Scheduling Inputs in Early Exit Neural Networks (Online Supplement)

Giuliano Casale, Manuel Roveri

Appendix A

We give here an example to illustrate the concepts using experimental data. This example concerns a custom 25-layers CNN endowed with ICs deployed at CNN layers $\{6, 12, 18, 25\}$, with the IC at the last layer simply being the final classifier. More specifically, the CNN comprises three 3×3 Conv2D layers with 64 filters, two 3×3 Conv2D layers with 128 filters, two 3×3 Conv2D layers with 256 filters, a dense layer with 2048 neurons and a final layer dense layer a number of neurons equal to the number of classes in the classification problem, which is the *CIFAR10* image classification dataset. Batch Normalization, Relu Activation, and 2×2 Max Pooling layers are also present. More details are given in Section VIII.

As mentioned, CNNs of this kind are treated for the sake of our notation as equal to an EEN with $L = 4$ layers. The intermediate ICs are characterized by the three thresholds, set to $\tau_{i,1}$, $\tau_{i,2}$ and $\tau_{i,3}$ for the ith served job, while $\tau_{i,4}$ is omitted since jobs always exit if they reach the final layer.

We illustrate in Table I a set of configurations measured on the considered EEN using its training set. The table shows the conditional accuracy of the EEN layers $l = \{1, \dots, 4\}$ with respect to thresholds $\tau_{i,l}$. The thresholds are assumed to be applied identically for every job i. The first four configurations $\{\gamma^{(1)}, \dots, \gamma^{(4)}\}$ represent single-exit configurations, while the remaining ones represent examples of multi-exit configurations. Note that, although in our notation conditional accuracies have a different symbol for single-exit configurations ($A^{(i)}$) than for multi-exit configurations (A_i), they are the same metric. That is, the $A^{(i)}$ and A_i symbols refer to average classification accuracies for jobs that exit at layer i.

The experiments in the table confirm that single-exit configurations $\gamma^{(i)}$ evidently do not introduce inter-dependencies among the $A^{(i)}$, since only one IC is active for any job. Instead, this does not hold when multi-exit configurations are considered where dependencies. This is visible, for example, in the changing values of the conditional accuracies A_4 at the final layer of τ_i^a, τ_i^b, τ_i^c, and τ_i^d. Characterizing such dependencies is challenging, as they are data-dependent, and the resulting model may also be difficult to use for searching an optimal schedule due to its dimensionality.

The data in Table I also confirm that the use of single exit configurations does not restrict the achievable accuracy. Table II shows the percentage of jobs exiting at the first three ICs q_1, q_2, q_3 and at the final classifier q_4 and the mean accuracy A of the EEN with respect to a multi-exit configuration τ^d, i.e., one of the multi-exit configurations considered in Table I, and the single-exit configurations $\{\gamma^{(1)}, \dots, \gamma^{(4)}\}$. The thresholds of the τ^d configuration have been selected as a representative example of a configuration that gives fairly high accuracy, in particular with high conditional accuracy at the first IC, which is a desirable property for a multi-exit EEN. Since A of τ^d lies between $A^{(2)}$ and $A^{(3)}$, and the input is i.i.d., we may devise a randomized policy that achieves the same accuracy using single-exit configurations alone. Assume layer $m = 2$ and $M = 3$ so that $A^{(m)} = 0.7543$ and $A^{(M)} = 0.8121$, respectively. This yields an exit probability

$$q_m = \frac{0.8121 - 0.7838}{0.8121 - 0.7543} = 0.4896 \tag{1}$$

Hence, according to Observation 1, we can define a non-preemptive policy with the same mean accuracy as configuration τ^d by choosing configuration $\gamma^{(2)}$ with probability $q_m = 0.4896$ or using $\gamma^{(3)}$ otherwise. We have carried out experimentally a test on this policy with constant exit vector $q = (0, 0.4896, 0.5104, 0)$ on two different partitions of the training set of about 8000 inputs each, obtaining a measured average accuracy of $A = 0.7838$ compared to the theoretical value of $A = 0.7832$, thus with negligible error due to the finite size of the input dataset.

Lastly, we compare the processing times of the different configurations. Table III introduces the processing times p_j^{Φ} and p_j^{Δ} appearing in (2), while Table IV details the values of p_l for the configuration τ and the single-exit configurations $\gamma_j^{(l)}$, $l \in G_L$. The overall processing time for the non-preemptive policy based on $\gamma^{(2)}$ and $\gamma^{(3)}$, is $65.05 \cdot 10^6 \times 0.4896 + 108.56 \cdot 10^6 \times (1 - 0.4896) = 87.25 \cdot 10^6$ MAC. This is lower than the overall processing time for configuration τ^d, which is $\sum_{i=1}^{4} q_i p_i = 99.22 \cdot 10^6$ MAC. Overall, this shows an example where a randomized non-preemptive policy that forces exit at layers 2 and 3 only can achieve the same mean classification accuracy, but lower processing time, than the original multi-exit configuration τ^d.

Appendix B

Given a reference hardware platform endowed with an off-the-shelf microcontroller (e.g., a board supporting the ARM Cortex M4 microcontroller), the typical energy consumption in RUN mode, with all the peripherals disabled, is approximately 50mA @ 200Mhz[1]. When the supply voltage is 3.3 V, the

G. Casale is with the Department of Computing, Imperial College London, London, UK, SW7 2AZ.
E-mail: g.casale@imperial.ac.uk
M. Roveri is with the Dipartimento di Elettronica, Informazione e Bioingengeria, Politecnico di Milano, 27100 Milan, Italy.
E-mail: manuel.roveri@polimi.it

[1] STMicroelectronics's AN4365 Application note: https://www.st.com/en/embedded-software/stsw-stm32142.html

TABLE I
THRESHOLDS AND CONDITIONAL ACCURACIES OF THE RUNNING CASE UNDER SINGLE-EXIT AND MULTI-EXIT CONFIGURATIONS

	$\tau_{i,1}$	$\tau_{i,2}$	$\tau_{i,3}$	$A^{(1)}$	$A^{(2)}$	$A^{(3)}$	$A^{(4)}$	A
$\gamma^{(1)}$	0.0	1.0	1.0	0.5973	-	-	-	0.5973
$\gamma^{(2)}$	1.0	0.0	1.0	-	0.7543	-	-	0.7543
$\gamma^{(3)}$	1.0	1.0	0.0	-	-	0.8121	-	0.8121
$\gamma^{(4)}$	1.0	1.0	1.0	-	-	-	0.8439	0.8439
	$\tau_{i,1}$	$\tau_{i,2}$	$\tau_{i,3}$	A_1	A_2	A_3	A_4	A
τ^a	0.9	1.0	1.0	0.8221	-	-	0.7795	0.7995
τ^b	0.9	0.9	1.0	0.8221	0.7252	-	0.6627	0.7637
τ^c	0.9	0.9	0.9	0.8221	0.7252	0.6737	0.4928	0.7611
τ^d	0.9	1.0	0.8	0.8221	-	0.7706	0.5267	0.7838

TABLE II
THRESHOLDS, EXIT PROBABILITIES AND ACCURACY AN EEN AS IN FIGURE 2 WITH $L = 4$, UNDER VARYING CONFIGURATIONS

	$\tau_{i,1}$	$\tau_{i,2}$	$\tau_{i,3}$	q_1	q_2	q_3	q_4	A
τ^d	0.9	1.0	0.8	0.468	0.0	0.487	0.045	0.7838
$\gamma^{(1)}$	0.0	1.0	1.0	1.0	0.0	0.0	0.0	0.5973
$\gamma^{(2)}$	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.7543
$\gamma^{(3)}$	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.8121
$\gamma^{(4)}$	1.0	1.0	1.0	0.0	0.0	0.0	1.0	0.8439

TABLE III
PROCESSING TIMES (IN $1 \cdot 10^6$ MACS) FOR THE RUNNING CASE

p_1^Φ	p_2^Φ	p_3^Φ	p_4^Φ	p_1^Δ	p_2^Δ	p_3^Δ	p_4^Δ
39.78	16.84	47.72	33.67	16.85	8.42	4.21	0.02

TABLE IV
PROCESSING TIME (IN $1 \cdot 10^6$ MACS) FOR THE SINGLE-EXIT AND MULTI-EXIT CONFIGURATIONS

	p_1	p_2	p_3	p_4
τ	56.63	81.90	133.84	167.54
$\gamma^{(1)}$	56.63	-	-	-
$\gamma^{(2)}$	-	65.05	-	-
$\gamma^{(3)}$	-	-	108.56	-
$\gamma^{(4)}$	-	-	-	138.04

TABLE V
COMPARISON OF REACTIVE POLICIES ON THE RUNNING CASE.

Policy	Slowdown R/p_{max}	Loss ratio D	Accuracy A
exit-last: $\gamma^{(4)}$	1.1911	0.0715	0.8439
reactive: $I = \{1,2,3,4\}$	0.6528	0.0150	0.8085
reactive: $I = \{1,4\}$	0.8861	0.0295	0.8021
exit-first: $\gamma^{(1)}$	0.2042	0.0001	0.5973

power consumption in RUN mode is 165 mW. Off-the-shelf microcontrollers typically require 4 cycles to compute floating point operations. Hence a MAC that comprises 2 floating point operations requires 8 cycles. Hence, the MAC clock is $25MHz$, meaning we can execute $25M$ MACs per second. Each MAC requires roughly $5 \cdot 10^{-5}ms$ to be executed. Considering the computed power consumption, each MAC requires approximately $0.0066\ uJ$ to be executed. Summing up, given the amount of MAC and the reference target hardware, the energy consumption can be directly computed by multiplying the amount of MAC operations per the energy consumption of a single MAC, as previously computed. For this reason, the amount of MAC operations represents a direct proxy of the energy consumption.

APPENDIX C

A. Running case for the reactive policy

We illustrate the list-based policy π_L on the example introduced in Section IV-B. We compare a four-level policy based on increasing conditional accuracies $I = \{1, 2, 3, 4\}$, a two-level policy with $I = \{1, 4\}$, and the single-exit configurations $\gamma^{(1)}$ and $\gamma^{(L)}$, referred to as *exit-first* and *exit-last* policies. We simulate 10^6 Poisson arrivals to an edge service that runs the EEN of the running case with arrival rate $\lambda = 0.007$ jobs/s and system capacity $K = 6$. Henceforth, we indicate with $p_{max} = \max_l p_l$ the maximum processing time for single-exit configurations, where the p_l values are computed as in Eq. (3). With these definitions, we study the *slowdown* metric, defined as R/p_{max}, which quantifies the effect of queueing; this is normalized by p_{max} for fairer comparison across neural network models with different processing times. We also consider the loss ratio D, which quantifies the probability that an incoming job is lost due to finite buffer capacity, and the mean accuracy A of completed jobs.

The results in the table suggest that the reactive policies provide a decrease in slowdown and loss ratio while retaining accuracy quite close to a complete pipeline process (γ_4) and much larger than an exit-first exit (γ_1). The result suggest that even very simple non-preemptive policies based on single-exit configurations can improve slowdown and loss ratios in EEN-based systems in return for marginal decreases in accuracy.

B. Running case for the model-based policy

Formula (11) is illustrated in Figure 1 for the 4-layer example introduced in Section IV-B. The edge service has a capacity $K = 6$. Recalling that in this example $m = 2$ and $M = 3$, a range of load factors $\rho = \lambda p \in \{1.2, 2.0\}$ is obtained by setting $q_2 = (p_3 - p)/(p_3 - p_2)$, $q_3 = 1 - q_2$. The SCV of the processing times resulting from this assignment varies with the load in $s^2 \in [0, 0.1265]$. The figure shows that the formula generally gives a close approximation for the loss ratio, which is tight to the exact value in higher loads.

Table VI illustrates the application of $\pi_M(D^*, \lambda)$ and $\pi_{MB}^{heur}(D^*, \lambda)$. As introduced before, the D metric describes a

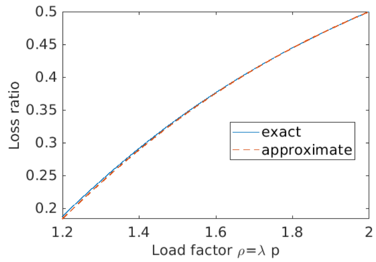


Fig. 1. Loss ratio approximation (11) applied to the running case.

TABLE VI
APPLICATION OF THE MODEL-BASED POLICIES BASED ON (10) AND (12) ON THE RUNNING CASE OF SECTION IV-C

Sched. Policy	Target D^*	Actual D	Accuracy A	EEN SCV s^2	Runtime (s)
(10)	0.0001	0.0001	0.7571	0.0193	15.6198
(12)	0.0001	0.0000	0.5973	0.0000	0.0060
(10)	0.0005	0.0005	0.7749	0.0673	40.2377
(12)	0.0005	0.0004	0.7640	0.0964	0.0051
(10)	0.0010	0.0010	0.7859	0.0597	60.0000
(12)	0.0010	0.0010	0.7792	0.0661	0.0048
(10)	0.0100	0.0100	0.8278	0.0144	60.0000
(12)	0.0100	0.0091	0.8234	0.0141	0.0049

job loss ratio, which the system designer seeks to match with the target value D^*, A is the mean accuracy of completed jobs, and s^2 quantifies the variability of processing time distribution resulting from the chosen policy. The Poisson arrival rate is expressed in terms of average arrivals per time to carry out a MAC operation, i.e., $\lambda = 0.03$ jobs/s. Due to the small size of the EEN, we restrict the execution of the global search used to evaluate the optimization program (10) to 60 seconds. The search is based on the GlobalSearch function available in MATLAB 2023a. In the solution of (12), we obtain $s^2_{max}(\rho)$ setting $s^2_+ = 1024$ and a sequence of 30 linear programs for equispaced values of $\rho \in [\lambda p_1, \lambda p_4]$, starting from $\rho = \lambda p_1$ and stopping the sequence as soon as the loss ratio is exceeded. The results shown in Table VI show that the heuristic solution based on the LP (12) is fairly close to the global solution. Running times for a single LP does not exceed 2 ms in all instances, resulting in different cumulative times based on the number of linear programs required to match the loss ratio. The LP has only 4 variables, 4 bounds, and 3 constraints. Hence, the policy based on this program may reasonably be implemented also on some edge devices for online use, while a global search is less likely to be acceptable for use directly on edge devices. The largest deviations are noticeable for $D^* = 0.0001$, for which the heuristic (11) is overly conservative, achieving a loss ratio of $2 \cdot 10^{-5}$ and setting all the exit probabilities in the first layer, which has accuracy $A_1 = 0.5973$. It should be noted that, if needed, using the exact $M/GI/1/K$ solution one may easily detect the overestimation and artificially inflate to some degree the target D^* supplied to the heuristic method to compensate for the error.

TABLE VII
SIMULATION RESULTS FOR THE RUNNING CASE IN SECTION C-C

Policy	Slowdown R/p_{max}	*Loss ratio* D	Accuracy A
exit-last	436.0020	0.0715	0.8439
reactive	334.6113	0.0337	0.8245
rate-based	215.4975	0.0020	0.7710
model-based	83.1035	0.0001	0.6183
exit-first	75.0948	0.0001	0.5973

C. Running case for the rate-based policy

We consider again the example of Section IV-B with capacity $K = 6$ and assume that the i-th service period starts with $k_i = 3$ jobs in the EEN. Assume the arrival rate estimate to be $\widehat{\lambda}_i = 0.0070$ job/s. The initial time budget is then $B_i = 428.57\ ms$ and $p = (56.63, 65.04, 108.55, 138.03)$. The π_R policy returns in this example a probability vector $q_i = (1/3, 0, 1/3, 1/3)$, since the optimal solution of the k-KP problem is $x = (x_2, x_3, x_4) = (0, 1, 1)$. Running a simulation with 10^5 samples and a lookup table with range $\widehat{\lambda}_i \in [\lambda/2, 2\lambda]$ and $J = 3$ intervals of equal width, we find the results shown in Table VII, which indicate that the rate-based policy delivers a low loss ratio and fairly high accuracy.

APPENDIX D

Without any loss of generality, we assume that the processing layers of the EEN correspond to convolutional layers. We can thus analytically compute the MACs requested by convolutional layers as follows. Assume that there are m convolutional filters of size $r \times s$, then

$$p_j^{\Phi} = e \times f \times r \times s \times c \times m,$$

where $e \times f \times m$ represent the dimensions of the output activation vector $\xi_{i,j}$ for input x_i, being

$$e = (h - r + u)/u \quad (2)$$

$$f = (w - s + u)/u \quad (3)$$

and where $h \times w \times c$ represent the dimensions of the input activation vector $\xi_{i,j-1}$, and u being the stride of the filter, i.e., the parameter controlling how the filter convolves around the input activation vector.

We also assume that the IC Δ_j are implemented through fully connected layers. Again, this is a common assumption since the final classification layers of CNNs rely on this type of layer to provide the classification output. Thanks to this assumption we can analytically compute the MACs requested by an IC Δ_j as $p_j^{\Delta} = h \times C$, where h is the size of input activation vector $\xi_{i,j-1}$ and C is the number of classes in the classification problem. Such a formulation can be easily extended by considering multiple dense processing layers as $p_j^{\Delta} = h \times h_1 \times C$ where h_1 is the dimension of an intermediate feature map.

APPENDIX E

Further sensitivity analysis of the experimental results with synthetic arrival processes is given in figures 2-5. Detailed numerical values for the data shown in the figures are given in the subsequent tables, which also include the model-based policy using the exact formulation (global optimization program), with and without the lookup table, to adapt to changing arrival rates. The sensitivity analysis is over the following factors:

- Sensitivity to the inter-arrival time distribution
- Sensitivity to the number of EEN layers L
- Sensitivity to the EEN used to process the arrivals
- Sensitivity to the load factor ρ
- Sensitivity to the buffer size

The results are shown in terms of average values across all the simulations and 95% confidence interval width. As indicated by the figures, the overall trends are qualitatively similar to those described in the paper. Upon reading the results, the following should be noted:

- Experiments that differ in the number of simulations may not be directly comparable. For example, the sensitivity analysis for the number of layers $L = 64$ and $L = 32$ relies on fewer simulations than the cases $L \leq 24$ because only a subset of the EENs are sufficiently deep to allow the execution of these scenarios. For example, $L = 64$ is available only for the VGG-19 EEN.
- The median exit layer value of each simulation is averaged across the simulations, thus resulting in non-integer values in the table.
- The results of the model-based policy that uses the exact formulation are very close to the model-based heuristic policy. In some cases, the latter is slightly more accurate than the exact formulation as we impose a timeout of 20 minutes on global search, which therefore may not always guarantee finding a global optimum.

The main insights arising from these experiments are discussed in the next subsections.

A. Overall results

These results are also shown in Figure 4(a) in the main paper, we thus point the reader to the commentary of that figure for a discussion.

B. Sensitivity to the inter-arrival time distribution

The three tables are also shown in figures 3(d), 3(f), and 3(e), we point to the comments on those figures in the main paper.

C. Sensitivity to the number of EEN layers L

The results for $L \in \{4, 8, 16, 24, 32\}$ display nearly identical relative performance for the considered methods, suggesting that the layering has a limited impact on the performance of the methods. Some magnitude deviations of the metrics are observed in the case $L = 64$, but the relative performance of the methods is unchanged. The deviation can be explained because the dataset includes fewer networks with 64 layers than in the other cases.

D. Sensitivity to the EEN used to process the arrivals

This case also shows consistent relative results across the methods in nearly all cases, with some magnitude deviations in the VGG-19 networks compared to the other EENs due to difference in the processing times of this network. In some instances, the model-based heuristic and the global optimization based methods invert their relative performance, however such deviations are well within confidence intervals and therefore appear negligible.

E. Sensitivity to the load factor ρ

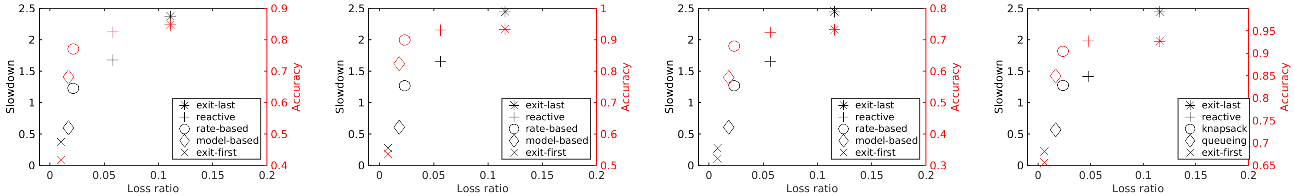
As in the main paper, the arrival rate is estimated for a given load factor ρ as $\lambda = \rho/p_{max}$. The results are fairly consistent across different utilization values, with the exception of the model-based policies, which appear to perform markedly better in higher load. This appears consistent with the fact that queueing approximations for systems with non-Poisson arrivals often converge in heavy load to those of Poisson-fed systems, making the model approximation increasingly accurate in heavier loads.

F. Sensitivity to the buffer size

The results indicate that the scheduling methods are more critical to use in scenarios where the buffers are very short, as very large job loss ratios can be incurred. Even when the loss ratio is low, while the accuracy is qualitatively similar across the different cases, there remain significant variations among the policies in terms of their slowdowns. Therefore, adopting the proposed heuristic can also remain beneficial if the loss ratio is not problematic.

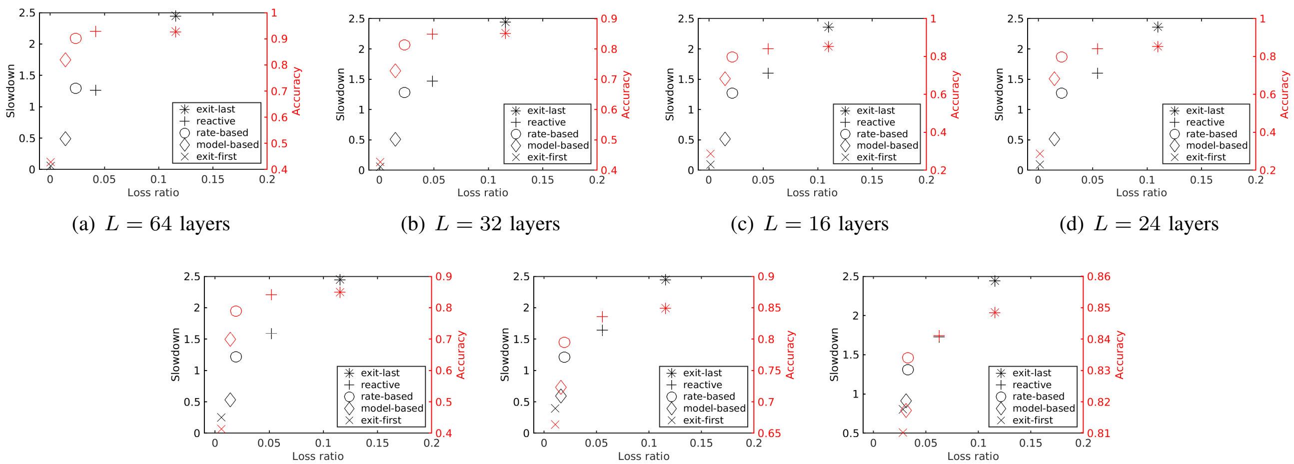
Appendix F

The processing times p_l and conditional accuracies $A^{(l)}$ for all the single-exit configurations of the EENs studied in Section VIII are given in tables XXIX-LII. Processing times are first calculated in MACs and then normalized so that $p_L = 1.0$.



(a) Custom VGG + CIFAR-10 (b) VGG16 + CIFAR-10 (c) VGG16 + CIFAR-100 (d) VGG19 + CIFAR-10

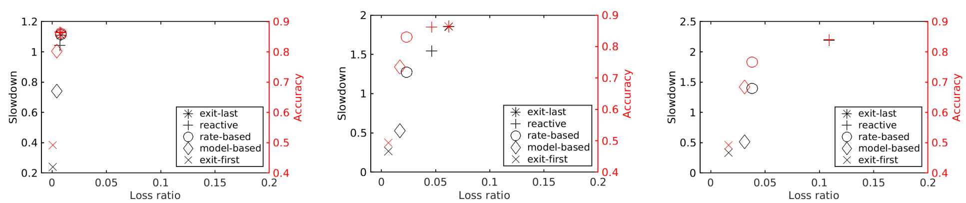
Fig. 2. Result sensitivity to EEN



(a) $L = 64$ layers (b) $L = 32$ layers (c) $L = 16$ layers (d) $L = 24$ layers

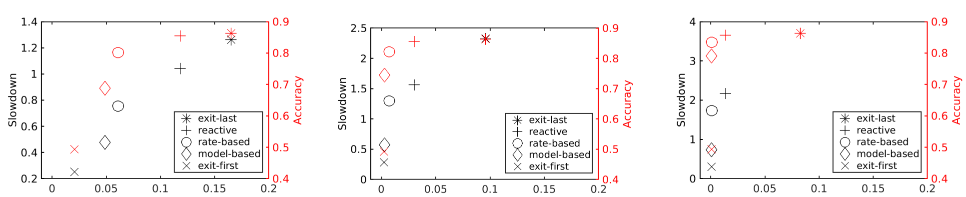
(e) $L = 8$ layers (f) $L = 4$ layers (g) $L = 2$ layers

Fig. 3. Result sensitivity to EEN number of early exit points.



(a) Load factor $\rho = 0.25$ (b) Load factor $\rho = 0.75$ (c) Load factor $\rho = 1.25$

Fig. 4. Result sensitivity to load factor $\rho = \lambda p$.



(a) Capacity $K = 2$ (b) Capacity $K = 5$ (c) Capacity $K = 9$

Fig. 5. Result sensitivity to system capacity, i.e., buffer size including the job in service.

TABLE VIII
OVERALL RESULTS: 648 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4319±0.1603	0.1145±0.0104	0.8636±0.0072	0.9941±0.0006	15.6667±1.1077
reactive	1.5937±0.0671	0.0540±0.0062	0.8565±0.0077	0.9422±0.0101	12.4815±0.8288
model-based	0.5946±0.0247	0.0174±0.0031	0.7408±0.0148	0.5623±0.0290	6.2670±0.5586
rate-based	1.2639±0.0435	0.0229±0.0034	0.8196±0.0094	0.8150±0.0212	13.1358±1.0278
exit-first	0.2820±0.0246	0.0077±0.0023	0.4930±0.0195	0.0492±0.0074	1.0000±0.0000

TABLE IX
EXPONENTIAL ARRIVALS, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.3395±0.2421	0.1117±0.0163	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.6342±0.1068	0.0633±0.0111	0.8574±0.0133	0.9460±0.0155	12.6250±1.4516
model-based	0.6594±0.0318	0.0209±0.0052	0.7837±0.0185	0.6486±0.0437	6.9861±0.9868
rate-based	1.2929±0.0664	0.0285±0.0059	0.8194±0.0162	0.8152±0.0358	13.2917±1.7998
exit-first	0.2896±0.0412	0.0089±0.0038	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE X
HIGH VARIANCE ARRIVALS (a^2=10), 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.6652±0.2595	0.1650±0.0215	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.6994±0.1116	0.0823±0.0127	0.8497±0.0140	0.9098±0.0222	11.8148±1.3983
model-based	0.7422±0.0440	0.0312±0.0071	0.7928±0.0183	0.6682±0.0450	7.4167±1.0306
rate-based	1.3733±0.0804	0.0403±0.0075	0.8066±0.0173	0.7750±0.0388	12.7454±1.7710
exit-first	0.3236±0.0547	0.0140±0.0055	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XI
LOW VARIANCE ARRIVALS (a^2=0.1), 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.2911±0.3259	0.0666±0.0127	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.4475±0.1282	0.0164±0.0048	0.8623±0.0129	0.9708±0.0128	13.0046±1.4741
model-based	0.3823±0.0361	0.0000±0.0000	0.6458±0.0328	0.3700±0.0522	4.3981±0.8401
rate-based	1.1253±0.0758	0.0001±0.0000	0.8329±0.0155	0.8548±0.0355	13.3704±1.7971
exit-first	0.2329±0.0269	0.0000±0.0000	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XII
64 LAYERS, 27 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4460±0.8414	0.1155±0.0544	0.9265±0.0000	0.9908±0.0000	64.0000±0.0000
reactive	1.2655±0.2524	0.0417±0.0285	0.9297±0.0012	0.9971±0.0024	40.0370±1.8243
model-based	0.4890±0.0970	0.0138±0.0123	0.8194±0.0619	0.7774±0.1235	18.9259±4.4118
rate-based	1.2962±0.2295	0.0235±0.0182	0.9020±0.0130	0.9419±0.0259	52.4444±6.8494
exit-first	0.0595±0.0004	0.0003±0.0004	0.4295±0.0000	0.0000±0.0000	1.0000±0.0000

TABLE XIII
32 LAYERS, 81 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4462±0.4644	0.1155±0.0300	0.8518±0.0242	0.9932±0.0021	32.0000±0.0000
reactive	1.4714±0.1815	0.0485±0.0166	0.8494±0.0254	0.9903±0.0039	26.8642±1.3683
model-based	0.5097±0.0596	0.0145±0.0071	0.7278±0.0446	0.7321±0.0633	14.9630±2.0125
rate-based	1.2866±0.1275	0.0227±0.0096	0.8134±0.0294	0.9134±0.0212	28.2963±1.5361
exit-first	0.0601±0.0002	0.0004±0.0002	0.4276±0.0463	0.0857±0.0084	1.0000±0.0000

TABLE XIV
24 LAYERS, 108 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4458±0.3999	0.1155±0.0258	0.9092±0.0068	0.9937±0.0010	24.0000±0.0000
reactive	1.5742±0.1648	0.0529±0.0150	0.9073±0.0084	0.9922±0.0036	20.2778±0.5900
model-based	0.5135±0.0512	0.0145±0.0061	0.7713±0.0376	0.7962±0.0495	9.7778±1.1194
rate-based	1.2789±0.1100	0.0220±0.0080	0.8648±0.0151	0.9316±0.0164	20.4074±1.1825
exit-first	0.0803±0.0067	0.0007±0.0005	0.3541±0.0323	0.2101±0.0255	1.0000±0.0000

TABLE XV
16 LAYERS, 108 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.3622±0.3876	0.1096±0.0251	0.8538±0.0187	0.9944±0.0007	16.0000±0.0000
reactive	1.6043±0.1688	0.0545±0.0151	0.8424±0.0202	0.9783±0.0082	12.9630±0.4277
model-based	0.5146±0.0508	0.0146±0.0062	0.6823±0.0436	0.7012±0.0600	6.1204±0.7805
rate-based	1.2724±0.1096	0.0214±0.0079	0.7981±0.0239	0.9050±0.0183	13.0463±0.7997
exit-first	0.0957±0.0117	0.0011±0.0009	0.2873±0.0328	0.0207±0.0041	1.0000±0.0000

TABLE XVI
8 LAYERS, 108 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4456±0.3998	0.1153±0.0258	0.8503±0.0188	0.9920±0.0012	8.0000±0.0000
reactive	1.5900±0.1649	0.0518±0.0147	0.8423±0.0197	0.9690±0.0110	6.4074±0.2151
model-based	0.5289±0.0465	0.0138±0.0058	0.6996±0.0383	0.6382±0.0576	3.1389±0.3304
rate-based	1.2215±0.1029	0.0194±0.0071	0.7900±0.0257	0.8353±0.0442	6.3611±0.4762
exit-first	0.2523±0.0274	0.0055±0.0036	0.4139±0.0405	0.0000±0.0000	1.0000±0.0000

TABLE XVII
4 LAYERS, 108 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4453±0.3997	0.1154±0.0258	0.8496±0.0187	0.9918±0.0027	4.0000±0.0000
reactive	1.6435±0.1675	0.0555±0.0156	0.8362±0.0204	0.9278±0.0208	3.3241±0.1376
model-based	0.5936±0.0407	0.0160±0.0069	0.7233±0.0319	0.3222±0.0638	1.5000±0.1740
rate-based	1.2118±0.1028	0.0193±0.0070	0.7952±0.0266	0.7043±0.0648	3.0926±0.2576
exit-first	0.3972±0.0159	0.0104±0.0051	0.6643±0.0338	0.0000±0.0000	1.0000±0.0000

TABLE XVIII
2 LAYERS, 108 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4465±0.3999	0.1154±0.0258	0.8484±0.0185	1.0000±0.0000	2.0000±0.0000
reactive	1.7302±0.1768	0.0625±0.0176	0.8412±0.0193	0.7939±0.0454	1.7593±0.0819
model-based	0.9125±0.0795	0.0310±0.0121	0.8173±0.0219	0.1725±0.0528	1.1111±0.0602
rate-based	1.3095±0.1112	0.0327±0.0116	0.8341±0.0203	0.5933±0.0785	1.5741±0.0948
exit-first	0.8066±0.0769	0.0279±0.0113	0.8103±0.0226	0.0000±0.0000	1.0000±0.0000

TABLE XIX
CUSTOM-CIFAR10, 135 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.3796±0.3480	0.1108±0.0225	0.8480±0.0011	0.9994±0.0002	10.8000±1.3939
reactive	1.6818±0.1539	0.0579±0.0142	0.8255±0.0056	0.9189±0.0220	9.3259±1.2036
model-based	0.5995±0.0418	0.0168±0.0065	0.6822±0.0284	0.5005±0.0607	4.7111±0.8498
rate-based	1.2272±0.0891	0.0214±0.0070	0.7711±0.0136	0.7598±0.0504	8.6074±1.3431
exit-first	0.3786±0.0337	0.0102±0.0051	0.4182±0.0455	0.0000±0.0000	1.0000±0.0000

TABLE XX
VGG16-CIFAR10, 162 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4459±0.3247	0.1154±0.0210	0.9334±0.0001	0.9978±0.0004	14.3333±1.6899
reactive	1.6598±0.1387	0.0561±0.0127	0.9312±0.0013	0.9502±0.0209	12.6296±1.6277
model-based	0.6084±0.0542	0.0180±0.0064	0.8236±0.0258	0.6001±0.0607	6.5247±1.1652
rate-based	1.2716±0.0896	0.0232±0.0070	0.8996±0.0084	0.8377±0.0424	12.3333±1.6513
exit-first	0.2743±0.0547	0.0076±0.0047	0.5364±0.0394	0.0853±0.0183	1.0000±0.0000

TABLE XXI
VGG16-CIFAR100, 162 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4458±0.3247	0.1154±0.0210	0.7333±0.0139	0.9957±0.0009	14.3333±1.6899
reactive	1.6587±0.1382	0.0563±0.0127	0.7243±0.0148	0.9402±0.0211	12.7160±1.6477
model-based	0.6091±0.0542	0.0181±0.0064	0.5797±0.0288	0.5436±0.0577	6.6296±1.1950
rate-based	1.2714±0.0896	0.0232±0.0070	0.6804±0.0183	0.8147±0.0425	12.3148±1.6510
exit-first	0.2742±0.0546	0.0076±0.0047	0.3212±0.0283	0.0680±0.0181	1.0000±0.0000

TABLE XXII
VGG19-CIFAR10, 189 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.4455±0.3001	0.1154±0.0194	0.9267±0.0003	0.9857±0.0014	21.4286±2.8886
reactive	1.4184±0.1121	0.0475±0.0109	0.9278±0.0009	0.9537±0.0177	14.4074±1.8144
model-based	0.5668±0.0451	0.0166±0.0055	0.8496±0.0153	0.5900±0.0541	6.8466±1.1238
rate-based	1.2769±0.0819	0.0236±0.0066	0.9051±0.0048	0.8353±0.0372	17.7619±2.6275
exit-first	0.2264±0.0451	0.0059±0.0038	0.6565±0.0234	0.0372±0.0097	1.0000±0.0000

TABLE XXIII
LOAD FACTOR $\rho = 0.250$, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	1.1120±0.0122	0.0078±0.0020	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.0445±0.0163	0.0070±0.0018	0.8658±0.0126	0.9983±0.0010	13.4167±1.4837
model-based	0.7404±0.0354	0.0041±0.0011	0.8025±0.0231	0.7612±0.0451	9.4398±1.2973
rate-based	1.1159±0.0119	0.0079±0.0020	0.8615±0.0126	0.9912±0.0019	15.6250±1.9289
exit-first	0.2400±0.0280	0.0008±0.0004	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XXIV
LOAD FACTOR $\rho = 0.750$, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	1.8589±0.1225	0.0621±0.0115	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.5448±0.0914	0.0463±0.0096	0.8626±0.0128	0.9746±0.0079	13.2824±1.4783
model-based	0.5276±0.0291	0.0169±0.0044	0.7355±0.0253	0.5204±0.0492	5.1898±0.7281
rate-based	1.2759±0.0760	0.0231±0.0056	0.8305±0.0156	0.8326±0.0354	13.1944±1.7713
exit-first	0.2687±0.0347	0.0060±0.0027	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XXV
LOAD FACTOR $\rho = 1.250$, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	4.3249±0.3411	0.2735±0.0113	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	2.1917±0.1426	0.1087±0.0126	0.8410±0.0146	0.8537±0.0255	10.7454±1.3342
model-based	0.5158±0.0535	0.0310±0.0077	0.6843±0.0264	0.4052±0.0449	4.1713±0.5854
rate-based	1.3998±0.1027	0.0378±0.0080	0.7669±0.0180	0.6212±0.0400	10.5880±1.5882
exit-first	0.3374±0.0584	0.0161±0.0061	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XXVI
CAPACITY $K = 2$, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	1.2636±0.0249	0.1650±0.0201	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.0437±0.0163	0.1183±0.0142	0.8557±0.0135	0.9406±0.0154	12.2269±1.4311
model-based	0.4775±0.0327	0.0487±0.0075	0.6877±0.0308	0.4487±0.0519	4.8611±0.8465
rate-based	0.7552±0.0354	0.0608±0.0080	0.8022±0.0187	0.7099±0.0497	9.4815±1.4821
exit-first	0.2530±0.0305	0.0204±0.0063	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XXVII
CAPACITY $K = 5$, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	2.3237±0.1704	0.0958±0.0162	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	1.5661±0.0702	0.0300±0.0052	0.8565±0.0134	0.9410±0.0184	12.5602±1.4452
model-based	0.5715±0.0349	0.0027±0.0015	0.7442±0.0243	0.5502±0.0478	5.7778±0.8325
rate-based	1.2997±0.0277	0.0071±0.0019	0.8219±0.0154	0.8461±0.0285	14.9630±1.8537
exit-first	0.2874±0.0419	0.0020±0.0014	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XXVIII
CAPACITY $K = 9$, 216 SIMULATIONS

Policy	*Slowdown* R/p_{max}	*Loss ratio* D	*Accuracy* A	*Norm. Accuracy* A_{norm}	*Median exit layer*
exit-last	3.7084±0.3869	0.0826±0.0154	0.8636±0.0125	0.9941±0.0011	15.6667±1.9288
reactive	2.1712±0.1556	0.0137±0.0029	0.8572±0.0134	0.9450±0.0188	12.6574±1.4527
model-based	0.7348±0.0513	0.0006±0.0006	0.7904±0.0189	0.6879±0.0466	8.1620±1.1504
rate-based	1.7367±0.0803	0.0010±0.0008	0.8348±0.0146	0.8890±0.0226	14.9630±1.9013
exit-first	0.3056±0.0527	0.0006±0.0006	0.4930±0.0339	0.0492±0.0128	1.0000±0.0000

TABLE XXIX
CUSTOM-CIFAR10: 24 LAYERS

$p_1 - p_8$	0.1349	0.1354	0.4088	0.4093	0.4098	0.4103	0.4107	0.4112
$p_9 - p_{16}$	0.3202	0.4708	0.4710	0.4712	0.7116	0.7118	0.7120	0.6665
$p_{17} - p_{24}$	0.7864	0.7865	0.7866	0.9959	0.9960	0.9961	0.9733	1.0000
$A^{(1)} - A^{(8)}$	0.1000	0.5827	0.5693	0.6007	0.6066	0.5848	0.6138	0.6740
$A^{(9)} - A^{(16)}$	0.7043	0.7003	0.7344	0.7534	0.7507	0.7788	0.8045	0.7906
$A^{(17)} - A^{(24)}$	0.8050	0.8204	0.8341	0.8434	0.8401	0.8507	0.7540	0.8483

TABLE XXX
CUSTOM-CIFAR10: 16 LAYERS

$p_1 - p_8$	0.1895	0.1901	0.5742	0.5749	0.5755	0.5762	0.5769	0.5775
$p_9 - p_{16}$	0.4497	0.6613	0.6616	0.6619	0.9994	0.9997	1.0000	0.9145
$A^{(1)} - A^{(8)}$	0.0999	0.5941	0.5832	0.5998	0.6237	0.6137	0.6033	0.6688
$A^{(9)} - A^{(16)}$	0.7026	0.7022	0.7328	0.7585	0.7541	0.7868	0.8077	0.8573

TABLE XXXI
CUSTOM-CIFAR10: 8 LAYERS

$p_1 - p_8$	0.4088	0.4103	0.3202	0.4712	0.7120	0.7865	0.9960	1.0000
$A^{(1)} - A^{(8)}$	0.5584	0.6073	0.6961	0.7521	0.8053	0.8252	0.8415	0.8525

TABLE XXXII
CUSTOM-CIFAR10: 4 LAYERS

$p_1 - p_4$	0.4103	0.4712	0.7865	1.0000
$A^{(1)} - A^{(4)}$	0.5973	0.7543	0.8121	0.8439

TABLE XXXIII
CUSTOM-CIFAR10: 2 LAYERS

$p_1 - p_2$	0.4712	1.0000
$A^{(1)} - A^{(2)}$	0.7355	0.8381

TABLE XXXIV
VGG16-CIFAR10: 32 LAYERS

$p_1 - p_8$	0.0593	0.0595	0.0597	0.0599	0.1800	0.1802	0.1804	0.1404
$p_9 - p_{16}$	0.2139	0.2140	0.2141	0.2142	0.3343	0.3344	0.3345	0.3145
$p_{17} - p_{24}$	0.3813	0.3813	0.3814	0.3814	0.5015	0.5016	0.5016	0.5017
$p_{25} - p_{32}$	0.6218	0.6218	0.6219	0.6119	0.6753	0.6753	0.6753	1.0000
$A^{(1)} - A^{(8)}$	0.5714	0.6004	0.5676	0.5138	0.6274	0.6623	0.6726	0.6982
$A^{(9)} - A^{(16)}$	0.7516	0.7735	0.7754	0.7706	0.7950	0.8487	0.8520	0.8609
$A^{(17)} - A^{(24)}$	0.8821	0.8873	0.8958	0.8919	0.8999	0.8986	0.8989	0.8958
$A^{(25)} - A^{(32)}$	0.9098	0.9223	0.9217	0.9193	0.9276	0.9298	0.9272	0.9323

TABLE XXXV
VGG16-CIFAR10: 24 LAYERS

$p_1 - p_8$	0.0595	0.0599	0.1802	0.1404	0.2140	0.2142	0.3344	0.3145
$p_9 - p_{16}$	0.3813	0.3814	0.5016	0.5017	0.6218	0.6119	0.6753	0.6754
$p_{17} - p_{24}$	0.7955	0.7955	0.9157	0.9107	0.9407	0.9408	0.9708	1.0000
$A^{(1)} - A^{(8)}$	0.3713	0.1002	0.2800	0.6528	0.7363	0.6886	0.8693	0.8670
$A^{(9)} - A^{(16)}$	0.9048	0.9000	0.9075	0.9018	0.9289	0.9245	0.9321	0.9346
$A^{(17)} - A^{(24)}$	0.9328	0.9345	0.9347	0.9345	0.9341	0.9336	0.9352	0.9325

TABLE XXXVI
VGG16-CIFAR10: 16 LAYERS

$p_1 - p_8$	0.0597	0.1802	0.2139	0.2142	0.3345	0.3813	0.5015	0.5017
$p_9 - p_{16}$	0.6219	0.6753	0.7955	0.7955	0.9157	0.9407	0.9708	1.0000
$A^{(1)} - A^{(8)}$	0.3028	0.2695	0.7547	0.7048	0.8678	0.9007	0.9125	0.9100
$A^{(9)} - A^{(16)}$	0.9285	0.9365	0.9352	0.9378	0.9371	0.9387	0.9376	0.9345

TABLE XXXVII
VGG16-CIFAR10: 8 LAYERS

$p_1 - p_8$	0.1802	0.2142	0.3813	0.5017	0.6753	0.7955	0.9407	1.0000
$A^{(1)} - A^{(8)}$	0.2507	0.6885	0.8937	0.9072	0.9334	0.9346	0.9362	0.9338

TABLE XXXVIII
VGG16-CIFAR10: 4 LAYERS

$p_1 - p_4$	0.3343	0.6218	0.9157	1.0000
$A^{(1)} - A^{(4)}$	0.7907	0.9293	0.9328	0.9334

TABLE XXXIX
VGG16-CIFAR10: 2 LAYERS

$p_1 - p_2$	0.6218	1.0000
$A^{(1)} - A^{(2)}$	0.9313	0.9340

TABLE XL
VGG16-CIFAR100: 32 LAYERS

$p_1 - p_8$	0.0593	0.0595	0.0597	0.0599	0.1800	0.1802	0.1804	0.1404
$p_9 - p_{16}$	0.2139	0.2140	0.2141	0.2142	0.3343	0.3344	0.3345	0.3145
$p_{17} - p_{24}$	0.3813	0.3813	0.3814	0.3814	0.5015	0.5016	0.5016	0.5017
$p_{25} - p_{32}$	0.6218	0.6218	0.6219	0.6119	0.6753	0.6753	0.6753	1.0000
$A^{(1)} - A^{(8)}$	0.1334	0.2037	0.1894	0.1036	0.2044	0.1968	0.2734	0.3158
$A^{(9)} - A^{(16)}$	0.3192	0.2941	0.4023	0.3484	0.3970	0.4813	0.5255	0.5196
$A^{(17)} - A^{(24)}$	0.5616	0.5677	0.5968	0.5797	0.5759	0.5815	0.6069	0.5802
$A^{(25)} - A^{(32)}$	0.5942	0.6593	0.6562	0.6585	0.6737	0.6744	0.6794	0.6978

TABLE XLI
VGG16-CIFAR100: 24 LAYERS

$p_1 - p_8$	0.0595	0.0599	0.1802	0.1404	0.2140	0.2142	0.3344	0.3145
$p_9 - p_{16}$	0.3813	0.3814	0.5016	0.5017	0.6218	0.6119	0.6753	0.6754
$p_{17} - p_{24}$	0.7955	0.7955	0.9157	0.9107	0.9407	0.9408	0.9708	1.0000
$A^{(1)} - A^{(8)}$	0.3713	0.1002	0.2800	0.6528	0.7363	0.6886	0.8693	0.8670
$A^{(9)} - A^{(16)}$	0.9048	0.9000	0.9075	0.9018	0.9289	0.9245	0.9321	0.9346
$A^{(17)} - A^{(24)}$	0.9328	0.9345	0.9347	0.9345	0.9341	0.9336	0.9352	0.9325

TABLE XLII
VGG16-CIFAR100: 16 LAYERS

$p_1 - p_8$	0.0597	0.1802	0.2139	0.2142	0.3345	0.3813	0.5015	0.5017
$p_9 - p_{16}$	0.6219	0.6753	0.7955	0.7955	0.9157	0.9407	0.9708	1.0000
$A^{(1)} - A^{(8)}$	0.1905	0.1731	0.3140	0.3426	0.5312	0.5664	0.5843	0.5752
$A^{(9)} - A^{(16)}$	0.6612	0.6749	0.6692	0.6813	0.6903	0.6971	0.6952	0.6933

TABLE XLIII
VGG16-CIFAR100: 8 LAYERS

$p_1 - p_8$	0.1802	0.2142	0.3813	0.5017	0.6753	0.7955	0.9407	1.0000
$A^{(1)} - A^{(8)}$	0.1677	0.3733	0.5692	0.5774	0.6805	0.6777	0.6972	0.6892

TABLE XLIV
VGG16-CIFAR100: 4 LAYERS

$p_1 - p_4$	0.3343	0.6218	0.9157	1.0000
$A^{(1)} - A^{(4)}$	0.4082	0.6565	0.6916	0.6919

TABLE XLV
VGG16-CIFAR100: 2 LAYERS

$p_1 - p_2$	0.6218	1.0000
$A^{(1)} - A^{(2)}$	0.6559	0.6948

TABLE XLVI
VGG19-CIFAR10: 64 LAYERS

$p_1 - p_8$	0.0585	0.0587	0.0589	0.0592	0.1778	0.1780	0.1782	0.1387
$p_9 - p_{16}$	0.2113	0.2114	0.2115	0.2116	0.3303	0.3304	0.3305	0.3107
$p_{17} - p_{24}$	0.3767	0.3767	0.3768	0.3768	0.4955	0.4955	0.4956	0.4956
$p_{25} - p_{32}$	0.6143	0.6143	0.6144	0.6144	0.6143	0.6143	0.6144	0.6045
$p_{33} - p_{40}$	0.6671	0.6672	0.6672	0.6672	0.7265	0.7266	0.7266	0.7266
$p_{41} - p_{48}$	0.7859	0.7859	0.7859	0.7859	0.9046	0.9046	0.9047	0.8997
$p_{49} - p_{56}$	0.9294	0.9294	0.9294	0.9294	0.9591	0.9591	0.9591	0.9591
$p_{57} - p_{64}$	0.9591	0.9591	0.9591	0.9591	0.9887	0.9888	0.9888	1.0000
$A^{(1)} - A^{(8)}$	0.4295	0.5681	0.5650	0.5273	0.6218	0.6364	0.6666	0.6838
$A^{(9)} - A^{(16)}$	0.7394	0.7521	0.7682	0.7683	0.7886	0.8525	0.8587	0.8614
$A^{(17)} - A^{(24)}$	0.8810	0.8863	0.8946	0.8924	0.8946	0.9095	0.9107	0.9007
$A^{(25)} - A^{(32)}$	0.9141	0.9212	0.9189	0.9104	0.9203	0.9265	0.9219	0.9226
$A^{(33)} - A^{(40)}$	0.9248	0.9294	0.9226	0.9209	0.9280	0.9297	0.9273	0.9228
$A^{(41)} - A^{(48)}$	0.9284	0.9311	0.9307	0.9207	0.9286	0.9273	0.9292	0.9301
$A^{(49)} - A^{(56)}$	0.9278	0.9289	0.9289	0.9245	0.9293	0.9302	0.9285	0.9213
$A^{(57)} - A^{(64)}$	0.9290	0.9286	0.9284	0.9226	0.9287	0.9307	0.9281	0.9265

TABLE XLVII
VGG19-CIFAR10: 32 LAYERS

$p_1 - p_8$	0.0587	0.0592	0.1780	0.1387	0.2114	0.2116	0.3304	0.3107
$p_9 - p_{16}$	0.3767	0.3768	0.4955	0.4956	0.6143	0.6144	0.6143	0.6045
$p_{17} - p_{24}$	0.6672	0.6672	0.7266	0.7266	0.7859	0.7859	0.9046	0.8997
$p_{25} - p_{32}$	0.9294	0.9294	0.9591	0.9591	0.9591	0.9591	0.9888	1.0000
$A^{(1)} - A^{(8)}$	0.5781	0.5517	0.6208	0.6736	0.7435	0.7193	0.8491	0.8545
$A^{(9)} - A^{(16)}$	0.8783	0.8750	0.9059	0.8990	0.9147	0.9038	0.9215	0.9189
$A^{(17)} - A^{(24)}$	0.9252	0.9276	0.9330	0.9150	0.9291	0.9216	0.9306	0.9298
$A^{(25)} - A^{(32)}$	0.9318	0.9220	0.9307	0.9037	0.9279	0.9246	0.9293	0.9252

TABLE XLVIII
VGG19-CIFAR10: 24 LAYERS

$p_1 - p_8$	0.0587	0.0592	0.1780	0.1387	0.2114	0.2116	0.3304	0.3107
$p_9 - p_{16}$	0.3767	0.3768	0.4955	0.4956	0.6143	0.6144	0.6143	0.6045
$p_{17} - p_{24}$	0.6672	0.6672	0.7266	0.7266	0.7859	0.7859	0.9046	1.0000
$A^{(1)} - A^{(8)}$	0.5737	0.4895	0.6470	0.6718	0.7542	0.7536	0.8389	0.8469
$A^{(9)} - A^{(16)}$	0.8898	0.8891	0.9026	0.9068	0.9113	0.9047	0.9204	0.9199
$A^{(17)} - A^{(24)}$	0.9240	0.9272	0.9287	0.8882	0.9304	0.9202	0.9281	0.9236

TABLE XLIX
VGG19-CIFAR10: 16 LAYERS

$p_1 - p_8$	0.0592	0.1387	0.2116	0.3107	0.3768	0.4956	0.6144	0.6045
$p_9 - p_{16}$	0.6672	0.7266	0.7859	0.8997	0.9294	0.9591	0.9591	1.0000
$A^{(1)} - A^{(8)}$	0.5561	0.6935	0.7604	0.8580	0.8893	0.9031	0.9140	0.9247
$A^{(9)} - A^{(16)}$	0.9289	0.9260	0.9285	0.9315	0.9336	0.9269	0.9315	0.9302

TABLE L
VGG19-CIFAR10: 8 LAYERS

$p_1 - p_8$	0.1387	0.3107	0.4956	0.6045	0.7266	0.8997	0.9591	1.0000
$A^{(1)} - A^{(8)}$	0.6787	0.8511	0.8969	0.9232	0.9105	0.9291	0.9268	0.9257

TABLE LI
VGG19-CIFAR10: 4 LAYERS

$p_1 - p_4$	0.3107	0.6045	0.8997	1.0000
$A^{(1)} - A^{(4)}$	0.8611	0.9239	0.9315	0.9292

TABLE LII
VGG19-CIFAR10: 2 LAYERS

$p_1 - p_2$	0.6045	1.0000
$A^{(1)} - A^{(2)}$	0.9184	0.9267