

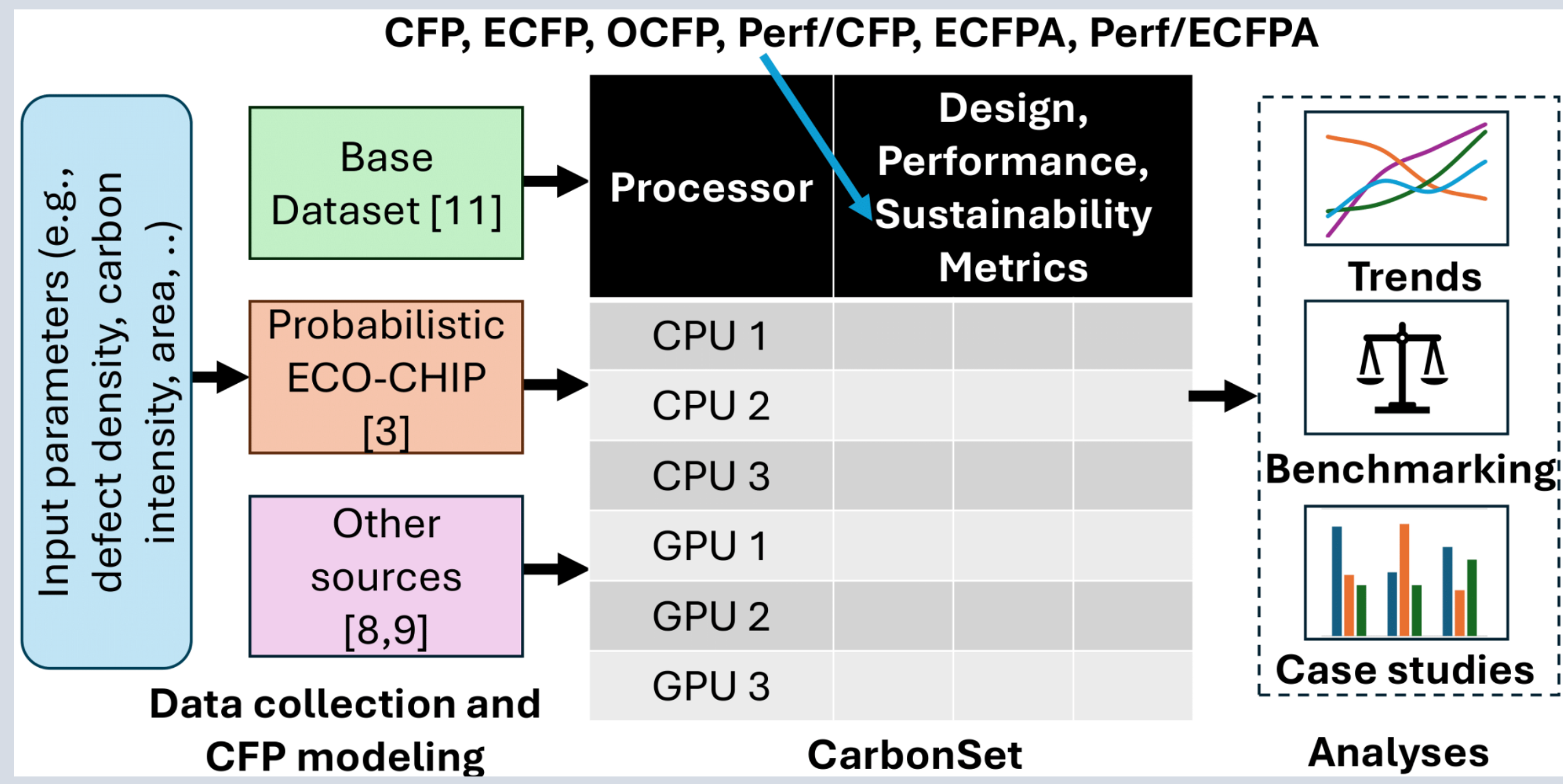
CarbonSet: A Dataset to Analyze Trends and Benchmark the Sustainability of CPUs and GPUs



Jiajun Hu, Chetan Choppali Sudarshan, Maxwell Clifford, Vidya A. Chhabria, Aman Arora

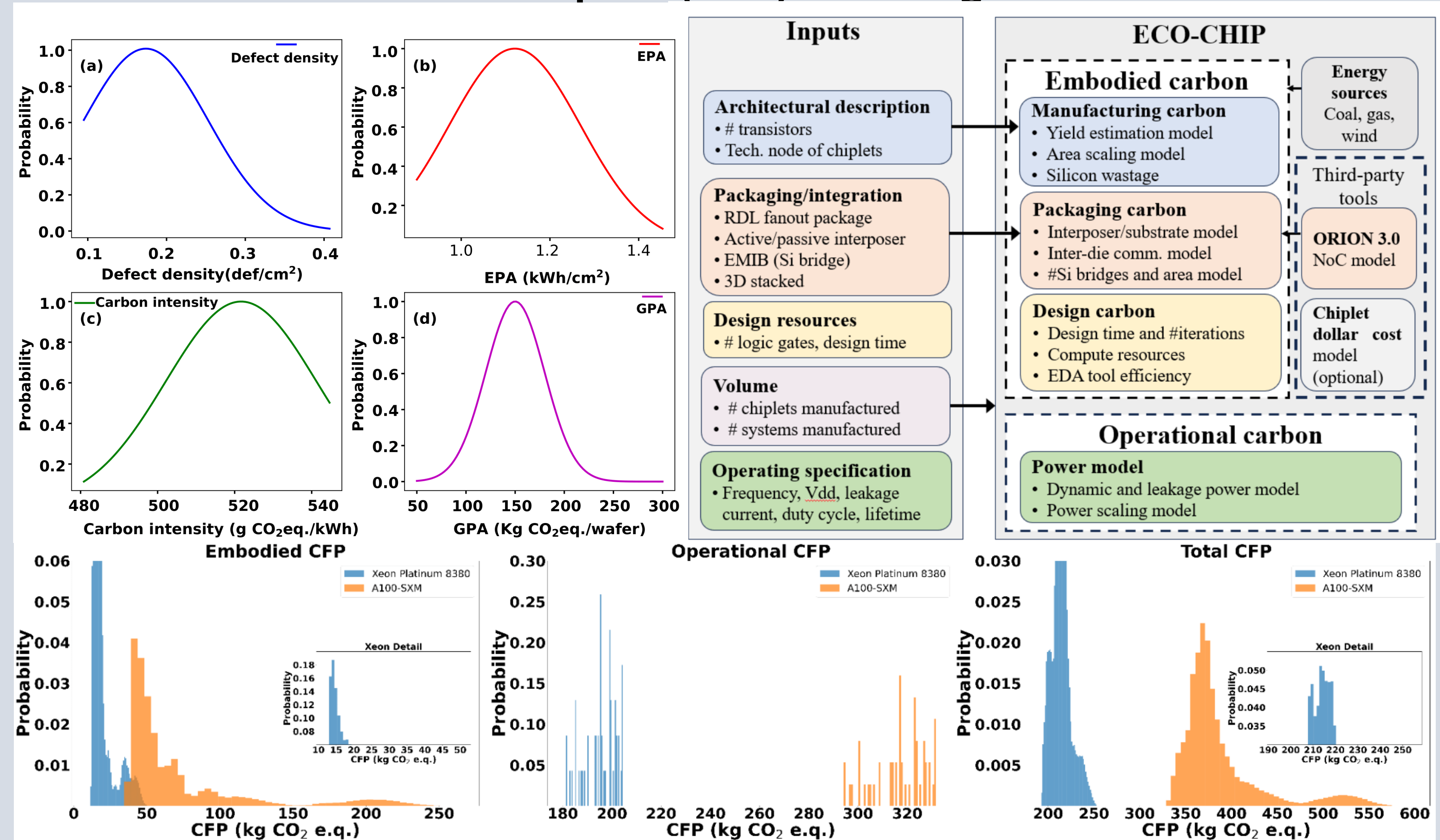
School of Augmented Intelligence and Computing, Arizona State University, Tempe, AZ, USA

Introduction

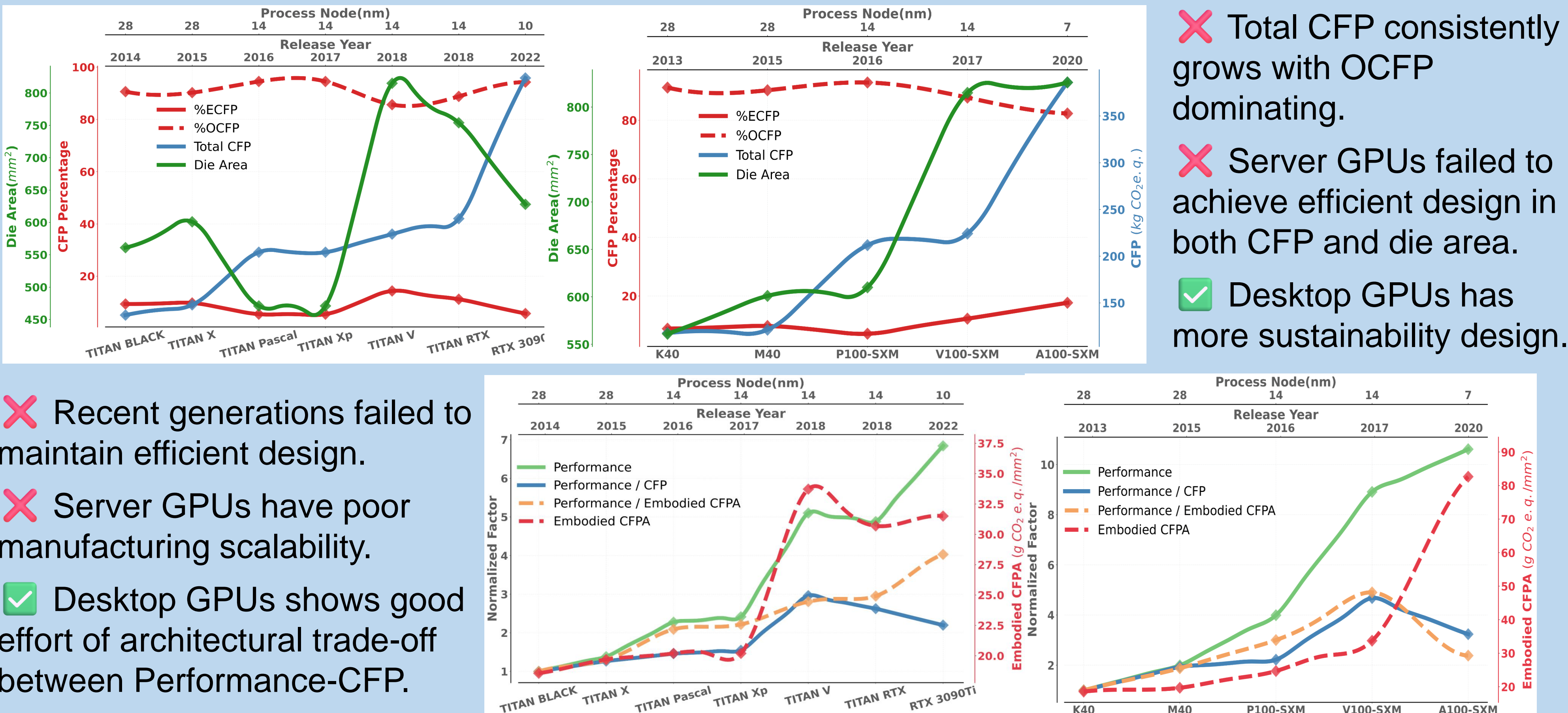


- **Carbon Footprint(CFP)** Total carbon emission, total CFP = OCFP + ECFP.
 - **Operational CFP(OCFP)** Operational carbon emission from runtime energy.
 - **Embodied CFP(ECFP)** Manufacturing and design carbon emission.
 - **Performance per CFP** Performance-Sustainability tradeoff
 - **ECFP per Area (ECFPA)** Measure embodied carbon emission per unit area
 - **Performance per ECFPA** Performance gain relative to ECFP and Area.
- We derived the **probabilistic distributions** of each manufacturing parameter across different process nodes and applied **Monte-Carlo Simulation** for each processor to generate a range of CFP values and use the **mean of them** as the final reference CFP value.

Probabilistic Carbon Footprint (CFP) Modeling



Monolithic GPU Sustainability Trend

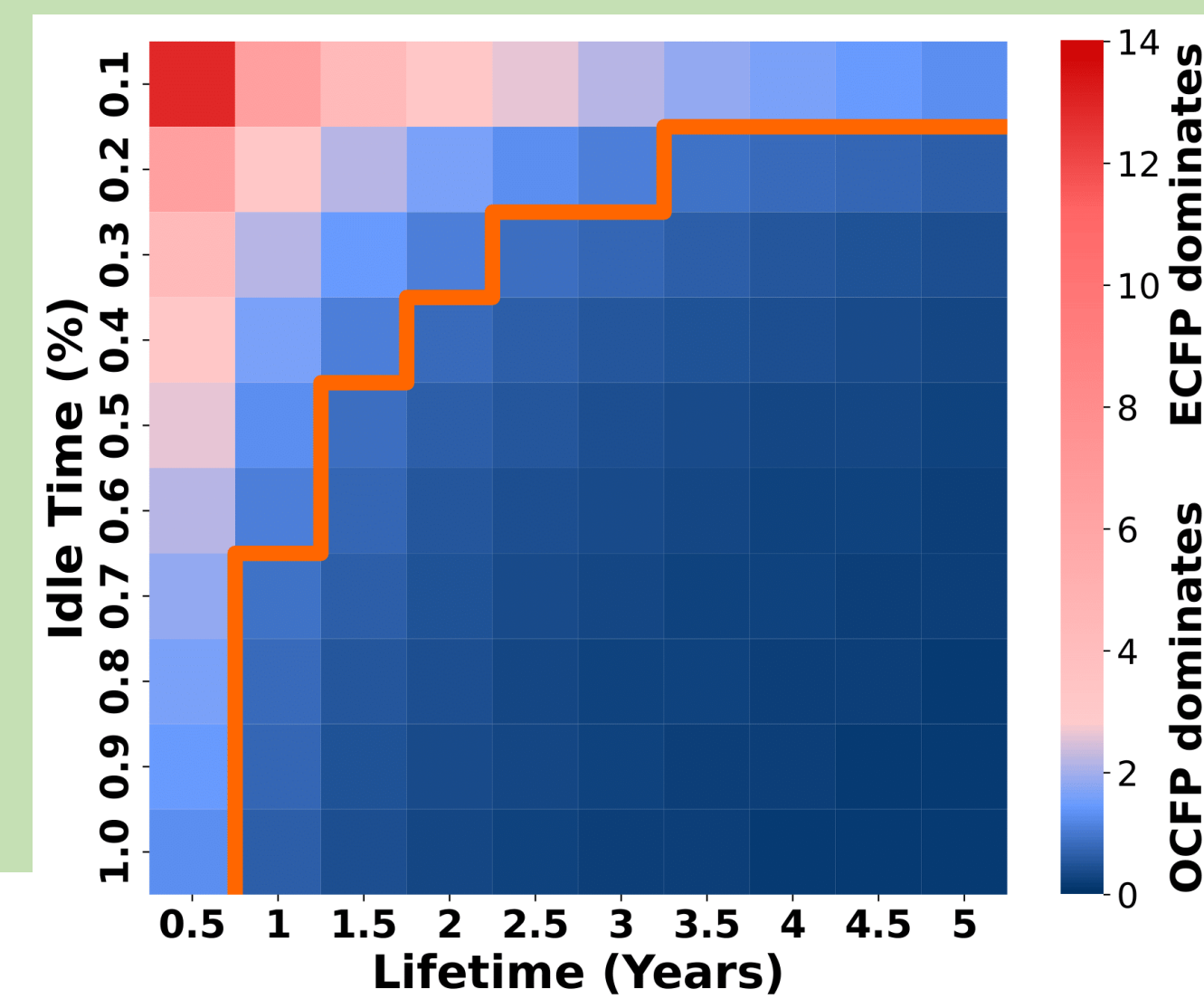


- ✗ Recent generations failed to maintain efficient design.
- ✗ Server GPUs have poor manufacturing scalability.
- ✓ Desktop GPUs shows good effort of architectural trade-off between Performance-CFP.

Insights from Case Studies

How much must chip lifetime increase to effectively amortize ECFP?

Users must either extend the device's lifetime or reduce idle time to shift below the orange line.

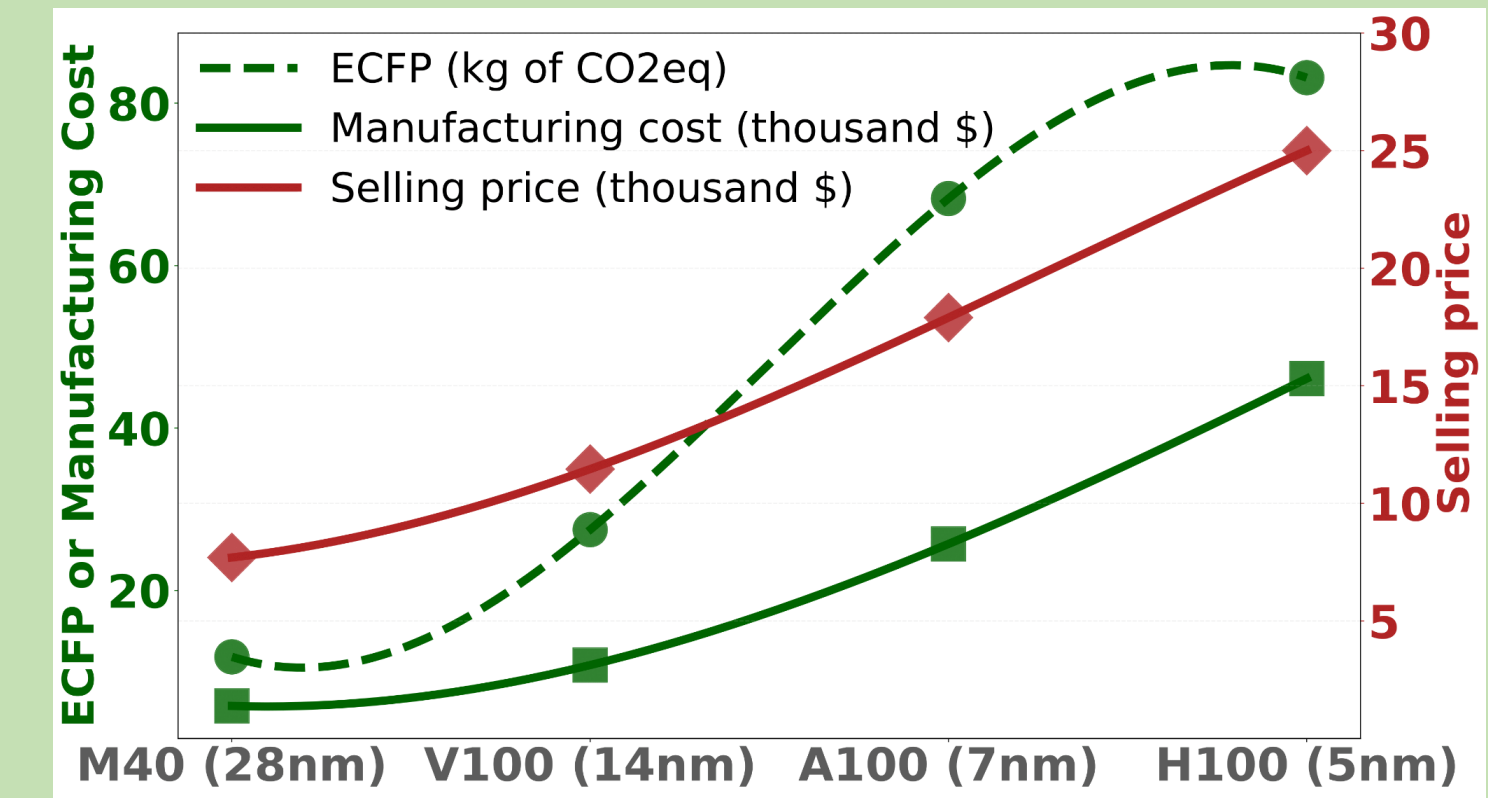


How has the AI boom impacted CFP?

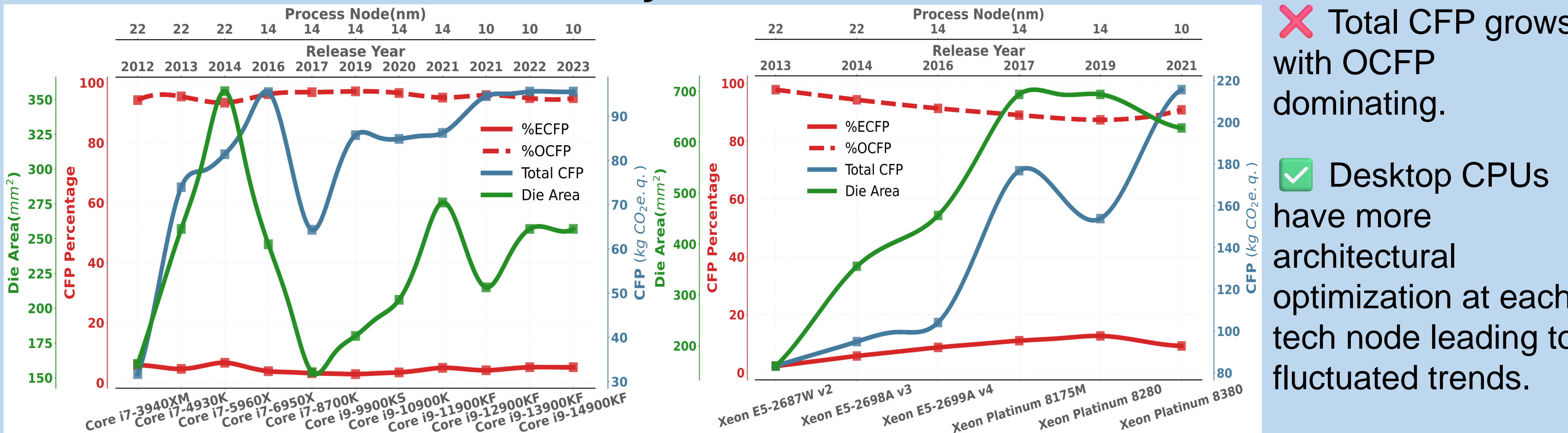
AI chip demand increased chip shipments. Combined with increased single chip CFP, this resulted in massive total CFP increase.

Is manufacturing cost a proxy for ECFP?

With the advancement of process node, the trend of manufacturing cost (\$) no longer aligns with ECFP trend.



Monolithic CPU Sustainability Trend



- ✗ Server CPUs have less efficient design with increased ECFP even in same tech node.
- ✓ Desktop CPUs maintained high manufacturing scalability as well as performance efficiency.

