Best Validation Loss vs Compute (Non-embedding)
ademamix | eryngii_sweeps

Legend:
- Depth 2
- Depth 3
- Depth 4
- Depth 5
- Depth 6
- Depth 7
- Depth 8
- No offset: $1.62e+01 \times C^{-0.044}$ ($R^2=1.000$)
- With offset: $8.60e-06 + 1.62e+01 \times C^{-0.044}$ ($R^2=1.000$)

X-axis: Compute (FLOPs)
Y-axis: Best Validation Loss