

**ACADEMIC AFFAIRS OFFICE
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE**

No. Acd./394/IAPC-105

Dated: June 14, 2021

Head, Department of Electronics & Communication Engg.

The IAPC in its 105th meeting held on 09.06.2021 vide Item No. 105.2.7(2) considered and approved the proposal of Department of Electronics & Communication Engg. to offer under mentioned existing PEC in both semesters.

1. ECN-525: Hardware Architecture for Deep-Learning (**Appendix-A**)



Assistant Registrar (Curriculum)

Encl: as above

Copy to (through e mail):-

1. All faculty
2. Head of all Departments/ Centres
3. Dean, Academic Affairs
4. Associate Dean of Academic Affairs (Curriculum)
5. Channel I/ Acad portal/ Academic webpage of iitr.ac.in

INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NAME OF DEPARTMENT/CENTRE: Department of Electronics and Communication Engineering

1. **Subject Code:** ECN-525 **Course Title:** Hardware Architecture for Deep-Learning
2. **Contact Hours:** **L:** 3 **T:** 1 **P:** 0
3. **Examination Duration (Hrs.):** **Theory:** 3 **Practical:** 0
4. **Relative Weightage:** **CWS:** 20-35 **PRS:** 0 **MTE:** 20-30 **ETE:** 40-50 **PRE:** 0
5. **Credits:** 4 6. **Semester:** Both 7. **Subject Area:** PEC
8. **Pre-requisite:** None
9. **Objective:** To learn the design of hardware architectures and accelerators for deep-learning/artificial-intelligence. This course is at the intersection of deep-learning and computer-architecture/embedded-system/VLSI.

10. Details of the Course

S.No.	Contents	Contact hours
1.	Background topics: Approximate computing and storage, Roofline Model, Cache tiling (blocking), GPU architecture, CUDA programming, understanding GPU memory hierarchy, FPGA architecture, Matrix multiplication using systolic array	8
2.	Convolutional strategies: Direct, FFT-based, Winograd-based and Matrix-multiplication based.	3
3.	Deep Learning on various hardware platforms: Deep learning on FPGAs and case study of Microsoft's Brainwave, Deep learning on Embedded System (especially NVIDIA's Jetson Platform), Deep learning on Edge Devices (smartphones), Deep learning on an ASIC (especially Google's Tensor Processing Unit.), Deep-learning on CPUs and manycore processor (e.g., Xeon Phi), Memristor-based processing-in-memory accelerators for deep-learning.	15
4.	Model-size aware Pruning of DNNs, Hardware architecture-aware pruning of DNNs, Understanding soft-errors. Understanding reliability of deep learning algorithms and accelerators	6
5.	Comparison of memory technologies (SRAM, DRAM, eDRAM, STT-RAM, PCM, Flash) and their suitability for designing memory-elements in DNN accelerator, Neural branch predictors and their applications	4
6.	Hardware/system-challenges in autonomous driving, Distributed training of DNNs and addressing memory challenges in DNN training	6
Total		42

11. Suggested Books:

S.No.	Name of Authors/Book/Publisher	Year of Publication / Reprint
1.	Computer Architecture: A quantitative approach (Sixth Edition), Hennessy, J. L., & Patterson, D. A., Elsevier https://www.google.co.in/books/edition/Computer_Architecture/cM8mDwAAQBAJ	2017

2.	Deep Learning for Computer Architects Brandon Reagen, Robert Adolf, Paul Whatmough, Gu-Yeon Wei, and David Brooks Synthesis Lectures on Computer Architecture, August 2017, Vol. 12, No. 4, Pages 1-123 (https://doi.org/10.2200/S00783ED1V01Y201706CAC041)	2017
3.	General-Purpose Graphics Processor Architectures Tor M. Aamodt, Wilson Wai Lun Fung, and Timothy G. Rogers, Synthesis Lectures on Computer Architecture, May 2018, Vol. 13, No. 2, Pages 1-140 (https://doi.org/10.2200/S00848ED1V01Y201804CAC044)	2018
4.	Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1, No. 2) . Cambridge: MIT press.	2016
5.	Selected research papers	