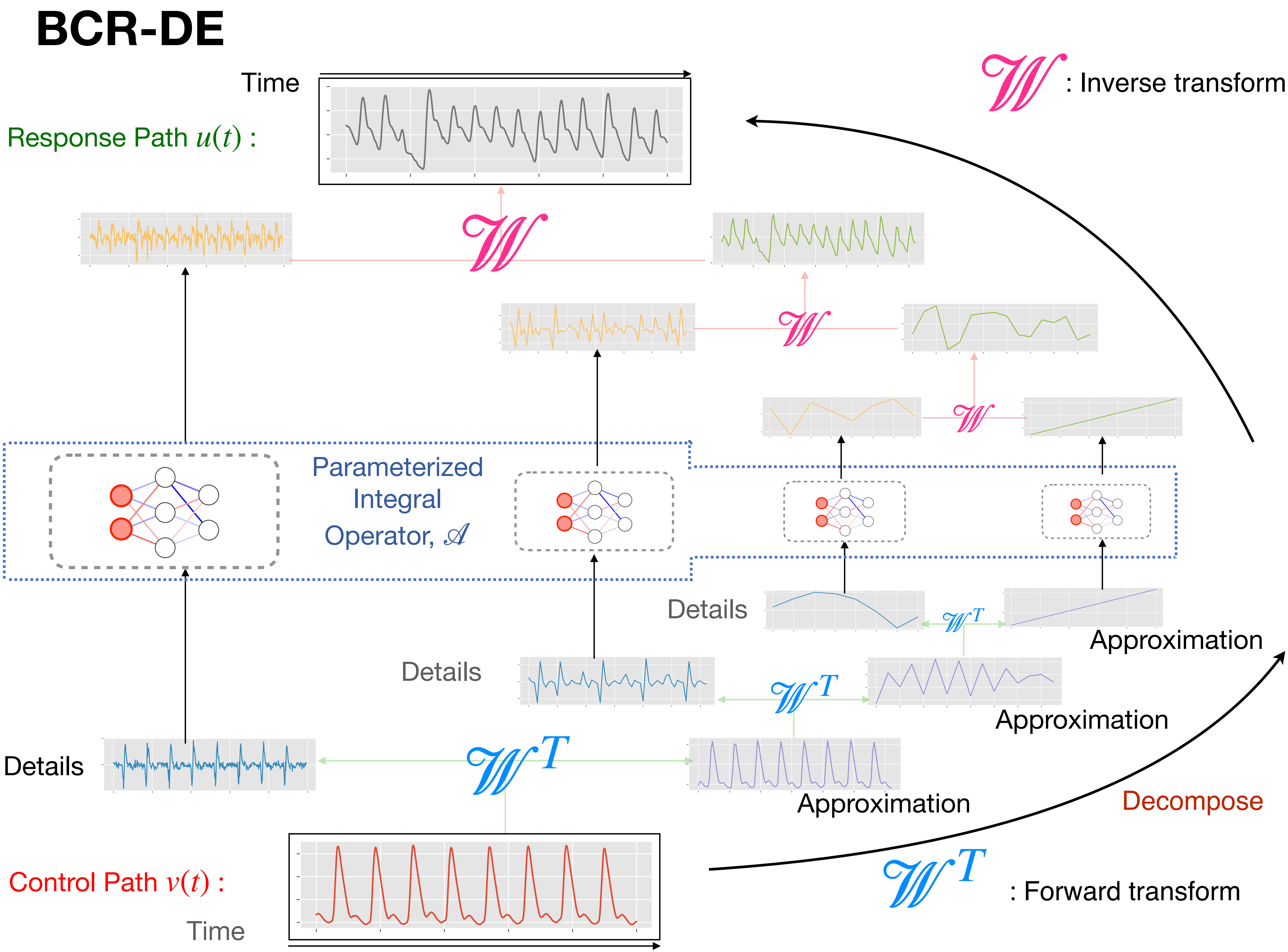
Fast Matrix-Vector Product

$$u^l = A^l v^l = \mathscr{W}\left(\begin{bmatrix} \alpha^{l+1} & \beta^{l+1} \\ \gamma^{l+1} & 0 \end{bmatrix} \begin{bmatrix} d^{l+1} \\ v^{l+1} \end{bmatrix} + \begin{bmatrix} 0 \\ u^{l+1} \end{bmatrix} \right)$$

Recursion!!



Partially Unshared Convolution (PUC)



BCR-DE
Time
Response Path $u(t)$:
: Inverse transform
Parameterized
Integral
Operator, $\mathscr{A}$
Details
Details
Details
Approximation
Approximation
Approximation
Decompose
: Forward transform
Control Path $v(t)$:
Time

Evaluation

How does BCR-DE compare to NCDE and NRDE on standard benchmarks?

Does BCR-DE provide an efficient sequence to sequence model for long sequences?

BCR-DE models coupled differential equations?

Evaluation

How does BCR-DE compare to NCDE and NRDE on standard benchmarks?

Does BCR-DE provide an efficient sequence to sequence model for long sequences?

BCR-DE models coupled differential equations?

Evaluation

Physiological Measurements (Regression)

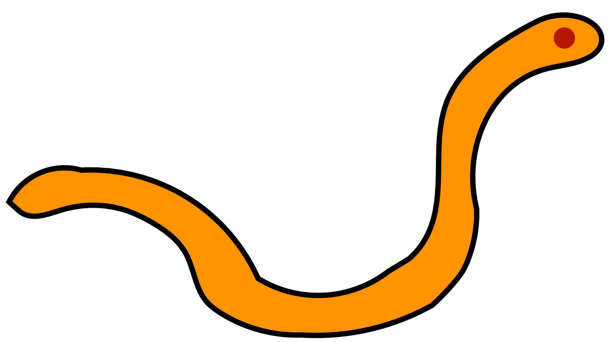
Model	RMSE			Time (hrs)		
	RR	HR	SpO_2	RR	HR	SpO_2
ODE-RNN (s512)	1.66 ± 0.06	6.75 ± 0.9	1.98 ± 0.31	0.0	0.1	0.1
NCDE (s1)	2.79 ± 0.04	9.82 ± 0.34	2.83 ± 0.27	23.8	22.1	28.1
NCDE (s512)	2.53 ± 0.03	12.22 ± 0.11	2.98 ± 0.04	0.1	0.0	0.1
NRDE (d3s8)	2.42 ± 0.19	7.67 ± 0.40	2.55 ± 0.13	2.9	3.2	3.1
NRDE (d3s128)	**1.51 ± 0.08**	**2.97 ± 0.45**	1.37 ± 0.22	0.5	1.7	1.7
NRDE (d3s512)	**1.49 ± 0.08**	3.46 ± 0.13	**1.29 ± 0.15**	0.3	0.4	0.4
BCR-DE	**1.53 ± 0.09**	**3.27 ± 0.16**	**1.18 ± 0.15**	0.4	0.5	0.9

Sequence Length: 4000

Dataset: BIDMC32

Evaluation

5-Class Classification



Model	Accuracy (%)	Time (hrs)
ODE-RNN (s128)	47.9 ± 5.3	0.01
NCDE (s4)	66.7 ± 11.8	5.5
NCDE (s128)	48.7 ± 2.6	0.1
NRDE (d2s4)	**83.8 ± 3.0**	2.4
NRDE (d3s128)	68.4 ± 8.2	0.1
BCR-DE	77.8 ± 1.2	0.01
BCR-DE (Noise)	78.7 ± 2.4	0.01

Sequence Length: 17000

Dataset: Eigenworms

Evaluation

How does BCR-DE compare to NCDE and NRDE on standard benchmarks?

Does BCR-DE provide an efficient sequence to sequence model for long sequences?

BCR-DE models coupled differential equations?

Evaluation

Auto-encoding (Medium Length sequence)

Task	Dataset	NCDE		NRDE		BCR-DE	
		MSE	Time (hrs)	MSE	Time (hrs)	MSE	Time (hrs)
AE	PPG	**6.05e-5**	3.63	0.014	0.67	0.012	0.2
	ECG	**6.06e-5**	3.03	0.014	0.57	0.024	0.19
DAE	PPG	**0.008**	4.1	0.023	0.92	**0.009**	0.18
	ECG	**0.008**	3.04	0.023	0.73	0.02	0.18
MAE	PPG	0.28	2.23	0.106	5.47	**0.024**	0.22
	ECG	0.29	1.5	0.106	3.76	**0.097**	0.23

Sequence Length: 4000

Dataset: BIDMC32

Evaluation

How does BCR-DE compare to NCDE and NRDE on standard benchmarks?

Does BCR-DE provide an efficient sequence to sequence model for long sequences?

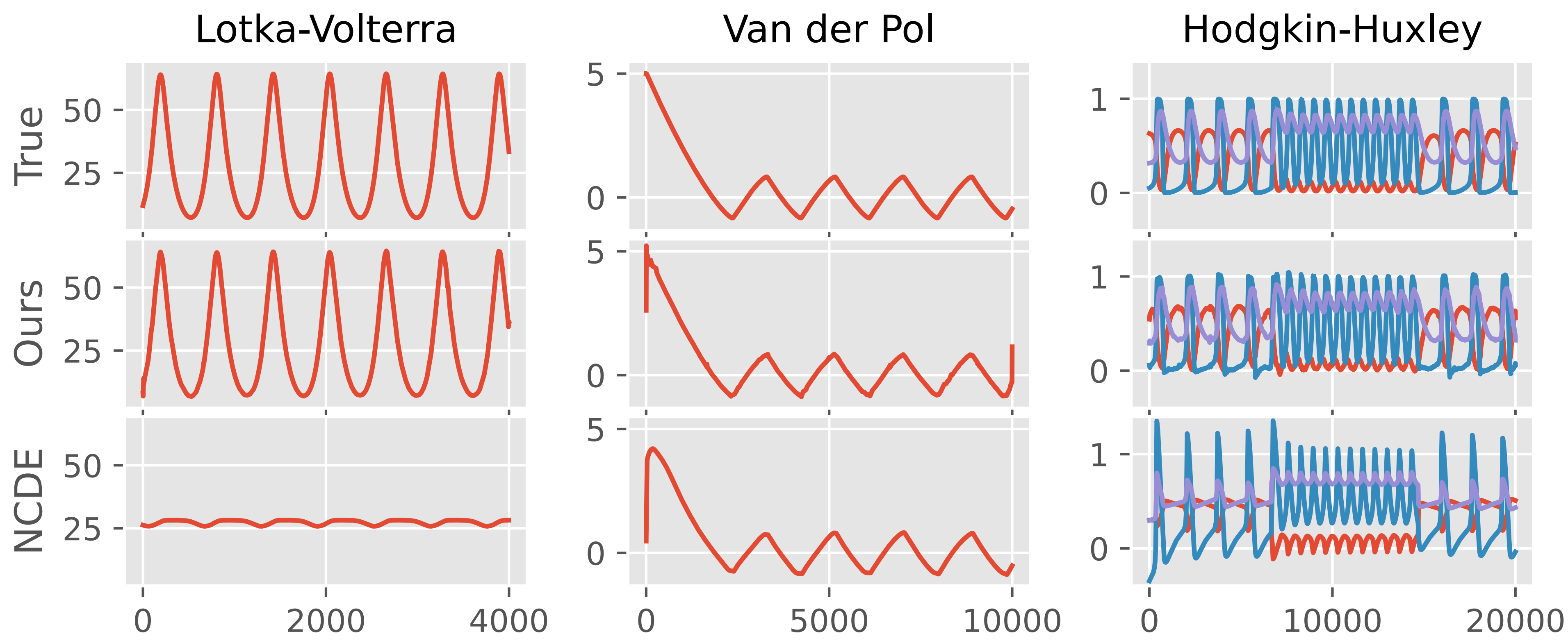
BCR-DE models coupled differential equations?

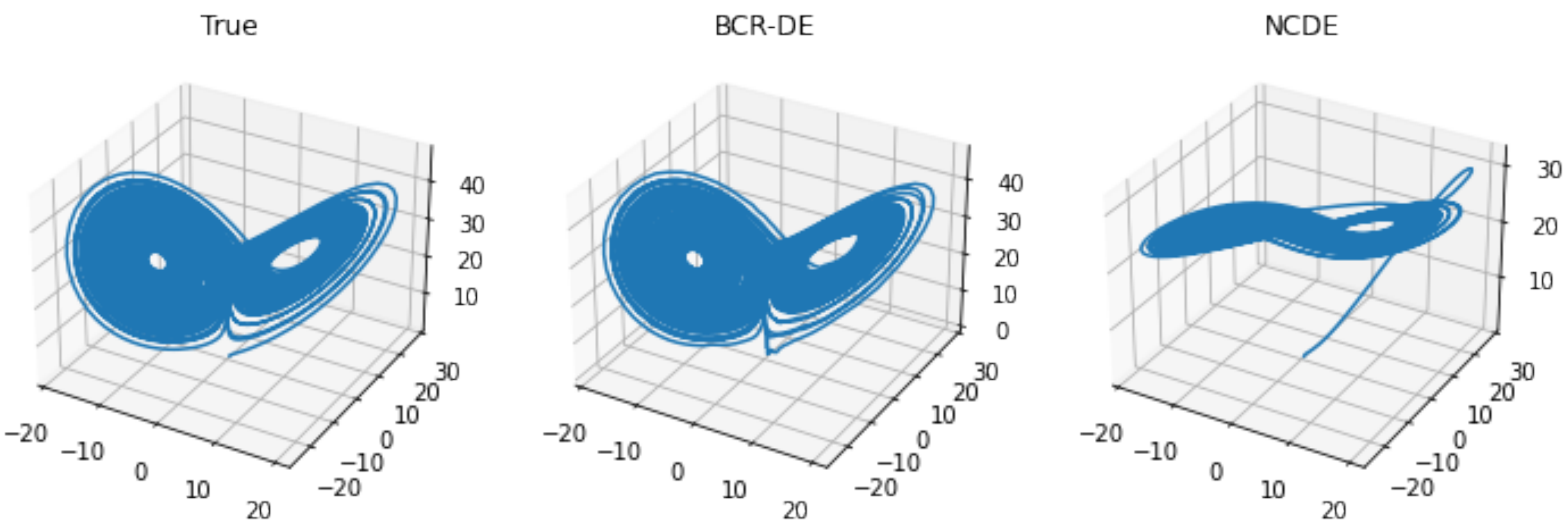
Evaluation

Coupled Differential Equations

Setting (Seq Len)	NCDE		NRDE		BCR_DE	
	MSE	Time (hrs)	MSE	Time (hrs)	MSE	Time (hrs)
Toy Coupled DE (4k)	1e-4	0.62	**6e-5**	0.02	3e-4	0.009
Lotka-Volterra (4k)	377.9	43.74	365.4	3.04	**0.19**	0.134
Van der Pol (10k)	0.023	43.6	0.94	7.43	**1e-3**	0.34
Chaotic Lorenz (10k)	66.15	42.9	133.3	3.7	**0.05**	0.35
Hodgkin-Huxley (20k)	0.02	45.35	1.24	4.28	**4e-4**	0.35
Benzene Conc. (240)	250.3	3.64	725.4	1.17	**212.9**	0.046

Evaluation


Lotka-Volterra
Van der Pol
Hodgkin-Huxley
True
Ours
NCDE
50
25
5
0
1
0
2000
4000
5000
10000
20000


True
BCR-DE
NCDE
−20
−10
0
10
20
30
40

Conclusion

BCR-DE to model controlled differential equations via **integral transform (operator)** and **Multi-Resolution Analysis (MRA)**

Efficient strategy to model **long** but *fixed* length sequences, by **unrolling** the dynamics

Best when number of levels of decomposition is large leading to **small** but effective coarse representation of the signal.

Efficient way to handle **coupled differential equations**

Thank you

Poster Session:

Exhibit Hall 1 #307

Wed 26 Jul 2 p.m. HST — 3:30 p.m. HST

https://github.com/sourav-roni/BCR-DE

spal9@wisc.edu