CoDeC: Communication-Efficient Decentralized Continual Learning

Sakshi Choudhary, ***Sai Aparna Aketi, Gobinda Saha, Kaushik Roy***

Purdue University

NEURAL INFORMATION PROCESSING SYSTEMS

PURDUE UNIVERSITY

INTRODUCTION

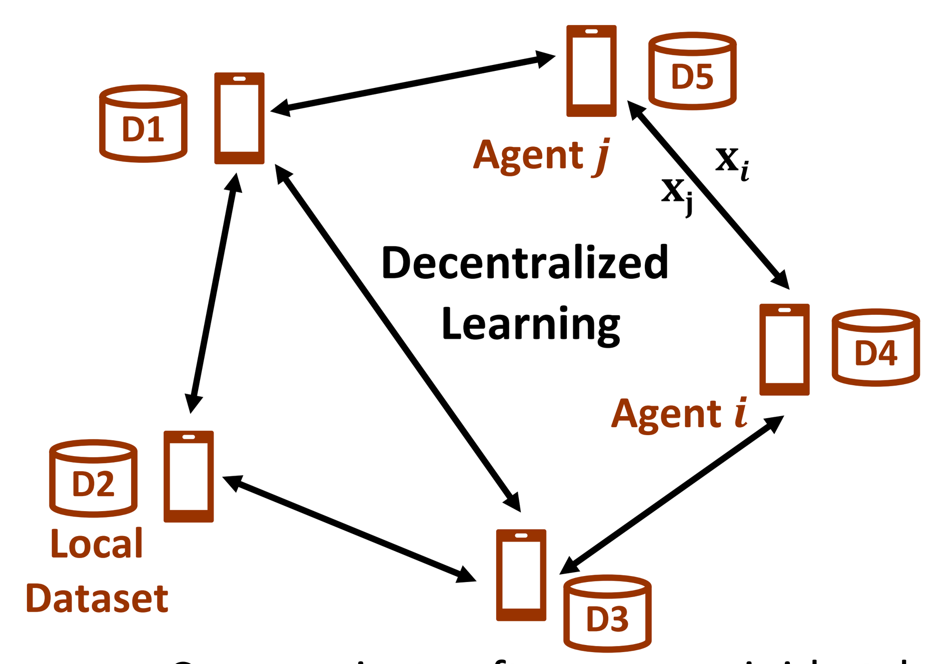
MOTIVATION

- Training at the edge utilizes spatially as well as temporally distributed private data.
- Hence, training algorithms that enable efficient continual learning over decentralized data become crucial.

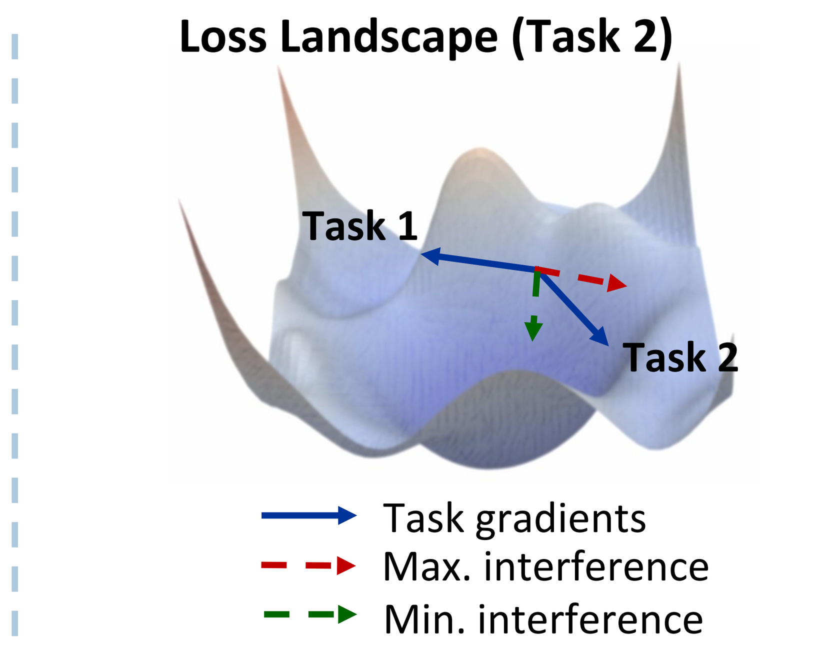
SOLUTION

- **CoDeC**, a novel communication-efficient decentralized continual learning algorithm
- Catastrophic forgetting mitigated with orthogonal gradient projection
- A lossless communication compression scheme based on gradient subspaces

CHALLENGES OF DECENTRALIZED CONTINUAL LEARNING

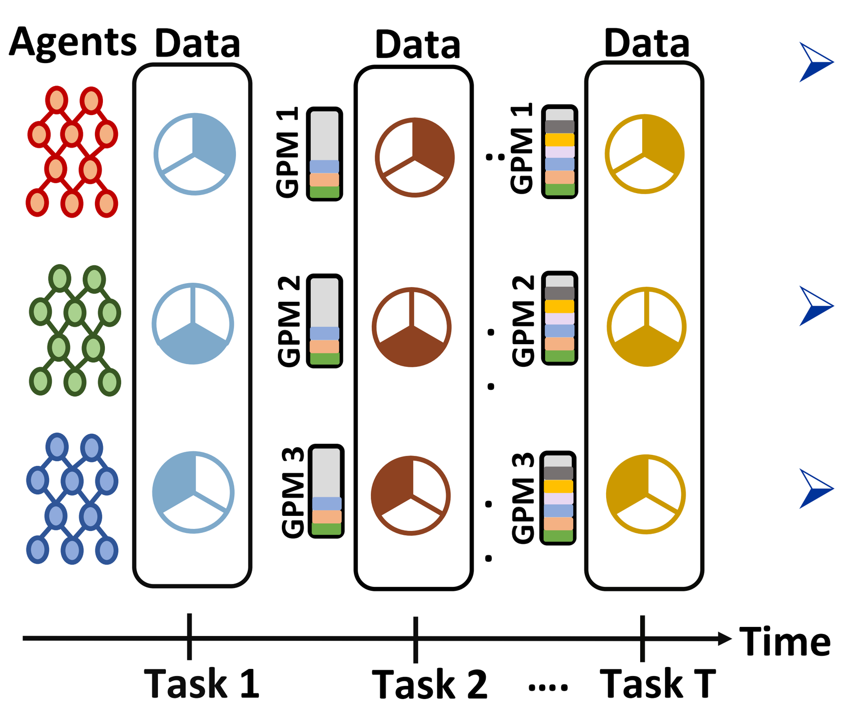


Communicate after every mini-batch update → high communication cost

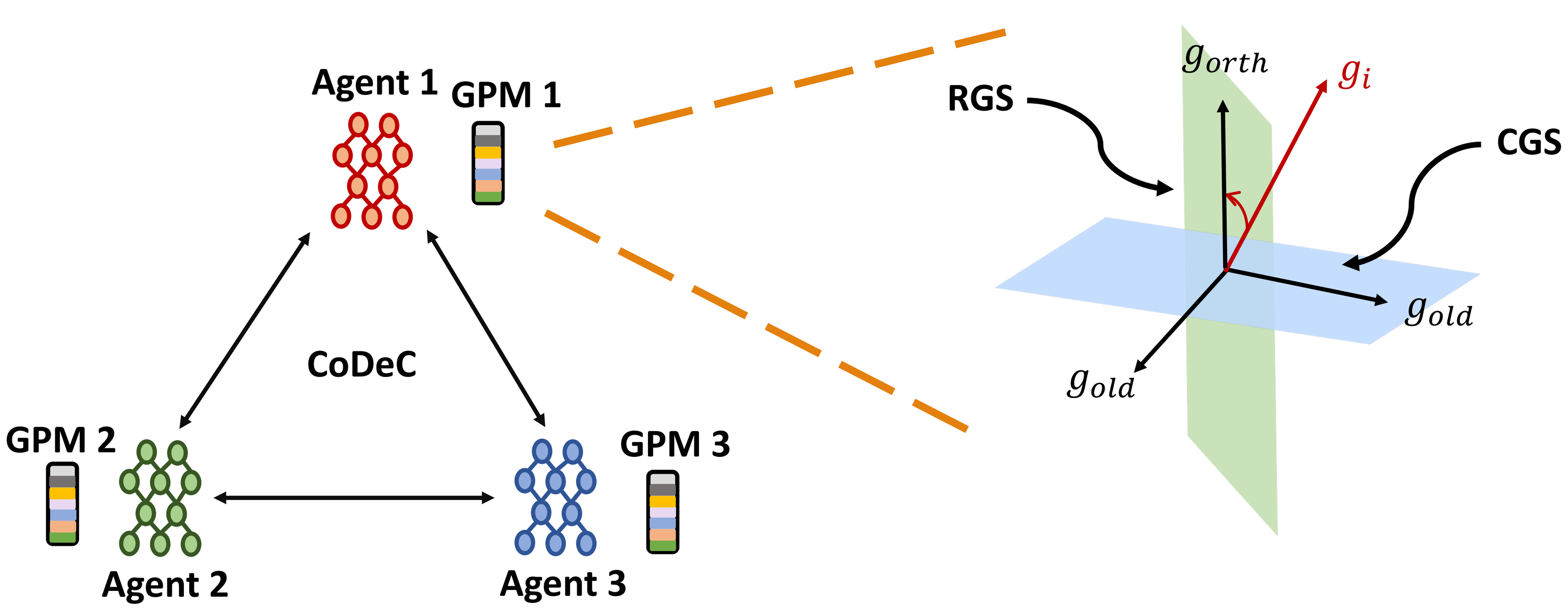


Interference → catastrophic forgetting

SETUP



- Independent and identical distribution (IID) of data for each task across the agents
- The agents communicate coefficients associated with the model updates
- Each layer's subspace partitioned into two subspaces: Core Gradient Space (CGS) & Residual Gradient Space (RGS)



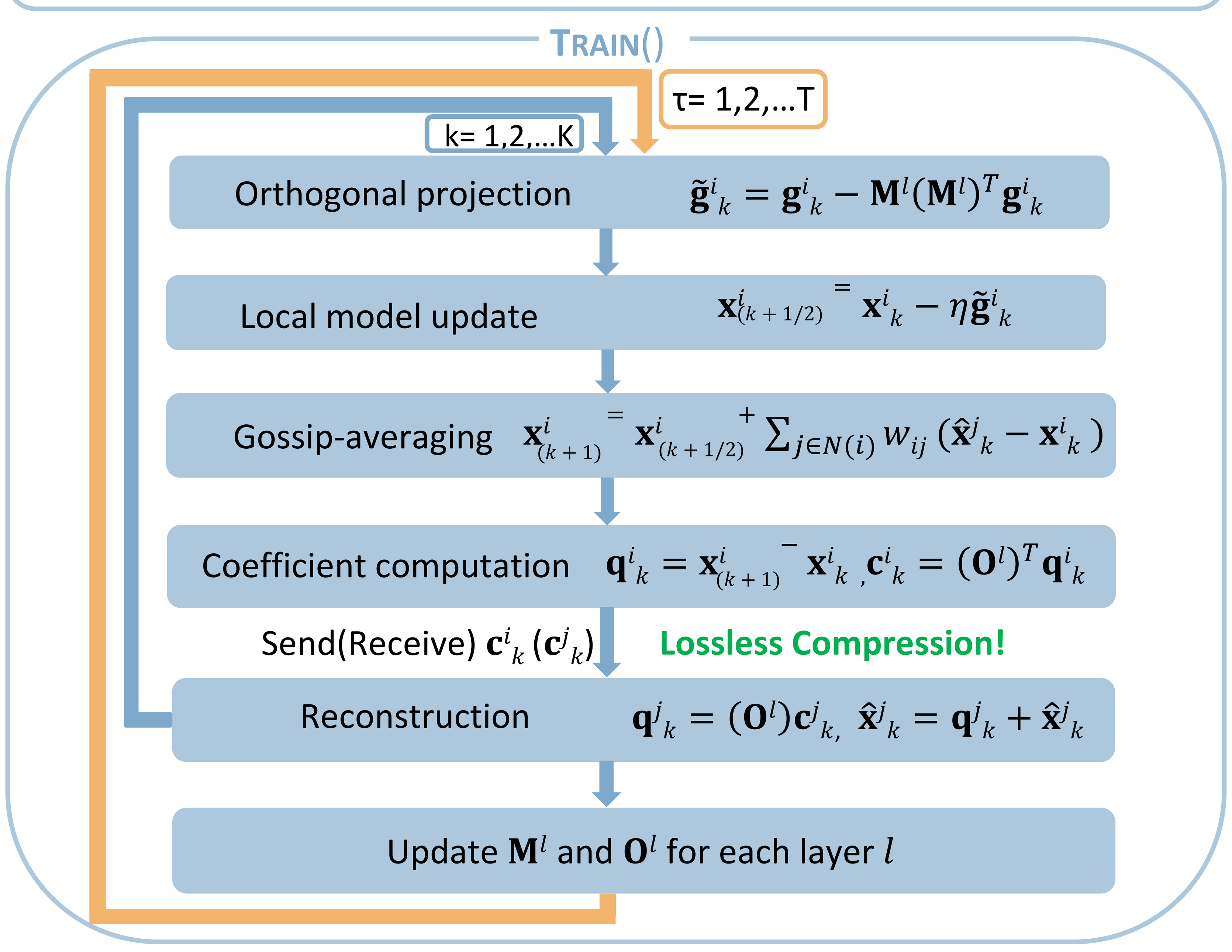
METHODOLOGY

NOTATIONS

N: no. of agents, K: total training iterations, T: total tasks, $[w_{ij}]$: mixing matrix, η: learning rate, $N(i)$: neighbors of agent $i \in [1, N]$

$\mathbf{M}^l$: CGS Matrix, $\mathbf{O}^l$: RGS Matrix

TRAIN()



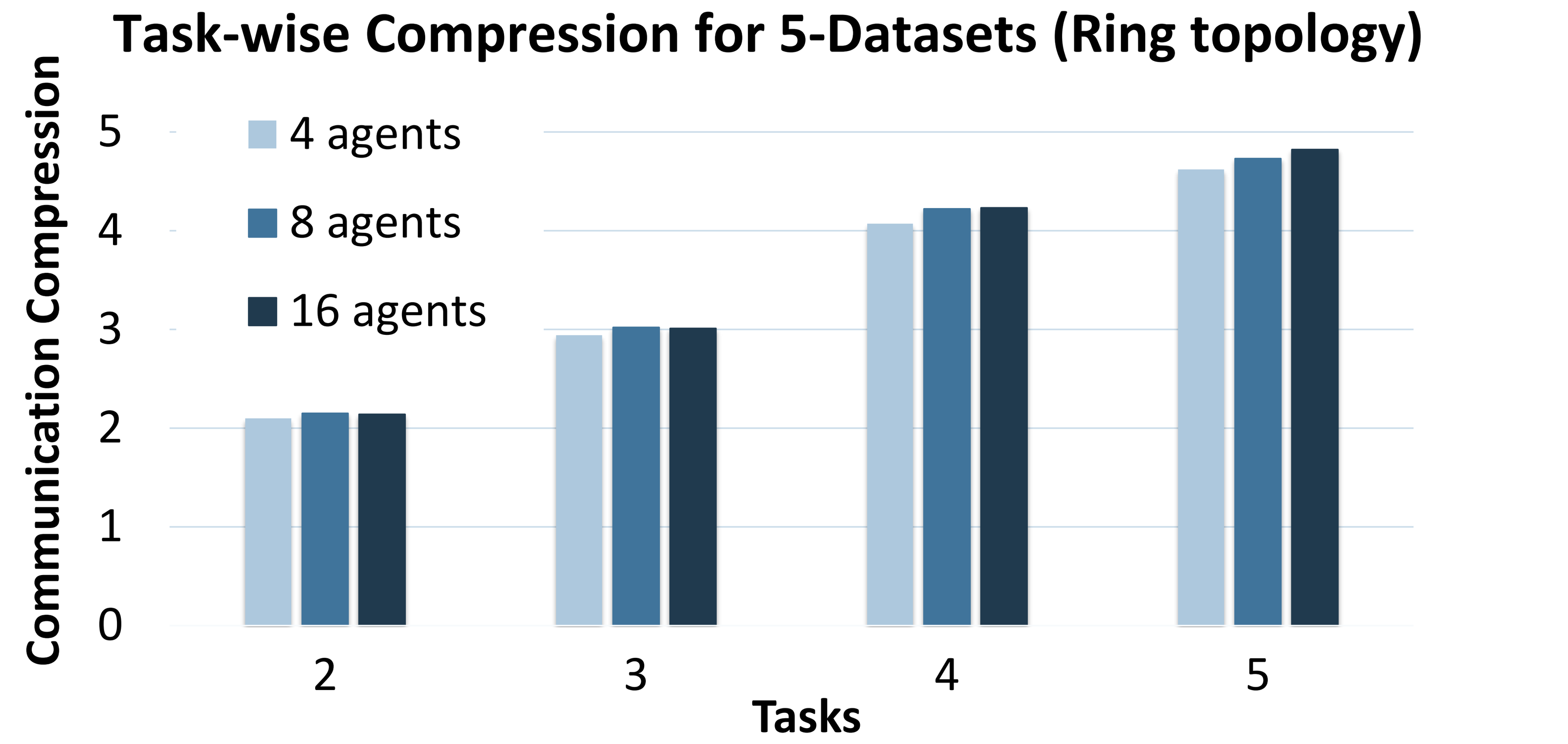
RESULTS

- **Theoretical analysis** of *CoDeC* yields a convergence rate of $O\left(\frac{1}{\sqrt{NK}}\right)$
- D-EWC & D-SI: Elastic Weight Consolidation (EWC) and Synaptic Intelligence (SI) adapted to a decentralized setting to establish baselines

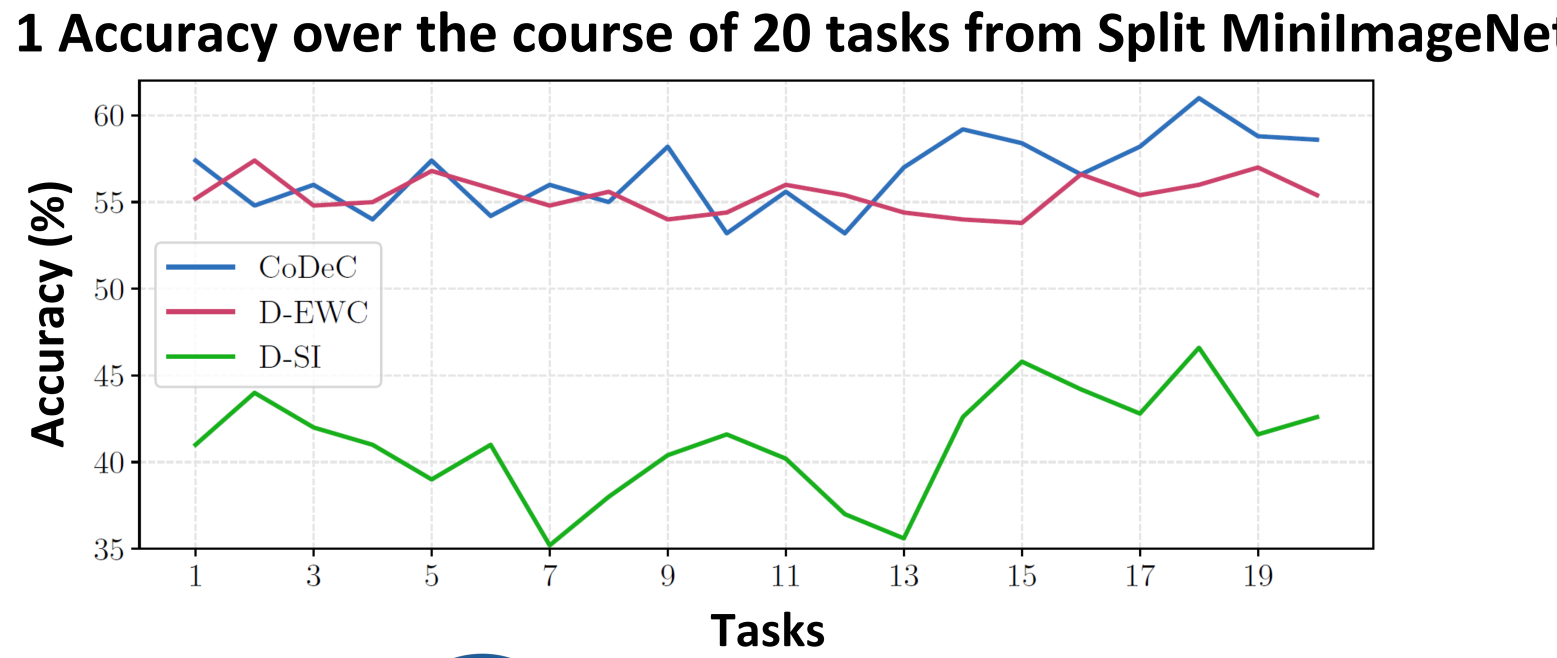
Dataset	Agents	Setup	Directed Ring ACC(%)	Directed Ring BWT(%)	Directed Ring CC	Torus ACC(%)	Torus BWT(%)	Torus CC
Split CIFAR-100	8	STL	64.99 ± 0.41	-	-	65.17 ± 0.44	-	-
		D-SI	39.54 ± 0.16	-1.08 ± 0.85	1x	39.36 ± 0.40	-1.25 ± 0.65	1x
		D-EWC	50.52 ± 0.58	0.51 ± 0.09	1x	49.41 ± 0.88	0.29 ± 0.27	1x
		CoDeC(f)	53.57 ± 0.38	-0.65 ± 0.52	1x	53.54 ± 0.35	-1.15 ± 0.41	1x
		CoDeC	53.63 ± 0.25	-0.43 ± 0.33	1.85x	53.62 ± 0.29	-0.64 ± 0.36	1.86x
	16	STL	58.31 ± 0.49	-	-	59.29 ± 0.12	-	-
		D-SI	34.66 ± 1.15	-1.23 ± 0.4	1x	34.86 ± 0.68	-1.16 ± 0.54	1x
		D-EWC	45.52 ± 0.60	0.22 ± 0.34	1x	44.53 ± 0.77	-0.20 ± 0.56	1x
		CoDeC(f)	48.05 ± 0.45	-0.38 ± 0.12	1x	48.19 ± 0.27	-0.29 ± 0.11	1x
		CoDeC	48.16 ± 0.33	-0.18 ± 0.28	1.84x	48.36 ± 0.04	-0.26 ± 0.31	1.84x
Split miniImageNet	8	STL	63.13 ± 0.86	-	-	66.27 ± 1.47	-	-
		D-SI	45.58 ± 1.24	-3.67 ± 1.27	1x	46.00 ± 0.73	-3.21 ± 0.51	1x
		D-EWC	46.39 ± 1.54	-1.64 ± 1.11	1x	48.23 ± 3.14	-1.02 ± 1.16	1x
		CoDeC(f)	53.22 ± 1.82	0.08 ± 0.45	1x	59.90 ± 0.48	0.37 ± 0.24	1x
		CoDeC	53.30 ± 1.25	-0.46 ± 0.48	1.37x	59.97 ± 0.87	-0.19 ± 0.98	1.53x
	16	STL	57.09 ± 1.55	-	-	63.51 ± 0.61	-	-
		D-SI	39.55 ± 0.87	-2.03 ± 0.69	1x	39.96 ± 0.47	-1.74 ± 0.96	1x
		D-EWC	39.67 ± 1.37	-1.32 ± 1.18	1x	45.14 ± 0.18	-0.64 ± 0.23	1x
		CoDeC(f)	45.29 ± 3.58	-0.99 ± 1.40	1x	51.03 ± 2.51	-0.01 ± 0.67	1x
		CoDeC	45.68 ± 0.77	0.61 ± 0.79	1.42x	51.32 ± 1.05	0.26 ± 0.56	1.39x
5-Datasets	8	STL	92.31 ± 0.06	-	-	92.32 ± 0.15	-	-
		D-SI	80.36 ± 0.15	-2.64 ± 0.07	1x	79.55 ± 0.33	-3.07 ± 0.13	1x
		D-EWC	85.69 ± 0.19	-0.92 ± 0.14	1x	82.99 ± 3.25	-2.10 ± 1.60	1x
		CoDeC(f)	86.54 ± 0.04	-4.37 ± 0.17	1x	85.92 ± 0.18	-5.10 ± 0.17	1x
		CoDeC	86.23 ± 0.22	-4.61 ± 0.32	2.17x	86.15 ± 0.17	-4.85 ± 0.26	2.19x
	16	STL	92.16 ± 0.16	-	-	91.76 ± 0.09	-	-
		D-SI	78.53 ± 0.62	-5.2 ± 0.56	1x	77.00 ± 0.11	-5.76 ± 0.22	1x
		D-EWC	82.19 ± 0.45	-0.18 ± 0.05	1x	81.48 ± 0.12	-0.56 ± 0.14	1x
		CoDeC(f)	86.36 ± 0.15	-4.36 ± 0.19	1x	84.91 ± 0.20	-5.48 ± 0.22	1x
		CoDeC	86.41 ± 0.16	-4.37 ± 0.24	2.16x	85.00 ± 0.55	-5.52 ± 0.35	2.23x

- Dimensionality of RGS reduces as the task sequence progresses.
- Consequently, it suffices for agents to communicate less with their peers.
- Compression ratios range from 2.1x (task 2) to 4.8x (task 5).

Task-wise Compression for 5-Datasets (Ring topology)



Task 1 Accuracy over the course of 20 tasks from Split MiniImageNet



Semiconductor Research Corporation

CoCoSys CENTER FOR THE CO-DESIGN OF COGNITIVE SYSTEMS