模型的内容性解释（鲍成）

- **功耗模型**：$P^{(c)}(f_{cpu}, f_{gpu}, f_{ddr}) = \kappa_{cpu}^{(c)} f_{cpu}^3 + \kappa_{gpu}^{(c)} f_{gpu}^3 + \kappa_{ddr}^{(c)} f_{ddr}^3 + \sigma^{(c)}.$
- **帧时/迭代时间模型**：$\hat{T}^{(c)}(f_{cpu}, f_{gpu}, f_{ddr}) = \max\left(\hat{T}_{cpugpu}^{(c)}(f_{cpu}, f_{gpu}),\ \hat{T}_{ddr}^{(c)}(f_{ddr})\right).$ FPS ≈ 1/T

帧率模型解释：

CPU和GPU为关键路径，CPU 负责逻辑/调度/命令提交，GPU 负责渲染执行，为串行关系

$$\hat{T}_{cpugpu}^{(c)}(f_{cpu}, f_{gpu}) = D_{\min}^{(c)}\left(\alpha_{cpu}^{(c)} \frac{f_{cpu,\max}}{f_{cpu}} + \alpha_{gpu}^{(c)} \frac{f_{gpu,\max}}{f_{gpu}}\right), \quad \alpha_{cpu}^{(c)} + \alpha_{gpu}^{(c)} = 1.$$

DDR为吞吐约束，与CPU、GPU为并行关系，T_{ddr}呈现roofline形态

$$\hat{T}_{ddr}^{(c)}(f_{ddr}) \approx \frac{M^{(c)}}{\min\left(C^{(c)} f_{ddr},\ BW_{\max}^{(c)}\right)} = \max\left(\frac{M^{(c)}}{C^{(c)} f_{ddr}},\ \frac{M^{(c)}}{BW_{\max}^{(c)}}\right).$$

DDR 模型的闭式近似：

方案A：有理函数（渐近线视角）$T(f) = \dfrac{af^2 + bf + c}{f} = af + b + \dfrac{c}{f}$

方案B：指数饱和（显式饱和下界）$T_{ddr}(f_{ddr}) = \tau_0^{(c)} + \tau_1^{(c)} \exp\left(-\lambda^{(c)} f_{ddr,\mathrm{GHz}}\right)$

模型的负载选取与拟合效果（任德上 & 蔡东辰）

- **功耗模型**

  负载选取：2个精致渲染场景（抖音播放视频/进出图库+唤醒小艺）和2个AI推理场景（相机录像和备忘录摘要）

  拟合效果：基于华为功耗版测得的数据，功耗模型的平均相对误差均小于8%

- **帧率模型**

  负载选取：1个精致渲染场景（抖音播放视频+唤醒小艺）和1个AI推理场景（相机录像）

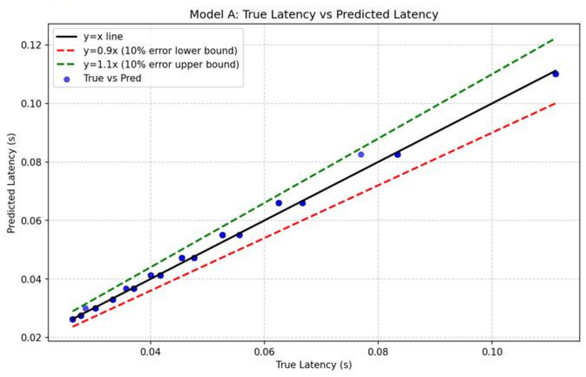
  拟合效果：抖音视频播放：相对平均误差4.82%（指数模型拟合），4.34%（对勾函数拟合）

  相机录像：相对平均误差2.82%（指数模型拟合），2.83%（对勾函数拟合）

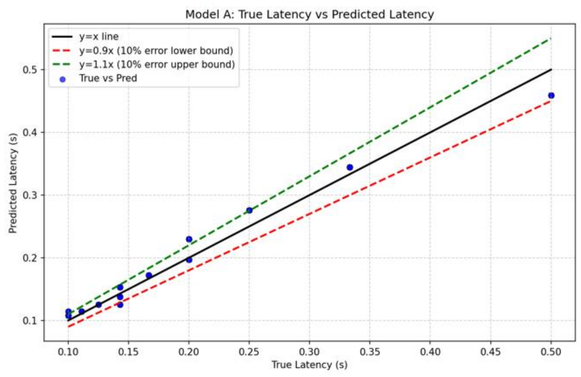
基于jetson的渲染场景补充验证（蔡东辰）

针对多组高精度渲染负载场景，拟合后的一致性图如下，即便是极高渲染负载的平均相对误差也较低。

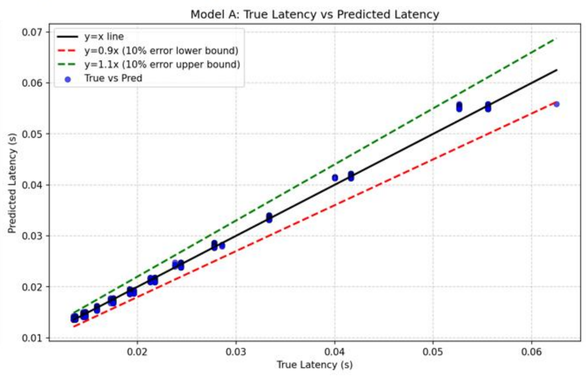
极高渲染负载
[3帧, 15帧]，平均相对误差6.75%



较高渲染负载
[9帧, 38帧]，平均相对误差1.44%



普通渲染负载
[16帧, 74帧]，平均相对误差1.61%



制作人：程诗淇　　审阅人：徐飞　吴旭东