Appendix

1 ENERGY OPTIMIZATION

Training Completion Time Model. For a specific training task (e.g., a particular DNN model), the time required to finish an training epoch on different devices is totally different. The training completion time is highly correlated with the following factors. First, since different devices can have different number of data objects to be trained, the training completion time can be determined by the amount of data objects d_i. Second, different clients can have totally different training platforms (e.g., CPU, GPU), thus the training completion time can also depend on different training platforms. Third, for a certain training platform, the hardware configuration can be different (e.g., number of cores, core frequency). Thus, we model the training completion time for one epoch as follows:

$$t_i = \frac{\theta * d_i}{c^\alpha f^\beta} \tag{1}$$

θ represents the coefficient which is different for different training platforms and devices. d_i represents the number of data objects required to be trained in this training epoch. c represents the number of cpu cores adopted during the training process. f represents the core frequency. α and β are the corresponding coefficients that can be obtained through the system identification process. We can find that the more data objects need to be trained, the longer time it requires to complete the training. The more number of cpu cores and high core frequency can help accelerate the training process.
DVFS Model. During the training process, the power consumption (e.g., different devices, different core frequency) can be different. Thus, we model the training power from CPU as follows:

$$P_i = P_{BL,N_c} + \sum_i^{N_c} P_{core,U_i,f_i} \tag{2}$$

where N_c is the number of cores enabled, P_{BL,N_c} is the baseline CPU power with N_c enabled cores, and P_{core,U_i,f_i} is power increment of core i when it is working at frequency f_i with utilization U_i. Moreover, the single core power P_{core,U_i,f_i} can be modeled as follows:

$$P_{core,U_i,f_i} = \gamma * f_i * U_i + \gamma_{base} \tag{3}$$

in which, γ is the corresponding power coefficient. At last, the energy required for a specific training epoch on client i can be modeled as follows:

$$E_i = P_i * t_i \tag{4}$$

2 POWER MODEL

Power modeling of smartphone components, especially those major energy-consuming ones, has been extensively studied in previous research. We adopt the commonly used power models for CPU, screen and WIFI, and integrate them with the power model of TEC. The four power models are listed in Table 1. The CPU power consumption is linearly related to its utilization given a specific frequency level [2]. The screen power consumption depends on the brightness [1].

Table 1: Power models.

Component	Model
CPU $\mu:\ utilization, 0 \leqslant \mu \leqslant 100$ $freq:\ frequency\,index, freq = 0, 1, 2 \cdots, n$	$\text{c}P^{CPU} = \gamma_{freq}^{CPU} \times \mu + C_{CPU}$
Screen $\alpha_\text{b}, \alpha_\text{w}, C_{Screen}:\ power\,coefficients$ $B_\text{level}:\ brightness\,level,\ 0 \leqslant B_\text{level} \leqslant 255$	$\text{c}\ P^{Screen} = \left(\frac{\alpha_\text{b}+\alpha_\text{w}}{2} \times B_\text{level}\right) + C_{Screen}$
WIFI $p:\ packet\,rate,\ t:\ threshold$	$\text{c}\ P^{WiFi} = \begin{cases} \gamma_l^{WIFI} \times p + C_l & if\,p \leqslant t \\ \gamma_h^{WIFI} \times p + C_h & if\,p > t \end{cases}$
eMMC where I can be calculated from Equation.	$\text{c}\ P^{TEC} = \alpha I \Delta T + I^2 R$

REFERENCES

[1] Farhan Azmat Ali, Pieter Simoens, Tim Verbelen, Piet Demeester, and Bart Dhoedt. 2016. Mobile device power models for energy efficient dynamic offloading at runtime. *Journal of Systems and Software* 113 (2016), 173–187.

[2] FL Tan and SC Fok. 2008. Methodology on sizing and selecting thermoelectric cooler from different TEC manufacturers in cooling system design. *Energy conversion and management* 49, 6 (2008), 1715–1723.