RotateAI Simulator Derivations

Maxence Morel Dierckx

February 26, 2026

1 Disclaimer

The simulator runs x86 binaries and measures x86 instruction counts. All derived metrics use these counts as proxies for ARM Cortex-M33 execution. Important disclaimers:

1. x86 instructions are not the same as ARM instructions. Different Instruction Set Architectures (ISAs) produce different instruction counts for identical computation.

2. x86 vectorises float32s to perform 4-8 floating point operations in a single instruction. The M33 architecture does not implement this "Single Instruction, Multiple Data" (SIMD) technology, so simulator results will heavily undercount the FLOPS of an actual tag.

3. Power characteristics (e.g. µA / MHz) are taken from STM's AN5652 datasheet and assume a specific voltage range. Actual consumption varies with peripheral activity, temperature, and process variation. These assumptions are configurable by design, with defaults in `config/stm32u5.toml`.

The derived values should be interpreted as **fermi estimates** of eventual characteristics. They have utility as estimates of feasibility and for comparison of different model pipelines. They should not be used as engineering specifications.

2 Variables

Symbol	Description	Source
I	Instructions per inference	`perf stat`
n	Number of samples in batch	Data
f_s	Sample rate (Hz)	Config
$f_{\max}$	Maximum operating frequency (MHz)	Config
k	Safety margin (≥ 1)	Config
D	DMIPS per MHz	Config
μ	Current per MHz (µA/MHz)	Config
V	Supply voltage (V)	Config

3   Minimum Operating Frequency

The Minimum Operating Frequency is the minimum CPU frequency for which the required instructions per inference can be completed in the given sample period (e.g., 200 ms at 5 Hz).

- The sample period is equal to the reciprocal of the sample rate f_s.

- The Dhrystone Million Instructions Per Second (DMIPS) per MHz, D, determines the amount of instructions our processor can perform in a given time. We convert to Hz first.

$$f_{\min} = \frac{I \cdot f_s}{D \cdot 10^6} \tag{1}$$

4   Energy Per Inference

Energy is the product of voltage, current, and time.

- At a given frequency f, the MCU draws $\mu \cdot f$ microamps of current.

- One inference takes $I/(f \cdot D \cdot 10^6)$ seconds to complete.

$$E = V \cdot (\mu \cdot f) \cdot \frac{I}{f \cdot D \cdot 10^6}$$

The frequency f cancels. Running faster draws more current for less time, and the product is constant within a voltage range. This simplifies to:

$$E = \frac{V \cdot \mu \cdot I}{D \cdot 10^{12}} \tag{2}$$

5   Duty Cycle

The duty cycle is the fraction of time the processor spends running inference versus sleeping. At the maximum operating frequency $f_{\max}$, this is the ratio of the minimum required frequency to the available frequency:

$$\delta = \frac{f_{\min}}{f_{\max}} \tag{3}$$

A duty cycle above 1 indicates the inference cannot complete within the sample period at f_{max}.

6 Power Consumption

Average power is the energy consumed per unit time. With f_s inferences per second:

$$P = E \cdot f_s \qquad (4)$$