

## Contents

|                                                           |          |
|-----------------------------------------------------------|----------|
| <b>BRAD DEVICES — Complete Unified Specification</b>      | <b>2</b> |
| How to Read This Spec                                     | 3        |
| Brand Pillar: Built in Africa, for the World              | 3        |
| The Five Supporting Pillars                               | 3        |
| Honesty Guardrails                                        | 4        |
| Existing Africa References (consolidated)                 | 4        |
| PART 0: ECOSYSTEM BLUEPRINT                               | 4        |
| Directory Structure                                       | 4        |
| Architecture Stack                                        | 6        |
| The Defining Architectural Insight: The Data Pipeline     | 6        |
| Generation Naming                                         | 6        |
| PART 1: CPU ARCHITECTURE                                  | 6        |
| 1.1 CPU Philosophy: Two Families (+ Datacenter Engine)    | 6        |
| 1.2 BradCore — The Ubiquitous Workhorse [Gen1]            | 7        |
| 1.3 HyperCore & Atlas Architecture — Mobile SoC [Gen1]    | 7        |
| 1.4 BradChip — Unified SoC Package [Gen1]                 | 8        |
| 1.5 BradXon — The Flagship Adaptive Processor             | 9        |
| PART 2: GPU ARCHITECTURE                                  | 9        |
| 2.1 Product Lineup                                        | 9        |
| 2.1a BradVector Engines — ISA & Compute Reference Docs    | 10       |
| 2.2 GPU Architecture Diagram [Gen1]                       | 11       |
| 2.3 BradRT Cores — Ray Tracing [Gen1]                     | 12       |
| 2.4 NeuroSense AI Engine (On-GPU) [Gen1]                  | 12       |
| 2.5 BradSense-Render — Neural Rendering Deep-Dive [Gen1]  | 12       |
| 2.6 BradFusion Fabric Integration                         | 13       |
| 2.7 BradGfx Max Dual Role                                 | 14       |
| 2.7a BradGfx Max Xtensive (Hybrid-UMA Edition)            | 14       |
| 2.8 BradSec Display Pipeline [Gen1]                       | 14       |
| PART 3: NPU ARCHITECTURE                                  | 14       |
| 3.1 BradNPU — Neural Processing Unit [Gen1]               | 14       |
| 3.2 Neural Compute Cluster (NCC) Architecture             | 14       |
| 3.3 HB-AIM (High-Bandwidth AI Memory)                     | 15       |
| 3.4 AI-ISA (Dedicated Instruction Set)                    | 15       |
| 3.5 MoE (Mixture of Experts) Engine                       | 16       |
| 3.6 Workload Scheduler                                    | 16       |
| 3.7 Use Case Performance                                  | 16       |
| 3.8 Fabrication & Power                                   | 16       |
| PART 4: MEMORY & STORAGE                                  | 17       |
| 4.1 Memory Architecture: BradRAM + SPMP [Gen1]            | 17       |
| 4.2 Storage Architecture: BradDrive + BradCompress [Gen1] | 19       |
| PART 5: VISION & CAMERA                                   | 21       |
| 5.1 BradVision — Complete Camera & Vision System [Gen1]   | 21       |
| 5.2 Sensor Technology                                     | 21       |
| 5.3 ISP Pipeline (14-Stage Hardware)                      | 22       |
| 5.4 Optics                                                | 22       |
| 5.5 BradVision Fusion Engine (Software on BradNPU)        | 22       |
| 5.6 Video Capabilities                                    | 23       |
| 5.7 Vision-as-Sensor                                      | 23       |
| 5.8 Modular Bayonet Optics [Vision]                       | 23       |
| PART 6: PLATFORM & MOTHERBOARDS                           | 24       |
| 6.1 SoC Platforms                                         | 24       |
| 6.2 BradBoard Motherboard Series [Gen1]                   | 24       |
| 6.3 Brad-CAMM Memory Slots                                | 26       |
| 6.4 Brad-M.3 Storage Slots                                | 26       |
| 6.5 BradLink Device Tethering [Gen2]                      | 26       |

|                                                                    |    |
|--------------------------------------------------------------------|----|
| 6.6 Target Device → Platform Mapping . . . . .                     | 26 |
| PART 7: SOFTWARE & OS . . . . .                                    | 27 |
| 7.1 BradOS Family [Gen1] . . . . .                                 | 27 |
| 7.2 Core OS Features . . . . .                                     | 28 |
| 7.2.5 Software Support & Update Policy (Committed) . . . . .       | 28 |
| 7.3 Brad Mobile Services (BMS) [Gen1] . . . . .                    | 30 |
| 7.4 EROE — Eco-Render Optimization Engine [Gen1] . . . . .         | 30 |
| 7.5 Brad-CVT & Unified Driver Suite [Gen1] . . . . .               | 31 |
| 7.6 BradDev Kit . . . . .                                          | 31 |
| PART 8: CONSOLE & PRODUCTS . . . . .                               | 32 |
| 8.1 BradAxis Console [Gen1] . . . . .                              | 32 |
| 8.2 Brad Ecosystem vs Apple Ecosystem (Short Form) . . . . .       | 34 |
| PART 9: HORIZON & VISION . . . . .                                 | 34 |
| 9.1 BradSkynet — AI Platform / AGI Infrastructure [Gen2] . . . . . | 34 |
| 9.2 BradAstra — VR Ecosystem & Headset [Gen2-Vision] . . . . .     | 35 |
| 9.3 BradLabs — Fabrication & R&D Division . . . . .                | 35 |
| 9.4 Roadmap Chronology . . . . .                                   | 35 |
| 9.5 Generation Timeline . . . . .                                  | 35 |
| PART 10: APPENDICES . . . . .                                      | 36 |
| Appendix A: Memory Hierarchy . . . . .                             | 36 |
| Appendix B: Security Boundaries . . . . .                          | 36 |
| Appendix C: Data Pipeline Walkthrough — Game Level Load . . . . .  | 37 |
| Appendix D: Bandwidth Budget . . . . .                             | 37 |
| Appendix E: Naming Conventions . . . . .                           | 37 |
| Appendix F: Credibility Tiers . . . . .                            | 38 |
| Appendix G: Open Questions for Brad . . . . .                      | 38 |
| Appendix H: Key Architectural Relationships . . . . .              | 39 |
| Appendix I: Rise of the Streets . . . . .                          | 39 |
| Appendix J: Complete Component Index . . . . .                     | 39 |
| End of Brad Devices Complete Unified Specification . . . . .       | 41 |



# BRAD DEVICES

*Built in Africa, for the World*

## BRAD DEVICES — Complete Unified Specification

*A full-stack silicon/hardware/software universe by Brad (age 15) Documents span July 2025 - July 2026. This spec book consolidates all 78 reference .md docs + Info.txt into one*

*document. Last updated: September 2026*

## How to Read This Spec

**Generation Badges** appear throughout to denote technology readiness:

| Badge              | Generation  | Years     | Meaning                                                         |
|--------------------|-------------|-----------|-----------------------------------------------------------------|
| <b>[Gen1]</b>      | Kinetic     | 2026-2029 | Shipping now. In production or ready for tape-out.              |
| <b>[Gen2]</b>      | Synaptic    | 2030-2033 | Stretch goal. Plausible with process shrink and iteration.      |
| <b>[Gen3]</b>      | Quantum     | 2034-2037 | Long-term R&D. Requires materials/physics breakthroughs.        |
| <b>[Gen4]</b>      | [Vision]    | 2038-2041 | Deep materials research, no known production path.              |
| <b>[Gen5]</b>      | Singularity | 2042+     | Theoretical. Self-evolving gates, quantum fabric, bio-synaptic. |
| <b>[Aesthetic]</b> | —           | —         | Marketing language, not technical claims.                       |

**Release cadence:** Brad Devices ships on a **4-year generation cycle** — each generation is a complete architectural leap, not an annual refresh. Within a generation, mid-cycle SKUs (process shrinks, memory/capacity upgrades, new form factors) are allowed, but the core architecture stays stable for the full 4-year window. Rationale and detail: ROADMAP.md §Release Cadence.

**Cross-references** in code font point to the source .md file for deeper detail.

**Credibility tiers** by document area: NPU, RAM, Drive, OS, Console, Datacenter (BradApex), BradXon platform = grounded engineering. Quantum Fabric = visionary. BradSense Render is the split case: the *architecture* (§2.5) is visionary, the *software reference upscaler* behind it now has a shipped M0 with receipts — see §2.5 and Brad Silicon/GPU/BRADSENSE\_RENDERER\_PLAN.md. (BradXon's ANF unit-morph is the one [Vision] shelf — see §1.5.)

## Brand Pillar: Built in Africa, for the World

Brad Devices is positioned as **Africa's first vertically-integrated technology company** — custom silicon, operating system, and consumer devices designed, engineered, and integrated under one roof, on the continent. The claim is scoped precisely: no African company today designs its own CPU/GPU silicon, its own OS, and its own devices at integrated scale.

### The Five Supporting Pillars

1. **Founder story** — rooted in Kisian; authenticity no competitor can copy.
2. **African engineering hubs** — Johannesburg data center (strategic, Brad Labs, Brad Software/OS/BMS\_SPEC.md), with Lagos/Nairobi/Cairo talent hubs planned; African engineers build the silicon, not just consume it.

3. **BradLabs foundry ambition** — a captive fab headquartered in Africa (2029+). Honest interim: **Samsung Foundry for all of Kinetic**, unified on a tiered node ladder (7LPP budget → SF4 → SF4P → SF4X flagship — cost + memory bundling), with a BradLabs custom PDK. Full tiering: BradLabs/BRADLABS\_FAB\_STRATEGY.md.
4. **Efficiency-first design** — low-power-first engineering (EROE, Falcon E-cores, DMIPS/W leadership) because power infrastructure is scarce across much of Africa — authentic to the continent *and* a product advantage everywhere.
5. **Developer ecosystem** — BradLabs Program (African game developers), 70/30 publishing split, BradStudios Originals (sibling company under the BradVerse Kenyan parent, not a Brad Devices division).

### Honesty Guardrails

- The claim is **integrated scope only** — not “Africa’s first tech company,” which would be false.
- The fab is partner-run (**Samsung unified across all Kinetic tiers**, Intel optional US hedge — SMIC excluded) until BradLabs reaches 2029+; no premature “made in Africa” silicon claims.
- Every pillar has a concrete program behind it, never a slogan.

### Existing Africa References (consolidated)

- Brad Software/OS/BMS\_SPEC.md:123 — Johannesburg data center (strategic, BradLabs)
- BRAD\_DEVICES\_COMPLETE\_SPEC.md §8.1, §9.3 — BradLabs Program (African devs)
- Brad Devices/Console/BRADAXIS\_SPEC.md — African-centric developer pillar

## PART 0: ECOSYSTEM BLUEPRINT

### Directory Structure

```

Brad Devices/
├── BRAD_DEVICES_COMPLETE_SPEC.md    ← You are here
├── MASTER_INDEX.md                 ← Ecosystem blueprint (375 lines)
├── VISION.md                       ← Sci-fi catalog with badges (172 lines)
├── ARCHITECTURE.md                 ← ASCII diagrams (196 lines)
├── ROADMAP.md                      ← Chronological evolution (240 lines)
├── NAMING_GUIDE.md                 ← Naming resolutions (327 lines)
├── BRAD_VS_APPLE.md                ← Apple comparison (167 lines)
├── HISTORY.md                      ← Shipped-receipts ledger (235 lines)
├── TASKS.md                        ← The live furnace (90 lines)
├── AGENT.md                        ← Agent rulebook (447 lines)
├── README.md                       ← Repo guide (57 lines)
├── BRADISA_CI_REMAINING.md         ← BradISA CI ledger (138 lines)
├── Brad Devices/
│   ├── BRADCUBE_SPEC.md            ← BradCube mini-PC line (105 lines)
│   ├── VR and BradSkynet/BRADSKYNET_ASTRA_LABS.md ← Horizon concepts (170 lines)
│   ├── Display/SENTINELVIEW.md     ← Display tech (157 lines)
│   ├── Console/BRADAXIS_SPEC.md    ← Console spec (305 lines)
│   ├── Storage/BRADDRIVE_COMPRESS_ARCH.md ← Storage architecture (186 lines)
│   └── Vision/BRADVISION_COMPLETE.md ← Camera & vision system (398 lines)
├── BradLabs/
│   └── BRADLABS_FAB_STRATEGY.md    ← Foundry & process ladder (87 lines)
├── Brad Silicon/
│   ├── BRADSILICON_STRATEGY.md     ← Silicon strategy (194 lines)
│   ├── PLATFORM_SPEC.md            ← BradBoard + SoC platforms (311 lines)
│   ├── PIPELINE_SPEC.md            ← Data pipeline (263 lines)
│   ├── Platform/BRADBOARD_DEEP_DIVE.md ← PCB stackups, Fabric, HSM (638 lines)
│   └── BradISA/

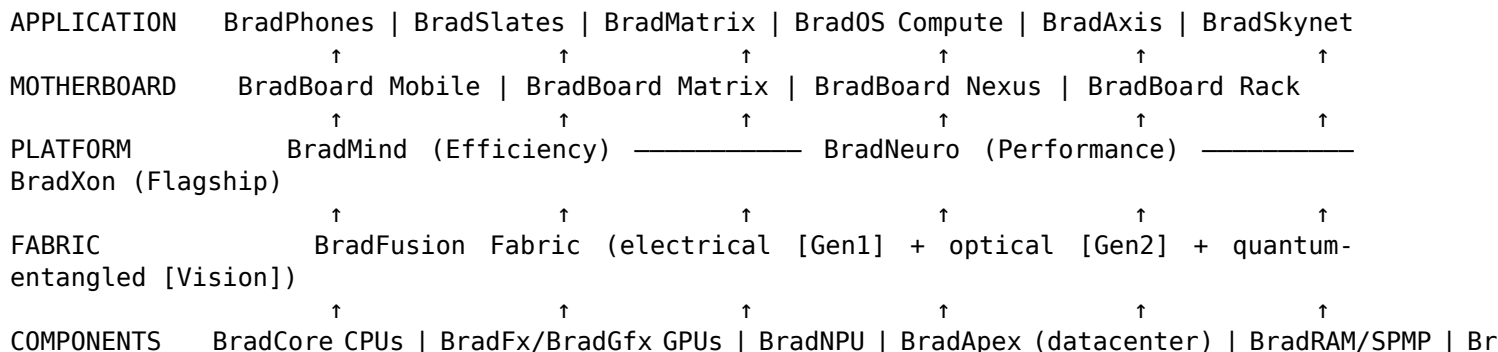
```

|                                   |                                                                                      |
|-----------------------------------|--------------------------------------------------------------------------------------|
| — ARCHITECTURE_GUIDE.md           | ← BradISA overview (105 lines)                                                       |
| — ARCHITECTURE_REFERENCE_MODEL.md | ← Reference model (93 lines)                                                         |
| — ISA_QUICK_REFERENCE.md          | ← Encode/decode tables (287 lines)                                                   |
| — BRADISA_VS_RISCV.md             | ← ISA vs RISC-V + the six fences (156 lines)                                         |
| — TOOLCHAIN_GUIDE.md              | ← bradasm/asm workflow (180 lines)                                                   |
| — TOOLCHAIN_STRATEGY.md           | ← Toolchain roadmap (150 lines)                                                      |
| — CPU/                            |                                                                                      |
| — HYPERCORE_ATLAS.md              | ← HyperCore mobile SoC (243 lines)                                                   |
| — HYPERCORE_API_REF.md            | ← SoC API (289 lines)                                                                |
| — HYPERCORE_REGMAP.md             | ← Register map (244 lines)                                                           |
| — BRADCHIP.md                     | ← SoC package (124 lines)                                                            |
| — EROE.md                         | ← Power/thermal governor (51 lines)                                                  |
| — Compute/                        |                                                                                      |
| — BRADAPEX_ACCELERATOR_SPEC.md    | ← Datacenter AI & HPC accelerator (237 lines)                                        |
| — BRAD_COMPUTE_ARCHITECTURE.md    | ← BradFx/BradGfx + BradApex (311 lines)                                              |
| — GPU/                            |                                                                                      |
| — GPU_ARCH.md                     | ← GPU arch + die names/registry (311 lines)                                          |
| — GPU_ISA.md                      | ← BradVector ISA (293 lines)                                                         |
| — GPU_COMPUTE.md                  | ← Vector compute model (331 lines)                                                   |
| — GPU_DRIVER.md                   | ← GPU driver suite (282 lines)                                                       |
| — GPU_MEMORY.md                   | ← GPU memory model (281 lines)                                                       |
| — GPU_REGMAP.md                   | ← Register map (264 lines)                                                           |
| — BRADSENSE_RENDER_SPEC.md        | ← BradSense-Render deep-dive (279 lines)                                             |
| — BRADSENSE_RENDERER_PLAN.md      | ← Reference-upscaler rung plan (148 lines)                                           |
| — NPU/                            |                                                                                      |
| — BRADNPU_ARCH.md                 | ← NPU architecture (181 lines)                                                       |
| — Memory/                         |                                                                                      |
| — BRADRAM_SPMP_ARCH.md            | ← Memory architecture (174 lines)                                                    |
| — BRADRAM_DDS.md                  | ← Data sensing (234 lines)                                                           |
| — BRADRAM_POWER.md                | ← Power model (191 lines)                                                            |
| — BRADRAM_RAS.md                  | ← Reliability (220 lines)                                                            |
| — BRADRAM_TIMING.md               | ← Timing (187 lines)                                                                 |
| — BRADRAM_VMEM.md                 | ← Virtual memory (220 lines)                                                         |
| — BRADRAM_REGMAP.md               | ← Register map (242 lines)                                                           |
| — MVRAM_LPGDDR_PARTNERSHIP.md     | ← LPDDR partnership (121 lines)                                                      |
| — Brad Software/                  |                                                                                      |
| — OS/                             |                                                                                      |
| — BRADOS_COMPLETE.md              | ← OS + BMS (226 lines)                                                               |
| — BRADOS_KERNEL_STRATEGY.md       | ← Kernel roadmap (167 lines)                                                         |
| — BMS_SPEC.md                     | ← Brad Mobile Services (664 lines)                                                   |
| — BRADCVT_DRIVERS.md              | ← Drivers suite (205 lines)                                                          |
| — BRADVECTOR_PLATFORM_MOAT.md     | ← BradVector platform & moat (295 lines)                                             |
| — CUSTOM_TECH.md                  | ← Custom-stack tech notes (89 lines)                                                 |
| — COMPANY/                        |                                                                                      |
| — COMPANY_STRUCTURE.md            | ← BradVerse Group → Brad Devices (116 lines)                                         |
| — DIVISIONS.md                    | ← Division map (22 lines)                                                            |
| — IP_REGISTRY.md                  | ← Unified IP registry (208 lines)                                                    |
| — EQUITY_PLAN.md                  | ← Founder-first cap table (104 lines)                                                |
| — NDA_TEMPLATE.md                 | ← Non-disclosure agreement (157 lines)                                               |
| — POLICIES.md                     | ← Company policies (212 lines)                                                       |
| — Peripherals/                    |                                                                                      |
| — Info.txt                        | ← AI co-architect conversation (2,167 lines)                                         |
| — src/                            | ← Reference toolchain: BVRT, braddev, bvlibs, bms, tifa, rtl + 8 READMEs (736 lines) |
| — Converted/                      | ← Archival .txt/.md source documents (3 files)                                       |

**Total ecosystem: 78 .md text files + Info.txt = 79 docs, 16,727 lines (excluding**

Converted/ originals; 81 files, 17,619 lines with them). New since prior revision: BradISA/ suite (ISA\_QUICK\_REFERENCE, BRADISA\_VS\_RISCV, TOOLCHAIN\_GUIDE, TOOLCHAIN\_STRATEGY), GPU ISA suite (GPU\_ISA/COMPUTE/DRIVER/MEMORY/REGMAP), BradSense-Render pair (BRADSENSE\_RENDER\_SPEC + BRADSENSE\_RENDERER\_PLAN), Memory suite (DDS/POWER/RAS/TIMING/VMEM/REGMAP/PARTNERSHIP), COMPANY/, BradLabs/, BRADSILICON\_STRATEGY, BRAD\_COMPUTE\_ARCHITECTURE, BRADVECTOR\_PLATFORM\_MOAT, BRADOS\_KERNEL\_STRATEGY, BRADCUBE\_SPEC, SENTINELVIEW, HYPERCORE\_API\_REF/REGMAP, BRADBOARD\_DEEP\_DIVE. This spec book merges all dedicated docs.

## Architecture Stack



## The Defining Architectural Insight: The Data Pipeline

BradDrive → BradCompress HDB → SPMP → BradFusion Fabric → Any Processor

Zero-copy, zero-CPU-overhead, direct-to-pool streaming. This is the ecosystem's core differentiator — Apple-style vertical integration where every component talks directly to every other via the Fabric, with no CPU involvement in data movement.

## Generation Naming

Brad Devices ships on a **4-year generation cycle** — a complete architectural leap per generation, with mid-cycle SKU/process refinements in between. See ROADMAP.md §Release Cadence for full rationale.

|                    |             |                                                |
|--------------------|-------------|------------------------------------------------|
| Gen 1 (2026–2029): | KINETIC     | – "Data in Motion. Intelligence in Real-Time." |
| Gen 2 (2030–2033): | SYNAPTIC    | – "Learning at the Speed of Thought."          |
| Gen 3 (2034–2037): | QUANTUM     | – "Beyond Binary. Into Entanglement."          |
| Gen 4 (2038–2041): | [VISION]    | – Experimental node (BradLabs, sub-1nm).       |
| Gen 5 (2042+):     | SINGULARITY | – Theoretical (self-evolving, quantum fabric). |

**Note on BradApex timing:** BradApex is a standalone datacenter accelerator, not a consumer generation. Its [Gen1] design → [Gen2] ramp 2027-28 occurs *within* the Gen1 Kinetic window (2026-2029) as a product ramp, complementing the consumer 4-year cadence.

## PART 1: CPU ARCHITECTURE

## 1.1 CPU Philosophy: Two Families (+ Datacenter Engine)

|                     | BradCore                                                                                          | BradXon                                                                                                         | BradApex                                                                         |
|---------------------|---------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| <b>Role</b>         | Ubiquitous workhorse — mass-market                                                                | Bleeding-edge flagship — apex predator                                                                          | Datacenter engine — AI & HPC accelerator                                         |
| <b>Markets</b>      | Phones, tablets, laptops, compute, automotive, health, education, BradAxis console (standard SKU) | VR ecosystem servers (BradAstra), BradOS Compute Pro, BradSkynet controller                                     | BradSkynet nodes (AI training/inference + scientific FP64)                       |
| <b>Architecture</b> | BradISA V1 custom P/E cores                                                                       | BradISA P/E cores + BradGfx Max + Neuro-Stream; ANF = compute partitions + reconfigurable NPU + fabric governor | BA-ISA: tensor engine (FP8/FP16/BF16/INT8) + first-class IEEE FP64 vector engine |
| <b>Process</b>      | Samsung 7LPP→SF4X tiered ladder [Gen1]                                                            | SF4X (4nm-class) chiplet [Gen1]; sub-1nm graphene/BN [Vision]                                                   | SF4X (4nm-class) [Gen2 ramp 2027–28]                                             |
| <b>GPU pairing</b>  | BradFx (integrated)                                                                               | BradGfx Max (integrated via ANF)                                                                                | None (not a GPU; memory-centric by design)                                       |
| <b>Node</b>         | Brad Sili-con/CPU/BRADCHIP.md                                                                     | Brad Sili-con/PLATFORM_SPEC.md                                                                                  | Brad Sili-con/Compute/BRADAPEX_ACCELERAT                                         |

## 1.2 BradCore — The Ubiquitous Workhorse [Gen1]

BradCore is the mass-market CPU core complex — asymmetric BradISA V1 with Phoenix (P) performance cores and Falcon (E) efficiency cores. Fabricated on the Samsung tiered ladder (7LPP→SF4X) at Samsung (unified fab; see BradLabs/BRADLABS\_FAB\_STRATEGY.md).

**Core types:** - **Phoenix (P):** High-performance out-of-order cores for compute-heavy workloads - **Falcon (E):** Power-efficient in-order cores for background tasks and system maintenance

**Integration:** BradCore ships as part of the BradChip SoC package (see 1.4), paired with BradFx GPU and BradNPU.

## 1.3 HyperCore & Atlas Architecture — Mobile SoC [Gen1]

HyperCore is the mobile silicon line — purpose-built for the pocket, not a cut-down PC. Targets: BradPhones and BradSlates.

**Philosophy:** BradCore is for sustained throughput. HyperCore is for **context persistence** — seamless app switching, instant resume, aggressive power gating.

|                      | BradCore (Desktop)              | HyperCore (Mobile)                            |
|----------------------|---------------------------------|-----------------------------------------------|
| <b>Primary Goal</b>  | Throughput (speed)              | Context (persistence)                         |
| <b>UMA Style</b>     | SPMP — wide-bus, high-bandwidth | AUC — state persistence, instant intelligence |
| <b>Compute Logic</b> | Phoenix & Falcon                | Phoenix & Sentinel                            |
| <b>GPU Engine</b>    | BradFx / BradGfx                | HyperCore GPU (Atlas-tuned)                   |
| <b>Memory Buffer</b> | 2GB SRAM (Gfx Max)              | 950MB MVRAM (LPDDR5X)                         |
| <b>Cooling</b>       | Active micro-fluidics           | BradCool-Nano passive vapor                   |
| <b>Node</b>          | 4nm→2nm→sub-1nm                 | 7LPP/SF4 (mature, stable)                     |

## HyperCore Product Lineup

| Tier            | Model    | NPU     | MVRAM | Target              | Key Advantage                        |
|-----------------|----------|---------|-------|---------------------|--------------------------------------|
| <b>Lite</b>     | HC-100   | 16-core | 128MB | BradPhone Essential | 3-day battery, GTX 1050-class GPU    |
| <b>Standard</b> | HC-300   | 32-core | 256MB | BradPhone Base      | Balanced AI-agent performance        |
| <b>Ultra</b>    | HC-500   | 48-core | 450MB | BradPhone Ultra     | Atlas Pro engine                     |
| <b>Elite</b>    | HC-Elite | 64-core | 950MB | BradSlate Prime     | Matches MacBook M2-class performance |

## Atlas Architecture Components

- **Atlas Unified Context (AUC):** Shared context, not just shared memory. CPU and GPU share context with zero-latency context switch (<100µs). GPU power-gates completely during text reading.
- **Sentinel Cores:** Ultra-low-leakage always-on cores (microwatts, not milliwatts) — distinct from Falcon cores. Handle clock, notifications, TIFA AI listening, FaceLock. Enable 48h+ standby.
- **MVRAM (Management Video RAM):** Dedicated on-die buffer using LPDDR5X, 3D-stacked tile on logic. Max 950MB on HC-Elite, 256 GB/s bandwidth, ~500mW active / ~50µW standby. Permanently caches: BradOS kernel, TIFA AI model weights, UI asset library, frozen app contexts (10+ apps).
- **HyperCore GPU:** Same BradVector Engines and Neuro-Stream Units as desktop BradFx/Gfx. Same shader math — desktop code runs on mobile without port. Scaled core count, same architecture.
- **BradCool-Nano:** 0.2mm vapor-chamber/graphene sandwich. Peak <45°C surface sustained gaming.

## Competitive Positioning (2026)

| Metric        | Apple A19 Pro      | HyperCore HC-Elite       |
|---------------|--------------------|--------------------------|
| Node          | 3nm (N3P)          | SF4X (4nm-class)         |
| On-Die Buffer | ~32MB System Cache | 950MB MVRAM (LPDDR5X)    |
| Architecture  | Legacy hybrid      | Atlas Unified Context    |
| AI Focus      | Photography / Siri | Agentic reasoning (TIFA) |
| Battery       | 18-24 hours        | 36-48 hours (Sentinel)   |

### 1.4 BradChip — Unified SoC Package [Gen1]

BradChip combines BradCore CPU + BradFx GPU + BradNPU + memory controller + BradFusion Fabric into one monolithic or chiplet-based die. It is the compute module that powers BradMatrix, BradSlates, and BradOS Compute Base systems. Slots into BradBoard Matrix socket.

| Tier            | CPU   | GPU       | NPU     | Fabric           | Process    | TDP    | Target                           |
|-----------------|-------|-----------|---------|------------------|------------|--------|----------------------------------|
| <b>Lite</b>     | P2+E4 | BradFx 10 | 256 NCC | Fabric           | 7LPP (7nm) | 10-15W | BradMatrix Essential             |
| <b>Standard</b> | P4+E4 | BradFx 30 | 384 NCC | Fabric           | SF4 (4nm)  | 15-25W | BradMatrix Base, BradSlate Prime |
| <b>Ultra</b>    | P6+E4 | BradFx 50 | 512 NCC | Fabric + optical | SF4P (4nm) | 25-40W | BradMatrix Prime 14"             |

| Tier         | CPU   | GPU                     | NPU         | Fabric              | Process    | TDP    | Target                    |
|--------------|-------|-------------------------|-------------|---------------------|------------|--------|---------------------------|
| <b>Elite</b> | P8+E4 | BradFx<br>70            | 768<br>NCC  | Fabric +<br>optical | SF4X (4nm) | 40-65W | BradMatrix<br>Ultra 16"   |
| <b>Max</b>   | P8+E8 | BradFx<br>90<br>(Max-i) | 1024<br>NCC | Fabric +<br>optical | SF4X (4nm) | 65-95W | BradOS<br>Compute<br>Base |

All tiers include: BradFusion Fabric, dual-channel memory controller (Brad-CAMM LPDDR/DDR), BradSec root-of-trust, BradCompress HDB interface, BradLink tethering.

**BGA/LGA packages** ranging from 25×30mm (Lite, soldered) to 45×50mm (Max, LGA socket + backplate).

**Competitive comparison with Apple M4 Max** — BradChip Max offers similar CPU config (16-core vs 16-core), competitive GPU (48 VE vs 40-core), stronger NPU (1024 NCC + HB-AIM vs 38 TOPS), upgradable Brad-CAMM LPDDR/DDR memory, and BradFusion Fabric + optical interconnect.

## 1.5 BradXon — The Flagship Adaptive Processor

BradXon is the flagship processor — a standalone chiplet CPU (not a mobile SoC) that integrates directly into BradBoard Nexus and Rack over the BradFusion fabric. Its defining feature is the **Adaptive Neural Fabric (ANF)**, Gen 1 [DESIGNED].

**ANF — what “adaptive” means [Gen1, DESIGNED]:** - **Compute Partitions (CPs):** the die’s execution arrays are grouped into secure hardware partitions (CPU / GPU / NPU sides), composed at boot under zone isolation (MIG-class). Partitions are re-allocated at domain boundaries, never per instruction. - **Reconfigurable Dataflow Array (RDA):** the NPU slice is a CGRA-class dataflow fabric, rebuilt at kernel launch — the AI math adapts to the model rather than to a fixed systolic pipe. - **Fabric Governor (FG):** the PMU + OS scheduler continually re-allocate clock, power, and fabric bandwidth between the CPU/GPU/NPU partitions (DVFS + heterogeneous scheduling). Budget reallocation is real; silicon reallocation is not.

**ANF — what Gen 2 keeps as [Vision]:** - **Unit morphing:** re-forming execution-unit tiles into a different domain category at runtime. The honest path is tile composition at integration: fabric tiles floorplanned CPU/GPU/NPU, mated into a package at build time — not per-instruction. - **Self-evolving logic gates:** [Vision] transistors that physically re-wire — requires fundamentally new device physics. - **Photonic / quantum fabric:** hybrid optical chiplet links [Gen2]; on-die photonic logic [Vision]; quantum-entangled pathways [Vision] — pure fiction by 2026 standards.

**Integrated & companion silicon:** - **Integrated BradGfx Max:** [Gen1] same silicon as the discrete card, attached via the ANF compute-partition layer - **BradGfx Max Xtensive:** [Gen1] discrete edition — 256 BradVector / 128 RT / 512 Neuro-Stream, 32 GB GDDR7 on a 192-512-bit bus, hybrid-UMA (dedicated VRAM + SPMP overflow) - **NeuroSense X-Units:** [Gen1] dedicated AIMP (AI Model Processors) - **BradSec-X Quantum Fortification:** [Vision] post-quantum crypto with self-healing integrity

**Targets:** BradSkynet Core Nodes, BradAstra Servers, BradOS Compute Pro Workstations. (*BradAxis uses standard BradCore — see §8.1.*)

## PART 2: GPU ARCHITECTURE

### 2.1 Product Lineup

#### BradGfx — Discrete-Package, Unified-Memory GPUs

*BradGfx is a **UMA GPU** for Lite→Max (discrete package, unified SPMP memory, Grace-Hopper class — no private VRAM). Only **Max Xtensive** carries a GDDR7 tank and is the line’s one true **hybrid dGPU**. See Brad Silicon/GPU/GPU\_ARCH.md §1.*

| SKU                    | Shader Cores | RT Cores | AI Cores | FP32 TFLOPS | TDP  | Target                                                     |
|------------------------|--------------|----------|----------|-------------|------|------------------------------------------------------------|
| <b>Lite</b>            | 512          | 16       | —        | 4-6         | 75W  | Budget gaming                                              |
| <b>Standard</b>        | 1024         | 32       | 8        | 10-14       | 130W | Mainstream (~RTX 3050/3060)                                |
| <b>Ultra</b>           | 2048         | 64       | 16       | 20-28       | 220W | Prosumer                                                   |
| <b>Max</b>             | 4096         | 128      | 32       | 40-55       | 350W | Flagship (discrete-package UMA or ANF-integrated)          |
| <b>Max Xn-ten-sive</b> | 4096         | 128      | 32+      | 40-55+      | 225W | Flagship hybrid dGPU — GDDR7 tank + SPMP (see GPU_ARCH §1) |

### BradFx — Integrated GPUs

| SKU                   | Shader Cores | RT Cores | AI Cores | FP32 TFLOPS | TDP    | Target                          |
|-----------------------|--------------|----------|----------|-------------|--------|---------------------------------|
| <b>Lite</b>           | 128          | —        | —        | 1-2         | 5-10W  | Budget devices                  |
| <b>Standard</b>       | 256          | —        | 4        | 2-4         | 10-20W | Laptops/tablets (~Apple M2 GPU) |
| <b>Ultra</b>          | 512          | 16       | 8        | 5-8         | 20-35W | Premium ultrabooks              |
| <b>Max</b><br>(Max-i) | 1024         | 32       | 16       | 10-15       | 35-65W | BradChip, powerful slates       |

### 2.1a BradVector Engines — ISA & Compute Reference Docs

The full instruction set, compute model, and verification story for the BradVector engines (all shader cores above) live in the reference documentation in Brad Silicon/GPU/ and Brad Silicon/Compute/:

| Reference Doc                | Covers                                                                                                                                                             |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Brad Silicon/GPU/GPU_ARCH.md | Torox G1 microarchitecture — BradVector / BradRT / Neuro-Stream units, full SKU lineup, ray-tracing cores, <b>die naming (BGT/BFT/BAT) + Torox G1 die registry</b> |
| Brad Silicon/GPU/GPU_ISA.md  | BradVector ISA — 64-bit dual-issue SIMT encoding, .bvbc portable bytecode vs .brsh native AOT, register file, shared memory                                        |

| Reference Doc                                     | Covers                                                                                                        |
|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| Brad Silicon/GPU/GPU_COMPUTE.md                   | Compute model, kernel launch, warp scheduling, memory model (SPMP / Infinity Buffer overlap)                  |
| Brad Silicon/GPU/GPU_REGMAP.md                    | Register map & programming model — register-level control of all Bradley tiers                                |
| Brad Silicon/GPU/GPU_MEMORY.md                    | Memory model & cache hierarchy — SPMP access, coherency, bandwidth allocation                                 |
| Brad Silicon/GPU/GPU_DRIVER.md                    | Kinetic Unified Driver — Vulkan 1.4, DirectX 12 Ultimate, Metal 3, Brad-Link API                              |
| Brad Silicon/GPU/BRADSENSE_RENDER_SPEC.md         | BradSense-Render neural reconstruction deep-dive (see §2.5)                                                   |
| Brad Silicon/GPU/BRADSENSE_RENDERER_PLAN.md       | The software reference upscaler — M0-M4 rungs, acceptance gates, receipt chain                                |
| Brad Silicon/Compute/BRAD_COMPUTE_ARCHITECTURE.md | BradVector as the CUDA-analog software platform; SPMP → CUDA-shared-equivalent, Infinity Buffer → CUDA-shared |

**Reference toolchain (released):** bradc (assembler → .bvbc), BVRT (interpreter), bradgdb (debugger), BradTimeline (profiler/trace). Full details in Brad Software/OS/BRADVECTOR\_PLATFORM\_MOAT.md.

## 2.2 GPU Architecture Diagram [Gen1]

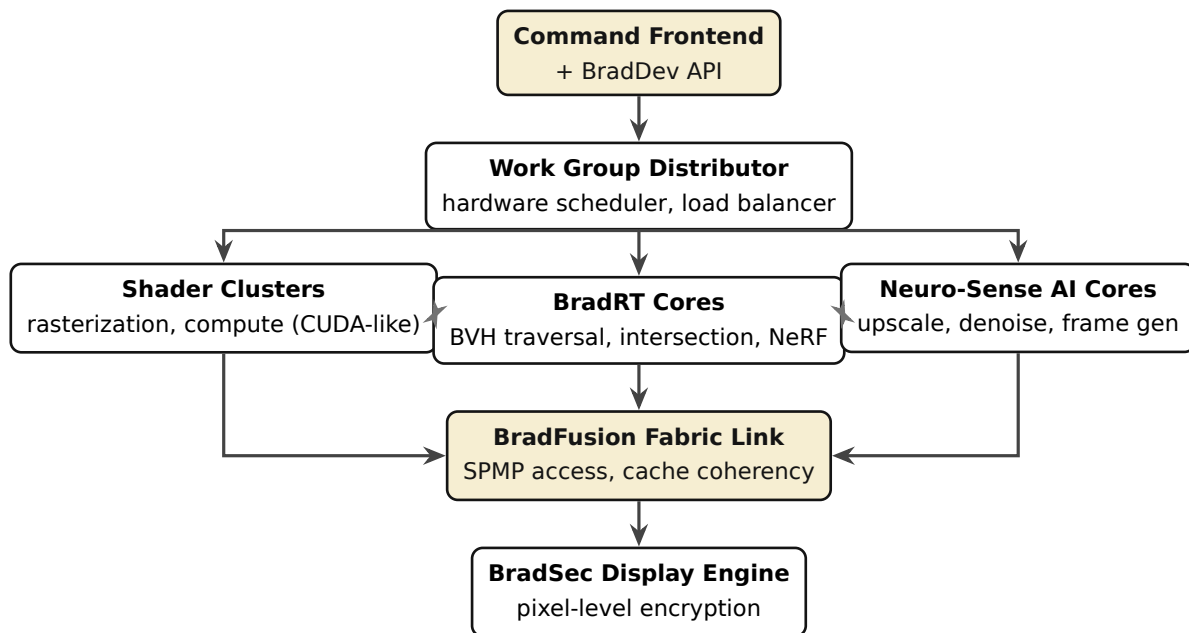


Figure 1: Gen1 unified GPU architecture — BRADGFX / BRADFX

## 2.3 BradRT Cores — Ray Tracing [Gen1]

Each BradRT Core contains: - **Ray Traversal Unit (RTU)**: Navigates BVH acceleration structures with proprietary cache-optimized tree traversal - **Intersection Processing Unit (IPU)**: Massively parallel ray-primitive intersection (triangles, spheres, line segments) - **Programmable Ray Pipeline (PRP)**: Custom ray behaviors, material interactions, secondary ray gen — supports Opacity Micromaps - **Triangle Cluster Accelerator**: Optimized ray tracing of complex clustered geometry

**RT scaling**: 16 cores / 4 GigaRays/s (Lite) → 128 cores / 40+ GigaRays/s (Max).

**Neural Rendering Acceleration [Gen1]**: Hardware optimized for NeRF (Neural Radiance Fields) evaluation — real-time digital doubles and volumetric effects (fog, smoke).

## 2.4 NeuroSense AI Engine (On-GPU) [Gen1]

| Feature                   | Description                                                                                   |
|---------------------------|-----------------------------------------------------------------------------------------------|
| <b>BradSense-Render</b>   | AI upscaling + frame generation. Context-aware predictive rendering, not just DLSS/FSR.       |
| <b>NeuroSense-Denoise</b> | AI denoising for ray-traced output — fewer rays per pixel while maintaining quality.          |
| <b>TIFA AI</b>            | Real-time workload monitoring — dynamically adjusts clock, voltage, rendering technique.      |
| <b>AI Physics</b>         | Dedicated AI cores accelerate physics simulation, character behavior, environmental dynamics. |

**Division of labor**: On-GPU AI cores handle real-time rendering inference (denoising, upscaling, frame gen). BradNPU handles heavy AI (model training, complex inference, NeuroSense assistant).

## 2.5 BradSense-Render — Neural Rendering Deep-Dive [Gen1]

BradSense-Render is not an upscaler — it is a **reconstruction engine** that uses the BradFusion Fabric to pull semantic object data from the game engine, then applies per-object neural weights to reconstruct each frame at higher-than-native fidelity.

### The Intelligence Pipeline:

1. **Semantic Feed (via Fabric)**: Game engine sends compact semantic packet — object ID map, material type (8 classes), depth buffer, motion vectors, albedo/normal — ~50MB/frame uncompressed, <5MB via BradCompress.
2. **Per-Object Neural Reconstruction**: Each object class has a dedicated lightweight neural network (expert). Loading expert weights from on-die SRAM in <1µs per type switch.
3. **Temporal Stabilization**: Per-object accumulation rates — static objects accumulate 16+ frames, fast-moving objects 1-2 frames. Zero ghosting (system knows *which* pixels belong to *which* object).
4. **Frame Generation**: Interpolated frames between real renders. 1 real → 1 AI (2×) or 1:2 (3×). No input lag increase (generated retroactively between real frames).

### Neuro-Stream Units (hardware requirements):

| Tier                | Neuro-Stream Units | On-die SRAM | Object Classes | Features             |
|---------------------|--------------------|-------------|----------------|----------------------|
| BradFx<br>10 (Lite) | 32                 | 8MB         | 4              | Basic reconstruction |

| Tier                                     | Neuro-Stream Units | On-die SRAM | Object Classes         | Features    |
|------------------------------------------|--------------------|-------------|------------------------|-------------|
| BradFx 30<br>(Standard)                  | 64                 | 16MB        | 6                      | —           |
| BradFx 50<br>(Ultra)                     | 96                 | 24MB        | 8 (all)                | —           |
| BradFx 70<br>(Elite)                     | 128                | 32MB        | 8 + temporal           | Full        |
| BradFx 90<br>(Max-i)                     | 192                | 48MB        | 8 + frame gen          | Full + FG   |
| BradGfx 100<br>(Lite)                    | 128                | 256MB       | 8 + FG + temporal      | Full        |
| BradGfx 300<br>(Standard)                | 256                | 512MB       | 8 + variable precision | Full        |
| BradGfx 500<br>(Ultra)                   | 384                | 1GB         | 8 + NeRF               | Full        |
| BradGfx 700<br>(Elite)                   | 448                | 1.5GB       | Full stack             | Full        |
| <b>BradGfx 192<br/>Max</b>               | <b>192</b>         | <b>2GB</b>  | <b>Full + MoE</b>      | <b>Full</b> |
| <b>BradGfx 512<br/>Max<br/>Xntensive</b> | <b>512</b>         | <b>2GB</b>  | <b>Full + MoE</b>      | <b>Full</b> |

**Quality Modes:** - **Quality:** 1440p→4K, 60% Neuro-Stream load - **Balanced:** 1080p→4K, 40% load (default) - **Performance:** 720p→4K, 25% load (competitive MP) - **Ultra Performance:** 540p→4K, 15% load (path tracing) - **Native AA:** 4K→4K, AI anti-alias only, 10% load

**Frame Generation:** 1× (double) or 2× (triple). No input lag penalty.

**Software reference status (receipts, not projections):** the pipeline above is the design intent. What runs today is the **M0 reference upscaler** in the bradsense driver — bradsense\_upscale\_rgba() (bilinear + Catmull-Rom bicubic, arbitrary ratio, luma-adaptive sharpen with an anti-ringing clamp), fed by bradsense\_supply\_frame() into bradsense\_reconstruct(), which scales to the quality mode's target and returns -ENODATA rather than inventing pixels. Receipt: src/tests/test\_bradsense\_pixels.c, 15/15 ctest green under the CI configuration (-DBRAD\_BUILD\_BVLIB=ON; 12 in the default local build), 2× super-resolution PSNR 39.43 dB (bilinear) / 39.44 (bicubic) / 39.57 (bicubic + sharpen). Temporal accumulation (M1) is queued, the metrics corpus and golden CI gate are M2, BradVector kernel mapping is M3, and the trained per-object experts are a separate M4 track. No frame generation, expert inference, or throughput number is claimed as shipped. Rungs and gates: Brad Silicon/GPU/BRADSENSE\_RENDERER\_PLAN.md.

## 2.6 BradFusion Fabric Integration

All BradGfx and BradFx GPUs are direct native peers on the Fabric: - **Zero-copy access to SPMP:** No separate VRAM. GPU reads textures/meshes/shaders directly from the unified LPDDR pool. -

**CPU-GPU-NPU coherency:** Ultra-low-latency heterogeneous compute communication. - **Fabric bandwidth reservation:** 64 GB/s reserved for semantic data path (non-blocking, separate from pixel data).

## 2.7 BradGfx Max Dual Role

Both: discrete card (for BradOS Compute workstations) AND integrated into BradXon via ANF (for bleeding-edge flagship systems). Same silicon, different packaging.

### 2.7a BradGfx Max Xtensive (Hybrid-UMA Edition)

Discrete-only edition aimed at Windows Pros, workstations, and AI/MoE training. Dedicated **32 GB GDDR7** VRAM on a **192-512-bit bus** (0.75-2.0 TB/s) — acts as its own VRAM tank on Windows, and as a high-speed cache with SPMP overflow on BradOS (~100 GB+ effective). SF4X 4nm-class (Brad-Gfx Max tier), 225W nominal, ~98 TFLOPS FP32 / ~786 TFLOPS FP16 AI.

## 2.8 BradSec Display Pipeline [Gen1]

| Feature                   | Description                                                    |
|---------------------------|----------------------------------------------------------------|
| Secure framebuffer        | Encrypted render target, decrypted only at display output      |
| Pixel-level encryption    | HDCP 2.3+ with hardware root of trust                          |
| Asset protection          | Textures/geometry stream encrypted from SPMP, decrypted on-GPU |
| Content tamper prevention | Integrity checks on display pipeline                           |

# PART 3: NPU ARCHITECTURE

## 3.1 BradNPU — Neural Processing Unit [Gen1]

BradNPU is the dedicated AI accelerator powering NeuroSense AI across the entire ecosystem. Purpose-built from the bottom up for neural workloads — matrix multiply, convolution, transformer attention, and mixture-of-experts routing.

| Tier               | Purpose                   | Deployment                                    |
|--------------------|---------------------------|-----------------------------------------------|
| BradCore + BradNPU | Pervasive AI everywhere   | BradChip (Lite→Max), HyperCore (HC-100→Elite) |
| BradXon + X-Units  | Bleeding-edge AI training | BradSkynet, BradAstra, BradOS Compute Pro     |
| Standalone BradNPU | Edge AI acceleration      | BradEco sensors, BradVision ISP co-processor  |

## 3.2 Neural Compute Cluster (NCC) Architecture

Each NCC is a **512-wide SIMD processor**:

| Sub-Unit                   | Width   | Function                                 |
|----------------------------|---------|------------------------------------------|
| Systolic array (INT4/INT8) | 512×512 | Matrix multiply, convolution (inference) |
| SIMD vector (FP16/BF16)    | 32-wide | Activations, layer norm, softmax         |
| SIMD vector (FP32)         | 16-wide | Training, high-precision fallback        |

| Sub-Unit           | Width  | Function                                                 |
|--------------------|--------|----------------------------------------------------------|
| Sparse engine      | —      | 2:4 structured pruning — skip zero weights at wire speed |
| Aggregation buffer | 256 KB | Local reduce/scatter for attention, MoE routing          |

### NCC Cluster Topology:

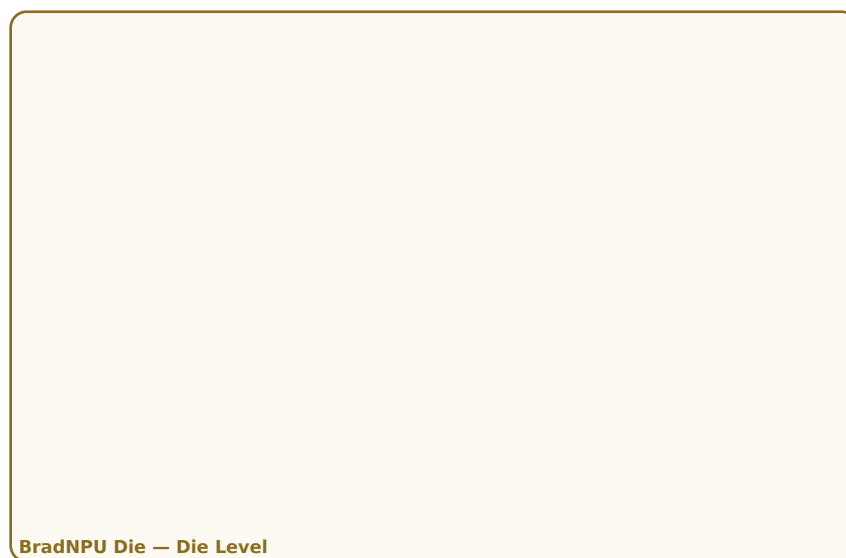


Figure 2: NCC cluster topology — 2×4 NCC mesh feeding HB-AIM, the AI-ISA scheduler, and the fabric

### NCC counts by tier:

| Tier              | NCCs | INT8 TOPS | FP16 TFLOPS | HB-AIM | TDP (NPU only) |
|-------------------|------|-----------|-------------|--------|----------------|
| BradChip Lite     | 16   | 32        | 8           | 16 MB  | 1.5 W          |
| BradChip Standard | 24   | 48        | 12          | 32 MB  | 2.5 W          |
| BradChip Ultra    | 32   | 64        | 16          | 64 MB  | 4 W            |
| BradChip Elite    | 48   | 96        | 24          | 128 MB | 6 W            |
| BradChip Max      | 64   | 128       | 32          | 256 MB | 9 W            |
| HyperCore HC-300  | 8    | 16        | 4           | 8 MB   | 0.8 W          |
| HyperCore Elite   | 24   | 48        | 12          | 32 MB  | 2.5 W          |

### 3.3 HB-AIM (High-Bandwidth AI Memory)

On-die SRAM dedicated to model weights and activations: - **Type:** Custom 6T SRAM (high-density, low-voltage) - **Bandwidth:** 2 TB/s per 16 MB tile (512-bit @ 2 GHz) - **Tiling:** 8×8 MB tiles per 64 MB region, independently powered - **Retention:** Near-threshold mode (0.4 V) — 5 μW/MB standby - **MoE spooler:** When experts exceed HB-AIM, cold experts spilled to SPMP via Fabric. MoE Router prefetches next N experts based on attention pattern prediction.

### 3.4 AI-ISA (Dedicated Instruction Set)

Custom single-cycle instructions on NCC hardware:

| Category      | Instructions                                            |
|---------------|---------------------------------------------------------|
| Matrix        | MATMUL, CONV, BATCHED_GEMM                              |
| Activation    | RELU, GELU, SILU, TANH, SIGMOID                         |
| Normalization | LAYER_NORM, BATCH_NORM, RMS_NORM (4 cycles fused)       |
| Attention     | QKV_PROJ, ATTN_MASK, SOFTMAX, ATTN_OUT (6 cycles fused) |
| MoE           | TOP_K_ROUTER, EXPERT_COMBINE, SPARSE_GATHER             |
| Memory        | HB_LOAD, HB_STORE, HB_SCATTER, HB_GATHER                |
| Sparse        | PRUNE_MASK, SPARSE_MATMUL, DENSE_RECOMPOSE              |
| Quant         | FP16_TO_INT8, INT8_TO_FP16, FP16_TO_FP4                 |

**Compiler target:** PyTorch/TensorFlow models graph-rewritten to AI-ISA.

### 3.5 MoE (Mixture of Experts) Engine

- **Top-2 routing:** Each token selects 2 of N experts
- **Router capacity:** 1024 experts per NCC cluster, 8192 per die
- **Expert size:** 1M–500M parameters (configurable)
- **Prefetch pipeline:** Router runs N tokens ahead of compute — expert prefetch overlaps with current inference, hiding latency

### 3.6 Workload Scheduler

| Priority | Workload                             | Latency Budget | NCC Allocation       |
|----------|--------------------------------------|----------------|----------------------|
| Critical | FaceLock, voice auth                 | <10ms          | 4 NCCs dedicated     |
| High     | BradVision Fusion, real-time AI      | <50ms          | 8-16 NCCs            |
| Medium   | TIFA inference, BradCoach            | <200ms         | Remaining NCCs       |
| Low      | Background indexing, predictive text | >500ms         | 1-4 NCCs (low power) |
| Bulk     | Offline training                     | Best effort    | All idle NCCs        |

### 3.7 Use Case Performance

| Use Case                     | Model               | NCCs | Latency           |
|------------------------------|---------------------|------|-------------------|
| Face unlock                  | MobileFaceNet       | 2    | 3ms               |
| Real-time depth (BradVision) | MiDaS v3 small      | 4    | 8ms @ 60fps       |
| Speech-to-text               | Whisper tiny        | 4    | 12ms per 5s audio |
| TIFA chat (7B model)         | LLaMA-7B 4-bit      | 48   | 25 tokens/ms      |
| SDXL image gen               | Stable Diffusion XL | 64   | 2s per 1024×1024  |

### 3.8 Fabrication & Power

| Parameter | Specification                                            |
|-----------|----------------------------------------------------------|
| Process   | 7LPP (Gen1 budget), SF4/SF4P (Gen1)                      |
| Frequency | 1.2 GHz (turbo), 400 MHz (efficient), 50 MHz (always-on) |

| Parameter        | Specification                            |
|------------------|------------------------------------------|
| Power states     | P0-P5 + OFF with 5µs wake latency        |
| Voltage          | 0.4 V (retention) - 0.9 V (turbo)        |
| Always-on domain | 2 MB SRAM + lightweight scheduler — 2 mW |

## PART 4: MEMORY & STORAGE

### 4.1 Memory Architecture: BradRAM + SPMP [Gen1]

**Two-tier unified architecture:** - **BradRAM** = the brand for our memory subsystem: standard JEDEC LPDDR5X/DDR5 dies + a Brad-designed controller (DDS predictive prefetch, L4 SRAM system-level cache, inline encryption, EROE telemetry) - **SPMP** = logical unified pool formed by the BradRAM subsystem + NeuroSense AI controller + Fabric integration

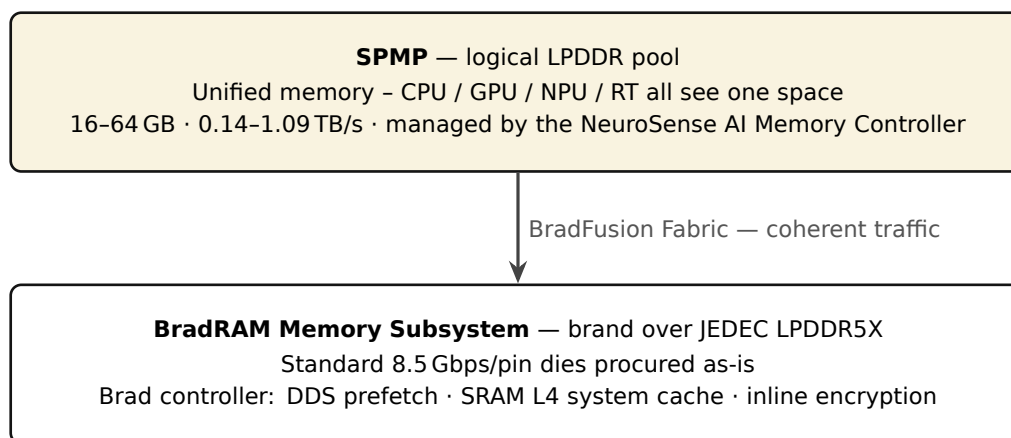


Figure 3: BradRAM + SPMP — two-tier unified memory: logical pool above the subsystem, joined by the fabric

### BradRAM Subsystem Specs

| Parameter         | Value                                                         |
|-------------------|---------------------------------------------------------------|
| Base standard     | JEDEC LPDDR5X-8533 (JESD209-5B), off-the-shelf dies           |
| Pin speed         | 8.5 Gbps per pin (Gen1) / 14.4 Gbps (Gen2 LPDDR6)             |
| Bus width         | 16-bit per die (128-1024-bit system bus)                      |
| Bandwidth per die | 17 GB/s (LPDDR5X) / 38.4 GB/s (LPDDR6)                        |
| Capacity per die  | 8 Gb (1 GB) - 32 Gb (4 GB)                                    |
| Stack height      | 4-8 dies per stack (wide-bus UMA)                             |
| Stack capacity    | 4 GB - 32 GB per stack                                        |
| Voltage           | VDD 1.05 V (active), VDDQ 0.5 V, 0.6 V (retention)            |
| Process           | Industry DRAM process (dies) + Brad memory-controller die     |
| Brad content      | Controller + L4 SRAM + encryption — <b>no custom DRAM die</b> |

**Dynamic Data-Sensing (DDS)** DDS is implemented on the **Brad memory-controller die** (0.5 mm<sup>2</sup>, ~50 µW), not inside the DRAM (standard JEDEC parts cannot host custom logic): - Monitors access patterns from CPU, GPU, NPU - Learns per-application access signatures - Proactively prefetches from BradDrive into L4 SRAM cache - Communicates prefetch hints to Drive controller via Fabric

**Effectiveness:** 85% hit rate for predicted workloads. Reduces SPMP demand by 30%.

### Memory Hierarchy

| Level | Type                         | Size per Stack | Access                | Location                      |
|-------|------------------------------|----------------|-----------------------|-------------------------------|
| L4    | SRAM<br>(system-level cache) | 64 MB          | 2 ns                  | On Brad memory-controller die |
| L3    | LPDDR<br>(fast)              | 25% of stack   | 10 ns                 | TSV-connected chiplets        |
| L2    | LPDDR<br>(main)              | 75% of stack   | 20 ns                 | Stacked dies                  |
| SPMP  | Logical pool                 | 16-64 GB       | 20-40 ns (via Fabric) | Aggregated                    |

The L4 SRAM is an Apple-SLC-style system-level cache on the controller die; the LPDDR tiers below it are standard JEDEC dies.

### SPMP Configurations

| Tier              | Bus Width | Capacity  | Bandwidth (LPDDR5X) | Bandwidth (LPDDR6) | Fabric Paths         | Used In                         |
|-------------------|-----------|-----------|---------------------|--------------------|----------------------|---------------------------------|
| <b>Lite</b>       | 128-bit   | 8-16 GB   | 136 GB/s            | 230 GB/s           | 2× 64-bit            | BradMatrix Essential            |
| <b>Standard</b>   | 256-bit   | 16-32 GB  | 272 GB/s            | 461 GB/s           | 4× 64-bit            | BradMatrix Base, BradSlate Base |
| <b>Pro</b>        | 384-bit   | 32-48 GB  | 409 GB/s            | 691 GB/s           | 6× 64-bit            | BradMatrix Prime                |
| <b>Max</b>        | 512-bit   | 48-64 GB  | 545 GB/s            | 922 GB/s           | 8× 64-bit            | BradOS Compute Pro              |
| <b>Enterprise</b> | 624-bit   | 64-128 GB | 1.09 TB/s           | 1.84 TB/s          | 16× 64-bit + optical | BradSkynet server               |

**NeuroSense AI Memory Controller** Dedicated RISC-V core cluster + neural accelerator between BradRAM and Fabric: - **Access prediction:** LSTM-based sequence model per active process (+5ns pipelined) - **Data placement:** Hot → L4 SRAM, warm → L3 fast LPDDR, cold → L2 main - **QoS arbitration:** Weighted fair queuing with urgency scoring - **Encryption:** AES-256-GCM inline (+3ns per 64B line) - **Integrity:** Merkle tree (SHA-3-256) verification concurrent with read

**Unified access model:**

Processor Read Request → BradFusion Fabric (1–4 cycles)  
 → SPMP Controller  
   ├─ L4 SRAM hit → return (2 ns)  
   ├─ L3 fast LPDDR hit → return (10 ns)  
   ├─ L2 main LPDDR hit → return (20 ns)  
   └─ Miss → Fabric → BradDrive HDB → decompress → SPMP write → return (<1ms)

**Zero-copy principle:** Data written once. CPU reads pixel buffer, GPU renders from same address, NPU infers from same. No copies.

**BradSec-RAM [Gen1]** Hardware-level memory encryption in the **Brad memory controller data path** (between CPU/GPU/NPU and the standard JEDEC dies): - AES-256-GCM per memory channel, per-session keys from PUF root-of-trust - Per 64-byte cache line with unique IV - Merkle tree integrity verification - +3ns per access, +5% power (always-on, no bypass) - Key scrub on chassis-open detect

### EROE Thermal Integration

- Per-die thermal sensors (0.1°C resolution) → BradOS EROE reads via Fabric
- Adaptive DRAM refresh rate (2× at 85°C+, 0.5× at <40°C)
- Hot page migration to cooler dies
- Bandwidth capping per application
- Sustain: 95% peak bandwidth indefinitely under 85°C

## 4.2 Storage Architecture: BradDrive + BradCompress [Gen1]

### Data Pipeline:

Application → File Request  
 ↓  
 BradDrive Controller (BDC)  
   ├─ NeuroSense-Storage AI predicts asset need  
   ├─ Read from BradNAND (compressed)  
   └─ Hardware Decompression Block (HDB)  
     ↓ Decompressed (zero-CPU)  
 SPMP (unified memory pool) ← direct DMA, no RAM staging  
 ↓  
 Any processor: CPU / GPU / NPU / RT

**Key differentiator:** CPU never touches compressed or decompressed data. Entire path from NAND to SPMP is hardware-accelerated.

### BradDrive Physical Drive

| Model                | Form Factor     | Capacity    | NAND Type | Seq. Read | Seq. Read (compressed) | Target                           |
|----------------------|-----------------|-------------|-----------|-----------|------------------------|----------------------------------|
| <b>BradDrive</b>     | Brad-M.3 (2230) | 512 GB-2 TB | TLC       | 14 GB/s   | 28 GB/s (2:1)          | BradMatrix Base, BradS-late Base |
| <b>BradDrive Pro</b> | Brad-M.3 (2280) | 1 TB-4 TB   | MLC       | 16 GB/s   | 40 GB/s (2.5:1)        | BradMatrix Prime                 |

| Model                       | Form Factor     | Capacity   | NAND Type      | Seq. Read | Seq. Read (compressed) | Target                                          |
|-----------------------------|-----------------|------------|----------------|-----------|------------------------|-------------------------------------------------|
| <b>BradDrive Elite</b>      | Brad-M.3 (2580) | 2 TB–8 TB  | SLC            | 18 GB/s   | 54 GB/s (3:1)          | Workstation (console uses 1/2 TB TLC, see §8.1) |
| <b>BradDrive Enterprise</b> | EDSFF E3        | 8 TB–32 TB | SLC/MLC hybrid | 28 GB/s   | 84 GB/s (3:1)          | BradSkynet server                               |

**Interface:** PCIe 6.0 x4 (64 GB/s uni, 128 GB/s bi). Brad-M.3 form factor (12mm wide, keyed for BradDrive only).

**BradDrive Controller (BDC)** Multi-core RISC-V processor + dedicated accelerators: - **RISC-V core cluster:** 4× in-order @ 1 GHz — command processing, FTL, scheduling - **BradCompress HDB:** Fixed-function decompression — 64 GB/s throughput - **NAND controller:** 16 channels @ 2400 MT/s per channel — ONFI 5.1 - **NeuroSense AI:** Tiny neural net (0.5 TOPS) — learns access patterns, predicts pre-fetch - **DRAM cache:** 2 GB LPDDR5X for hot data + FTL mapping table - **PCIe 6.0 PHY:** x4 lanes @ 64 GT/s PAM4 - **BradSec AES-256:** Inline encrypt-on-write, decrypt-on-read - **ECC:** LDPC + RAID-like cross-die parity — corrects up to 2Kbit errors/sector

### BradNAND [Gen1]

| Type       | Bits/Cell | Program | Read  | Endurance | Use                                 |
|------------|-----------|---------|-------|-----------|-------------------------------------|
| <b>SLC</b> | 1         | 20 μs   | 15 μs | 100K P/E  | Elite, Pro (Brad Soft-ware/OS/data) |
| <b>MLC</b> | 2         | 40 μs   | 20 μs | 30K P/E   | Pro (bulk)                          |
| <b>TLC</b> | 3         | 80 μs   | 25 μs | 5K P/E    | Mass-market                         |
| <b>QLC</b> | 4         | 150 μs  | 35 μs | 1K P/E    | Cold storage                        |

**3D layer count:** 238 layers (Gen1), 350 layers (Gen2 — Vision).

**NeuroSense-Storage AI** On-controller neural network predicting data access: - Input: File system metadata, application context, user behavior, time-of-day - Output: Probability scores for next-N-second block access - Pre-fetch: Predicted blocks loaded from NAND → HDB decompression → SPMP before request - Rise of the Streets example: Learns map coordinate (x,y) → triggers zone texture load → preloads 500 MB before player arrives

### BradCompress Compression Codec Three-stage lossless compression:

Input Data Block

- ↓ Stage 1: Dictionary Coding (64 KB sliding window, domain-tuned)
- ↓ Stage 2: Run-Length Encoding (collapses repeated tokens, adaptive threshold 3+)
- ↓ Stage 3: Entropy Coding (range encoding, 8× 2D context models)

Output: bitstream

**Domain Profiles:**

| Profile  | Avg Ratio | Decomp Speed | Use                           |
|----------|-----------|--------------|-------------------------------|
| Mesh     | 3.5:1     | 56 GB/s      | 3D vertex data, index buffers |
| Texture  | 2.8:1     | 62 GB/s      | BCn textures, raw RGBA        |
| Audio    | 2.2:1     | 48 GB/s      | WAV, raw PCM, synths          |
| Video    | 2.5:1     | 45 GB/s      | Intra-frames, motion vectors  |
| AI Model | 2.0:1     | 64 GB/s      | FP16/INT8 weights             |
| General  | 1.8:1     | 35 GB/s      | Binary blobs, archives        |
| Mixed    | 2.5:1 avg | 52 GB/s      | Asset bundles (game levels)   |

**Hardware Decompression Block (HDB)**

| Parameter        | Spec                                               |
|------------------|----------------------------------------------------|
| Architecture     | 256-wide SIMD pipeline (custom, not GPU)           |
| Clock            | 1.2 GHz                                            |
| Throughput       | 64 GB/s peak, 40 GB/s sustained mixed              |
| Latency          | 128 cycles (first byte), pipelined                 |
| L0 cache         | 256 KB — dictionary tables + recently decoded data |
| Fabric interface | Dedicated SPMP DMA engine — 64 GB/s direct write   |

**Direct-to-SPMP zero-copy API:**

```
void* addr = bradDirectLoad("/game/city_block_05.bcm", SPMP_TIER_FAST);
// No malloc, no memcpy, no CPU-side decompression
```

**Security**

- **Data-at-rest:** AES-256-XTS — encrypts on NAND write, decrypts on read
- **Data-in-transit:** AES-256-GCM — Fabric packets encrypted between HDB and SPMP
- **Anti-rollback:** Monotonic counter in BradSec prevents firmware downgrade
- **Self-encrypting drive:** TCG Opal 2.0 compliant

**PART 5: VISION & CAMERA****5.1 BradVision — Complete Camera & Vision System [Gen1]**

BradVision is the perceptual nervous system — a vertically integrated stack spanning sensor optics, hardware ISP, AI computational photography, display integration, and security sensing.

Light → Optics → Sensor → ISP Hardware → BradNPU AI → SPMP → Display

**5.2 Sensor Technology**

Custom Quad-Bayer CMOS sensors co-engineered with BradLabs:

| Spec        | Lite                 | Standard          | Pro               | Elite                   |
|-------------|----------------------|-------------------|-------------------|-------------------------|
| Resolution  | 48MP (12MP eff.)     | 108MP (27MP eff.) | 200MP (50MP eff.) | 200MP + 48MP dual-layer |
| Pixel size  | 1.0µm (2.0µm binned) | 0.8µm (1.6µm)     | 0.64µm (1.28µm)   | 0.64µm + 1.0µm          |
| Sensor size | 1/2"                 | 1/1.4"            | 1/1.12"           | 1" type                 |

| Spec          | Lite                    | Standard        | Pro                  | Elite            |
|---------------|-------------------------|-----------------|----------------------|------------------|
| HDR           | Staggered               | Staggered + DCG | DCG + Multi-Sampling | Dual-gain + QBC  |
| Readout       | 12-bit                  | 14-bit          | 16-bit               | 16-bit dual-path |
| FPS           | 30fps full / 60fps 12MP | 30fps / 120fps  | 24fps / 180fps       | 30fps / 240fps   |
| Dynamic range | 76dB                    | 82dB            | 88dB                 | 96dB             |

**Pixel technologies:** Dual Pixel PDAF (100%), Smart-ISO Pro (triple-gain), Staggered HDR, DCG HDR, QBC Re-mosaic (AI-trained), adaptive pixel-binning, dual-layer pixel (Elite only).

### 5.3 ISP Pipeline (14-Stage Hardware)

Sensor RAW → Black Level → Linearization → Defective Pixel → Crosstalk Comp  
 → White Balance → AI Demosaic → Color Correction → Gamma → Tone Mapping  
 → Edge Enhancement → Chroma Denoise → Lens Shading → Distortion Correct  
 → Output (YUV/RGB/RAW) → SPMP

**AI-Assisted Demosaic:** Lightweight CNN trained on 10M+ RAW patches — 2× resolution recovery vs traditional binning, 80% false color reduction at high-ISO, <2ms per 48MP frame (dedicated CNN accelerator).

**ISP Hardware Specs:** 7nm dedicated ISP dielet, 200MP @ 30fps, 12 GPixel/s, 16 pixels/clock @ 800MHz, 4 TOPS AI accelerator, 8MB on-die SRAM, Fabric native (64 GB/s to SPMP), up to 4 concurrent streams.

### 5.4 Optics

| Tier            | Wide (Main) | Ultrawide  | Telephoto                                                          | ToF/Depth       |
|-----------------|-------------|------------|--------------------------------------------------------------------|-----------------|
| <b>Lite</b>     | 26mm f/1.8  | 13mm f/2.2 | —                                                                  | —               |
| <b>Standard</b> | 24mm f/1.7  | 12mm f/2.0 | 52mm f/2.4 (2×)                                                    | VCSEL ToF       |
| <b>Pro</b>      | 23mm f/1.5  | 12mm f/1.8 | 70mm f/2.0 (3×) + 120mm f/2.8 (5×)                                 | dToF            |
| <b>Elite</b>    | 22mm f/1.4  | 10mm f/1.8 | 85mm f/1.8 (3.5×) + 170mm f/2.8 (7×) + 280mm f/3.5 (11× periscope) | dToF + IR flood |

**Technologies:** 9-layer nano-coated AR, hybrid glass-polymer elements, BradSense piezoelectric micro-actuators (10µs response, 0.1µm precision), sensor-shift OIS (Elite, 5-axis, ±3°), liquid lens (Pro/Elite).

**BradClarity UDC Display [Gen2]:** Under-display camera with lower pixel density (200 PPI vs 460 PPI), transparent cathode (gallium-doped zinc oxide), AI compensation for diffraction pattern. Target: 32MP UDC indistinguishable from surrounding display.

### 5.5 BradVision Fusion Engine (Software on BradNPU)

| Mode          | Frames              | Purpose             | Compute      |
|---------------|---------------------|---------------------|--------------|
| Smart HDR     | 3–7 staggered EV    | Extended DR         | 2ms (16 NCC) |
| Night Fusion  | 8–15 short+long     | Low-light recovery  | 5ms (32 NCC) |
| Action Fusion | 4 ultra-short burst | Motion freeze + HDR | 3ms (24 NCC) |

| Mode             | Frames            | Purpose               | Compute       |
|------------------|-------------------|-----------------------|---------------|
| Portrait Fusion  | 2 wide+tele       | Depth-aware isolation | 4ms           |
| Super-Res Fusion | 4-9 micro-shifted | 2×-3× resolution      | 8ms (64 NCC)  |
| Focus Fusion     | 3-7 focus stack   | Extended DoF (macro)  | 6ms (48 NCC)  |
| Multi-Cam Fusion | All cameras       | Unified optical zoom  | 10ms (64 NCC) |

**NeuroSense AI Weighting:** Face priority (3× weight in portraits), edge detection, motion detection, texture analysis.

**Continuum Flow:** Capture session state preserves across Brad devices — start on phone, continue on slate, edit on BradMatrix Prime.

## 5.6 Video Capabilities

| Tier     | Max Resolution | Max FPS | Codecs                                          |
|----------|----------------|---------|-------------------------------------------------|
| Lite     | 8K             | 30fps   | H.265, AV1                                      |
| Standard | 8K             | 60fps   | H.265, AV1, BradCompress                        |
| Pro      | 8K             | 120fps  | H.265, AV1, BradCompress, ProRes RAW            |
| Elite    | 8K             | 240fps  | H.265, AV1, BradCompress, ProRes RAW, CinemaDNG |

**Features:** Cinematic Mode (real-time DoF racking, adjustable after capture), Action Mode (4K120 with BradSense OIS + horizon lock), Night Video (temporal fusion 3-frame sliding window), Super Slow-Mo (720p960 / 1080p960, AI interpolation to 1920fps), BradCompress Video (~30% smaller than AV1, hardware decode in BradVision Engine).

## 5.7 Vision-as-Sensor

**BradSec FaceLock:** VCSEL + DOE (30,000 dots), 940nm IR flood, dedicated 2MP IR sensor → direct BradNPU, 3D face mesh, sub-100ms unlock, spoof detection (liveness + micro-expression + pulse on Elite).

**BradVerify:** Multimodal biometrics (face + iris + fingerprint), liveness challenge (random micro-expressions), on-device only.

**XR / AR Interface:** 6-DoF head tracking (1000Hz IMU + 60Hz camera), hand tracking (ultrawide 120° FOV, 60fps, 21-point skeleton), room mapping (ToF + UW real-time mesh via BradNPU), plane detection (<10ms per plane).

## 5.8 Modular Bayonet Optics [Vision]

Hot-swappable lens modules via magnetic bayonet mount: Portrait 50mm f/1.4 (28g), Macro 90mm f/2.8 (35g), Telephoto 200mm f/3.5 (65g), Wide 12mm f/2.0 (32g), Anamorphic 35mm f/1.8 (45g). Standardized MIPI connector + calibration EEPROM per module.

## PART 6: PLATFORM & MOTHERBOARDS

### 6.1 SoC Platforms

**BradMind — The Heart (Efficiency) [Gen1]** **Philosophy:** “Smart, Cool, Enduring.” Optimized for performance-per-watt and silent operation.

| Feature    | Lite                           | Standard             | Pro                                   | Elite                                 |
|------------|--------------------------------|----------------------|---------------------------------------|---------------------------------------|
| CPU        | P1+E4                          | P2+E4                | P4+E4                                 | P6+E4                                 |
| GPU        | BradFx 10 (8 VE)               | BradFx 30 (16 VE)    | BradFx 50 (24 VE)                     | BradFx 70 (32 VE)                     |
| NPU        | 256 NCC                        | 384 NCC              | 512 NCC                               | 768 NCC                               |
| Memory I/F | LPDDR5X                        | LPDDR6 / Brad-CAMM   | Brad-CAMM DDR                         | Brad-CAMM LPDDR6                      |
| Process    | 7LPP (7nm)                     | SF4 (4nm)            | SF4P (4nm)                            | SF4X (4nm)                            |
| TDP        | 2-5W                           | 5-10W                | 10-25W                                | 25-45W                                |
| Target     | BradPhone Essential, wearables | BradPhone, BradSlate | BradSlate Prime, BradMatrix Essential | BradPhone Ultra, BradMatrix Essential |

**BradNeuro — The Brain (Performance) [Gen1-Gen2]** **Philosophy:** “See, Sense, Simulate.” Sensory fusion at silicon level.

| Feature    | Ultra                             | Elite                   | X                                                          |
|------------|-----------------------------------|-------------------------|------------------------------------------------------------|
| CPU        | P4+E8                             | P8+E4                   | BradXon (ANF)                                              |
| GPU        | BradFx 90 (48 VE)                 | BradFx 90 + BradGfx 100 | BradGfx Max (ANF-integrated)                               |
| NPU        | 768 NCC + HB-AIM                  | 1024 NCC + HB-AIM       | X-Units (AIMP)                                             |
| Fabric     | Fabric + optical [Gen2]           | Fabric + optical        | Quantum-entangled [Vision]                                 |
| Memory I/F | Brad-CAMM LPDDR6                  | Brad-CAMM LPDDR6        | SPMP (LPDDR UMA)                                           |
| Process    | SF4X (4nm)                        | SF2 (2nm) [Gen2]        | 1.4nm→sub-1nm                                              |
| TDP        | 45-65W                            | 65-125W                 | 125-350W+                                                  |
| Target     | BradMatrix Prime, BradSlate Ultra | BradOS Compute WS       | BradAstra headset (XR SKU), BradAstra servers, Brad-Skynet |

### 6.2 BradBoard Motherboard Series [Gen1]

**BradBoard Mobile — Ultra-Compact** Embedded PCB, no user-accessible slots, SoC + memory soldered.

| Variant         | Size    | Target                           | SoC                    | Memory        |
|-----------------|---------|----------------------------------|------------------------|---------------|
| Mobile Nano     | 35×35mm | BradPhone Essential, wearables   | Mind Lite              | LPDDR5X       |
| Mobile Standard | 45×45mm | BradPhone Base, BradSlate Lite   | Mind Standard          | LPDDR6        |
| Mobile Pro      | 55×55mm | BradSlate, BradMatrix Essential  | Mind Pro               | Brad-CAMM DDR |
| Mobile Elite    | 65×65mm | BradPhone Ultra, BradSlate Prime | Mind Elite / BradNeuro | Brad-CAMM     |

**Construction:** 8-12 layer PCB with embedded micro-fluidic channels, Direct-Path Trace Geometry (SoC-to-memory in microns), BradCool passive thermal, BradSec root-of-trust fused at fabrication.

**BradBoard Matrix — Modular Plate** Modular SoC socket + Brad-CAMM + Brad-M.3 slots for user-upgradable memory and storage.

| Variant       | Size      | SoC            | CAMM   | M.3 | Target               |
|---------------|-----------|----------------|--------|-----|----------------------|
| Matrix Air    | 150×100mm | Mind Pro/Elite | 1× DDR | 1×  | BradMatrix Essential |
| Matrix Pro    | 180×120mm | Neuro Ultra    | 2× DDR | 2×  | BradMatrix Prime 14" |
| Matrix Studio | 200×140mm | Neuro Elite    | 2× DDR | 2×  | BradMatrix Ultra 16" |

**Key technologies:** SoC Module Socket (gold-plated compression pins, hot-swappable), Brad-CAMM flat screw-mounted modules, Brad-M.3 keyed slots, BradLightSpeed optical lane [Gen2], BradCool vapor chamber integrated into PCB stack.

**BradBoard Nexus — E-ATX / Desktop** Full-length expansion, dual-Xon support, complete Fabric implementation.

| Variant       | Size               | SoC                        | CAMM     | M.3 | Target             |
|---------------|--------------------|----------------------------|----------|-----|--------------------|
| Nexus Pro     | 305×244mm          | Neuro Elite + 1× Xon       | 4× DDR   | 4×  | BradOS Compute     |
| Nexus Elite   | 305×267mm          | 2× Xon                     | 8× DDR   | 6×  | BradOS Compute Pro |
| Nexus Console | 280×200mm (custom) | BradCore (8P+8E) + Gfx Max | 2× GDDR7 | 2×  | BradAxis           |

**Key technologies:** Dual SoC Socket [Gen1], SPMP Controller Hub, full-length BradGfx slot (x16 PCIe 6.0 + ANF direct-lane), BradLightSpeed Fabric Backplane [Gen2] (embedded optical waveguides, 12+ layer PCB), BradCool X-Liquid ports [Gen2], BradSec-X HSM soldered on-board.

**BradBoard Rack — Data Center** Composable compute sleds for BradSkynet and BradAstra servers.

| Variant     | Factor  | CPU's  | CAMM    | M.3 | GPU's         | Target           |
|-------------|---------|--------|---------|-----|---------------|------------------|
| Rack Sled   | 1U half | 1× Xon | 2× DDR  | 2×  | —             | BradSkynet node  |
| Rack Quad   | 2U full | 4× Xon | 8× DDR  | 4×  | 4× Gfx Max    | BradAstra server |
| Rack Matrix | 4U      | 8× Xon | 16× DDR | 8×  | N+1 redundant | Core cluster     |

### 6.3 Brad-CAMM Memory Slots

| Feature     | Brad-CAMM DDR             | Brad-CAMM LPDDR6         |
|-------------|---------------------------|--------------------------|
| Pin count   | 384                       | 512                      |
| Data rate   | 8.8 Gbps (DDR6)           | 14.4 Gbps (LPDDR6)       |
| Bus width   | 64-bit per module         | 128-bit per module       |
| Form factor | Flat plate, screw-mounted | Flat plate + backplate   |
| Thermal     | Integrated heat spreader  | Integrated heat spreader |

### 6.4 Brad-M.3 Storage Slots

Custom 12mm-wide (vs M.2's 22mm), keyed for BradDrive only. Lite (x2, 32 GB/s, 42mm) and Pro (x4, 64 GB/s, 60mm).

### 6.5 BradLink Device Tethering [Gen2]

- **BradLink Local:** Wireless fabric extension (60 GHz, sub-mm wave) — phone offloads AI to nearby BradMatrix
- **BradLink Fabric Bridge:** Wired — two boards act as single shared-memory system
- **Latency:** <1μs local, <100μs tethered

### 6.6 Target Device → Platform Mapping

| Device               | SoC Platform  | Motherboard     | CPU   | GPU                  |
|----------------------|---------------|-----------------|-------|----------------------|
| BradPhone Essential  | Mind Lite     | Mobile Nano     | P1+E4 | BradFx 10            |
| BradPhone Base       | Mind Standard | Mobile Standard | P2+E4 | BradFx 30            |
| BradPhone Ultra      | Mind Elite    | Mobile Elite    | P6+E4 | BradFx 70            |
| BradSlate Base       | Mind Standard | Mobile Standard | P2+E4 | BradFx 30            |
| BradSlate Prime      | Mind Elite    | Mobile Elite    | P6+E4 | BradFx 70            |
| BradSlate Ultra      | Neuro Ultra   | Mobile Elite    | P8+E4 | BradFx 90            |
| BradMatrix Essential | Mind Pro      | Matrix Air      | P4+E4 | BradFx 50            |
| BradMatrix Prime 14" | Neuro Ultra   | Matrix Pro      | P8+E4 | BradFx 90            |
| BradMatrix Ultra 16" | Neuro Elite   | Matrix Studio   | P8+E8 | BradFx 90 + Gfx 500M |

| Device             | SoC Platform       | Motherboard         | CPU                  | GPU                    |
|--------------------|--------------------|---------------------|----------------------|------------------------|
| BradOS Compute     | Neuro Elite        | Nexus Pro           | P8+E8                | BradGfx 300            |
| BradOS Compute Pro | BradXon            | Nexus Elite         | BradXon ×2           | BradGfx Max            |
| BradCube Base      | BradChip Max       | Matrix Studio-class | P8+E8                | BradFx 90 (Max-i)      |
| BradCube Prime     | BradChip Max (bin) | Matrix Studio-class | P8+E8                | BradFx 90 + eGPU lane  |
| BradCube X Prime   | BradXon (single)   | Studio X-class      | BradXon              | BradGfx                |
| BradCube X Ultra   | BradXon (1-2)      | Studio X-class      | BradXon ×1-2         | BradGfx Max            |
| BradAxis Console   | BradCore           | Nexus Console       | BradCore 8P+8E       | BradGfx Max (fabric)   |
| BradAstra Headset  | Neuro X (XR)       | Mobile Elite        | P8+E8                | BradFx 90 (integrated) |
| BradAstra Server   | BradXon            | Rack Quad           | BradXon ×4           | BradGfx Max ×4         |
| BradSkynet Node    | BradApex           | Rack Sled           | BradXon (controller) | — (AIMP focused)       |

## PART 7: SOFTWARE & OS

### 7.1 BradOS Family [Gen1]

| SKU                   | Target                 | Kernel                               | UI                      | Min RAM | Storage |
|-----------------------|------------------------|--------------------------------------|-------------------------|---------|---------|
| <b>BradOS Mobile</b>  | Smartphones            | Hybrid (Zircon + Linux)              | Touch-first, gesture    | 4 GB    | 64 GB   |
| <b>BradOS Slate</b>   | Tablets, foldables     | Hybrid (Zircon + Linux)              | Touch + Pen + Desktop   | 8 GB    | 128 GB  |
| <b>BradOS Compute</b> | Laptops, desktops      | Full Linux (mainline + Brad patches) | Desktop (KDE-based)     | 16 GB   | 256 GB  |
| <b>BradOS Nano</b>    | Wearables, IoT, BradTV | Zircon microkernel (no Linux)        | Card-based, voice-first | 512 MB  | 8 GB    |

**Performance Tiers (Cross-SKU):** - **Nano:** Ultra-light, smooth on 2GB RAM — comparable to Android Go - **Standard:** Balanced hero — comparable to Pixel OS meets Ubuntu - **Pro:** Unlocked, desktop-class multitasking, encrypted storage — comparable to macOS meets Fedora

**Specialized Editions:** - **BradOS Dev:** Compute + preloaded compilers, simulators, SDKs (Brad-Code Studio, BradCloud CLI) - **BradOS XR:** Compute x Nano — mixed reality for AR/VR headsets,

spatial UI, 6-DoF - **BradOS Gaming:** Compute Microkernel — BradAxis console OS, minimal overhead, instant-on - **BradOS Education:** Slate/Nano — lockdown mode, curriculum management, offline-capable

## 7.2 Core OS Features

**NeuroSense AI Integration:** - Predictive UI (app launch prediction based on time/location/context) - Contextual app launching (opening photo → camera shortcut) - Adaptive performance (EROE shifts power based on foreground task) - AI wallpaper (generative, adapts to time/weather/mood) - Smart notifications (priority sorting by learned behavior)

**BradSec Integration:** - Verified boot chain (PUF → bootROM → kernel signature) - Biometric unlock (FaceLock + fingerprint) - Per-file AES-256-GCM with keys in secure enclave - Zircon capability-based app sandboxing - Isolated secure enclave for biometric data

**BradSuite Pre-installed Apps:** BradBrowse (privacy-first WebKit), BradSearch (no profiling), BradMail (BradID + PGP), BradCloud (file sync, Continuum Flow-aware), BradOffice (local-first, BradCompress-native), BradPhotos (AI org, RAW editing), BradCode (terminal + IDE with BradDev Kit integration).

**OTA Updates:** Dual-slot A/B seamless updates, BradCompress-aware binary delta (~200 MB typical), firmware bundle updates (OS + Sec + Drive) in single reboot.

### 7.2.5 Software Support & Update Policy (Committed)

Brad Devices offers **long-lived software support** across every platform — a deliberate counter to planned obsolescence and a natural fit with the 4-year generation cadence.

**OS & Firmware Updates — 7 Years** Every device — **BradPhones, BradSlates, BradMatrix, BradOS Compute desktops, and the BradAxis console** — receives **7 years of OS and feature updates** from the product's launch date. Within those 7 years users get:

- **Major OS releases** (feature + architecture updates)
- **Security patches** (critical/zero-day fixes shipped promptly)
- **Stable, non-breaking updates** (never brick the device, always rollback-capable via A/B)

A 7-year window comfortably spans **~1.75 generations** of the 4-year cadence, so every device is covered for its own generation plus the majority of the next.

| Platform                     | OS + Feature Updates | Notes                                          |
|------------------------------|----------------------|------------------------------------------------|
| BradPhones /<br>BradSlates   | 7 years              | Matches/exceeds the<br>flagship-phone<br>norm  |
| BradMatrix (laptops)         | 7 years              | Laptop-class<br>long-support                   |
| BradOS Compute<br>(desktops) | 7 years              | Desktop<br>long-support                        |
| BradAxis Console             | 7 years              | Console-<br>generation-<br>spanning<br>support |

**CPU & GPU Driver Support — 9 Years (Game-Ready) + 4-Year Legacy** Graphics and compute drivers follow a **two-stage driver policy**:

1. **Game-Ready driver support — 9 years.** After a Brad Silicon CPU/GPU part launches, Brad ships **game-ready drivers for 9 years** — day-0 game optimizations, feature updates, and performance tuning for new titles and workloads.
2. **Legacy status — 4 further years.** After the 9-year game-ready window, the part enters **legacy status** for 4 additional years: security fixes, stability patches, and **bug-fix-only** updates (no new feature/game-day-0 optimization).

Net driver lifetime: **13 years** per Brad silicon part (9 game-ready + 4 legacy).

| Stage             | Duration    | Includes                                                                     |
|-------------------|-------------|------------------------------------------------------------------------------|
| <b>Game-Ready</b> | years 0-9   | Day-0 game optimizations, feature updates, perf tuning, new-workload support |
| <b>Legacy</b>     | years 10-13 | Security fixes, stability patches, bug-fix-only (no new features / day-0)    |

This policy is anchored in the **Kinetic Unified Driver** architecture (one driver stack across all Brad silicon, Brad Software/OS/BRADCVT\_DRIVERS.md), which makes long-lived, cross-SKU driver support structurally feasible. It is a core part of the “built to last, not built to be replaced” brand promise: a 4-year architecture that you’d otherwise feel pressured to swap is genuinely kept alive with software for a decade-plus.

**Why Brad Can Sustain This — and Why It’s a Deliberate Differentiator** This is not charity and not a marketing line. Long-lived support is **structurally possible for Brad in a way it is not for the incumbents**:

- **No legacy code burden.** NVIDIA, AMD, and Intel carry decades of installed silicon, ISA baggage, and compatibility obligations. Every game-ready driver must be tested and tuned against hardware their engineers may no longer even have on a bench. Brad starts with **one clean driver architecture (Kinetic Unified Driver) across a small, coherent silicon family** — so “optimizing for older parts” is the same effort as optimizing for new ones. We are not dragging a 20-year tail; we are tuning a single, young, well-understood stack.
- **No annual-release treadmill.** Because Brad ships a **4-year generation**, the engineering team is not forced to drop old parts the moment a new SKU launches. The cadence itself frees the headcount that incumbents burn on perpetual refresh. A slower hardware pace means every architecture and every driver gets *more* years of love, not fewer.
- **No corporate-bloat incentive to drop support.** Dropping software support early is how incumbents push new hardware and defend margins. Brad’s model — fewer platforms, longer life, an upgradeable Brad-CAMM memory standard — does not depend on forcing churn. Obsolescence here is not a business lever; it is an accident to be avoided.

The result is a flywheel that builds **brand prestige and customer love**:

No legacy baggage + no annual refresh churn

- engineers keep optimizing every part
- old products stay genuinely good, not just “supported”
- customers feel respected, upgrade on merit not pressure
- brand trust compounds across generations

The **software policy is a deliberate differentiator**, not a compliance checkbox. It says: buy a Brad device and it will be made *better* over time, for longer than anyone else is willing to promise — because we are the only vendor whose architecture, cadence, and incentives are actually aligned to honor that promise. Where rivals treat “age of support” as a cost to minimize, Brad treats it as the core of the brand.

### 7.3 Brad Mobile Services (BMS) [Gen1]

BMS is the sovereign services layer — no dependency on Google, Apple, or any third-party.

**Duality Strategy (launch):** Ship with GMS as default for user familiarity. BMS as opt-in parallel. When ready — flip the switch.

| Service                            | Description                          | Replaces                 |
|------------------------------------|--------------------------------------|--------------------------|
| <b>BradMail</b>                    | Secure email with BradID             | Gmail                    |
| <b>BradHub</b>                     | Curated app store                    | Play Store               |
| <b>BradPay</b>                     | Fintech backbone, BradSec encrypted  | Google Pay               |
| <b>BradConnect</b>                 | Messaging + community platform       | WhatsApp / Discord       |
| <b>BradCloud</b>                   | Storage + sync                       | iCloud / Google Drive    |
| <b>BradBrowse &amp; BradSearch</b> | Privacy-first web + search           | Chrome / Google Search   |
| <b>BradGeo</b>                     | Mapping + location, smart city ready | Google Maps              |
| <b>BradNotify</b>                  | Notification service (no Firebase)   | Firebase Cloud Messaging |
| <b>BradSphere</b>                  | Social platform [Future]             | Instagram / Facebook     |
| <b>BradPass</b>                    | Premium subscription tier            | Xbox Game Pass           |

**Reference implementation (shipped):** the REST v1 spec and a stdlib-only reference server ship at `src/bms/(api_v1.md + reference_server.py)`, no third-party dependencies) — BradID auth, BradMail, BradHub, BradPay ledger (blake2b digests, 30-cent transfer fee), BradCloud, BradGeo, BradNotify. All 14 request/response cases verified against the running server; demo credentials and the error envelope are in the spec.

**Privacy principles:** Encryption by default, anonymization by architecture, local-first AI (all NeuroSense on-device BradNPU, not cloud), on-device consent, auditable open-source privacy client.

### 7.4 EROE — Eco-Render Optimization Engine [Gen1]

System-wide power and thermal governor embedded in BradOS.

#### Workload Profiles:

| Profile       | CPU | GPU | NPU | Memory | Example             |
|---------------|-----|-----|-----|--------|---------------------|
| Compute       | 80% | 10% | 10% | 60%    | Compiling           |
| Gaming        | 30% | 70% | 20% | 90%    | Rise of the Streets |
| AI Inference  | 10% | 10% | 80% | 70%    | TIFA, BradVision    |
| Video         | 10% | 30% | 0%  | 40%    | AV1 decode          |
| Idle          | 5%  | 0%  | 0%  | 10%    | Lock screen         |
| Battery Saver | 30% | 20% | 20% | 30%    | Low battery         |

**Cooling Integration:** - <60°C: Max performance, no limits - 60-80°C: Gradual power cap, fan curve increase - 80-95°C: Hard limit 80% max power, aggressive fan - >95°C: Minimum throttle, display warning

#### API:

```
bradERoeSetProfile(EROE_PROFILE_GAMING);
bradERoeDeclareWorkload("render", EROE_UNIT_GPU, 0.7f);
```

**Reference implementation (shipped):** `src/drivers/eroe.c` implements the thermal governor for real — junction/board/battery telemetry, cooling thresholds, and a workload governor — exercised by the driver test suite and the BradDev Kit self-test.

## 7.5 Brad-CVT & Unified Driver Suite [Gen1]

**Philosophy:** “Don’t fight CUDA, absorb it.” Every CUDA kernel, PyTorch model, game engine runs on Brad silicon without modification.

### Brad-CVT (CUDA-to-Vector Translator)

CUDA Binary / PTX → Brad-CVT JIT/AOT (LLVM-based)

- ISA Emulation Layer + Thread Remapper + Memory Redirector
- BradVector-Optimized Kernel → BradGfx/BradFx

**Components:** - **AOT:** Offline compiler — CUDA PTX/SASS → BradVector ISA at install time - **JIT:** Runtime compiler — dynamically-generated kernels, sub-ms translation - **ISA Emulation Layer:** Dedicated silicon on every BradGfx/BradFx — translates legacy CUDA calls at wire speed - **Thread Remapper:** Maps CUDA warps (size 32) onto dual-issue BradVector lanes - **Memory Redirector:** CUDA shared → SRAM Infinity Buffer, global → SPMP LPDDR or HB-AIM

**Supported CUDA Libraries (zero-modification):** > **Honesty note:** these are *design targets* on the reference architecture, **not shipped-chip benchmark claims** — Brad silicon has not been built or measured against RTX 5090. Expected bandwidth/storage advantages drive the direction listed here, but real numbers await silicon. - cuBLAS: +15-30% vs RTX 5090 *target* (Infinity Buffer advantage) - cuDNN: +20-40% *target* (HB-AIM > HBM) - TensorRT: +25-50% *target* (MoE sparsity win) - NCCL: +30% *target* (on-die fabric > PCIe)

**Native open standard support:** SYCL 2020, Intel OneAPI, Vulkan 1.4, DirectX 12 Ultimate, Metal 3.

**Kinetic Unified Driver** Single driver stack managing SPMP, AUC, Fabric, power, and memory across all Brad silicon.

**Driver Modules:** - **SPMP Manager:** Allocates LPDDR, zero-copy HDB→SPMP→Fabric, pool coalescing - **AUC Manager:** Manages 950MB MVRAM SRAM cache, cache-line granularity, TIFA AI prefetch - **Fabric Scheduler:** Gang scheduling across chiplets, NUMA-aware, optical lane steering - **Power Governor:** Per-cluster DVFS, 6 p-states, temperature-aware via BradCool telemetry - **Memory Allocator:** 4KB/64KB/2MB pages, huge page hints, zero-copy DMA - **Queue Manager:** Multi-queue (64 compute + 32 copy + 16 direct), HW priority lanes

**Zero-Overhead Wrappers:** Drop-in replacements for NVIDIA runtime — `#include <brad_runtime.h>` replaces `#include <cuda_runtime.h>`. No source changes needed.

**Reference drivers (shipped):** `src/drivers/` holds real reference implementations of the Kinetic Unified Driver’s modules — SPMP first-fit pool allocator with coalescing + fragmentation stats, Fabric DMA/atomics/bind-memory, EROE thermal governor — and a new `bradlib` (`src/include/brad/bradlib.h`) exposes the platform as one C API: `compile` → `pack` → `session` → `launch` → `run` → `inspect` → `breakpoints`.

**TIFA-SDK** Developer tool that makes CUDA irrelevant for next-gen: - Python-first (`import tifa`), C++ backend - Auto-NeuroStream: model auto-partitioned across BradNeuro, BradVector, BradFx - MoE Spooler: expert prefetching to HB-AIM, cold spill to SPMP - One-line PyTorch: `model = tifa.optimize(model)` - FP8/FP4/INT4 native quantization, no calibration data needed

**Reference package (shipped):** `src/tifa/` is a working Python-first SDK — `import tifa`, AutoNeuroStream partitioner, MoE expert spooler, and the exact FP8 E4M3FN / E5M2 / FP4 E2M1 / INT4 digit formats (spot-checked bit-identical against the spec values: 240, 57344, 3.0, -5). Stdlib-only, zero dependencies.

## 7.6 BradDev Kit

| Tool          | Purpose                                                               |
|---------------|-----------------------------------------------------------------------|
| Brad-CVT CLI  | brad-cvt compile kernel.cu                                            |
| brad-gpu      | GPU monitoring (info, top)                                            |
| Brad-Link API | C ABI (bradLinkAlloc, bradLinkLaunch, etc.)                           |
| Debugger      | gdb-lldb hybrid with BradVector register inspection                   |
| Profiler      | Per-kernel cycle/active/occupancy + memory timeline                   |
| TIFA Studio   | GUI for model optimization and deployment                             |
| BFL           | Brad Fusion Language — DSL for custom BradVector kernels (SYCL-based) |

**Reference implementations (released, src/bradvector/):** bradc (basm → .bvbc assembler), BVRT (.bvbc interpreter), bradgdb (breakpoints, register/PC inspect), bradtimeline (per-cycle trace → CSV). These model the ISA at SIMT-VM level and run unattended in CI. See Brad Software/OS/BRADVECTOR\_PLATFORM\_MOAT.md.

**The one-binary Kit (shipped):** all of the above — plus the SPMP/EROE/Fabric drivers — are wrapped by the unified braddev CLI at src/braddev/:

| Subcommand                 | What it really is                                                                                     |
|----------------------------|-------------------------------------------------------------------------------------------------------|
| braddev cvt                | bradc assembler (.bvbs → .bvbc, stdin OK, -d disasm)                                                  |
| braddev run / dbg          | BVRT launch + bradgdb breakpoints; --arg seeds S16..S31; --dump S#/M<addr> scalar + memory inspection |
| braddev gpu info / gpu top | device profile + single-sample BradTimeline perf snapshot                                             |
| braddev timeline           | per-kernel cycle profile + CSV export                                                                 |
| braddev test               | self-test linking the drivers AND the whole toolchain pipeline                                        |

Verified end-to-end: SAXPY computes  $y[i] = 4 \cdot x[i] + y[i]$  over eight elements with zero traps ( $y[0] = 6.0$ ,  $y[7] = 48.0$  in float terms),  $\text{fib}(8) = 21$ , and all self-test checks pass. The reference ISA's honest rules are exercised in the examples: S0 is a read-only zero register, structural ALUs and MOV are float-domain, the integer ops (IADD/ISUB/IMUL) take registers only, and integer constants therefore enter via launch args — exactly the semantics silicon implements.

## PART 8: CONSOLE & PRODUCTS

### 8.1 BradAxis Console [Gen1]

**Target launch:** Holiday 2027

**Philosophy:** Player-centric ecosystem — unrivaled performance, revolutionary immersion, sustainable ecosystem, accessibility.

#### Hardware

| Component | Specification                                                           |
|-----------|-------------------------------------------------------------------------|
| CPU       | BradCore console SKU (standard, 7LPP) — 8× Phoenix P + 8× Falcon E      |
| GPU       | BradGfx Max (fabric) — 1024 BradVector, 256 RT, 192 Neuro-Stream cores  |
| NPU       | BradNPU — 64 NCCs, 128 TOPS INT8                                        |
| Memory    | 16/24 GB GDDR7 (CPU+GPU shared UMA) — ~1.0 TB/s base / ~1.15 TB/s Elite |

| Component | Specification                                                          |
|-----------|------------------------------------------------------------------------|
| Storage   | 1/2 TB TLC BradNAND (SLC-cache), PCIe 6.0 x4, HDB (direct-to-GDDR)     |
| Cooling   | Vapor chamber + liquid metal TIM + active BradCool fan (whisper-quiet) |
| PSU       | 450 W internal, 80+ Platinum. Peak draw: 350 W. Standby: <1 W.         |

**Connectivity:** Wi-Fi 7, Bluetooth 6.0, 3× USB-C (front) + 2× USB-C (rear), HDMI 2.2 (8K@120Hz VRR), DP 2.0 (8K@120Hz DSC), 10 Gigabit Ethernet, 2× internal storage expansion.

**Nexus Controller BradSense Adaptive Haptics:** - 4× high-definition LRA actuators (20 Hz–10 kHz) - Texture feel (sand, ice, rain), environmental cues (tremors, wind), directional feedback, force/resistance

**Dynamic Adaptive Triggers:** - Programmable resistance, tension, tactile feedback - Variable pull weight (bowstring, weapon recoil), tactile clicks and hard stops

**Biometric Feedback (opt-in):** - PPG pulse sensor (right grip), GSR electrodes (both grips) - Game difficulty scaling based on stress/engagement - All data processed on-controller (RISC-V secure core) — raw data never leaves

**NexusMods — Modular System:** - Swappable D-pad (cross → disc → 8-way) - Swappable thumbsticks (convex → concave → extended precision) - Mini-OLED display (1.2" 480×480 — quick stats, inventory) - Precision trackball, custom macro pad (6 buttons), flight stick grip, back paddle expansion (2-4 paddles)

**Wireless:** BradLink proprietary (2.4 GHz + 5 GHz adaptive), <1ms latency, 2000 Hz polling, 15m range, 3000 mAh Li-Po (40h typical), USB-C fast charge (15min → 8h play).

**BradOS Gaming** Custom microkernel variant: - 3-second boot (power-on to game library) - <1 second game resume (instant state save/restore to GDDR) - Zero background processes — game gets 100% of CPU/GPU/NPU - Hardware-accelerated UI compositor (1% performance cost) - System updates applied during rest mode, never interrupt gameplay

## Services

| Tier                | Price   | Features                                                                                             |
|---------------------|---------|------------------------------------------------------------------------------------------------------|
| Free                | \$0     | Online multiplayer, basic cloud saves, store access                                                  |
| BradAstra Unlimited | \$15/mo | All first-party games day-one, 200+ library, cloud streaming, enhanced saves                         |
| BradPass            | \$25/mo | Unlimited + 7-day early access, \$10/mo store credit, cosmetics, priority support, Experimental Labs |

**Developer Ecosystem:** BradDev Kit (unified-memory-aware SDK — SPMP on mobile, GDDR on console; BradSense haptics, NexusMods API), BradLabs Program (funding for indie/emerging devs, esp. African), BradStudios Originals (sibling company under BradVerse Kenya — first-party titles, Rise of the Streets leads), 70/30 revenue split publishing.

## 8.2 Brad Ecosystem vs Apple Ecosystem (Short Form)

Full analysis: BRAD\_VS\_APPLE.md. Positioning (per Info.txt:83): “the Apple + NVIDIA baby” — Apple-style vertical integration + NVIDIA-style discrete rendering and open compute.

| Layer        | Apple                             | Brad                                                               |
|--------------|-----------------------------------|--------------------------------------------------------------------|
| Silicon      | A-series / M-series               | HyperCore,<br>BradChip/BradCore,<br>BradXon                        |
| Memory       | LPDDR UMA + 32MB SLC,<br>soldered | SPMP 16-256GB + MVRAM<br>950MB + Brad-CAMM<br>(upgradeable)        |
| AI           | Apple Intelligence, ~38 TOPS      | NeuroSense AI, 32-128<br>TOPS, TIFA                                |
| GPU          | Integrated only (MetalFX)         | BradFx (integrated) +<br>BradGfx (discrete, RT,<br>NeRF)           |
| Software/API | iOS/macOS, Metal only             | BradOS; Vulkan 1.4 +<br>DX12U + Metal 3 + SYCL;<br>Brad-CVT (CUDA) |
| Services     | iCloud, Apple Pay, Apple One      | BMS: BradCloud, BradPay,<br>BradPass                               |
| Display      | Liquid Retina, ProMotion, XDR     | BradLumin Display,<br>SentinelView                                 |

**Scorecard: - Beats Apple:** memory flexibility (Brad-CAMM vs soldered), MVRAM freeze (~30× the system cache), bandwidth ceiling (1.84-2 TB/s), battery (36-48h), openness (4 GPU APIs + CUDA), discrete RT GPU tier, sovereign services. - **Trails Apple:** peak IPC (ROB 128 vs >150, lower frequency), node (7nm/4nm-class vs 3nm N3E), compiler maturity, software polish — and Apple ships while Brad is a spec. - **On par:** unified-memory philosophy, vertical integration, security posture (BradSec vs Secure Enclave).

## PART 9: HORIZON & VISION

### 9.1 BradSkynet — AI Platform / AGI Infrastructure [Gen2]

**Target launch:** 2027-2028

Distributed AI compute platform running across the Brad ecosystem — from individual BradNPUs in edge devices to massive BradApex clusters in data centers.

#### Architecture:

BradSkynet Node

- └─ BradApex × 4–8                    ← compute engine (see `Brad Silicon/Compute/BRADAPEX\_ACCELERATOR\_SPEC.md`)
- └─ BradXon CPU                      ← node controller / orchestration
- └─ SPMP Enterprise (512 GB–1 TB pool)
- └─ BradDrive Enterprise (HDB direct-to-pool)
- └─ BradFusion Fabric (optical, sub-μs, inter-node)

BradApex (Gen1 design, Gen2 ramp 2027-28) replaced the original “BradGfx Max × N GPU cluster” concept as the node’s primary compute — a ground-up datacenter accelerator (tensor engine for AI + first-class IEEE FP64 for HPC), memory-centric by design. BradGfx Max remains a graphics product; BradXon orchestrates but does not compute.

**Proposed capabilities:** Distributed MoE (experts sharded across nodes, sub-μs inter-node latency via optical Fabric), federated learning (edge NPUs train local models, gradient updates only to

Skynet), auto-scaling AI (workloads scale from phone to datacenter without re-partitioning), continuous learning (models update from live data streams).

## 9.2 BradAstra — VR Ecosystem & Headset [Gen2-Vision]

**Target launch:** 2030+ (Gen2 Synaptic window) — aligned to the 4-year generation cadence

BradAstra is Brad's **VR ecosystem and VR headset** — the answer to Apple Vision Pro / visionOS. Three layers under one brand: the **headset device** (BradNeuro X, BradOS XR), the **ecosystem** (persistent metaverse world, cross-device presence, creator economy), and the **infrastructure** (Brad-Skynet nodes / Rack Quad servers that render cloud scenes).

**Headset:** BradNeuro X (XR SKU, SF4X 4nm-class) mobile SoC — integrated BradFx Max-class + BradNPU; BradVision XR tracking (6-DoF head + hand, room mesh); 4K per-eye micro-OLED (UDC BradClarity); SPMP LPDDR unified pool; BradOS XR (spatial UI, hand tracking). Heavy scenes offload to the BradAstra server over BradLink/edge — mobile SoC + cloud server compute model.

**Ecosystem features:** Persistent world state in SPMP Enterprise streamed via BradCompress Video, neural semantic scene reconstruction, real-time AI NPCs (TIFA-powered), cross-device presence (Continuum Flow — walk in on the headset, continue on BradAxis), creator economy on BradDev Kit + BradLabs.

## 9.3 BradLabs — Fabrication & R&D Division

**Current Role (2026) [Gen1]:** - Process partnership with Samsung (unified across all Kinetic tiers — 7LPP→SF4X) — BradLabs provides custom PDK - BradNAND fabrication (238-layer 3D NAND) - Developer program: SDK preview, custom kernel workshops, expert weight trainer (NeRF training cloud tool) - Funding + mentorship for indie game developers (esp. African creators)

**Future Vision (2029+) [Vision]:** - Captive foundry: sub-1nm graphene/BN channels - BradNAND Gen2: 350-layer 3D NAND, 5+ GB/s per die - Optical interconnect: on-chip photonic communication - Bio-synaptic NPU: neuromorphic spiking neural networks - Ambient energy: BradWear powered by body heat + RF harvesting

## 9.4 Roadmap Chronology

**Phase 1 — Foundations (Jul-Aug 2025):** BradNPU, SPMP, BradGfx v1, BradXon original concept, BradOS lineup.

**Phase 2 — Expansion (Sep 2025):** BradRT Cores, BradVision, BradVerse CPU Lineup, BMS, BradGfx & BradFx Official Series, BradAxis Console, Additional Details.

**Phase 3 — Refinement (Oct-Nov 2025):** BradVision ISP, BradRAM, BradDrive + BradCompress, BradGfx v2.

**Proposed Product Release Order:** 1. Foundation IP: BradNPU, SPMP, BradCompress 2. Consumer Silicon: BradCore, BradFx, BradRAM, BradDrive 3. Platform Launch: BradOS, BMS, BradPhones/Slates 4. Performance Tier: BradGfx (Lite-Ultra), BradRT Cores, BradVision ISP 5. Flagship: BradAxis Console, BradGfx Max, BradAstra Unlimited 6. Apex: BradXon, BradSkynet, BradAstra, BradOS Compute Pro

## 9.5 Generation Timeline

**Cadence:** Every **4 years** Brad ships a new generation — a complete architectural leap, not an annual refresh. Within a generation, mid-cycle SKUs (node shrinks, memory upgrades, new form factors) refine the architecture without changing it.

| Years          | Gen   | Name        | Key Nodes                                                                                                                                                            |
|----------------|-------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>2026-29</b> | Gen 1 | Kinetic     | BradMind/Neuro SoCs, BradBoard Mobile/Matrix/Nexus, Brad-CAMM, BradCool micro-fluidics, BradLink local (60 GHz)                                                      |
| <b>2030-33</b> | Gen 2 | Synaptic    | Node shrink (SF3/SF2 → 1.4nm), BradLightSpeed optical to Matrix tier, BradBoard Rack, BradLink Fabric Bridge (wired), BradSec-X to Neuro Elite, BradCool X-Liquid    |
| <b>2034-37</b> | Gen 3 | Quantum     | Sub-1nm fabrication (vision), quantum-entangled Fabric to Nexus tier, graphene channel transistors, self-evolving logic, bio-synaptic NPU, ambient energy harvesting |
| <b>2038-41</b> | Gen 4 | [Vision]    | BradLabs captive fab scaling, deep on-die photonic logic, BN/2D-material dielectrics                                                                                 |
| <b>2042+</b>   | Gen 5 | Singularity | Self-evolving gates, quantum-entangled fabric, bio-synaptic                                                                                                          |

**Note:** BradApex is a standalone datacenter accelerator, not a consumer generation — its [Gen1] design → [Gen2] ramp 2027-28 occurs *within* the Gen1 Kinetic window as a product ramp, complementing the 4-year consumer cadence.

## PART 10: APPENDICES

### Appendix A: Memory Hierarchy

REGISTER / L1 (per core) – sub-ns, KB

↓

L2 CACHE (per cluster) – ~1ns, hundreds of KB

↓

L3 CACHE (shared, on BradFusion Fabric) – ~5-10ns, MB

↓

L4 CACHE (on Brad memory-controller die – SRAM) – 2ns, 64 MB per stack

↓

SPMP (LPDDR Unified Pool – main memory) – ~20-40ns, 8-128 GB

↓

HDB L0 CACHE (on BradDrive Controller) – dictionary tables, recently decompressed

↓

BradDrive STORAGE (BradNAND) – PCIe 6.0, TB scale

↓

BradCloud / External (network, cold storage)

### Appendix B: Security Boundaries

Layer 0: Fabrication (BradLabs) – hardware root of trust (PUF)

Layer 1: BradDrive Controller (BradSec) – encrypt at rest (AES-256-XTS)

Layer 2: HDB data path – transparent encryption going in, decrypted going out

Layer 3: SPMP (BradSec) – encrypt in memory (AES-256-GCM), integrity checks (Merkle)

Layer 4: BradRAM die (BradSec-RAM) – physical memory encryption per cache line  
 Layer 5: CPU/GPU (BradSec-X on Xon, BradSec on Core) – compute zone isolation  
 Layer 6: GPU Display Pipeline (BradSec at Pixel Level) – encrypted framebuffer, HDCP 2.3+  
 Layer 7: Software (BradOS, BMS) – user authentication, DRM, secure enclave

### Appendix C: Data Pipeline Walkthrough — Game Level Load

1. Player triggers level load (or NeuroSense-Storage AI predicts it based on map position)
2. BradDrive Controller receives request → NeuroSense AI identifies required assets → BDC fetches from
3. Data passes through HDB: adaptive clocking ramps for burst → domain-specific decompression (geometry to-SPMP pipeline)
4. Data arrives in SPMP: NeuroSense AI Memory Controller organizes by priority (render > AI > game logic)
5. Processors access via BradFusion Fabric: GPU reads textures/meshes/shader, CPU reads game state/physics
6. BradRAM DDS learns access patterns → prefetches next likely assets → caches in L4 SRAM → loop continues

**Result:** Loading screens eliminated. Zero CPU tax on any data path.

### Appendix D: Bandwidth Budget

| Stage              | Bandwidth              | Notes                             |
|--------------------|------------------------|-----------------------------------|
| BradDrive → HDB    | 64 GB/s                | Raw NAND (PCIe 6.0 x4 uni)        |
| HDB decompression  | Multiplies by 2-4×     | Hardware-accelerated, no CPU cost |
| HDB → SPMP         | Same, zero-copy        | Direct pipeline                   |
| SPMP → Fabric      | Exceeds LPDDR standard | 3D-stacked TSVs reduce distance   |
| Fabric → Processor | Sub-ns per hop         | Mesh NoC topology                 |

### Appendix E: Naming Conventions

| Correct Term                    | Wrong Term(s)                                                   | Type                                                |
|---------------------------------|-----------------------------------------------------------------|-----------------------------------------------------|
| <b>BradFusion Fabric</b>        | BradFusion Platform & Chipset                                   | On-chip mesh NoC                                    |
| <b>BradFusion Platform</b>      | —                                                               | Superset (Fabric + I/O + controllers)               |
| <b>BradFx Standard</b>          | Balanced                                                        | GPU tier naming                                     |
| <b>BradSlates</b>               | BradPads                                                        | Tablet product name                                 |
| <b>BradCompress HDB</b>         | Hardware Decompression block / BradCompress Decompression Block | Full formal name                                    |
| <b>BradGfx Max (Discrete)</b>   | —                                                               | Standalone card                                     |
| <b>BradGfx Max (Integrated)</b> | —                                                               | In BradXon via ANF                                  |
| <b>BradGfx Max Xtensive</b>     | —                                                               | Discrete GDDR7 hybrid-UMA edition (192-512-bit bus) |
| <b>BradSec</b>                  | —                                                               | Base hardware security                              |

| Correct Term                    | Wrong Term(s) | Type                                                                                                              |
|---------------------------------|---------------|-------------------------------------------------------------------------------------------------------------------|
| <b>BradSec-RAM</b>              | —             | Inline memory encryption in the memory-controller data path (AMD-SME/MKTME-class, protecting standard JEDEC dies) |
| <b>BradSec-X</b>                | —             | Quantum-fortified, self-healing                                                                                   |
| <b>BradSec Display Pipeline</b> | —             | Pixel-level content protection                                                                                    |
| <b>SPMP</b>                     | —             | Logical unified pool from BradRAM dies + controller                                                               |
| <b>BradRAM</b>                  | —             | Physical LPDDR5X die technology powering SPMP                                                                     |

## Appendix F: Credibility Tiers

**Grounded [Gen1] — real products, real engineering:** BradNPU, BradRAM, SPMP, BradDrive, BradCompress, BradVision ISP, BradAxis Console, BradOS, BMS, EROE, BradCore, BradFx/BradGfx (excluding 2GB SRAM), BradRT Cores, HyperCore/Atlas, BradChip, Brad-CVT/Drivers

**Mixed — real core + visionary extensions:** BradXon (ANF is [Gen1] — compute partitions, RDA, fabric governor; the unit-morph is [Vision]), BradGfx Max 2GB SRAM Infinity-Buffer ([Vision] — 3D-stacked tile under compute die), BradSense-Render (core is [Gen1], full neural rendering is [Vision]), BradClarity UDC ([Gen2]), BradFusion Fabric optical interconnects ([Gen2]), SPMP LPDDR UMA ([Gen2])

**Visionary [Vision] — requires fundamental breakthroughs:** Sub-1nm graphene channel transistors, quantum-entangled Fabric, self-evolving logic gates, bio-synaptic NPU, ambient energy harvesting, 2GB monolithic SRAM, on-die photonic logic

**Aesthetic — marketing language, not technical:** “Living compute substrate,” “hardware embodiment of consciousness,” “synthesize reality,” “beyond von Neumann,” “Sovereign Visuals,” “soul of sight”

## Appendix G: Open Questions for Brad

1. **BradBoard vs chassis integration:** Should high-end Matrix/Nexus boards integrate into chassis or remain swappable? Info.txt suggests “cable-less” is the future direction.
2. **Brad-CAMM standard:** Lock in pin count and keying for Gen 1? Need to ensure LPDDR6 14.4+ Gbps signal integrity.
3. **BradLink security model:** When a BradPhone tethers to a BradMatrix, should the laptop trust the Phone’s BradSec, or require independent authentication?
4. **Rack sled form factor:** Standard 1U half-width, or custom BradLabs width for better thermals?

5. **Nexus Console vs standard Nexus:** Should BradAxis use a custom board or modified Nexus Pro with Xon-only? **Resolved:** custom 280×200mm board, soldered BradCore 8P+8E + GDDR7 (no Xon involvement).
6. **Full custom kernel timeline:** BradOS 3.0 with BradKernel (no Linux dependency) targeted for 2028. Realistic?

## Appendix H: Key Architectural Relationships

1. **Data Pipeline (the real differentiator):** BradDrive → BradCompress HDB → SPMP → BradFusion Fabric → any processor. Zero-copy, zero-CPU-overhead, direct-to-pool streaming.
2. **BradFusion Fabric:** Mesh NoC interconnecting all components at hardware level. Extends from die → package → board → rack (optical in [Gen2], quantum-entangled in [Vision]).
3. **SPMP is the nexus:** Every compute element pulls from the same unified LPDDR pool via the Fabric.
4. **CPU-GPU integration:** BradCore (workhorse) has integrated BradFx. BradXon (flagship) has integrated BradGfx Max via ANF. BradGfx Max Xtensive is the discrete GDDR7 edition (hybrid-UMA).
5. **HyperCore - BradCore relationship:** HyperCore/Atlas is the mobile specialization — context persistence over throughput. Same shader math, same architecture patterns, different optimization target.
6. **BradRAM → SPMP relationship:** BradRAM is the physical LPDDR die. SPMP is the logical pool. Multiple BradRAM dies + controller = SPMP. They are the same physical silicon at different abstraction layers.

## Appendix I: Rise of the Streets

The ecosystem's showcase reference game. **Still in development** — no dedicated documentation (beyond SDK references in Brad Software/OS/BRADCVT\_DRIVERS.md). Referenced in BradAxis, BradDrive, and BradSense-Render docs as the title that demonstrates the pipeline — seamless open-world streaming, no loading screens, NeuroSense-AI-predicted asset pre-fetching, per-object neural reconstruction. A reference title, not the product itself. Mobile-compatible: same shader math on HyperCore, runs on BradSlate without a port.

## Appendix J: Complete Component Index

| Component                                 | Type                   | Badge                           | Defined In                 |
|-------------------------------------------|------------------------|---------------------------------|----------------------------|
| BradCore                                  | CPU                    | [Gen1]                          | BRADCHIP.md                |
| BradXon                                   | CPU (Flagship)         | [Gen1]; ANF unit-morph [Vision] | PLATFORM_SPEC.md           |
| HyperCore                                 | SoC (Mobile)           | [Gen1]                          | HYPERCORE_ATLAS.md         |
| HC-100/300/500/Elite                      |                        |                                 |                            |
| Atlas Architecture (AUC, Sentinel, MVRAM) | Architecture           | [Gen1]                          | HYPERCORE_ATLAS.md         |
| BradChip Lite→Max                         | SoC Package            | [Gen1]                          | BRADCHIP.md                |
| BradGfx Lite→Max                          | GPU (Dedicated)        | [Gen1]                          | GPU_ARCH.md                |
| BradFx Lite→Max                           | GPU (Integrated)       | [Gen1]                          | GPU_ARCH.md                |
| BradVector ISA + Engines                  | GPU Vector ISA/Compute | [Gen1]                          | GPU_ISA.md, GPU_COMPUTE.md |
| BradRT Cores                              | RT Accelerator         | [Gen1]                          | GPU_ARCH.md                |

| Component                                           | Type                     | Badge         | Defined In                                           |
|-----------------------------------------------------|--------------------------|---------------|------------------------------------------------------|
| BradSense-Render                                    | Neural Rendering         | [Gen1]        | BRADSENSE_RENDER_SPEC.md                             |
| Neuro-Stream Units                                  | GPU AI Cores             | [Gen1]        | BRADSENSE_RENDER_SPEC.md                             |
| BradNPU (+NCC, HB-AIM, MoE, AI-ISA)                 | NPU                      | [Gen1]        | BRADNPU_ARCH.md                                      |
| BradRAM (+L4 SRAM)                                  | Memory Die               | [Gen1]        | BRADRAM_SPMP_ARCH.md, BRADRAM_DDS.md/BradRAM_POWER/R |
| SPMP (Lite→Enterprise)                              | Unified Pool             | [Gen1]        | BRADRAM_SPMP_ARCH.md, BRADRAM_VMEM.md                |
| NeuroSense AI Memory Controller                     | Controller               | [Gen1]        | BRADRAM_SPMP_ARCH.md                                 |
| BradSec-RAM                                         | Security (Memory)        | [Gen1]        | BRADRAM_SPMP_ARCH.md, BRADRAM_RAS.md                 |
| BradDrive (Standard/Pro/Elite/Enterprise)           | SSD                      | [Gen1]        | BRADDRIVE_COMPRESS_ARCH.md                           |
| BradCompress (+HDB)                                 | Codec + HW               | [Gen1]        | BRADDRIVE_COMPRESS_ARCH.md                           |
| BradVision (Lite→Elite, ISP, Fusion, Optics, Coach) | Camera/Vision            | [Gen1]        | BRADVISION_COMPLETE.md                               |
| BradOS (Mo-<br>bile/Slate/Compute/Nano + Editions)  | OS                       | [Gen1]        | BRADOS_COMPLETE.md                                   |
| BMS (10 pillars)                                    | Cloud Services           | [Gen1]        | BMS_SPEC.md, BRADOS_COMPLETE.md                      |
| EROE                                                | Power Governor           | [Gen1]        | EROE.md, BRADOS_COMPLETE.md                          |
| Brad-CVT (+ AOT/JIT, ISA Emulation)                 | CUDA Translator          | [Gen1]        | BRADCVT_DRIVERS.md                                   |
| Kinetic Unified Driver                              | Driver Stack             | [Gen1]        | BRADCVT_DRIVERS.md                                   |
| TIFA-SDK                                            | AI SDK                   | [Gen1]        | BRADCVT_DRIVERS.md                                   |
| BradVector Platform (bradc/BVRT/bradgdb)            | SW Toolchain (reference) | [Gen1 ref]    | BRADVECTOR_PLATFORM_MOAT.md, GPU_ISA.md              |
| BradAxis Console (+Nexus Controller, NexusMods)     | Console                  | [Gen1]        | BRADAXIS_SPEC.md                                     |
| BradMind Lite→Elite                                 | SoC Platform             | [Gen1]        | PLATFORM_SPEC.md                                     |
| BradNeuro Ultra→Elite/X                             | SoC Platform             | [Gen1-Vision] | PLATFORM_SPEC.md                                     |
| BradBoard                                           | Motherboard              | [Gen1]        | PLATFORM_SPEC.md                                     |
| Mobile/Matrix/Nexus/Rack                            |                          |               |                                                      |
| Brad-CAMM LPDDR/DDR                                 | Memory Module            | [Gen1]        | PLATFORM_SPEC.md                                     |
| Brad-M.3                                            | SSD Form Factor          | [Gen1]        | PLATFORM_SPEC.md                                     |
| BradCool (micro-fluidics, X-Liquid, Nano)           | Cooling                  | [Gen1-Gen2]   | PLATFORM_SPEC.md                                     |
| BradLink (Local, Fabric Bridge)                     | Tethering                | [Gen2]        | PLATFORM_SPEC.md                                     |
| BradSkynet                                          | AI Infrastructure        | [Gen2]        | BRADSKYNET_ASTRALABS.md                              |
| BradAstra                                           | VR Ecosystem + Headset   | [Gen2-Vision] | BRADSKYNET_ASTRALABS.md                              |
| BradLabs                                            | Fab/R&D                  | [Gen1-Vision] | BRADSKYNET_ASTRALABS.md                              |
| BradClarity UDC                                     | Display                  | [Gen2]        | BRADVISION_COMPLETE.md                               |
| Modular Bayonet Optics                              | Lens System              | [Vision]      | BRADVISION_COMPLETE.md                               |
| BradLightSpeed (hybrid photonic)                    | Interconnect             | [Gen2]        | PLATFORM_SPEC.md                                     |
| Quantum-entangled Fabric                            | Interconnect             | [Vision]      | ARCHITECTURE.md                                      |

| Component                       | Type     | Badge    | Defined In |
|---------------------------------|----------|----------|------------|
| Self-evolving logic gates       | CPU Tech | [Vision] | VISION.md  |
| Bio-synaptic NPU                | NPU Tech | [Vision] | VISION.md  |
| Ambient energy harvesting       | Power    | [Vision] | VISION.md  |
| BradSec-X Quantum Fortification | Security | [Vision] | VISION.md  |

---

## End of Brad Devices Complete Unified Specification

*Compiled from 78 supplementary .md reference documents + Info.txt (2,167-line AI co-architect conversation). Total ecosystem: 79 text files, 16,727 lines (excluding Converted/ originals; 81 files, 17,619 lines with them). This compendium: 1,733 lines, 41 pages (merged/deduplicated from the supplementary docs). Created by Brad (age 15), Jul-Nov 2025. Consolidated Jun-Jul 2026. Refreshed Sept 2026. "Data in Motion. Intelligence in Real-Time." — Kinetic (Gen 1)*