Hybrid Electro-Optical Intra-Data Center Networks Tailored for Different Traffic Classes

Speaker: Lin Wang

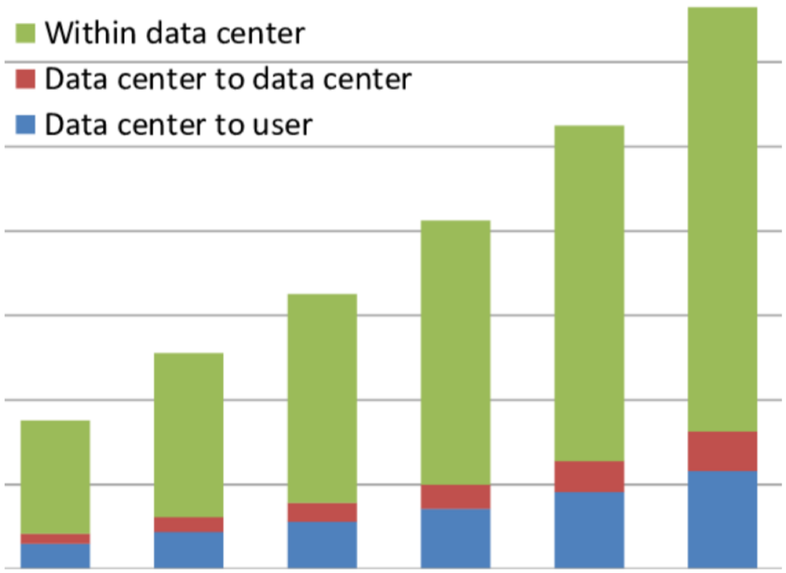
M. Balanici, et. al. “Hybrid Electro-Optical Intra-Data Center Networks Tailored for Different Traffic Classes,” JOCN, 2018.

UCDAVIS

ETRI
한국전자통신연구원
Electronics and Telecommunications Research Institute

Motivation

- **Traffic demand increasing in datacenter networks**
  - Cloud-service, parallel-computing, etc., lead to huge amount of intra datacenter traffic growth.
  - Cisco forecasts 31% increase per year of datacenter traffic by 2021



Datacenter traffic loads is growing

UCDAVIS
UNIVERSITY OF CALIFORNIA

Contributions of this work

- Apply long-range dependence (LRD) properties to model intra-data center network traffic.
- Based on the study of traffic in intra-data center, propose a new concept of traffic classes by defining five different applications typically hosted by cloud data centers.
- Evaluate performance of different services running on traditional and hybrid electro-optical Fat-Tree architectures.

Traffic in intra-data center

- **Long- range dependence (LRD):**

  Machine-generated data, display a high burstiness and an extreme variability over a wide range of time scales, characterized by intense spurts of activity and traffic bursts, which can last from milliseconds to hours and days.

- **Short-range dependence (SRD)**

  Voice traffic or web service

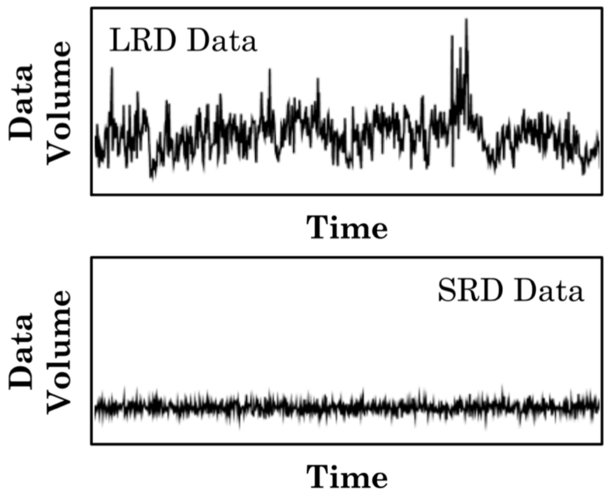


Fig. 1. Comparison of long-range dependence (LRD) versus short-range dependence (SRD) behaviors and their corresponding traffic burstiness. Notice the nature and the variation range of the generated data volumes (data rates) from the mean value of the two data types.

Traffic in intra-data center

- **Long- range dependence (LRD):**

  Machine-generated data, display a high burstiness and an extreme variability over a wide range of time scales, characterized by intense spurts of activity and traffic bursts, which can last from milliseconds to hours and days.

- **Traffic Model**

  Chaotic Map Model

$$x_{n+1} = \begin{cases} x_n + (1-d)\left(\frac{x_n}{d}\right)^{m_1}, & 0 < x_n \le d, \\ x_n - d\left(\frac{1-x_n}{1-d}\right)^{m_2}, & d < x_n < 1, \end{cases}$$

d: decision threshold in the data generation/non-generation procedure. (ON/OFF)

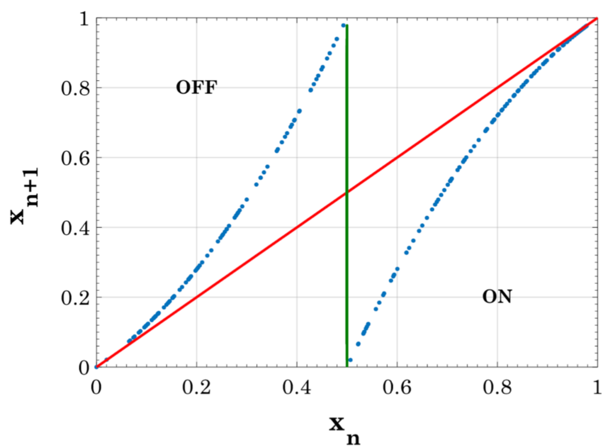


Fig. 2. LRD traffic generation model based on chaotic mapping introduced in Eq. (3). Notice the blue dots corresponding to non-generation instances (OFF) in the range [0, 0.5] and the blue dots representing data generation patterns (ON) within the range (0.5, 1). For this mapping, the total number of simulated states is $N = 200$, whereas the intermittency coefficients are $m_1 = m_2 = 2$. The threshold is set to $d = 0.5$ (green normal) for an equiprobable distribution of generation and non-generation instances.

Traffic Generation in intra-data center

- **Apply Chaotic Map Model to generate traffic**

- **Elephant flows**

  LRD characteristic brings the heavy-tailed distribution to the table and result in high traffic intensities (green circle).

- **Mice flows**

  Self-similarity property results in a high variability of the data rate around the mean value, which comes in the form of traffic bursts and fast changing spurts of activity (red circle)

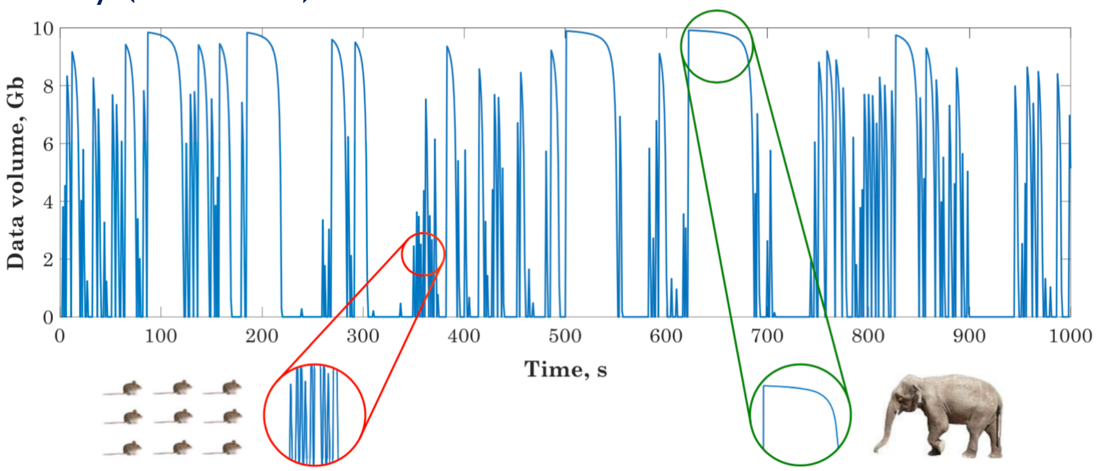


Fig. 3. Server traffic generation as an LRD process ($m_1 = m_2 = 2$, $d = 0.5$) mapped into Data volume $= f(t)$. The burstiness of data traffic is clearly visible, with two traffic patterns being distinguishable. The *mice* (red circle) appear as traffic bursts and spikes with a high variability, which are typically characterized by data streams intended for different destinations. The second pattern is an *elephant* (green circle) defined as a data transfer with a high payload and a much longer duration in time, normally addressed for one single destination (P2P).

Traffic Generation in intra-data center

- **Apply Chaotic Map Model to generate traffic**

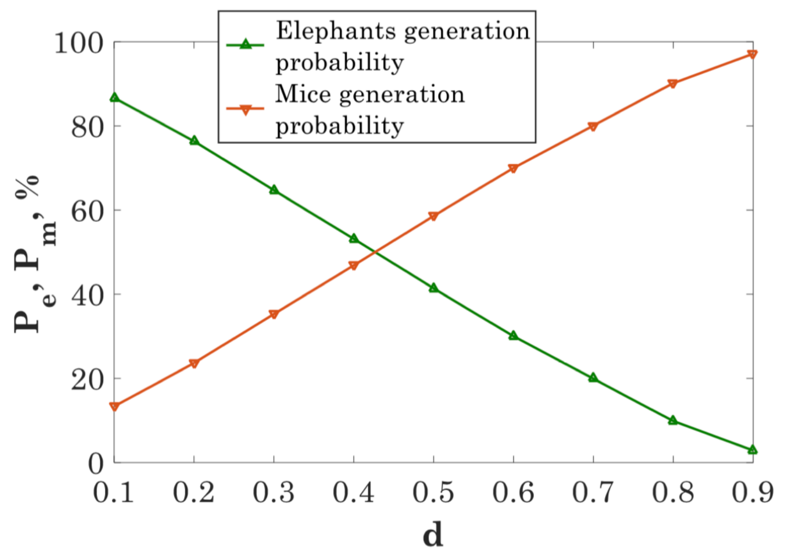


Fig. 4. Mice versus elephants generation probabilities (P_e and P_m) as a function of d. With the increase of the threshold parameter d, the traffic generation probability decreases, which leads to a lower probability for elephant occurrence and the prevalence of mice over elephants. The opposite is valid for lower values of d.

Traffic Classes in intra-data center

- **Five different traffic classes in intra-data center**

TABLE I
TRAFFIC CLASSES AND THEIR GENERATION PARAMETERS

Application Class	Traffic Nature	Inter-Rack Traffic [28,29]	d	R_{app}/R_{tot}
Web	mice $\gg$ eleph	68.0%	0.9	19%
Data storage	mice $\approx$ eleph	65.2%	0.5	18%
Data mining	eleph $\gg$ mice	84.2%	0.3	23%
Social networking	mice $>$ eleph	88.6%	0.7	24%
Video streaming	eleph $\gg$ mice	60.0%	0.2	16%

Intra-data Center architectures

- Traditional electrical Fat-tree intra-data center architecture
- Hybrid electro-optical Fat-Tree architecture

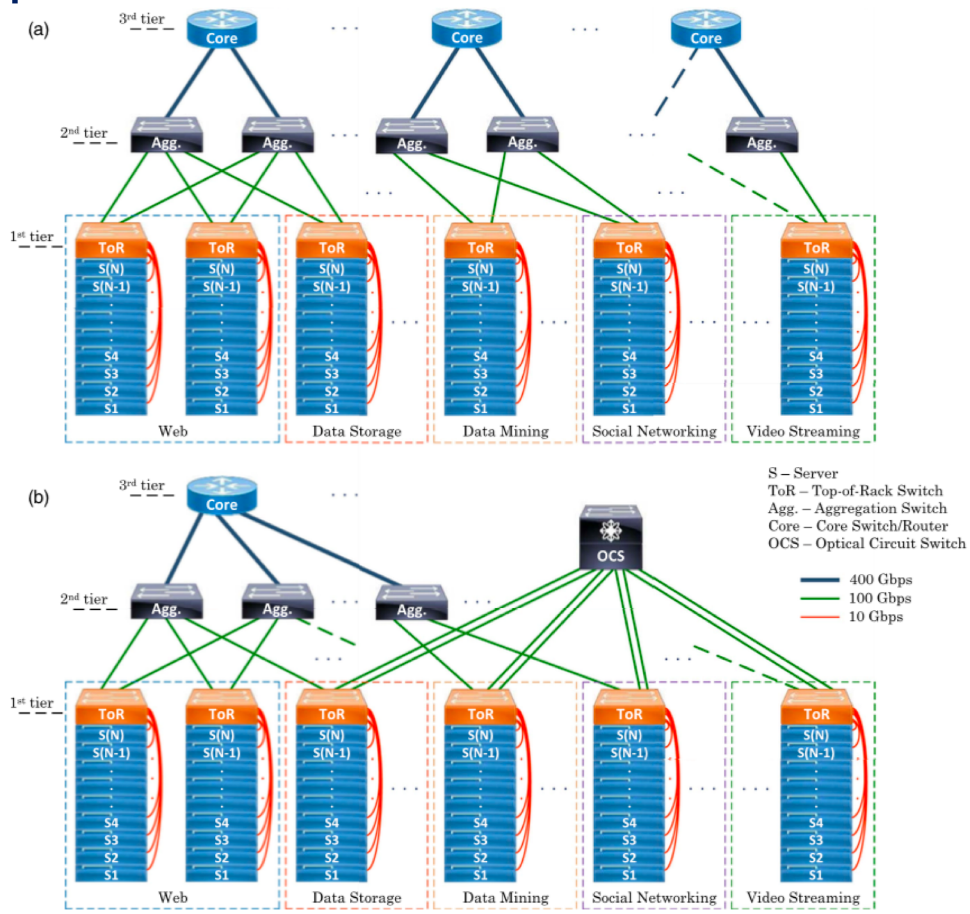


Fig. 5. Electrical Fat-Tree intra-data center network (a) before and (b) after optimization. Notice the partial substitution of the EPS grid with the OCS layer.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Intra-data Center architectures

- **Traditional electrical Fat-tree intra-data center architecture**
- **Hybrid electro-optical Fat-Tree architecture**

TABLE II
DCN ARCHITECTURE BEFORE (E) AND AFTER (H) OPTIMIZATION

Device	Interface/Throughput	$\sum$, E/H
Server	SFP+/10 Gb/s	32,000
ToR	SB: SFP+/32 × 10 Gb/s	1,000
	NB: QSFP28/2 × 100 Gb/s (E)	
	NB: QSFP28/3 × 100 Gb/s (H)	
Aggregation	SB: QSFP28/32 × 100 Gb/s	63/**38**
	NB: CFP8/5 × 400 Gb/s	
Core	SB: CFP8/15 × 400 Gb/s	21/**13**
OCS	800 ports	0/1

Simulations

- **Evaluate blocking probability performance of five different traffic classes in intra-data center.**

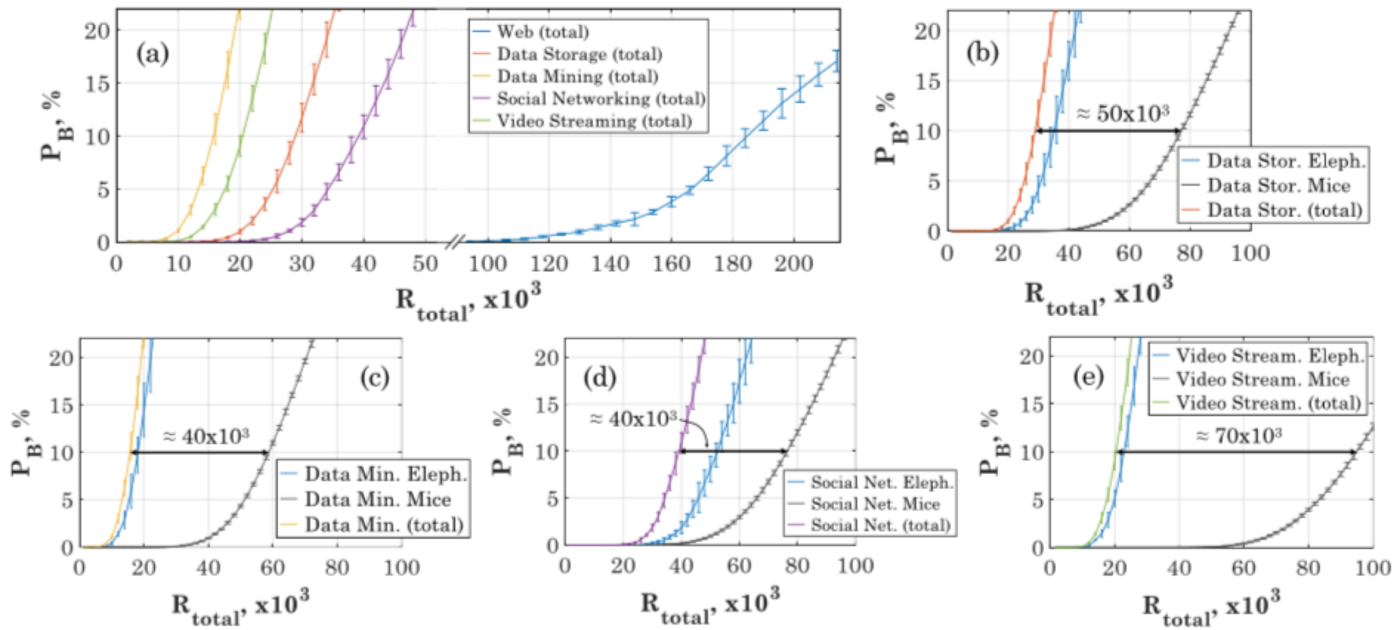


Fig. 6. Simulation of the five different services at the ToR layer. Due to offloading of elephants through the OCS layer, a noticeable relaxation in the EPS grid follows (black curves). The interface serving the elephants remains highly busy as in the pure electrical architecture due to their aggregated intensity.

UC DAVIS
UNIVERSITY OF CALIFORNIA

Simulations

- **Evaluate throughput performance of five different traffic classes in intra-data center.**

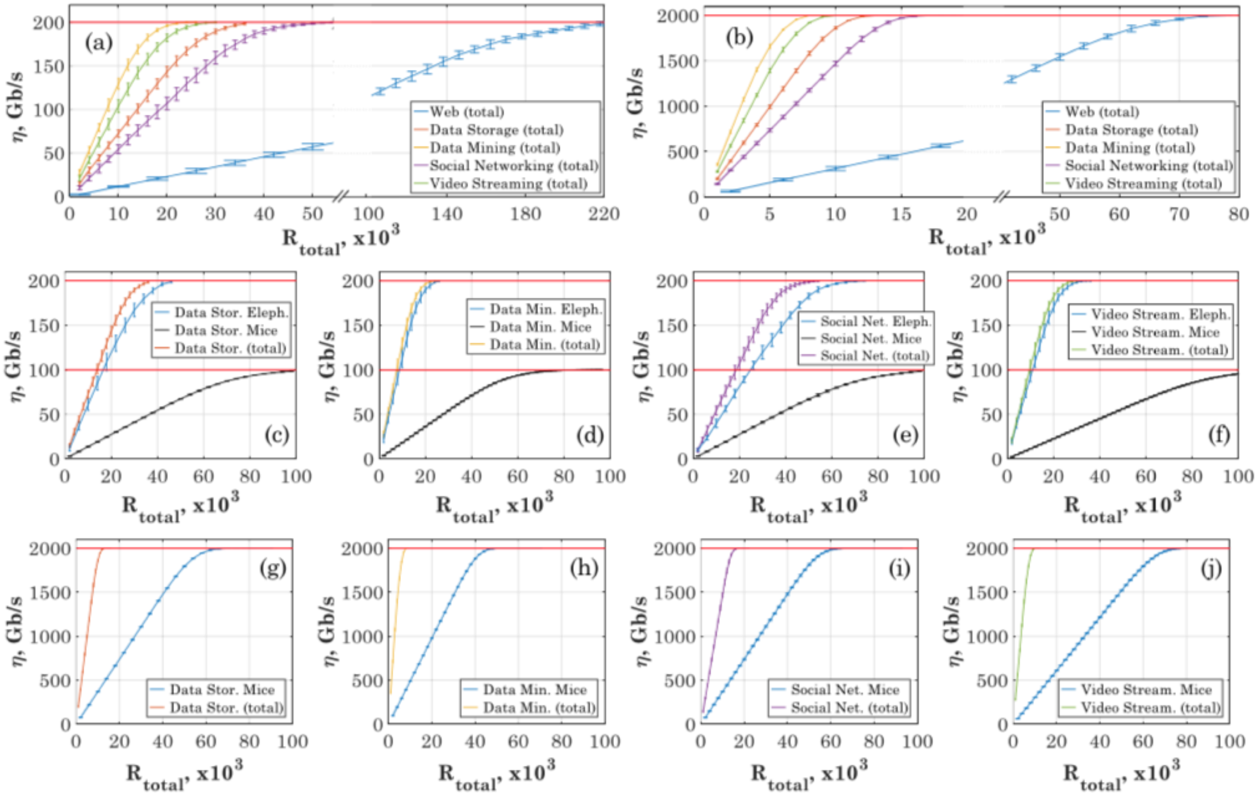


Fig. 8. (a), (c)–(f) Simulation of the five different services at the ToR and (b), (g)–(j) aggregation layers. The mean throughput utilization η in the traditional Fat Tree (a), (b) reaches saturation fastest for elephant-dominated applications. With migration to the hybrid Fat Tree (c)–(j) and offloading of elephants through the OCS layer, a significant alleviation of throughput utilization in both layers is noticed. The red horizontal lines indicate the maximal available bandwidth.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Simulations

- **Evaluate cost for traditional and hybrid electro-optical Fat-Tree architectures**

TABLE III
RELATIVE PRICES PER ONE UNIT OF EQUIPMENT

Component	Cost, p.u.	Energy, W
ToR switch (2/3 NB ports)	1/1.1	190/200
Aggr. switch	8	750
Core switch	8	750
OCS	40	100
SFP+	1	1
QSFP28	10	3.3
CFP8	50	12

TABLE IV
CAPEX AND OPEX BEFORE (LEFT) AND AFTER (RIGHT) OPTIMIZATION

	Electric		Hybrid	
Components	Total Cost, p.u.	Total Energy, kW	Total Cost, p.u.	Total Energy, kW
ToR	1000	190	1100	200
Aggregation	504	47.2	304	28.5
Core	168	15.7	104	9.7
OCS	—	—	40	0.1
$\sum$, **Switches**	**1672**	**252.9**	**1548**	**238.3**
SFP+	32000	32	32000	32
QSFP28	40160	13.2	40160	13.2
CFP8	31500	7.5	19250	4.6
$\sum$, **Pluggables**	**103660**	**52.7**	**91410**	**49.8**

Simulations

- **Evaluate cost for traditional and hybrid electro-optical Fat-Tree architectures**

TABLE IV
CAPEX AND OPEX BEFORE (LEFT) AND AFTER (RIGHT) OPTIMIZATION

	Electric		Hybrid	
Components	Total Cost, p.u.	Total Energy, kW	Total Cost, p.u.	Total Energy, kW
ToR	1000	190	1100	200
Aggregation	504	47.2	304	28.5
Core	168	15.7	104	9.7
OCS	—	—	40	0.1
$\sum$, **Switches**	**1672**	**252.9**	**1548**	**238.3**
SFP+	32000	32	32000	32
QSFP28	40160	13.2	40160	13.2
CFP8	31500	7.5	19250	4.6
$\sum$, **Pluggables**	**103660**	**52.7**	**91410**	**49.8**

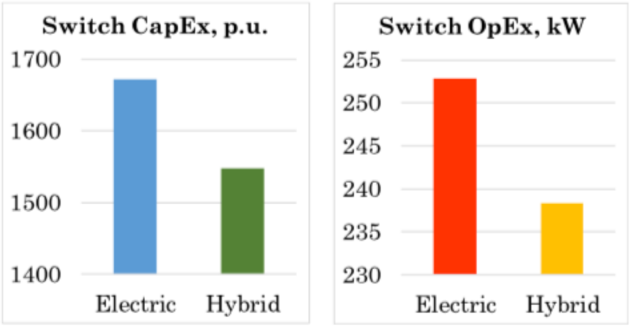


Fig. 9. CAPEX and OPEX of switching hardware before and after optimization.

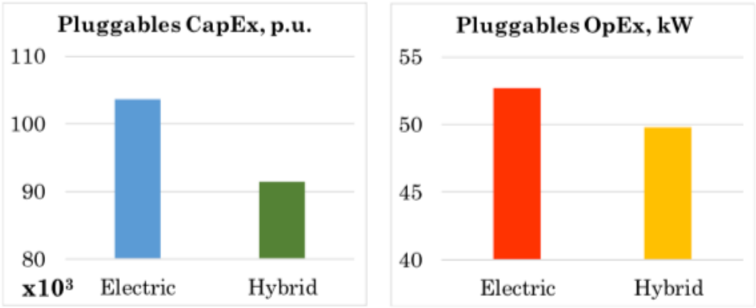


Fig. 10. CAPEX and OPEX of pluggable components before and after optimization.

Thank You !

- **amlwang@ucdavis.edu**