

International Studies
in the Field of
**Electrical Electronics
and Communications
Engineering**

— October 2026 —

EDITOR

Doç. Dr. Abdurrahman GÜNDAY

gece
kitaplığı

İmtiyaz Sahibi / Yaşar Hız
Yayına Hazırlayan / Gece Kitaplığı

Birinci Basım / Ekim 2026 - Ankara
ISBN / 978-625-321-330-5

© copyright

Bu kitabın tüm yayın hakları Gece Kitaplığı'na aittir.

Kaynak gösterilmeden alıntı yapılamaz, izin almadan hiçbir yolla çoğaltılamaz.

Gece Kitaplığı

Kızılay Mah. Fevzi Çakmak 1. Sokak

Ümit Apt No: 22/A Çankaya/ANKARA

0312 384 80 40

www.gecekitapligi.com / gecekitapligi@gmail.com

Baskı & Cilt

Bizim Büro

Sertifika No: 42488

**INTERNATIONAL STUDIES
IN THE FIELD OF
ELECTRICAL ELECTRONICS AND
COMMUNICATIONS ENGINEERING**

October 2026

EDITOR

Doç. Dr. Abdurrahman GÜNDAY

CONTENTS

CHAPTER 1

LIGHT-EMITTING MEMORY: MEMRISTIVE DEVICES WITH OPTICAL READ-OUT FOR ARTIFICIAL INTELLIGENCE AND CLOUD STORAGE

Mustafa ANUTGAN, Tamila ANUTGAN 7

CHAPTER 2

MULTILEVEL INVERTER (MLI) TOPOLOGIES FOR PHOTOVOLTAIC APPLICATIONS: FROM CLASSICAL STRUCTURES TO REDUCED-SWITCH DESIGNS

Kenan GÜRÇAM 37

CHAPTER 3

A-IGZO TFTS FOR LARGE-AREA ELECTRONICS: MATERIALS PHYSICS, DEVICE ARCHITECTURES, AND FABRICATION PROCESSES

*Mehmet İzzeddin GÜLER, Tamila ANUTGAN,
Mustafa ANUTGAN* 59

CHAPTER 1

LIGHT-EMITTING MEMORY: MEMRISTIVE DEVICES WITH OPTICAL READ-OUT FOR ARTIFICIAL INTELLIGENCE AND CLOUD STORAGE

Mustafa ANUTGAN¹, Tamila ANUTGAN²

¹ Prof. Dr., Thin Film Devices Laboratory (IFALAB), Materials Research and Development Center (MARGEM), Karabuk University, 78050, Karabuk, Turkiye, ORCID: 0000-0003-2535-2646

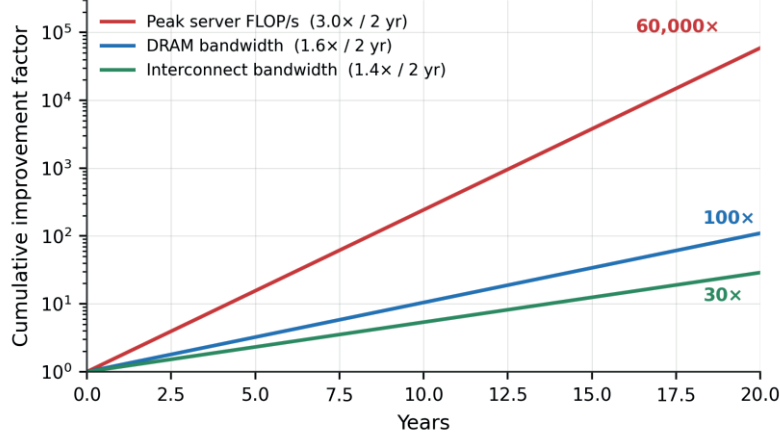
² Prof. Dr., Thin Film Devices Laboratory (IFALAB), Materials Research and Development Center (MARGEM), Karabuk University, 78050, Karabuk, Turkiye, ORCID: 0000-0002-0287-265X

1. Introduction: Why Memory Became the Problem

For most of the history of computing, performance gains came chiefly from faster logic, while memory was treated as a passive store. Artificial intelligence (AI) workloads have reversed this balance. Gholami et al. (2024) quantified the mismatch: peak floating-point operations per second (FLOP/s) improved by a factor of 3.0 every two years, dynamic random-access memory (DRAM) bandwidth by 1.6 and interconnect bandwidth by 1.4 (Figure 1). Demand moved in the opposite direction: between 2018 and 2022 large language model (LLM) training compute grew by about $750\times$ per two years, and Epoch AI places frontier training-compute growth at $4\text{--}5\times$ per year (Sevilla & Roldán, 2024).

Four of the six production neural-network workloads analysed on Google’s first tensor processing unit were memory-bandwidth limited (Jouppi et al., 2017, Fig. 5). Under Epoch AI’s component-cost model for AI accelerators, High-Bandwidth Memory (HBM) represented an estimated 52% of total component spending in Q1 2024, rising to 63% in Q4 2025 (Somala, 2026).

Figure 1
Divergence of Compute and Memory Improvement Rates



Note. Curves are generated from the growth factors reported by Gholami et al. (2024). Original figure.

This “memory wall” is being attacked on two fronts. The first moves memory closer to computation, as with HBM and memory attached through Compute Express Link (CXL; Section 2.2). The second introduces new non-volatile, dense memory devices able to compute where the data reside. Resistive random-access memory (RRAM or ReRAM) is a leading candidate: each cell stores information as a conductance state, and dense crossbars inserted into the back end of a logic process can perform analogue multiply–accumulate operations in place

(Sections 3, 4 and 6.4). If the stored state could be read optically rather than electrically, the read path might escape the limits of electrical interconnect (Section 5).

This chapter first quantifies what an emerging memory must deliver (Section 2) and examines RRAM against those requirements (Sections 3, 4 and 6), then one deliberately speculative variant, light-emitting memory (LEM; Section 5), before enumerating open research gaps (Section 7).

2. The Demand Side: A Quantitative Requirement Set

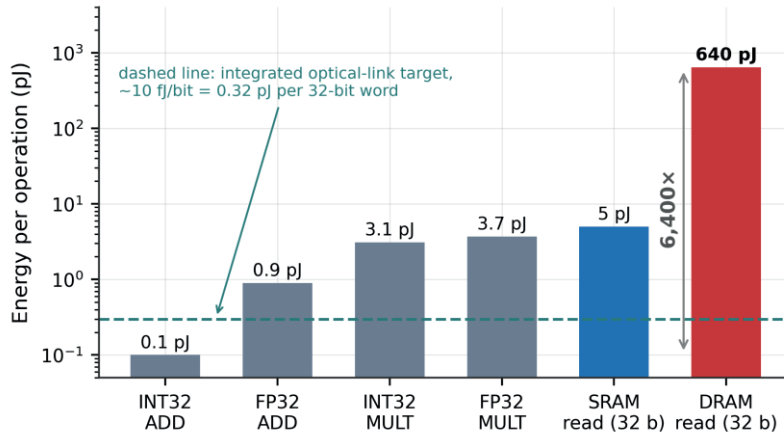
Before any candidate technology can be judged, the demands that AI and cloud workloads place on memory must be stated in quantitative terms.

2.1. The Energy of Moving Data

In Horowitz's (2014) 45 nm, 0.9 V reference budget, as tabulated by Han et al. (2016), a 32-bit integer addition costs 0.1 pJ, a 32-bit static random-access memory (SRAM) cache read 5 pJ, and a 32-bit DRAM read 640 pJ. Moving one word from DRAM therefore costs about 6,400 integer additions (Figure 2).

Figure 2

Energy per Operation at 45 nm, 0.9 V



Note. Values are those of Horowitz (2014) as tabulated by Han et al. (2016); the dashed line marks the ~10 fJ/bit target that Miller (2017) identifies for integrated optical links. Original figure.

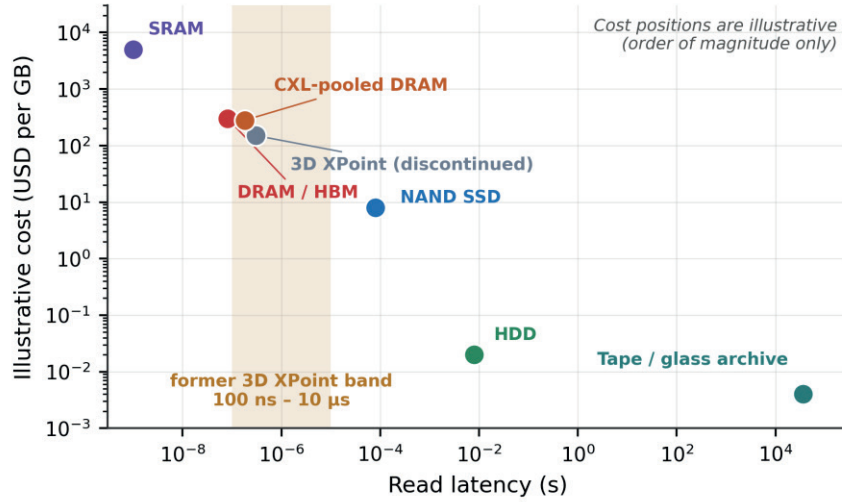
2.2. Storage Tiers, the Storage-Class Latency Band, and Disaggregation

Energy is one dimension; latency and cost per bit determine where a memory can sit in the hierarchy. In Figure 3, on-chip SRAM occupies the ~1 ns tier. Between the tens-of-nanoseconds DRAM tier and the tens-of-microseconds NOT-AND (NAND) flash tier lies the storage-class latency band (100 ns – 10 μ s); DRAM attached through CXL now serves part of this band (H. Li et al., 2023).

In Microsoft’s Pond study, DRAM can account for 50% of server cost, and under an emulated 182% latency increase 26% of 158 workloads slow by less than 1%, while 21% slow by more than 25% (H. Li et al., 2023). The target for an emerging memory is therefore not DRAM latency parity: sub-microsecond read latency at materially lower cost per bit is a plausible target for a quantified fraction of latency-tolerant cloud workloads, although latency-critical workloads such as LLM token generation remain constrained by memory bandwidth and latency.

Figure 3

Illustrative Cost and Latency Positions for Storage and Memory Tiers



Note. Latency values for DRAM and Optane are the measured figures of Izraelevitz et al. (2019); CXL-pooled latency is from H. Li et al. (2023); archival read characteristics from Anderson et al. (2023). Cost positions are illustrative order-of-magnitude indications, not measured prices. Original figure.

2.3. Energy and Sustainability Constraints

These costs accumulate at data-centre scale. The International Energy Agency (IEA) reports about 415 TWh of data-centre electricity consumption in 2024, projected at about 945 TWh in 2030 (IEA, 2025). Further efficiency gains must also come from inside the information technology stack, where Section 2.1 identifies data movement as the governing term.

2.4. Scaling Limits of the Incumbents

Each incumbent memory faces a distinct physical limit.

DRAM. Cells remain around $6F^2$; capacitor aspect ratio is a principal scaling challenge, and it is unclear whether DRAM can scale beyond 10 nm (LaPedus, 2019).

3D NAND. Layer counts have been reported at 321 in production (Shilov, 2025), but reducing the 40 nm word-line z-pitch degrades retention (Rachidi & Rosmeulen, 2025).

SRAM. TSMC's bitcell area was $0.021 \mu\text{m}^2$ at both N5 and N3E against $1.6\text{--}1.7\times$ chip-level transistor scaling (WikiChip Fuse, 2022). On-die cache can no longer absorb growing working sets.

2.5. The Resulting Requirement Set

Table 1 combines these constraints into a requirement set, against which the technologies reviewed in this chapter are assessed.

Table 1

Requirements an Emerging Memory Must Satisfy to Be Relevant to AI and Cloud Workloads

Metric	Requirement & Target	Classification & Anchor
Read latency	$<100 \text{ ns}$ for a DRAM-tier role; $<1 \mu\text{s}$ is a plausible architectural target for a pooled, CXL-attached tier serving latency-tolerant workload subsets	[Workload requirement] H. Li et al. (2023): $+70\text{--}90 \text{ ns}$ (8–16-socket pools); $<5\%$ slowdown for 43% of 158 workloads under emulated 182% latency
Write/read symmetry	Minimise write/read bandwidth asymmetry	[Architectural target] Izraelevitz et al. (2019; preprint): 2.9:1 read:write bandwidth on Optane
Endurance	$\geq 10^9$ for a main-memory tier; $10^5\text{--}10^6$ adequate for read-dominated AI weight storage (authors' estimate)	[Architectural target] 10^{12} cycles demonstrated (Lee et al., 2011)
Retention	10 years at 85°C	[Industry standard] Golonzka et al. (2019) for embedded RRAM
Read energy	Must beat DRAM's $\sim 5\text{--}20 \text{ pJ/bit}$; $\leq 1 \text{ pJ/bit}$ to matter	[Workload requirement] Horowitz (2014); Miller (2017); $\leq 1 \text{ pJ/bit}$ is the authors' target
Write energy	Sub-picojoule write energies are reported for selected RRAM cells	[Review-compiled] minimum energy per bit below 1 pJ reported for selected RRAM devices (Wang, Song, et al., 2024); condition-specific, energy scope unstated
Cost per bit	Likely to become the binding commercial constraint once workload-acceptable latency is achieved	[Market / Product constraint] Handy (2022) estimates peak 3D XPoint output at approximately 0.15 million wafers/year, at most $\sim 1/120$ of DRAM's; this is an industry estimate, not manufacturer-disclosed production data
Density	4F^2 crossbar and/or 3D stackable	[Industry benchmark] DRAM cells around 6F^2 (LaPedus, 2019); SRAM bitcell $0.021 \mu\text{m}^2$ at both N5 and N3E (WikiChip Fuse, 2022)
Multilevel	$\geq 2 \text{ bits/cell}$ for cost parity once cell area is fixed	[Architectural target] authors' estimate; accuracy versus levels in Mehonic et al. (2019)
Read bandwidth	Must scale with array area, not package perimeter	[Architectural target] Section 5.1; Miller (2017)

3. Resistive Switching Memory: Physical Foundations

Against this requirement set, resistive switching memory is considered first. Its physical basis determines which Table 1 requirements it can plausibly meet.

3.1. Taxonomy and Comparative Position

Redox-based resistive memory divides, in the nomenclature of Waser et al. (2009), into the valence change mechanism (VCM), electrochemical metallization (ECM) and the thermochemical mechanism (TCM). Distinct from these are phase-change memory (PCM), magnetoresistive random-access memory (MRAM) in its spin-transfer-torque (STT) and spin-orbit-torque variants, and ferroelectric capacitor (FeRAM) and field-effect transistor (FeFET) memories. Table 2 compares them against the requirement set of Table 1.

Table 2
Comparative Position of Emerging Non-Volatile Memories

Technology	Switching element	Write speed	Endurance (cycles)	Retention	MLC	Source
RRAM / OxRAM	HfO _x , TaO _x filament	10 ns; pulses to 250 ps	10 ⁴ (product) – 10 ¹² (lab)	10 yr at 85–125 °C	2–4 bit	Böttger et al. (2020); Lee et al. (2011); Golonzka et al. (2019)
CBRAM / ECM	Ag or Cu filament in SiO ₂ , GeS _x	1–5 ns	~10 ⁸	>10 ³ s	limited	Valov et al. (2011)
PCM	Ge ₂ Sb ₂ Te ₅ amorphous/crystalline	50–400 ns	10 ⁷ (chip)	10 yr at 85 °C, drift-limited	3 bit	Hellenbrand et al. (2024); Joshi et al. (2020)
STT-MRAM	MgO magnetic tunnel junction	~160 ns (28 nm product)	>10 ¹⁴	10 yr at 89 °C	no	Hellenbrand et al. (2024)
FeRAM	HfO ₂ or PZT capacitor	2 ns	10 ¹⁵	121 yr at 85 °C	no	Hellenbrand et al. (2024)
FeFET	HfO ₂ ferroelectric gate	20 ns	>5×10 ¹⁰	>10 ³ s at 25 °C (weak point)	2 bit	Dutta et al. (2022)

Note. Ranges span laboratory demonstrations and shipped products and should not be read as a single consistent operating point. MLC = multi-level cell; OxRAM = oxide RRAM; CBRAM = conductive-bridge random-access memory; HfO_x/HfO₂ = hafnium oxide; TaO_x = tantalum oxide; SiO₂ = silicon dioxide; GeS_x = germanium sulfide; Ge₂Sb₂Te₅ = germanium antimony telluride; MgO = magnesium oxide; PZT = lead zirconate titanate.

3.2. Switching Mechanisms

These memories store information in the conductance of a thin insulating layer. Filamentary switching is distinguished from interface-type switching by its area independence: the low-resistance state (LRS) does not scale with electrode area, whereas interface-type switching modulates a Schottky barrier and does (Sawa, 2008).

In the valence-change mechanism, electroforming generates oxygen vacancies; vacancy ordering forms the conductive filament, and reset opens a nanometre-scale gap rather than dissolving it (Figure 4a). For Ta₂O₅ cells, extrapolation indicates an intrinsic switching speed of about 10 ps at 3 V, although SET times below 770 ps could not be observed directly (Böttger et al., 2020).

Electrochemical metallization proceeds by anodic dissolution of an active electrode, cation drift through a solid electrolyte, and cathodic reduction into a metallic bridge (Valov et al., 2011). Long retention at read bias must coexist with fast switching at write bias, which requires strongly non-linear kinetics; Menzel et al. (2013) model these kinetics.

The compliance current imposed during set determines the filament cross-section, so LRS resistance scales approximately inversely with it; it is therefore both the primary multilevel-programming knob and the primary determinant of variability, and reporting it is non-negotiable (Lanza et al., 2019).

3.3. Materials, with Emphasis on CMOS-Compatible Silicon Compounds

Which mechanism dominates depends strongly on the switching material. Hafnium oxide (HfO₂) is the industry workhorse, integrated at 22 nm with 10⁴-cycle endurance and ten-year retention at 85 °C on 7.2 Mb arrays (Golonzka et al., 2019) and, at 130 nm, in a three-million-device compute-in-memory (CIM) chip (W. Wan et al., 2022).

The silicon-compound family combines resistive switching, luminescence and compatibility with complementary metal–oxide–semiconductor (CMOS) dielectric-deposition modules. Mehonic et al. (2018) document three distinct switching modes in silicon oxide (SiO_x). Two comparisons against HfO₂ are relevant to back-end-of-line insertion: a resistance window of about 50× against about 15×, and negligible resistance variation after 1 h at up to 260 °C, against a roughly fivefold LRS increase after 1 h at 250 °C in HfO₂.

Silicon nitride switching has four mechanisms under active debate, and the ambiguity is itself a research gap. Transient-current tracing in hydrogenated amorphous silicon nitride (a-SiN_x:H) estimates a Si dangling-bond density above 10¹⁸ cm⁻³ in the LRS (T. Chen et al., 2023). Analyses of stoichiometric and nitrogen-implanted silicon nitride (Si₃N₄) have instead been

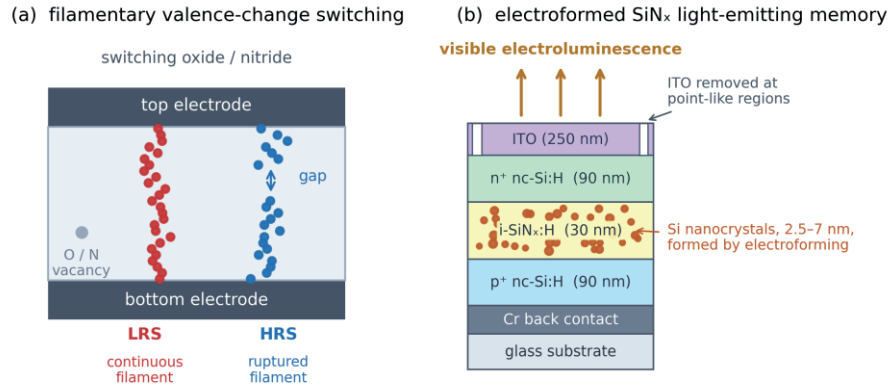
interpreted as consistent with nitrogen-vacancy-mediated switching (Mavropoulis et al., 2025). A third account invokes nanocrystal-assisted filament formation in Si-nanocrystal-embedded Si_3N_4 , with space-charge-limited conduction in the high-resistance state (HRS) (Vázquez Jiménez et al., 2026), and a fourth treats trap-assisted tunnelling, as modelled for oxide RRAM (Larcher & Padovani, 2017), as the dominant high-resistance conduction channel. SiN_x and SiO_x avoid the Ag and Cu mobile-ion concern of electrochemical cells, although memory-grade back-end compatibility requires stack-specific qualification.

3.4. Array Integration

Beyond the single cell, density and accuracy depend on how cells are assembled into arrays. The one-transistor-one-resistor (1T1R) cell is adopted in the industrial embedded RRAM macros surveyed by Wang, Song, et al. (2024) because the select transistor supplies accurate compliance current and eliminates sneak paths, at the cost of the $4F^2$ crossbar density.

Figure 4

(a) *Filamentary Valence-Change Switching* and (b) *Cross-Section of the Electroformed $\alpha\text{-SiN}_x\text{:H } p^+in^+$ Light-Emitting Memory*



Note. Original schematic; mechanism after Waser et al. (2009) and Larcher and Padovani (2017); device geometry from T. Anutgan et al. (2017), T. Anutgan et al. (2019) and M. Anutgan et al. (2020).

4. Memristive Arrays as an AI Compute Fabric

4.1. Principle and Demonstrated Silicon

The same arrays that store data can also compute with it. A crossbar of resistive cells performs a matrix–vector product in a single analogue settling time through Ohm’s law and Kirchhoff’s current law. The repeated movement of stationary weights is avoided at the crossbar core, though peripheral data transfers remain (Sebastian et al., 2020).

This principle has been demonstrated in silicon. NeuRRAM integrated three million $\text{HfO}_x/\text{TaO}_x$ cells and reported fully hardware-measured 85.7% on CIFAR-10 with ResNet-20

(W. Wan et al., 2022). On PCM, IBM's HERMES chip in 14 nm delivered 9.76 tera-operations per second per watt (TOPS/W) in its low-precision mode (Le Gallo et al., 2023), and a 34-tile, 35-million-device analogue chip reached 12.4 TOPS/W (Ambrogio et al., 2023).

4.2. Where the Efficiency Claims Are Optimistic

These efficiencies must, however, be placed in context. Against an NVIDIA H100 SXM at about 2.8 TOPS/W for dense 8-bit integer (INT8) operations (NVIDIA, 2022), HERMES represents about $3.5\times$ and the 34-tile analogue chip about $4.4\times$ — an illustrative comparison only, under non-iso-precision, non-iso-technology and non-iso-system conditions, and far short of the hundredfold gains implied by isolated crossbar metrics. The dominant reason is the converter periphery: in the ISAAC analysis, analogue-to-digital converters (ADCs) take 58% of tile power (Shafiee et al., 2016).

4.3. Non-Idealities and Their Measured Cost

Device non-idealities further erode the advantage measured at the array core. Phase-change conductance drift follows a power law with exponent near 0.06; a million-device prototype kept 93.75% on ResNet-32/CIFAR-10 at 25 s (Joshi et al., 2020). Parasitic wire resistance and capacitance limit crossbar cores to several hundred to about a thousand rows or columns (M. Jiang et al., 2025).

4.4. The Large-Language-Model Era

Large language models sharpen the question of capacity, because their weights must be held close to the compute. Under a 14 nm cell-area model, which excludes peripheral circuitry, storing all 175 billion GPT-3 parameters with four-bit memristive cells is estimated to require about 274 mm^2 , against a modelled area of about $1,646\text{ mm}^2$ for DRAM; these are architectural estimates, not fabricated-chip measurements (Wang, Luo, et al., 2025).

5. Light-Emitting Memory

The preceding sections treated resistive memory as an electrically read device. If the resistance state of a cell also controlled its light emission, the stored information could be read optically. This section sets out the case for optical read-out (Section 5.1), light emission from silicon-based materials (Section 5.2), the authors' electroformed silicon nitride device (Section 5.3), related device classes (Section 5.4) and maturity (Section 5.5).

5.1. The Architectural Argument for Optical Read-Out

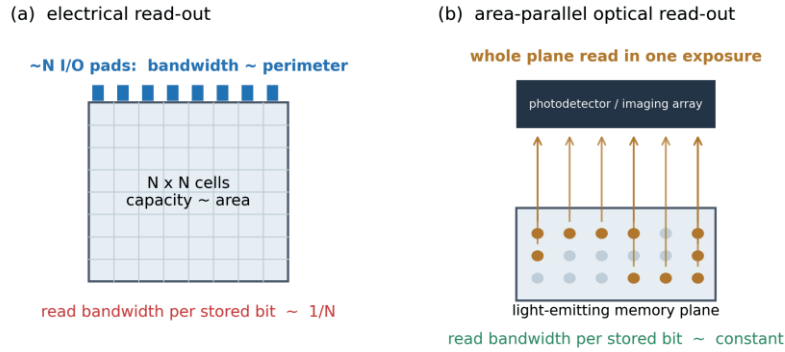
The case for optical read-out does not rest on latency but on a scaling asymmetry (Figure 5). The capacity of a memory die grows with its area, while the number of electrical connections available to read it grows with its perimeter; read bandwidth per stored bit therefore falls as arrays grow.

Optics can relax this asymmetry, although detector integration, collection efficiency and routing introduce separate limits. An imaging read-out addresses the array in two dimensions at once, so read bandwidth scales with area rather than perimeter; as a non-memristive existence proof, Project Silica reads sectors containing more than 100,000 spatially encoded voxels (Anderson et al., 2023), but whether comparable scaling is achievable in an electrically addressed, self-emissive LEM array remains untested.

An LEM element would be the missing transducer: the stored state is itself the optical signal.

Figure 5

The Read-Bandwidth Scaling Asymmetry



Note. (a) Electrical read-out is bounded by the number of input/output (I/O) connections, which scales with die perimeter while capacity scales with area. (b) Area-parallel optical read-out addresses the whole plane in one exposure. Original schematic; the underlying interconnect physics is that of Miller (2017).

5.2. Light Emission from Silicon-Based Structures

Whether this transducer can be realised depends on whether a CMOS-compatible memory material can emit light efficiently. Silicon quantum-dot (QD) light-emitting diode (LED) devices have reached 12.2% external quantum efficiency (EQE) in the red (Yamada et al., 2022), but these are large-area thin-film stacks, not nanoscale filaments.

Silicon-rich nitride occupies a favourable position: its photoluminescence quantum efficiency of 7% exceeds the 4.5% of a silicon-rich oxide reference (Dal Negro et al., 2006).

5.3. Electroformed Silicon Nitride Light-Emitting Memory

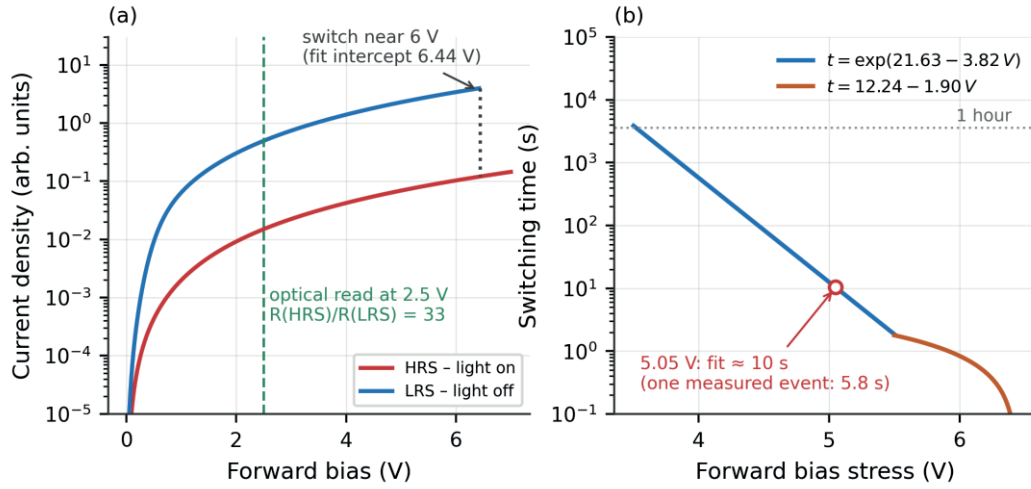
Silicon nitride is therefore an attractive host for combining memory and emission, as in the electroformed a-SiN_x:H heterojunction p⁺i n⁺ diode developed by the present authors' group and collaborators. The structure is glass / Cr / p⁺ hydrogenated nanocrystalline silicon (nc-Si:H, 90 nm) / intrinsic a-SiN_x:H (30 nm) / n⁺ nc-Si:H (90 nm) / indium tin oxide (ITO, 250 nm) with 1 mm diameter ITO dot contacts (Figure 4b). Electroforming is a single irreversible event: a forward bias of 14 V with the supply current limited to 50 mA converts the diode into a visible emitter whose electroluminescence intensity is enhanced by at least a factor of thirty (M. Anutgan et al., 2011).

Real-time optical microscopy establishes that light is emitted through the regions where residual ITO is present (M. Anutgan et al., 2020).

The memory behaviour was reported by T. Anutgan et al. (2017) and is summarised in Figure 6 and Table 3. Under forward bias the device sits in the HRS and emits light; a reverse sweep raises the current density into the LRS, which the device retains under forward bias without emission until, at about 6 V, the current collapses onto the original curve and emission resumes. Optical read requires sufficient photon flux, reached at 2.5 V, where the electrical resistance ratio of the two states is 33. Because self-emissive optical readout requires forward carrier injection (2.5 V), well above the read bias of 0.3 V or less recommended for a 2-bit hafnium oxide 1T1R array (Shim et al., 2020) and in the same forward polarity as reset kinetics, optical readout introduces an intrinsic read-disturb tension requiring pulse-level validation. Switching kinetics follow $t = \exp(21.63 - 3.82 V)$ below 5.50 V and $t = 12.24 - 1.90 V$ above it, with individual events scattering about these fits (e.g., 5.8 s at 5.05 V). Above about 6.44 V switching is immediate on the seconds scale; no sub-second time was measured. Extrapolating the exponential branch to zero bias gives an estimate exceeding 78 years; this is a kinetic extrapolation from room-temperature voltage stress, not a substitute for high-temperature retention bakes, which remain to be performed. Roughly 90 of 240 test devices were electroformed, with similar characteristics. We operationally define a dual-function LEM by three simultaneous criteria: (i) a non-volatile resistance state in the active element, (ii) state-dependent electroluminescence at a fixed, non-destructive read bias, and (iii) self-emission from the memory cell without a separate LED or laser.

Figure 6

(a) Schematic Bias Sequence and (b) Switching Time Kinetics of the Electroformed SiN_x Light-Emitting Memory



Note. (a) Schematic reconstruction of the reported bias sequence, showing the high-resistance light-on and low-resistance light-off branches, the 2.5 V optical read point and the transition near 6 V (6.44 V is the zero-time intercept of the linear kinetic fit, an empirical fit artifact of seconds-scale measurement resolution rather than an instantaneous physical switching threshold). Axes are qualitative. (b) Switching time against forward bias stress from the two fitted expressions; the open marker shows an individual measured switching event at 5.05 V (5.8 s). Original figure; data from T. Anutgan et al. (2017).

Table 3

Reported Parameters of the Electroformed $\alpha\text{-SiN}_x\text{:H}$ Light-Emitting Memory Placed Against the Requirements of Table 1

Parameter	Reported value	Status against Table 1
Electroforming	14 V forward, 50 mA compliance, single irreversible event	One-time; high variance operation (see G9)
Transition-voltage intercept	6.44 V (zero-time intercept of the linear kinetic fit); measured transition near 6 V	Not a measured write voltage; high relative to sub-1 V array periphery
Read voltage	2.5 V, set by photon detection not electrical margin	Far above the ≤ 0.3 V read bias recommended for a 2-bit HfO_2 1T1R array (Shim et al., 2020)
On/off ratio	33 at the optical read point	The ratio exceeds the dynamic-range threshold used in multilayer-perceptron simulations (on/off > 3; Mehonic et al., 2019), but no multilevel optical states have been demonstrated in this device
Switching time	Falls with bias: fitted ~ 1 h at 3.50 V, ~ 10 s at 5.05 V (one event 5.8 s); immediate on the seconds scale above ~ 6.44 V; no sub-second value measured	Seconds-scale resolution only; not comparable with the ns-scale RRAM entries of Table 2
Retention (zero-bias kinetic estimate)	>78 years at 25 °C, extrapolated from voltage-stress kinetics; elevated-temperature retention untested	Exceeds 10 years only as a 25 °C kinetic extrapolation; the 10 yr @ 85 °C target is untested

Parameter	Reported value	Status against Table 1
Emissive layer	i-SiN _x :H, 30 nm, Si nanocrystals 2.5–7 nm	CMOS-compatible deposition; emission lifetime not measured in this device (sub-ns–5 ns reported for Si-rich nitride films; Dal Negro et al., 2006)
Array status	Single cells; ~90 of 240 discrete devices formed	Single cells (~90 formed); no electroformed SiN _x array identified in the Box 1 search. (Multi-cell emissive arrays in that search: organic/QD light-emitting memristor, organic memory-computing LED and ferroelectric–lanthanide systems; see G18)

5.4. Other Classes of Light-Emitting and Optically Read Memory

The silicon nitride device is one of several approaches that couple a memory state to an optical signal. They differ in whether light is emitted by the memory cell itself, generated by a separate emitter, or used only to read or to program the cell, and are grouped accordingly below.

Series-integrated and multi-output light-emitting memories. Chang et al. (2014) described as a light-emitting memory a tandem Ag/SiO₂/graphene/SiO₂/p-GaN (p-type gallium nitride) structure with an electroluminescence threshold shifted from about 6 V in the LRS to about 8 V in the HRS.

Filament-localised electroluminescence. An Ag/a-SiO_x/Pt junction emits under 100 nA compliance, but photon counts track the current change ratio rather than absolute conductance (Cheng et al., 2022). Hamdad et al. (2023) found that volatile switching is dominated by hot-electron overbias emission, whereas a non-volatile gap state adds an inelastic-tunnelling contribution. Transient and stored-state regimes must therefore be separated experimentally.

Optical read-out without self-emission and emerging emissive arrays. An Ag/P3HT/ITO 4×4 crossbar (P3HT, poly(3-hexylthiophene)) is read at zero bias under 530 nm illumination (Z. Zhou et al., 2024, Fig. 4c); the 4×4 pattern that electrical read-out misidentifies because of sneak paths is read correctly by the optical method (Fig. 4d–f).

Emissive memristor arrays. Z. Chen et al. (2024) built an organic/QD light-emitting memristor (N-LEM) array of 50 physical devices; its states are pulse-programmed. Wen et al. (2026) reported an 8×8 ferroelectrically programmable lanthanide luminescent memristor array with 16 non-volatile upconversion photoluminescence levels under 980 nm excitation.

5.5. Maturity: An Explicit Gap Analysis

The available array demonstrations therefore belong to operationally distinct classes, and none of the records in Table 4 provides, in the material examined, the full metric set needed to evaluate self-emissive memory against electrical read-out. Electroformed SiN_x LEMs remain single cells (T. Anutgan et al., 2017), and ferroelectric-lanthanide arrays modulate 980 nm-pumped upconversion photoluminescence rather than self-emitting under bias (Wen et al., 2026).

Across the classes in Table 4, no record whose full text and relevant supplementary information were examined reports EQE or quantum yield, absolute optical power, collection efficiency, detector responsivity and total read energy together; fields that could not be checked are coded NV (not verifiable) in Table 4. A closed optical read-energy budget therefore cannot yet be constructed. In the electroformed cells the optical discriminator is on/off emission at a threshold, and optical retention under repeated reads, inter-emitter crosstalk and array-level variability remain uncharacterised. Figure 7 compares the reported physical device counts.

Three physical obstacles underlie this. The first is a current and area mismatch: the silicon-based emitters of Sections 5.2–5.3 are characterised on millimetre-scale contacts (T. Anutgan et al., 2017), whereas low-power filamentary switching operates at compliance currents from about 100 nA (Cheng et al., 2022) to tens of μ A (IEEE, 2023). The second is the correlation problem of Section 5.4. The third is detector integration: no monolithic device combining an LEM cell with its detector was identified among the records retrieved by our structured, non-exhaustive search (Box 1; completed 30 September 2026), and in the nearest hybrid the 120 μ m-diameter photodiode is some 57 times larger than the 200 μ m² memory cell (Pan et al., 2025).

6. Reliability, Manufacturing, and the Benchmarking Problem

Whether RRAM, and any optical variant of it, can progress from laboratory demonstration to product depends on reliability, reproducible benchmarking and manufacturing economics.

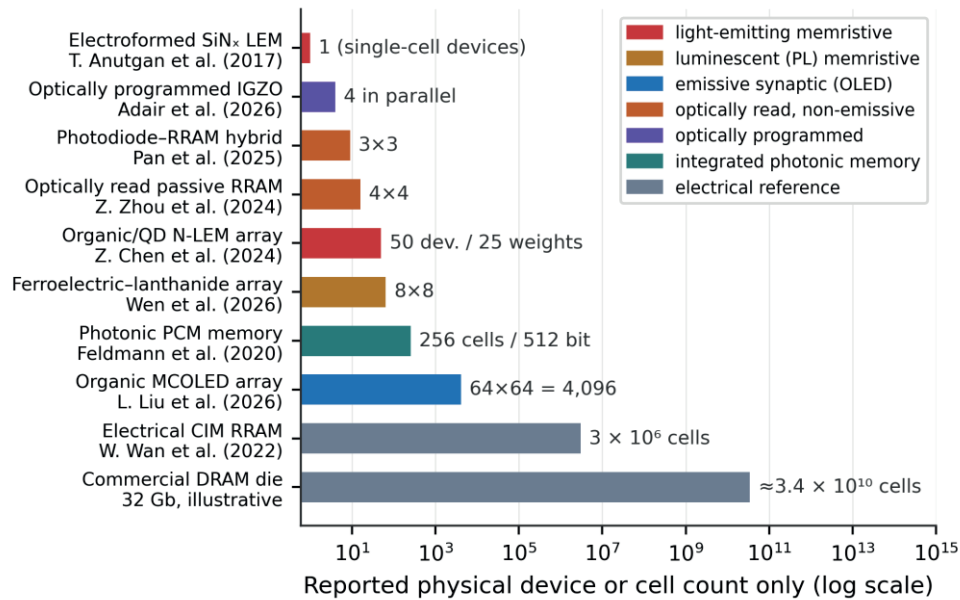
6.1. Retention, Read Disturb, and Relaxation

Retention in filamentary OxRAM is governed by thermally activated ion diffusion: 0.38 eV has been measured for TaO/HfO_x and TaO/AlO_x (tantalum, hafnium and aluminium oxide) self-rectifying stacks, with LRS retention extrapolated to more than five years at room temperature (Lin et al., 2017).

Read disturb is a second concern: Shim et al. (2020) conclude from conductance drift on a 2-bit HfO₂ 1T1R array that read voltage should be held at or below 0.3 V to maintain inference accuracy.

Figure 7

Reported Physical Devices or Cells Across Optoelectronic Memory Classes



Note. Counts are taken from the labelled sources: T. Anutgan et al. (2017), single-cell devices; Adair et al. (2026), 4 devices; Pan et al. (2025), 3×3; Z. Zhou et al. (2024), 4×4; Z. Chen et al. (2024), 50 physical devices; Wen et al. (2026), 8×8; Feldmann et al. (2020), 256 cells; L. Liu et al. (2026), 64×64 emissive synaptic array; W. Wan et al. (2022), 3 × 10⁶ cells; DRAM, an illustrative 32 Gb die (≈3.4 × 10¹⁰ cells). Original figure.

6.2. Variability and Endurance

Reliability is also shaped by device-to-device and cycle-to-cycle variation. Resistance variability in HfO₂ RRAM increases as programming current is reduced (Fantini et al., 2013); because lower compliance current also lowers array power, the two are in opposition.

Endurance shows a wide laboratory–product gap: laboratory devices have reached 10¹² cycles (Lee et al., 2011), while Intel’s embedded RRAM in 22FFL is specified at 10⁴ cycles with 10-year retention at 85 °C (Golonzka et al., 2019).

6.3. The Benchmarking and Reproducibility Problem

Comparing results across laboratories is itself a problem. Lanza et al. (2025) write that memristor performance ‘has been often misrepresented in articles that lack statistics and use non-standardized characterization and simulation protocols.’ No consensus standardisation or benchmarking framework was identified for optically read or light-emitting memristive devices in our structured, non-exhaustive literature search completed on 30 September 2026 (Box 1), where electrical protocols do not reach.

6.4. Manufacturing Economics and the Lesson of 3D XPoint

Even a reliable, well-benchmarked memory must be manufacturable at competitive cost. Resistive memory's decisive manufacturing advantage is that it can be inserted into an existing logic back-end with an additional mask, against ten or more for embedded flash (Wang, Song, et al., 2024). The production embedded resistive memory listed here is oxide-based: TSMC at 40 nm (mass production), 28 nm and 22 nm (Wang, Song, et al., 2024).

The commercial history of storage-class memory suggests that an intermediate latency tier is insufficient for a sustainable market position. Handy (2022) reports annual Optane revenue of approximately \$392 million in 2020; Intel attributes an operating loss of \$576 million to the business that year (Intel Corporation, 2022a); Intel began winding the business down in the second quarter of 2022, recognising a \$559 million inventory impairment (Intel Corporation, 2022b). Handy (2022) estimates DRAM output at approximately 18 million wafers per year and peak 3D XPoint output at approximately 0.15 million wafers per year; these are industry estimates rather than manufacturer-disclosed production data.

7. Open Research Gaps

The gaps identified in Sections 3–6 are collected below as twenty-three items (G1–G23). G18–G23 concern light-emitting and optically read memory, and their absence statements are scoped to the records retrieved by the search described in Box 1.

Materials and Device Physics

- G1.** State: an activation energy of 0.38 eV was extracted for retention (Lin et al., 2017). Target: a model linking deposition and anneal conditions to the activation-energy tail.
- G2.** State: lowering programming current broadens resistance distributions (Fantini et al., 2013). Target: sub- μ A programming with a log-resistance spread comparable to that at tens of μ A.
- G3.** State: compact reliability models have been proposed (Y. Liu et al., 2021). Target: a physical model that predicts the conductance distribution and long-term drift.
- G4.** State: in HfO₂-based RRAM, random telegraph noise and discrete defect interactions produce state-dependent current fluctuations that may reduce multilevel margins. Target: a protocol-independent scaling limit for these margins (see Puglisi et al., 2016).
- G5.** Four accounts of silicon nitride switching coexist. A discriminating experiment would combine area scaling, temperature-dependent transport and hydrogen-content variation.

- G6.** Target: a foundry-scale macro based on non-filamentary, area-dependent switching of the interface type described by Sawa (2008).

Array Architecture

- G7.** Target: a selector for dense passive arrays.
- G8.** Target: an experimentally demonstrated, selector-free passive RRAM crossbar above 1 Mb with sneak-path and IR-drop (current–resistance voltage drop) mitigation.
- G9.** State: forming has been addressed architecturally in a 40 nm macro (Chiu et al., 2019). Target: forming-free stacks with production-grade memory windows.

Reliability and Error Management

- G10.** Target: array-level bit-error-rate data to 10^{-9} across full endurance, retention bake and worst-case temperature.
- G11.** State: error correction for RRAM CIM has been reported in a 40 nm, 64 Kb macro (Crafton et al., 2021). Target: an error-correcting code framework with validation on a physical array.
- G12.** State: reprogramming only when drift exceeds a threshold has been evaluated by experiment-augmented circuit simulation with the Simulation Program with Integrated Circuit Emphasis (SPICE; Yaldagard et al., 2026). Target: array-scale silicon validation of such schemes.

Modelling and Design Automation

- G13.** Target: an openly shared measured dataset against which compact-model statistical parameters are jointly calibrated.
- G14.** State: open ReRAM design resources exist (Giacomin & Gaillardon, 2019; Mifsud & Constandinou, 2023; SkyWater Technology & Google, n.d.). Target: a process design kit that exposes the joint distribution of conductance state, programming energy and retention.

System Integration and Benchmarking

- G15.** Target: an iso-accuracy, iso-technology benchmark for memristive accelerators whose energy accounting includes converters (Section 4.2), forming circuitry and write energy.
- G16.** State: on-device self-calibration of a physical 32×32 memristor array has been reported (Jeong et al., 2025). Target: an autonomous toolchain that recalibrates a deployed RRAM array.
- G17.** State: standardised characterisation protocols have been published (Lanza et al., 2019, 2021). Target: their routine adoption.

- G18.** State: no electroformed SiN_x LEM array was identified among the records retrieved by our structured, non-exhaustive search (Box 1; completed 30 September 2026). Target: an electroformed SiN_x array with fixed-bias non-disturb optical read-out.
- G19.** State: no retained record for which the relevant full text and supplementary information of the cited source were accessible reports EQE or quantum yield, absolute optical output, collection efficiency, detector responsivity and total read energy together; fields that could not be checked are coded NV in Table 4. Target: these quantities measured together.
- G20.** State: electrically programmed synaptic states with optical output have been demonstrated (Z. Chen et al., 2024; L. Liu et al., 2026). Target: a non-volatile emission-intensity-versus-conductance transfer curve at a fixed, non-disturb read bias.
- G21.** State: for the 8×8 ferroelectric–lanthanide array, nearest-neighbour optical crosstalk is estimated to be below 10⁻⁵ (Wen et al., 2026). Target: measured crosstalk for a scaled, self-emissive LEM array.
- G22.** State: non-volatile photoluminescence-state retention exceeding 27 hours has been reported for ferroelectric–lanthanide devices (Wen et al., 2026). Target: electrical and optical retention measured together.
- G23.** State: technology-agnostic electrical benchmarking methodologies exist (Stathopoulos et al., 2019), but among the records retrieved by our structured, non-exhaustive search (Box 1; completed 30 September 2026) no extension specific to self-emissive or luminescence-read memory was identified. Target: standardised protocols for optical read-out.

Table 4

Metric Coverage for Representative Records of the Optoelectronic Memory Classes in the Structured Literature Search

Metric	Electroformed SiN _x LEM	Series-integrated Ag/SiO ₂ /graphene/GaN	CsPbBr ₃ QD LEM	Organic/QD N-LEM	Organic MCOLED array
Reference	T. Anutgan et al. (2017)	Chang et al. (2014)	Yen et al. (2021)	Z. Chen et al. (2024)	L. Liu et al. (2026)
Electrical state variable	Resistance (HRS/LRS)	Resistance	Resistance	Synaptic conductance	Synaptic conductance (5-bit)
Optical state variable	EL on/off	EL turn-on threshold	EL on/off	EL intensity	EL (circularly polarised)
Excitation source	Electrical	Electrical	Electrical	Electrical pulses	Electrical
EL or PL	EL	EL	EL	EL	EL

Metric	Electroformed SiN _x LEM	Series-integrated Ag/SiO ₂ /graphene/ GaN	CsPbBr ₃ QD LEM	Organic/QD N-LEM	Organic MCOLED array
Fixed-bias read	2.5 V detection point (Table I); repeated-read disturb NR	NR (threshold shift)	Parallel read, paired series devices; non-disturb NR	NR (pulsed)	NV
Retention	NR measured; >78 yr extrapolated	NR	>10 ⁵ s electrical, 0.1 V read (Yen et al., 2021, Suppl. Fig. 7a); optical NR	Volatile; decay over 71.6 s (Z. Chen et al., 2024, Suppl. Fig. S15)	>10 ⁴ s, quasi-non-volatile (L. Liu et al., 2026, Suppl. Fig. 25)
EQE / PLQY	NR	NR	NR	NR	max. EQE 0.94% (L. Liu et al., 2026, Suppl. Table 2)
Absolute optical power	NR	NR	NV	NV	NV (peak luminance 22,306 cd m ⁻²)
Array size	Single cells	Single cell	Single device pair	50 devices (25 weights)	64×64; 100×180 display
Crosstalk	NA	NA	NA	NR	NV
Metric (cont.)	Ferroelectric–lanthanide luminescent memristor	Filament photon source	Optically read RRAM	All-optical memlumor	
Reference	Wen et al. (2026)	Cheng et al. (2022)	Z. Zhou et al. (2024)	Marunchenko et al. (2024)	
Electrical state variable	Ferroelectric domain/polarisation state	Filament conductance	Resistance	NA	
Optical state variable	Upconversion PL intensity (16 levels)	Photon count rate	Photocurrent	PLQY	
Excitation source	External 980 nm pump	Electrical	External 530 nm light	External optical pump	
EL or PL	PL (upconversion)	EL	NA (photocurrent)	PL	
Fixed-bias read	Optical (980 nm pump, CMOS); no bias	0.5 V electrical; optical NR	0 V	NA	
Retention	>27 h	Volatile LRS	>10 ⁴ s electrical (Z. Zhou et al., 2024, Suppl. Fig. S2a–d)	ns–min; days	
EQE / PLQY	UCQY 4.38% → 6.25%	NR (sim. radiation eff.)	NA		PLQY ≈20%
Absolute optical power	NV (normalised PL in Wen et al., 2026, SI)	NR	NA		NV

Metric (cont.)	Ferroelectric–lanthanide luminescent memristor	Filament photon source	Optically read RRAM	All-optical memlumor
Array size	8×8	Single junction	4×4	No addressed array
Crosstalk	<10 ⁻⁵ (est., 400 μm pitch)	NA	Not quantified; 16 cells read correctly (Fig. 4d–f)	NA

Note. NR = not reported in the examined full text and supplementary information of the cited source; UCQY = upconversion quantum yield; EL = electroluminescence; PL = photoluminescence; PLQY = photoluminescence quantum yield; MCOLED = memory-computing organic light-emitting diode; CsPbBr₃ = caesium lead bromide; NV = not verifiable from the material accessible at the date of search; NA = not applicable to the device class; Suppl. = Supplementary; SI = Supplementary Information; all Supplementary Figures, Tables and Notes cited in this chapter belong to the cited primary articles, and no supplementary information accompanies this chapter; ‘preprint’ identifies values verified only against an author preprint. The reported UCQY values are source-specific and depend on excitation power density, optical collection and calibration conditions; they are not directly comparable with EQE or PLQY values measured under different protocols.

Box 1. Literature search protocol (scope: light-emitting and optically read memory only).

Date of last search: 30 September 2026. Sources: Crossref and OpenAlex metadata search, arXiv, general web search, and publisher full-text pages. Inclusion: devices with an electrically or optically addressable memory state and a state-dependent optical output, optical read-out or optical programming response. Exclusion: purely electrical memories, purely photonic interconnects, and simulations without fabricated devices. Classification used the operational definition of Section 5.3 (criteria i–iii). The search is structured and non-exhaustive; absence statements about light-emitting and optically read memory refer to the records it identified. Eighteen records were retained for screening; nine are tabulated in Table 4, and Pan et al. (2025) and Adair et al. (2026) are used in Figure 7 but are not tabulated; Feldmann et al. (2020) appears in Figure 7 only as an integrated photonic-memory comparator. Yang et al. (2026) was not included in Table 4 because fixed-bias, non-disturb optical storage read-out (criterion (ii) of Section 5.3) could not be verified. Per-query retrieval and de-duplication counts were not logged.

8. Conclusion

This chapter has examined whether resistive switching memory, and a light-emitting variant of it, can meet the memory requirements of AI and cloud computing. Resistive switching memory is in production as embedded non-volatile memory, for example at TSMC (Section 6.4), and has been integrated at the three-million-device scale for analogue CIM with fully hardware-measured accuracy. The evidence reviewed here does not yet establish resistive memory as an economically competitive replacement for standalone mass storage.

Multi-cell optoelectronic memristive arrays are realisable, but no electroformed SiN_x array was identified among the records retrieved by our structured, non-exhaustive search (Box 1; completed 30 September 2026).

Until a CMOS-compatible, self-emissive array demonstrates fixed-bias non-disturb optical read-out together with efficiency, retention, variability and crosstalk statistics, the system-level case for LEM will remain promising but unproven.

References

- Adair, A., Robertson, J., Tsiamis, A., Awadein, M., Stathopoulos, S., Herrnsdorf, J., Dawson, M. D., Prodromakis, T., & Hurtado, A. (2026). Parallel spatial photonic programming of optoelectronic IGZO RRAM with a compact μ LED array. arXiv preprint arXiv:2608.16807. <https://doi.org/10.48550/arXiv.2608.16807>
- Ambrogio, S., Narayanan, P., Okazaki, A., Fasoli, A., Mackin, C., Hosokawa, K., ... Burr, G. W. (2023). An analog-AI chip for energy-efficient speech recognition and transcription. *Nature*, 620(7975), 768–775. <https://doi.org/10.1038/s41586-023-06337-5>
- Anderson, P., Aranas, E. B., Assaf, Y., Behrendt, R., Black, R., Caballero, M., Cameron, P., Canakci, B., de Carvalho, T., Chatzieleftheriou, A., Clarke, R. S., Clegg, J., Cletheroe, D., Cooper, B., Deegan, T., Donnelly, A., Drevinskas, R., Gaunt, A., Gkantsidis, C., ... Winzeck, S. (2023). Project Silica: Towards sustainable cloud archival storage in glass. In *Proceedings of the 29th Symposium on Operating Systems Principles* (pp. 166–181). Association for Computing Machinery. <https://doi.org/10.1145/3600006.3613208>
- Anutgan, M., Anutgan, T., & Atilgan, I. (2020). SEM, EDX spectroscopy and real-time optical microscopy of electroformed silicon nitride-based light emitting memory device. *The European Physical Journal Applied Physics*, 89(1), 10303. <https://doi.org/10.1051/epjap/2020190300>
- Anutgan, M., Anutgan, T., Atilgan, I., & Katircioglu, B. (2011). Electroforming of amorphous silicon nitride heterojunction pin visible light emitter. *IEEE Transactions on Electron Devices*, 58(8), 2537–2543. <https://doi.org/10.1109/TED.2011.2156409>
- Anutgan, T., Anutgan, M., & Atilgan, I. (2019). Transmission electron microscope imaging of plasma grown electroformed silicon nitride-based light emitting diode for direct examination of nanocrystallization. *The European Physical Journal Applied Physics*, 88(3), 30102. <https://doi.org/10.1051/epjap/2020190298>
- Anutgan, T., Anutgan, M., Atilgan, I., & Katircioglu, B. (2017). Electroformed silicon nitride based light emitting memory device. *Applied Physics Letters*, 111(5), 053502. <https://doi.org/10.1063/1.4997029>

- Böttger, U., von Witzleben, M., Havel, V., Fleck, K., Rana, V., Waser, R., & Menzel, S. (2020). Picosecond multilevel resistive switching in tantalum oxide thin films. *Scientific Reports*, 10, 16391. <https://doi.org/10.1038/s41598-020-73254-2>
- Chang, C.-W., Tan, W.-C., Lu, M.-L., Pan, T.-C., Yang, Y.-J., & Chen, Y.-F. (2014). Electrically and optically readable light emitting memories. *Scientific Reports*, 4, 5121. <https://doi.org/10.1038/srep05121>
- Chen, T., Leng, K., Ma, Z., Jiang, X., Chen, K., Li, W., Xu, J., & Xu, L. (2023). Tracing the Si dangling bond nanopathway evolution in a-SiN_x:H resistive switching memory by the transient current. *Nanomaterials*, 13(1), 85. <https://doi.org/10.3390/nano13010085>
- Chen, Z., Lin, Z., Yang, J., Chen, C., Liu, D., Shan, L., Hu, Y., Guo, T., & Chen, H. (2024). Cross-layer transmission realized by light-emitting memristor for constructing ultra-deep neural network with transfer learning ability. *Nature Communications*, 15(1), 1930. <https://doi.org/10.1038/s41467-024-46246-3>
- Cheng, B., Zellweger, T., Malchow, K., Zhang, X., Lewerenz, M., Passerini, E., ... Leuthold, J. (2022). Atomic scale memristive photon source. *Light: Science & Applications*, 11, 78. <https://doi.org/10.1038/s41377-022-00766-z>
- Chiu, Y.-C., Hu, H.-W., Lai, L.-Y., Huang, T.-Y., Kao, H.-Y., Chang, K.-T., Ho, M.-S., Chou, C.-C., Chih, Y.-D., Chang, T.-Y., & Chang, M.-F. (2019). A 40nm 2Mb ReRAM macro with 85% reduction in FORMING time and 99% reduction in page-write time using auto-FORMING and auto-write schemes. In 2019 Symposium on VLSI Technology (pp. T232–T233). IEEE. <https://doi.org/10.23919/VLSIT.2019.8776540>
- Crafton, B., Spetalnick, S., Yoon, J.-H., Wu, W., Tokunaga, C., De, V., & Raychowdhury, A. (2021). CIM-SECDED: A 40nm 64Kb compute in-memory RRAM macro with ECC enabling reliable operation. In 2021 IEEE Asian Solid-State Circuits Conference (A-SSCC) (pp. 1–3). IEEE. <https://doi.org/10.1109/A-SSCC53895.2021.9634742>
- Dal Negro, L., Yi, J. H., Michel, J., Kimerling, L. C., Chang, T.-W. F., Sukhovatkin, V., & Sargent, E. H. (2006). Light emission efficiency and dynamics in silicon-rich silicon nitride films. *Applied Physics Letters*, 88(23), 233109. <https://doi.org/10.1063/1.2208378>

- Dutta, S., Ye, H., Aabrar, K. A., Kirtania, S. G., Khanna, A., Ni, K., & Datta, S. (2022). Logic compatible high-performance ferroelectric transistor memory. *IEEE Electron Device Letters*, 43(3), 382–385. <https://doi.org/10.1109/LED.2022.3148669>
- Fantini, A., Goux, L., Degraeve, R., Wouters, D. J., Raghavan, N., Kar, G., Belmonte, A., Chen, Y.-Y., Govoreanu, B., & Jurczak, M. (2013). Intrinsic switching variability in HfO₂ RRAM. In 2013 5th IEEE International Memory Workshop (pp. 30–33). IEEE. <https://doi.org/10.1109/IMW.2013.6582090>
- Feldmann, J., Youngblood, N., Li, X., Wright, C. D., Bhaskaran, H., & Pernice, W. H. P. (2020). Integrated 256 cell photonic phase-change memory with 512-bit capacity. *IEEE Journal of Selected Topics in Quantum Electronics*, 26(2), 8301207. <https://doi.org/10.1109/JSTQE.2019.2956871>
- Gholami, A., Yao, Z., Kim, S., Hooper, C., Mahoney, M. W., & Keutzer, K. (2024). AI and memory wall. *IEEE Micro*, 44(3), 33–39. <https://doi.org/10.1109/MM.2024.3373763>
- Giacomin, E., & Gaillardon, P.-E. (2019). A resistive random access memory add-on for the NCSU FreePDK 45 nm. *IEEE Transactions on Nanotechnology*, 18, 68–72. <https://doi.org/10.1109/TNANO.2018.2881109>
- Golonzka, O., Arslan, U., Bai, P., Bohr, M., Baykan, O., Chang, Y., Chaudhari, A., Chen, A., Clarke, J., Connor, C., Das, N., English, C., Ghani, T., Hamzaoglu, F., Hentges, P., Jain, P., Jezewski, C., Karpov, I., Kothari, H., ... Fischer, K. (2019). Non-volatile RRAM embedded into 22FFL FinFET technology. In 2019 Symposium on VLSI Technology (pp. T230–T231). IEEE. <https://doi.org/10.23919/VLSIT.2019.8776570>
- Hamdad, S., Malchow, K., Avetisyan, D., Dujardin, E., Bouhelier, A., Zhou, Y., Cheng, B., Zellweger, T., & Leuthold, J. (2023). Overbias and quantum tunneling in light-emitting memristors. *Physical Review Applied*, 20(2), 024057. <https://doi.org/10.1103/PhysRevApplied.20.024057>
- Han, S., Liu, X., Mao, H., Pu, J., Pedram, A., Horowitz, M. A., & Dally, W. J. (2016). EIE: Efficient inference engine on compressed deep neural network. In Proceedings of the 43rd International Symposium on Computer Architecture (pp. 243–254). IEEE. <https://doi.org/10.1109/ISCA.2016.30>

- Handy, J. (2022, September 16). Why ‘emerging’ memories have not succeeded — yet. eeNews Europe. <https://www.eenewseurope.com/en/why-emerging-memories-have-not-succeeded-yet/>
- Hellenbrand, M., Teck, I., & MacManus-Driscoll, J. L. (2024). Progress of emerging non-volatile memory technologies in industry. *MRS Communications*, 14(6), 1099–1112. <https://doi.org/10.1557/s43579-024-00660-2>
- Horowitz, M. (2014). Computing’s energy problem (and what we can do about it). In *IEEE International Solid-State Circuits Conference Digest of Technical Papers* (pp. 10–14). IEEE. <https://doi.org/10.1109/ISSCC.2014.6757323>
- IEEE. (2023). International Roadmap for Devices and Systems (IRDS) 2023 update: Beyond CMOS and emerging materials integration. IEEE. <https://irds.ieee.org/editions/2023>
- Intel Corporation. (2022a). Form 10-K: Annual report for the fiscal year ended December 25, 2021. U.S. Securities and Exchange Commission. <https://www.sec.gov/Archives/edgar/data/50863/000005086322000007/intc-20211225.htm>
- Intel Corporation. (2022b). Form 10-Q: Quarterly report for the quarterly period ended July 2, 2022. U.S. Securities and Exchange Commission. <https://www.sec.gov/Archives/edgar/data/50863/000005086322000030/intc-20220702.htm>
- International Energy Agency [IEA]. (2025). Energy and AI. IEA. <https://www.iea.org/reports/energy-and-ai>
- Izraelevitz, J., Yang, J., Zhang, L., Kim, J., Liu, X., Memaripour, A., ... Swanson, S. (2019). Basic performance measurements of the Intel Optane DC persistent memory module. arXiv preprint arXiv:1903.05714. <https://doi.org/10.48550/arXiv.1903.05714>
- Jeong, H., Han, S., Park, S.-O., Kim, T. R., Bae, J., Jang, T., Cho, Y., Seo, S., Jeong, H.-J., Park, S., Park, T., Oh, J., Park, J., Koh, K., Kim, K.-H., Jeon, D., Kwon, I., Yoon, Y.-G., & Choi, S. (2025). Self-supervised video processing with self-calibration on an analogue computing platform based on a selector-less memristor array. *Nature Electronics*, 8(2), 168–178. <https://doi.org/10.1038/s41928-024-01318-6>

- Jiang, M., Xu, Y., Li, Z., & Li, C. (2025). Current opinions on memristor-accelerated machine learning hardware. *Current Opinion in Solid State and Materials Science*, 37, 101226. <https://doi.org/10.1016/j.cossms.2025.101226>
- Joshi, V., Le Gallo, M., Haefeli, S., Boybat, I., Nandakumar, S. R., Piveteau, C., ... Eleftheriou, E. (2020). Accurate deep neural network inference using computational phase-change memory. *Nature Communications*, 11, 2473. <https://doi.org/10.1038/s41467-020-16108-9>
- Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., ... Yoon, D. H. (2017). In-datacenter performance analysis of a tensor processing unit. In *Proceedings of the 44th International Symposium on Computer Architecture* (pp. 1–12). ACM. <https://doi.org/10.1145/3079856.3080246>
- Lanza, M., Pazos, S., & Aguirre, F. (2025). The Achilles' heel of memristive technologies. *Nature Reviews Electrical Engineering*, 2(10), 654–656. <https://doi.org/10.1038/s44287-025-00207-0>
- Lanza, M., Waser, R., Ielmini, D., Yang, J. J., Goux, L., Suñé, J., ... Pazos, S. (2021). Standards for the characterization of endurance in resistive switching devices. *ACS Nano*, 15(11), 17214–17231. <https://doi.org/10.1021/acsnano.1c06980>
- Lanza, M., Wong, H.-S. P., Pop, E., Ielmini, D., Strukov, D., Regan, B. C., ... Shi, Y. (2019). Recommended methods to study resistive switching devices. *Advanced Electronic Materials*, 5(1), 1800143. <https://doi.org/10.1002/aelm.201800143>
- LaPedus, M. (2019, November 21). DRAM scaling challenges grow. *Semiconductor Engineering*. <https://semiengineering.com/dram-scaling-challenges-grow/>
- Larcher, L., & Padovani, A. (2017). Multiscale modeling of oxide RRAM devices for memory applications: From material properties to device performance. *Journal of Computational Electronics*, 16(4), 1077–1084. <https://doi.org/10.1007/s10825-017-1095-3>
- Le Gallo, M., Khaddam-Aljameh, R., Stanisavljevic, M., Vasilopoulos, A., Kersting, B., Dazzi, M., ... Sebastian, A. (2023). A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference. *Nature Electronics*, 6(9), 680–693. <https://doi.org/10.1038/s41928-023-01010-1>

- Lee, M.-J., Lee, C. B., Lee, D., Lee, S. R., Chang, M., Hur, J. H., Kim, Y.-B., Kim, C.-J., Seo, D. H., Seo, S., Chung, U.-I., Yoo, I.-K., & Kim, K. (2011). A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta₂O_{5-x}/TaO_{2-x} bilayer structures. *Nature Materials*, 10(8), 625–630. <https://doi.org/10.1038/nmat3070>
- Li, H., Berger, D. S., Hsu, L., Ernst, D., Zardoshti, P., Novakovic, S., ... Bianchini, R. (2023). Pond: CXL-based memory pooling systems for cloud platforms. In *Proceedings of ASPLOS '23* (Vol. 2, pp. 574–587). ACM. <https://doi.org/10.1145/3575693.3578835>
- Lin, Y.-D., Chen, P.-S., Lee, H.-Y., Chen, Y.-S., Rahaman, S. Z., Tsai, K.-H., Hsu, C.-H., Chen, W.-S., Wang, P.-H., King, Y.-C., & Lin, C. J. (2017). Retention model of TaO/HfO_x and TaO/AlO_x RRAM with self-rectifying switch characteristics. *Nanoscale Research Letters*, 12, 407. <https://doi.org/10.1186/s11671-017-2179-5>
- Liu, L., Chen, Z., Ma, X., Li, X., Cai, Y., Liu, S., Huang, W., Liu, D., Lin, Z., Zhang, J., Hu, Y., Wei, Z., & Chen, H. (2026). Circularly polarized bulk-heterojunction memory-computing organic light-emitting diodes for a neural radiance field-based 3D display. *Nature Photonics*, 20(10), 1182–1192. <https://doi.org/10.1038/s41566-026-01966-4>
- Liu, Y., Zhao, M., Gao, B., Hu, R., Zhang, W., Yang, S., Yao, P., Xu, F., Xi, Y., Zhang, Q., Tang, J., Qian, H., & Wu, H. (2021). Compact reliability model of analog RRAM for computation-in-memory device-to-system codesign and benchmark. *IEEE Transactions on Electron Devices*, 68(6), 2686–2692. <https://doi.org/10.1109/TED.2021.3069746>
- Marunchenko, A., Kumar, J., Kiligaridis, A., Tatarinov, D., Pushkarev, A., Vaynzof, Y., & Scheblykin, I. G. (2024). Memlumor: A luminescent memory device for energy-efficient photonic neuromorphic computing. *ACS Energy Letters*, 9(5), 2075–2082. <https://doi.org/10.1021/acsenergylett.4c00691>
- Mavropoulis, A. E., Karakolis, P., Vasileiadis, N., Sygellou, L., Stavroulakis, E., Ioannou-Sougleridis, V., Normand, P., Sirakoulis, G. Ch., & Dimitrakakis, P. (2025). Resistance switching properties of stoichiometric and nitrogen implanted silicon nitride nanolayers on n and p-type Si substrates. In *2025 IEEE 25th International Conference on Nanotechnology (NANO)* (pp. 477–481). IEEE. <https://doi.org/10.1109/NANO63165.2025.11113675>

- Mehonic, A., Joksas, D., Ng, W. H., Buckwell, M., & Kenyon, A. J. (2019). Simulation of inference accuracy using realistic RRAM devices. *Frontiers in Neuroscience*, 13, 593. <https://doi.org/10.3389/fnins.2019.00593>
- Mehonic, A., Shluger, A. L., Gao, D., Valov, I., Miranda, E., Ielmini, D., Bricalli, A., Ambrosi, E., Li, C., Yang, J. J., Xia, Q., & Kenyon, A. J. (2018). Silicon oxide (SiO_x): A promising material for resistance switching? *Advanced Materials*, 30(43), 1801187. <https://doi.org/10.1002/adma.201801187>
- Menzel, S., Tappertzhofen, S., Waser, R., & Valov, I. (2013). Switching kinetics of electrochemical metallization memory cells. *Physical Chemistry Chemical Physics*, 15(18), 6945–6952. <https://doi.org/10.1039/C3CP50738F>
- Mifsud, A., & Constandinou, T. G. (2023). Towards a CMOS-process-portable ReRAM PDK. In 2023 21st IEEE Interregional NEWCAS Conference (NEWCAS) (pp. 1–5). IEEE. <https://doi.org/10.1109/NEWCAS57931.2023.10198171>
- Miller, D. A. B. (2017). Attojoule optoelectronics for low-energy information processing and communications. *Journal of Lightwave Technology*, 35(3), 346–396. <https://doi.org/10.1109/JLT.2017.2647779>
- NVIDIA Corporation. (2022). NVIDIA H100 Tensor Core GPU datasheet. <https://resources.nvidia.com/en-us-hopper-architecture/nvidia-tensor-core-gpu-datasheet>
- Pan, W., Wang, L., Tang, J., Huang, H., Hao, Z., Sun, C., ... Luo, Y. (2025). Optoelectronic array of photodiodes integrated with RRAMs for energy-efficient in-sensor computing. *Light: Science & Applications*, 14, 48. <https://doi.org/10.1038/s41377-025-01743-y>
- Puglisi, F. M., Larcher, L., Padovani, A., & Pavan, P. (2016). Anomalous random telegraph noise and temporary phenomena in resistive random access memory. *Solid-State Electronics*, 125, 204–213. <https://doi.org/10.1016/j.sse.2016.07.019>
- Rachidi, S., & Rosmeulen, M. (2025, November 7). Unlocking z-pitch scaling for next-generation 3D NAND flash. imec. <https://www.imec-int.com/en/articles/unlocking-z-pitch-scaling-next-generation-3d-nand-flash>

- Sawa, A. (2008). Resistive switching in transition metal oxides. *Materials Today*, 11(6), 28–36. [https://doi.org/10.1016/S1369-7021\(08\)70119-6](https://doi.org/10.1016/S1369-7021(08)70119-6)
- Sebastian, A., Le Gallo, M., Khaddam-Aljameh, R., & Eleftheriou, E. (2020). Memory devices and applications for in-memory computing. *Nature Nanotechnology*, 15(7), 529–544. <https://doi.org/10.1038/s41565-020-0655-z>
- Sevilla, J., & Roldán, E. (2024). Training compute of frontier AI models grows by 4-5x per year. Epoch AI. <https://epoch.ai/blog/training-compute-of-frontier-ai-models-grows-by-4-5x-per-year>
- Shafiee, A., Nag, A., Muralimanohar, N., Balasubramonian, R., Strachan, J. P., Hu, M., ... Srikumar, V. (2016). ISAAC: A convolutional neural network accelerator with in-situ analog arithmetic in crossbars. In *Proceedings of the 43rd International Symposium on Computer Architecture* (pp. 14–26). IEEE. <https://doi.org/10.1109/ISCA.2016.12>
- Shilov, A. (2025, August 6). Inside the future of 3D NAND: The roadmap to 500 layers. Tom's Hardware. <https://www.tomshardware.com/pc-components/storage/inside-the-future-of-3d-nand-the-roadmap-to-500-layers>
- Shim, W., Luo, Y., Seo, J.-S., & Yu, S. (2020). Impact of read disturb on multilevel RRAM based inference engine: Experiments and model prediction. In *2020 IEEE International Reliability Physics Symposium (IRPS)* (pp. 1–5). IEEE. <https://doi.org/10.1109/IRPS45951.2020.9129252>
- SkyWater Technology & Google. (n.d.). SKY130 ReRAM and examples (SkyWater provided) (Version 2.0.3) [Process design kit library]. GitHub. Retrieved September 30, 2026, from https://github.com/google/skywater-pdk-libs-sky130_fd_pr_reram
- Somala, V. (2026, May 21). Memory has grown to nearly two-thirds of AI chip component costs. Epoch AI. <https://epoch.ai/data-insights/ai-chip-component-cost-shares>
- Stathopoulos, S., Michalas, L., Khiat, A., Serb, A., & Prodromakis, T. (2019). An electrical characterisation methodology for benchmarking memristive device technologies. *Scientific Reports*, 9, 19412. <https://doi.org/10.1038/s41598-019-55322-4>
- Valov, I., Waser, R., Jameson, J. R., & Kozicki, M. N. (2011). Electrochemical metallization memories — Fundamentals, applications, prospects. *Nanotechnology*, 22(25), 254003. <https://doi.org/10.1088/0957-4484/22/25/254003>

- Vázquez Jiménez, A., Moreno Moreno, M., Palacios Huerta, L., Rosales Quintero, P., & Morales Sánchez, A. (2026). Low-voltage and high-reliability resistive switching in $\text{Si}_3\text{N}_4\text{:Si-NCs}$ memristor structures. *Journal of Semiconductors*, 47(4), 042305. <https://doi.org/10.1088/1674-4926/25100015>
- Wan, W., Kubendran, R., Schaefer, C., Eryilmaz, S. B., Zhang, W., Wu, D., ... Cauwenberghs, G. (2022). A compute-in-memory chip based on resistive random-access memory. *Nature*, 608(7923), 504–512. <https://doi.org/10.1038/s41586-022-04992-8>
- Wan, Z., Crafton, B., Spetalnick, S., Yoon, J.-H., & Raychowdhury, A. (2022, May). RRAM-ECC: Improving reliability of RRAM-based compute in-memory [Paper presentation]. Non-Volatile Memories Workshop (NVMW 2022), La Jolla, CA, United States. <https://nvmw.ucsd.edu/onlineprogram/>
- Wang, Y., Wang, S., Parsons, R., Novick, A., Gopal, V., Jang, K., ... Bergman, K. (2024). Silicon photonics chip I/O for ultra high-bandwidth and energy-efficient die-to-die connectivity. In *IEEE Custom Integrated Circuits Conference* (pp. 1–8). IEEE. <https://doi.org/10.1109/CICC60959.2024.10529018>
- Wang, Z., Luo, T., Liu, C., Liu, W., Goh, R. S. M., & Wong, W.-F. (2025). Enabling energy-efficient deployment of large language models on memristor crossbar: A synergy of large and small. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(2), 916–933. <https://doi.org/10.1109/TPAMI.2024.3483654>
- Wang, Z., Song, Y., Zhang, G., Luo, Q., Xu, K., Gao, D., Yu, B., Loke, D., Zhong, S., & Zhang, Y. (2024). Advances of embedded resistive random access memory in industrial manufacturing and its potential applications. *International Journal of Extreme Manufacturing*, 6(3), 032006. <https://doi.org/10.1088/2631-7990/ad2fea>
- Waser, R., Dittmann, R., Staikov, G., & Szot, K. (2009). Redox-based resistive switching memories — Nanoionic mechanisms, prospects, and challenges. *Advanced Materials*, 21(25–26), 2632–2663. <https://doi.org/10.1002/adma.200900375>
- Wen, Y., Chen, Z., Cao, Y., Ma, W., Ren, H., Du, X., Hu, M., Feng, Y., Zhang, F., Li, X., Wu, Z., Liu, W., Hao, J., & Zhang, Y. (2026). Ferroelectrically programmable lanthanide luminescent memristor. *Nature Communications*, 17. Advance online publication. <https://doi.org/10.1038/s41467-026-77397-0>

- WikiChip Fuse. (2022, December 12). IEDM 2022: Did we just witness the death of SRAM? WikiChip Fuse. <https://fuse.wikichip.org/news/7343/iedm-2022-did-we-just-witness-the-death-of-sram/>
- Yaldagard, M. A., Bende, A., Diware, S., Rana, V., Hamdioui, S., & Bishnoi, R. (2026). Detection of read-disturb effects in RRAM-based computation-in-memory architectures for neural networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 73(8), 5648–5661. <https://doi.org/10.1109/TCSI.2026.3675336>
- Yamada, H., Watanabe, J., Nemoto, K., Sun, H.-T., & Shirahata, N. (2022). Postproduction approach to enhance the external quantum efficiency for red light-emitting diodes based on silicon nanocrystals. *Nanomaterials*, 12(23), 4314. <https://doi.org/10.3390/nano12234314>
- Yang, C., Ye, J., Zou, Y., Xu, Y., Yang, P., Yang, S., Chen, G., & Chen, H. (2026). Three-terminal reconfigurable volatile/nonvolatile light-emitting memristor for integrated neuromorphic display systems. *Advanced Materials*, 38(32), e73239. <https://doi.org/10.1002/adma.73239>
- Yen, M.-C., Lee, C.-J., Liu, K.-H., Peng, Y., Leng, J., Chang, T.-H., ... Lee, Y.-J. (2021). All-inorganic perovskite quantum dot light-emitting memories. *Nature Communications*, 12, 4460. <https://doi.org/10.1038/s41467-021-24762-w>
- Zhou, Z., Wu, Y., Pan, K., Zhu, D., Li, Z., Yan, S., Xin, Q., Wang, Q., Qian, X., Xiu, F., Huang, W., & Liu, J. (2024). A memristive-photoconductive transduction methodology for accurately nondestructive memory readout. *Light: Science & Applications*, 13(1), 175. <https://doi.org/10.1038/s41377-024-01519-w>

CHAPTER 2

MULTILEVEL INVERTER (MLI) TOPOLOGIES FOR PHOTOVOLTAIC APPLICATIONS: FROM CLASSICAL STRUCTURES TO REDUCED-SWITCH DESIGNS

Kenan GÜRÇAM¹

¹ Assist. Prof. Dr., Iğdır University, Vocational School of Technical Sciences, Department of Electricity and Energy, 76000 Iğdır, Türkiye.
ORCID: 0000-0001-9051-5953

1. Introduction

Conventional two-level inverters are of limited use in high-voltage applications that demand high power quality, because their switching devices are exposed to the full DC-link voltage and their output voltage contains substantial harmonics. Multilevel inverters mitigate these problems by synthesizing the output voltage from discrete steps: as the number of levels increases, the output approaches the sinusoidal reference, total harmonic distortion (THD) and dv/dt decrease, and the devices experience lower voltage stress (Kouro et al., 2010; Rodriguez et al., 2002). These advantages first established MLIs in medium-voltage industrial drives (Rodriguez et al., 2007) and have since extended their use to PV, electric vehicle, and microgrid applications.

The appeal of MLIs in PV systems does not stem from output quality alone. A PV source has a power–voltage characteristic that varies with irradiance and temperature, and it requires a conversion stage operating together with maximum power point tracking (MPPT). In transformerless structures, the leakage current flowing through the parasitic capacitance between the PV panel and ground directly affects topology selection (Kerekes et al., 2011; Kjaer et al., 2005). On the other hand, increasing the number of levels also increases the number of switches, drivers, capacitors, and diodes, which creates new trade-offs in cost, control complexity, and reliability. This trade-off is the main motivation behind recent research on reduced-switch topologies.

The aim of this chapter is to bring classical and reduced-switch MLI topologies together in a single comparative framework and to evaluate them against design criteria specific to PV applications. The method combines a literature review with an analytical comparison based on component count and conduction path; the THD values in Section 5.2 were calculated for this chapter under the ideal-switch assumption.

The remainder of the chapter is organized as follows. Section 2 covers the classical topologies, Section 3 the reduced-switch structures, and Section 4 the modulation, MPPT, and control approaches. Section 5 discusses performance metrics (losses, THD, voltage/current stress, PV-specific constraints, and economic assessment). Section 6 addresses topology selection by application scale, wide bandgap semiconductors, fault tolerance, and open research questions, and Section 7 summarizes the conclusions.

2. Classical Multilevel Inverter Topologies

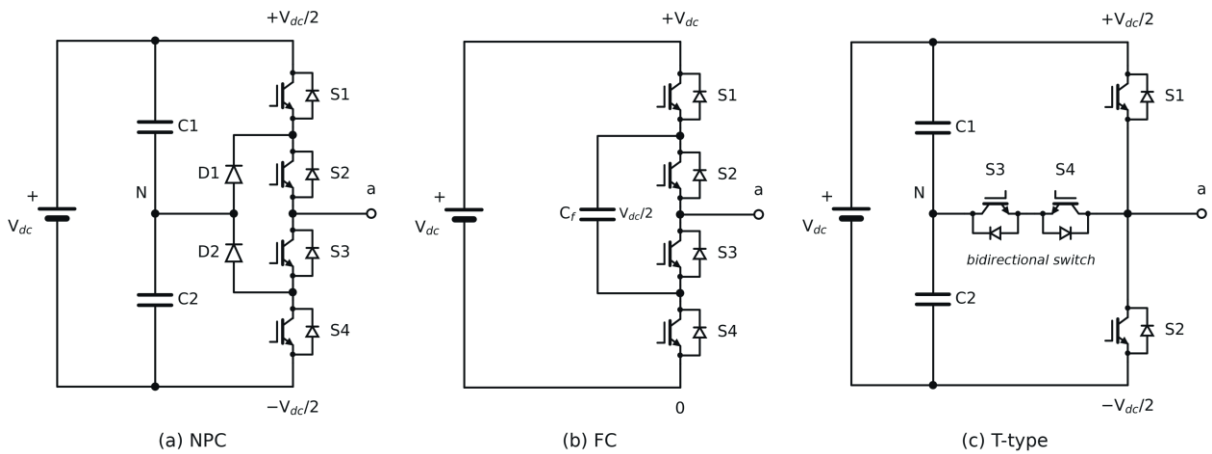
The multilevel inverter family rests on three classical structures: the neutral-point-clamped (NPC) (Nabae et al., 1981), flying capacitor (FC) (Meynard & Foch, 1992), and cascaded H-bridge (CHB) (Malinowski et al., 2010) topologies. Each structure is considered below for one phase leg, with N denoting the number of output voltage levels. In the figures, switches are drawn with the symbol of an IGBT with an antiparallel diode.

2.1. Neutral-Point-Clamped (NPC) Topology

In the NPC structure, the DC link is divided by two series capacitors, and the resulting midpoint is connected to the phase output through clamping diodes (Figure 1a). In the three-level structure, each switch blocks only half of the DC-link voltage, which is the main advantage in high-voltage applications (Nabae et al., 1981).

Figure 1

Single Phase Legs of the Three-Level (a) NPC, (b) FC, and (c) T-Type Topologies



The weakness of the NPC is that the number of clamping diodes grows rapidly with the number of levels, as $(N-1)(N-2)$; five levels require 12 diodes, and the voltage sharing among the diodes may be unequal. In addition, the neutral-point potential between the DC-link capacitors drifts, and this drift must be suppressed by an additional balancing loop (Celanovic & Boroyevich, 2000). For these reasons, the NPC is mostly limited to three levels in practice. Nevertheless, because it can operate from a single DC source and has a mature control base, it is widely used in medium- and large-scale PV plant inverters fed from a single DC link (Kouro et al., 2010).

2.2. Flying Capacitor (FC) Topology

The FC structure uses flying capacitors instead of clamping diodes (Figure 1b). Each capacitor is held at a specific voltage, and the same output level can be produced by more than one switching combination; this redundancy is used to balance the capacitor voltages (Meynard & Foch, 1992).

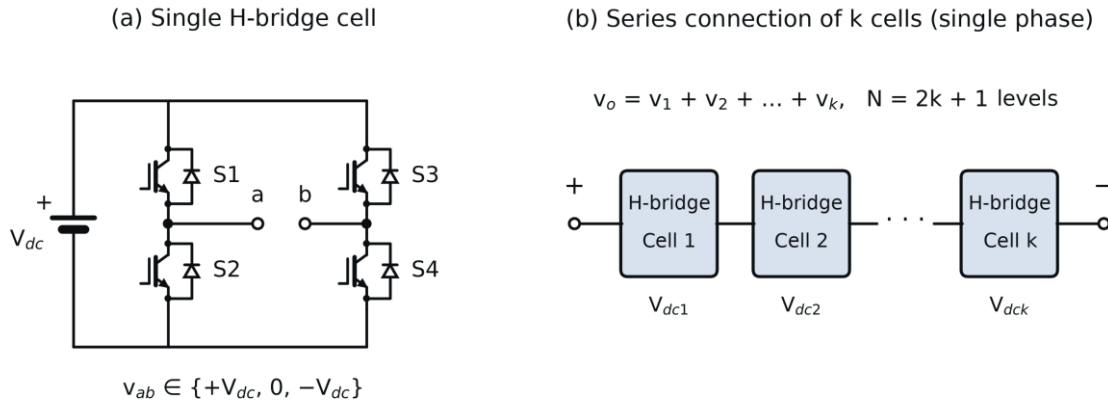
The number of flying capacitors grows as $(N-1)(N-2)/2$, which raises cost and volume, and precharging and balancing the capacitors increase control complexity. The redundant switching states could in principle provide fault tolerance, but this requires fault diagnosis and real-time recalculation of the switching state (Lezana et al., 2010), so it is rarely exploited in practice.

2.3. Cascaded H-Bridge (CHB) Topology

The CHB structure is formed by connecting in series single-phase H-bridge cells fed from independent DC sources (Figure 2). Since each cell can produce $+V_{dc}$, 0, and $-V_{dc}$, k cells yield a $(2k+1)$ -level output; the number of levels increases linearly with the number of cells, and the structure scales by reusing the same cell (Malinowski et al., 2010).

Figure 2

CHB Topology: (a) Single H-Bridge Cell, (b) Series Connection of k Cells



From a PV perspective, the main advantage of the CHB is that each cell can be fed by a separate PV string, enabling independent MPPT per cell. In return, each cell requires an isolated DC source, which increases the number of strings and the system cost. Power imbalances between the sources, particularly under partial shading, can distort the output waveform and require additional control for power distribution among the cells (Kouro et al., 2010). This constraint has motivated the development of switched-rectifier-based voltage multiplexers and hybrid

structures that can operate from a single DC source while partly preserving the modularity of the CHB (Hataş, 2025b).

2.4. Comparison of Classical Topologies

Table 1 gives the component counts of the three classical topologies for an N-level phase leg. The switch count is identical, but the auxiliary components differ: at five levels, the NPC needs 12 clamping diodes, the FC 6 flying capacitors, and the CHB 2 isolated DC sources. Because these counts grow with the square of N for the NPC and FC but linearly for the CHB, the CHB scales best at high level counts, at the price of requiring isolated sources.

Table 1

Component Counts of Classical MLI Topologies per Phase Leg (N: Number of Levels)

Topology	Switches	Clamping diodes	Capacitors	Isolated DC sources
NPC	$2(N-1)$	$(N-1)(N-2)$	$N-1$ (DC link)	1
FC	$2(N-1)$	0	$(N-1)(N-2)/2$ (flying) + $N-1$ (DC link)	1
CHB	$2(N-1)$	0	0	$(N-1)/2$

Note. Values are per phase leg; in three-phase applications, the switch count is multiplied by three, while the DC-link capacitors are shared among the phases. The relations are based on the basic structures in Nabae et al. (1981), Meynard and Foch (1992), and Malinowski et al. (2010).

3. Reduced-Switch-Count Topologies

In classical topologies, the rapid growth in the number of switches, diodes, or capacitors with the number of levels has led to structures that aim to produce the same number of levels with fewer components. There are three main motivations: cost, since each additional switch brings a driver circuit, cooling, and control complexity; reliability, since the probability of failure rises with component count; and efficiency, since each switch contributes conduction and switching losses. Within this framework, T-type and ANPC structures, PUC and asymmetric cascaded H-bridge structures, switched-capacitor topologies, and new voltage-level generation methods that reach level counts as high as 15 with a reduced number of components stand out (Hataş, 2025a).

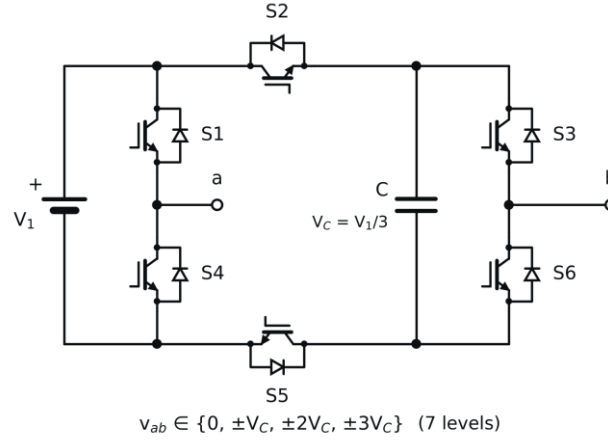
3.1. T-Type and ANPC Topologies

The T-type topology is a derivative of the NPC: instead of clamping diodes, it uses a bidirectional switch (S3–S4) that connects the DC-link midpoint to the output (Figure 1c). The outer switches (S1, S2) block the full DC-link voltage, whereas the middle switches block half of it; hence, the three-level T-type differs from the NPC in the voltage rating required per switch.

The efficiency advantage of the T-type comes from its conduction path: at the outer levels ($\pm V_{dc}/2$), the current flows through only one switch, whereas in the NPC it flows through two series switches at the same level. At the zero level, the current passes through one switch and one diode in both structures: in the T-type, one IGBT of the common-emitter bidirectional switch and the antiparallel diode of the other conduct, whereas in the NPC an inner switch and a clamping diode conduct. Consequently, the T-type is advantageous at high modulation indices, where the outer levels dominate, and in low-voltage applications, as demonstrated for a highly efficient three-level T-type converter designed for low-voltage applications (Schweizer & Kolar, 2013). The active NPC (ANPC) structure replaces the clamping diodes with active switches to provide more than one current path at the zero level, thereby allowing the loss distribution among the switches to be controlled (Brückner et al., 2005).

3.2. PUC Topology

The packed U-cell (PUC) topology produces a multilevel output with a single DC source and one capacitor. Its seven-level version operates with six switches and one capacitor; when the capacitor voltage is held at $V_c = V_1/3$, the levels 0, $\pm V_c$, $\pm 2V_c$, and $\pm 3V_c$ are obtained, and the output voltage is limited to at most V_1 (Vahedi & Al-Haddad, 2016). The structure consists of the S1/S4 leg connected to the V_1 source, the S2/S5 switches between the source and the capacitor, and the S3/S6 leg connected to the capacitor. Each switch pair is driven complementarily, and the output is taken across terminals a–b (Figure 3). Holding the capacitor voltage at $V_1/3$ requires a balancing control loop.

Figure 3*Seven-Level PUC Topology ($V_c = V_1/3$)*

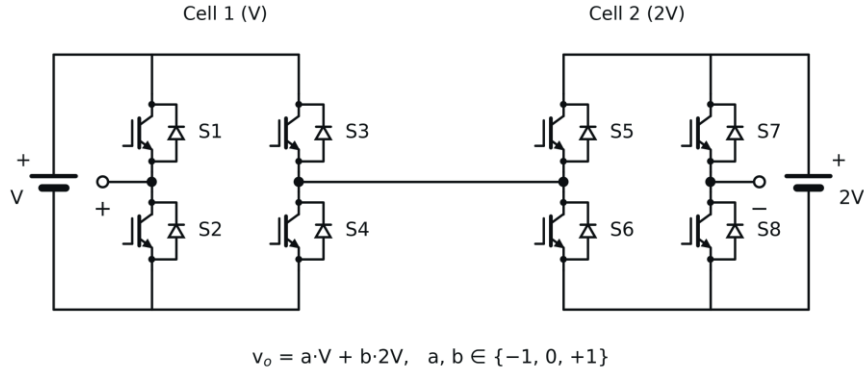
In the PUC, the output voltage cannot exceed the main source value; the V_c source increases the number of levels but not the amplitude. A separate boost stage may therefore be needed when the DC-link voltage can fall below the grid peak voltage. Because the zero level can be obtained with more than one switching state (redundant zero states), the number of commutations at polarity changes can be reduced (Vahedi & Al-Haddad, 2016). To overcome this amplitude and level constraint, hybrid PUC topologies that produce an 11-level output by using high-frequency-link (HFL) support and a dynamic switched rectifier have also been developed (Hataş et al., 2026). Operation from a single DC source is an important advantage for PV; however, in the basic PUC7, neither output terminal is permanently connected to a DC source terminal, so the common-mode voltage varies with the switching state, and leakage current must be addressed separately in transformerless use (Section 5.4).

3.3. Asymmetric Cascaded H-Bridge Topology

In the symmetric CHB, all cell sources are equal, and k cells give $2k+1$ levels (Section 2.3). By choosing a voltage ratio between the cell sources, more levels can be obtained with the same number of cells; this approach is known as the asymmetric (binary-ratio) CHB (Hataş, 2025c; Malinowski et al., 2010; Venkataramanaiah et al., 2017). For two cells with a $V : 2V$ ratio (Figure 4), each cell transfers its own source to the output with a multiplier of $\{-1, 0, +1\}$, so the output becomes $v_o = a \cdot V + b \cdot 2V$ ($a, b \in \{-1, 0, +1\}$), forming seven equally spaced levels: $0, \pm V, \pm 2V$, and $\pm 3V$. The same two-cell symmetric structure would provide only five levels; the asymmetric ratio raises the level count to seven with the same eight switches and increases the level-to-switch ratio to 0.88.

Figure 4

Two-Cell Asymmetric CHB Topology ($V : 2V$ Ratio, Seven Levels)

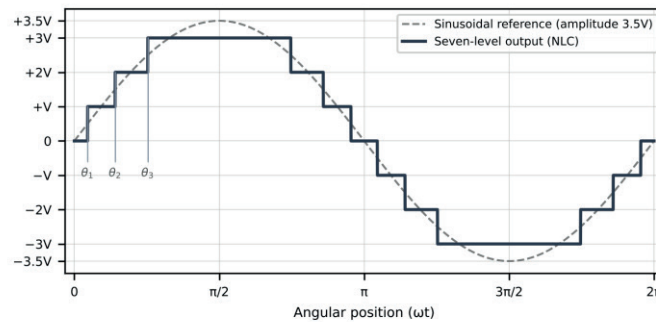


The $\pm V$ levels can each be produced by two cell states, for example, $+V$ by $(a, b) = (+1, 0)$ or $(-1, +1)$; these redundant states can be used to balance the power sharing between the cell sources, whereas the other levels have a single state.

The stepped output obtained by sequencing these states with NLC is shown in Figure 5; the step size is V . In NLC, the output moves to the next level at the angles $\theta_k = \arcsin((k - 0.5)/3.5)$, where the reference crosses the $(k - 0.5)V$ thresholds; for a reference amplitude of $3.5V$, $\theta_1 \approx 8.2^\circ$, $\theta_2 \approx 25.4^\circ$, and $\theta_3 \approx 45.6^\circ$. At this amplitude, the reference reaches the upper threshold of the highest level ($3V + V/2$) only at its peak; thus, the rounding error remains within $\pm V/2$ at all times, and the largest fundamental component is obtained without saturation.

Figure 5

Seven-Level Output Waveform With NLC (Reference Amplitude $3.5V$)



From a PV perspective, per-cell MPPT is possible because each cell can be fed by a separate PV string. However, the $V : 2V$ ratio must be maintained; when the maximum power point voltages of the two strings change differently with irradiance and shading, the ratio is disturbed

and the levels are no longer equally spaced. In practice, therefore, the ratio is fixed by a DC-DC stage, the cells are supplied through new switched-rectifier-based asymmetric CHB concepts (Hataş, 2025c), or the level mapping in the modulation is adapted to the actual cell voltages. Because the same load current flows through both cells, the power sharing between them is unequal; the higher-voltage cell generally carries more power. The switches of the larger cell must also withstand 2V. Owing to the requirement for two isolated DC sources and the parasitic capacitances of the cell sources to ground, leakage current must be addressed separately (Section 5.4).

Example 1. In a seven-level asymmetric CHB connected to a 230 V (rms) single-phase grid, the peak voltage is $230 \cdot \sqrt{2} \approx 325$ V. Since the maximum output is 3V, $V \geq 325/3 \approx 108$ V is required. If $V = 120$ V is chosen, the cell sources are 120 V and 240 V, totaling 360 V; the modulation index is $m = 325/360 \approx 0.90$, and the step size is 120 V. The switch voltage stress is 120 V in cell 1 and 240 V in cell 2 (excluding safety margin). For the same output, a two-level full-bridge inverter requires a DC link of at least 325 V, and all of its switches must withstand this voltage.

3.4. Switched-Capacitor-Based Topologies

Switched-capacitor structures embed voltage boosting in the switching pattern without a separate DC-DC boost stage: switching the capacitors between parallel (charging) and series (discharging) configurations yields multiples of the input voltage (Hinago & Koizumi, 2012), which can remove the need for an additional boost circuit for low-voltage PV strings. In return, inrush currents and voltage ripple must be handled carefully, so many designs seek self-balancing of the capacitors.

3.5. Comparison of Reduced-Switch Topologies

The choice among reduced-switch structures depends not only on the switch count but also on the number of independent sources, the complexity of capacitor balancing, voltage stress, and common grounding (Table 2). Since the level counts differ, the level-to-switch ratio is a fairer measure: at seven levels, the PUC reaches 1.17 and the asymmetric CHB 0.88, whereas the NPC and symmetric CHB remain at 0.58. A higher ratio is not necessarily better, as the PUC pays with capacitor balancing and the asymmetric CHB with isolated sources and unequal voltage ratings.

Table 2*Comparison of Classical and Reduced-Switch Topologies*

Topology	Levels	Switches	Levels / switch	Capacitors / sources	Voltage stress	Common grounding
NPC	7	12	0.58	30 diodes, 6 capacitors	Symmetric	Structure-dependent
CHB (symmetric)	7	12	0.58	3 isolated sources	Symmetric	No
T-type	5	8	0.63	Single source	Outer/middle switches differ	Yes, in half-bridge
PUC	7	6	1.17	1 capacitor	Asymmetric (V_1 , $2V_1/3$, $V_1/3$)	No (basic structure)
CHB (asymmetric, $V:2V$)	7	8	0.88	2 isolated sources	Asymmetric (V , $2V$)	No

Note. Classical and asymmetric CHB values are derived from Table 1 and Section 3.3, and PUC values from Vahedi and Al-Haddad (2016). The T-type row refers to the single-phase five-level arrangement in which two three-level T-type legs are connected as a full bridge; common grounding is achieved only in the single-leg (half-bridge) arrangement when the DC-link midpoint is connected to the grid neutral. The common grounding column is a qualitative assessment.

4. Modulation and Control Techniques

The performance of an MLI is determined not only by its topology but also by the applied modulation, which directly affects the switching frequency, harmonic content, switching losses, and dynamic response. Two approaches stand out in the literature: carrier-based sinusoidal pulse width modulation (SPWM) and nearest level control (NLC) (Holmes & Lipo, 2003; Rodriguez et al., 2002).

4.1. Sinusoidal Pulse Width Modulation (SPWM)

In SPWM, a sinusoidal reference is compared with multiple carriers. The carriers can be level-shifted (LS-PWM: in-phase, phase-opposition, and alternative phase-opposition dispositions) or phase-shifted (PS-PWM), and each arrangement distributes the harmonic energy differently across the frequency spectrum (Holmes & Lipo, 2003). Owing to its high switching frequency, SPWM suppresses low-order harmonics and allows a smaller output filter; however, because it increases switching losses, it is preferred mainly in low- and medium-power applications, where these losses remain acceptable.

4.2. Nearest Level Control (NLC)

In NLC, the instantaneous value of the reference is rounded to the nearest discrete level the inverter can produce, and the corresponding switching state is selected directly. Since no carrier comparison is needed, the computational burden is low, and the switching frequency can be kept close to the fundamental frequency. As the number of levels increases, the steps become smaller, so acceptable THD is achieved even at a low switching frequency; at low level counts or low modulation indices, however, the low-order harmonics of NLC remain pronounced. This behavior is shown numerically in Section 5.2.

4.3. MPPT Integration

In PV applications, the modulation layer must work in harmony with the MPPT algorithm. Perturb and observe (P&O) and incremental conductance (INC) are widespread owing to their low computational burden, but under partial shading they may miss the global maximum among multiple local peaks (Esram & Chapman, 2007), which has motivated particle swarm optimization and fuzzy logic approaches for multilevel systems.

In MLIs, balancing secondary capacitors or auxiliary sources may conflict with the main MPPT, depending on the topology. An NPC with a single DC link needs only one central MPPT; a CHB allows independent per-cell MPPT, which helps under partial shading but complicates inter-cell synchronization; in the PUC, the inner capacitor-balancing loop constrains the switching states while the outer MPPT loop sets the input voltage; and in the asymmetric CHB, MPPT must not disturb the cell-voltage ratio. Pursuing these objectives without mutual interference is a central issue in control design.

4.4. Control Architecture: From Classical PI Loops to Model Predictive Control

The classical approach uses cascaded PI controllers in the voltage and current loops; although simple and widespread, it may fall short for nonlinear requirements such as capacitor balancing. Model predictive control (MPC) uses a model to compute the effect of each possible switching state over the next sampling period and selects the state that minimizes a cost function (Cortés et al., 2008; Vazquez et al., 2014). Since objectives such as current tracking, capacitor balancing, and switching-frequency limitation can be optimized jointly in a single cost function, MPC can ease the conflict between separate PI loops in topologies with additional balancing constraints (e.g., the PUC). The price is a variable switching frequency and a high

computational burden, which constrains the choice of embedded processor (Vazquez et al., 2014).

In grid-connected operation, a phase-locked loop (PLL) synchronizes the current reference with the grid voltage; under distorted or unbalanced grid conditions, its performance directly affects the output-current THD and the power factor. The PLL should therefore be designed together with the modulation and MPPT within a single control hierarchy.

5. Performance Analysis and Comparison

The merit of an MLI is assessed in terms of efficiency, THD, switching and conduction losses, and the distribution of voltage/current stress. Specifically for PV, leakage current, DC-link ripple, and partial-load efficiency should be added to these metrics.

5.1. Loss Analysis Method

Loss analyses in tools such as PLECS, PSIM, and MATLAB/Simulink use datasheet on-state resistance, threshold voltage, and switching energy curves to compute the total loss and efficiency at a given operating point (input voltage, output power, switching frequency). Conduction losses arise from the on-state voltage drop and grow with the square of the current, whereas switching losses equal the energy per transition multiplied by the switching frequency.

The efficiency advantage of reduced-switch structures is often attributed to the lower switch count; what matters, however, is how many series components the current passes through in each switching state (conduction path length). A topology with fewer switches may therefore fail to deliver the expected gain if, in some states, it routes the current through more than one series switch (Section 3.1). Comparative studies should report the average and worst-case conduction path lengths, ratings, and switching frequency under identical conditions.

5.2. THD and Power Quality Assessment

THD summarizes the magnitude of all harmonics other than the fundamental; in grid-connected PV systems, it is framed by standards such as the IEEE 519 voltage distortion limits (Shelar et al., 2024) and the IEEE 1547 current distortion limits (Institute of Electrical and Electronics Engineers [IEEE], 2018). IEEE 519 specifies 8% THD and 5% individual harmonic distortion for voltage at 1 kV and below, whereas IEEE 1547 sets a 5% limit for total demand distortion. Increasing the number of levels generally lowers THD (Hataş, 2025a); however, the

relationship is not linear and depends on the modulation, switching frequency, and filter design. Hence, THD values can be compared only when the modulation method, frequency, and measurement band are specified.

To illustrate this dependence, the THD of an ideal equal-step phase voltage was calculated for this chapter (50 Hz fundamental; $m = 1.0$ for NLC; $m = 0.95$ and a 2 kHz carrier frequency for level-shifted phase-disposition SPWM (LS-PD)); lossless ideal switches; harmonic bands $h \leq 100$ and $h \leq 35$). The results are given in Table 3.

Table 3

Phase-Voltage THD Values (%) by Number of Levels and Harmonic Band (Calculated for This Chapter)

Levels (N)	NLC, $h \leq 100$	LS-PD SPWM, $h \leq 100$	NLC, $h \leq 35$	LS-PD SPWM, $h \leq 35$
3	30.5	49.7	29.6	6.3
5	17.1	26.4	16.1	6.5
7	11.7	18.2	10.5	4.4
9	8.8	13.6	7.8	3.3
11	7.1	10.9	6.1	2.5
15	5.0	7.4	3.3	2.1

Note. The calculation refers to an ideal-switch, unfiltered phase voltage; in three-phase systems, the line-voltage THD is lower because triplen harmonics cancel.

Three observations follow. First, THD decreases markedly with the number of levels for both methods; for NLC, it falls from 30.5% ($N = 3$) to 5.0% ($N = 15$). Second, in the $h \leq 100$ band, NLC yields lower THD than SPWM at every level count, because the harmonics around the 2 kHz carrier fall within this band. Third, in the low-order band ($h \leq 35$), the picture reverses: at $N = 3$, SPWM gives 6.3% and NLC 29.6%, and the gap narrows to 2.1% versus 3.3% at $N = 15$. NLC is thus attractive at seven or more levels, whereas SPWM is preferable at three and five levels in terms of low-order harmonics. Since SPWM harmonics cluster around the carrier and are easily filtered, a fair comparison requires equal switching frequency and filter conditions.

5.3. Voltage and Current Stress Distribution

Equal voltage ratings for all switches (symmetric stress) are advantageous for uniform component procurement and spare-part management; the NPC and FC are good examples. In

some reduced-switch structures, however, different switches are subjected to different voltages: in the PUC, the S1/S4 switches must withstand V_1 , the S2/S5 switches $V_1 - V_c = 2V_1/3$, and the S3/S6 switches $V_c = V_1/3$; in the asymmetric CHB, the switches of the larger cell must withstand $2V$ and those of the smaller cell V ; and in the T-type, the outer and middle switches also require different ratings. While this reduces the switch count, it may require procurement of more than one component type.

5.4. PV-Specific Design Criteria

Leakage current. In transformerless PV inverters, the common-mode voltage creates a leakage current through the parasitic capacitance between the panel and ground; this current is limited for safety and EMI reasons. DIN VDE V 0126-1-1 specifies an upper limit of 300 mA for continuous leakage current (VDE Verlag, 2013). Structures that keep the common-mode voltage constant or with low variation, such as topologies in which the negative PV terminal is common with the output, are therefore preferred (Kerekes et al., 2011). In this respect, the half-bridge T-type (and the NPC in the same arrangement), whose DC-link midpoint can be connected to the grid neutral, is advantageous. In the basic PUC7, the common-mode voltage varies with the switching state, so common-ground derivatives or modulation-level measures are needed; in CHB structures with isolated cell sources, each PV string has its own parasitic capacitance to ground, so leakage current must be specifically addressed. The use of multilevel structures in transformerless single-phase PV inverters has been examined in this respect (González et al., 2008).

DC-link ripple. In single-phase systems, the instantaneous power pulsates at twice the grid frequency (100 Hz), and if this ripple is transferred from the DC link to the PV source, it reduces MPPT efficiency (Kjaer et al., 2005). A DC-link capacitor or a power decoupling circuit is therefore required; in multilevel structures, the split DC link and secondary capacitors distribute this ripple differently, and capacitor sizing varies with the topology.

Partial-load efficiency. Since PV inverters operate below rated power most of the time, weighted metrics such as the European efficiency (the average of the efficiencies at 5%, 10%, 20%, 30%, 50%, and 100% load, weighted by 3%, 6%, 13%, 10%, 48%, and 20%, respectively) are used instead of a single peak efficiency value. Since switching losses dominate at low load, structures that improve this range (NLC with a low switching frequency, topologies with few switches) stand out.

Standards and grid codes. Safety requirements are defined by IEC 62109-1/-2 (International Electrotechnical Commission [IEC], 2010, 2011), grid interface characteristics by IEC 61727 (Wu et al., 2017), and anti-islanding tests by IEC 62116 (IEC, 2014). National grid codes additionally require reactive power support, low-voltage ride-through (LVRT), and active power curtailment; since reactive power delivery imposes ripple on the DC link, capacitor balancing in multilevel structures must remain effective at these operating points. The applicable limits vary by country and connection level and should be verified against current regulations.

5.5. Economic Assessment and Total Cost of Ownership

Topology selection should be justified economically as well as by technical performance. The total cost of ownership includes, in addition to the initial investment cost, the monetary value of the energy losses incurred over the operating life. A topology with fewer switches may offer an advantage in initial investment, but if its conduction path leads to higher losses, this advantage may be offset or even reversed by lifetime energy losses. Comparing the additional investment cost of a topology relative to a reference with the lifetime return of its additional efficiency can be expressed by the following condition:

$$\Delta C_{inv} \leq \sum_{t=1..T} (\Delta \eta \cdot E_t \cdot c_e) / (1 + r)^t \quad (1)$$

Here, ΔC_{inv} denotes the additional investment cost, $\Delta \eta$ the difference in the annual average (weighted) efficiency of the inverter, E_t the annual energy entering the inverter, c_e the unit price of energy, r the discount rate, and T the project lifetime. If the inequality is not satisfied, the more efficient topology cannot be justified economically. The same framework can also be built on the levelized cost of energy (LCOE); such an assessment provides a firmer basis for discussing whether a given topology is truly the best choice at a given scale (residential rooftop or MW-scale plant).

6. Applications and Current Trends

6.1. Topology Selection by Application Scale

The application field of MLIs has expanded beyond medium-voltage drives to PV plants, electric vehicle charging, wind energy, and microgrids (Kouro et al., 2010). Topology preference varies with application scale, as summarized in Table 4. The asymmetric CHB is an additional option for string-level systems with two isolated PV strings, whereas the FC is less common in PV because of its capacitor volume.

Table 4*Typical Topology Preference and Rationale by PV Application Scale*

Application scale	Typical topology	Rationale	Main constraint
MW-scale plant	NPC / ANPC (3 levels)	Single DC link, mature control	Neutral-point balance, diode count
String level (kW to tens of kW)	CHB (symmetric/asymmetric), T-type	Per-cell MPPT, high efficiency	Isolated sources, different switch ratings
Residential rooftop (low power)	T-type, PUC	Few switches, single DC source	Capacitor balance, output amplitude limit, leakage current (PUC)
Microgrid	CHB, modular structures	Modularity, grid and island mode flexibility	Control complexity

Note. The table is a qualitative summary compiled from Kouro et al. (2010), Rodriguez et al. (2002), and the assessments in this chapter.

6.2. Wide Bandgap Semiconductor Technologies (SiC/GaN)

Silicon carbide (SiC) and gallium nitride (GaN) devices offer higher switching frequencies, lower on-state resistance, and higher temperature capability than Si devices (Millán et al., 2014); combined with reduced-switch MLIs, they can shrink passive filters and raise efficiency. However, high dv/dt causes voltage ringing and EMI through parasitic inductances, so layout- and packaging-related parasitics must be considered in gate driver and filter design.

6.3. Fault-Tolerant Designs

Fault-tolerant MLI design aims to keep the system running at reduced performance (e.g., with fewer levels) after an open- or short-circuit switch fault (Lezana et al., 2010), which is particularly valuable in remote PV plants. Common strategies include redundant switches or cells, reconfigurable switching matrices, and real-time fault diagnosis; deep learning-based diagnosis has proven effective for asymmetric CHB inverters under dynamic loads (Hataş, 2026). Each strategy improves reliability at the expense of cost and control complexity.

In a low-power (200 V) comparison, the mean time to failure (MTTF) was 50,036 h for the NPC, 42,245 h for the FC, and 65,451 h for the CHB (Alavi et al., 2016). Although the absolute values depend on the assumed failure rates, the ranking shows that capacitor and diode counts matter as much as the switch count. Reduced-switch topologies have few redundant states, so

a single switch failure usually causes a greater loss of function than in the CHB or FC; this trade-off should be weighed in topology selection.

6.4. The Turkish Context

Installed solar capacity in Türkiye is growing rapidly: according to Ember data, 4.6 GW of new solar capacity was commissioned in 2025 (TaiyangNews, n.d.), and about one third of the 2035 solar–wind capacity target was reached in the same year (Enerdata, n.d.). This growth is increasing inverter demand, and the domestic manufacture of an international inverter brand has been announced (Ekonomist, n.d.).

However, no publicly available and independently verified data exist on the topologies of the inverters installed or manufactured in Türkiye. The comparison in this chapter is therefore a framework derived from the international literature, and the figures above, taken from news sources, should be verified against official data (Ministry of Energy and Natural Resources, TEİAŞ, EPDK). For domestic inverter design, topology selection in terms of efficiency, cost, and grid code compliance (e.g., T-type at residential scale, CHB at string scale) is a concrete R&D priority.

6.5. Open Research Questions and Limitations

Several issues remain open in the field. First, in reduced-switch topologies, capacitor or source balancing interferes with MPPT; operating these two loops without mutual disturbance has not yet been fully resolved. Second, the computational burden of modulation grows with the number of levels and must be kept at a manageable level on embedded controllers. Third, the high-frequency and low-loss benefits that SiC and GaN devices would bring when combined with reduced-switch structures have not yet been sufficiently demonstrated. Fourth, the cost–reliability balance of fault-tolerant designs should be modeled systematically for different application scenarios.

This chapter is a review: it provides no topology-specific efficiency or thermal results, only an analytical comparison of component counts and an ideal-switch THD calculation. The PUC is presented at a basic level (see Hataş et al., 2026, and Vahedi & Al-Haddad, 2016, for switching tables and balancing control), power sharing in the asymmetric CHB is discussed only qualitatively (Hataş, 2025c; Venkataramanaiah et al., 2017), and the reliability data rely on a single source and should not be used to rank topologies definitively.

7. Conclusion

This chapter has traced multilevel inverters for photovoltaic applications from classical to reduced-switch topologies. As the number of levels increases, the auxiliary component count of the NPC and FC and the isolated-source requirement of the CHB become binding constraints. The T-type and ANPC address efficiency by shortening the conduction path, whereas the PUC and asymmetric CHB address cost by producing more levels with the same switch count, at the price of capacitor balancing, unequal voltage ratings, or voltage-ratio maintenance. No single topology is therefore best for all applications: the NPC family suits large plants with a single DC link, the CHB and T-type suit string-level systems, and the T-type and PUC suit low-power rooftop systems.

The modulation and control layer is as decisive as the topology. The THD calculation showed that NLC achieves adequate output quality at high level counts with a low switching frequency, whereas SPWM better suppresses low-order harmonics at three and five levels. Switch count or THD alone is not a sufficient criterion; conduction path length, leakage current, DC-link ripple, partial-load efficiency, the interaction of MPPT with capacitor or source balancing, and lifetime cost must be assessed together. Wide bandgap devices and fault tolerance will shape future designs, and the rapid growth of PV in Türkiye makes topology selection a concrete R&D topic for domestic inverters. Future work should compare topologies under identical operating conditions through loss and thermal analysis with real PV profiles and should validate the MPPT–balancing interaction experimentally.

References

- Alavi, O., Viki, A. H., & Shamlou, S. (2016). A comparative reliability study of three fundamental multilevel inverters using two different approaches. *Electronics*, 5(2), Article 18. <https://doi.org/10.3390/electronics5020018>
- Brückner, T., Bernet, S., & Güldner, H. (2005). The active NPC converter and its loss-balancing control. *IEEE Transactions on Industrial Electronics*, 52(3), 855–868.
- Celanovic, N., & Boroyevich, D. (2000). A comprehensive study of neutral-point voltage balancing problem in three-level neutral-point-clamped voltage source PWM inverters. *IEEE Transactions on Power Electronics*, 15(2), 242–249.
- Cortés, P., Kazmierkowski, M. P., Kennel, R. M., Quevedo, D. E., & Rodríguez, J. (2008). Predictive control in power electronics and drives. *IEEE Transactions on Industrial Electronics*, 55(12), 4312–4324.
- Ekonomist. (n.d.). ZES, Solis'i Türkiye'de üretecek [ZES will manufacture Solis in Türkiye]. Retrieved October 17, 2026, from <https://www.ekonomist.com.tr/enerji/zes-solis-i-turkiye-de-uretecek-54121>
- Enerdata. (n.d.). Türkiye reached one third of its 2035 solar and wind capacity target in 2025. Retrieved October 17, 2026, from <https://www.enerdata.net/publications/daily-energy-news/turkiye-reached-one-third-its-2035-solar-and-wind-capacity-target-2025.html>
- Esram, T., & Chapman, P. L. (2007). Comparison of photovoltaic array maximum power point tracking techniques. *IEEE Transactions on Energy Conversion*, 22(2), 439–449.
- González, R., Gubía, E., López, J., & Marroyo, L. (2008). Transformerless single-phase multilevel-based photovoltaic inverter. *IEEE Transactions on Industrial Electronics*, 55(7), 2694–2702.
- Hataş, H. (2025a). Design and analysis of 15-level inverter with a new voltage level generation method. *Ain Shams Engineering Journal*, 16(10), Article 103679.
- Hataş, H. (2025b). Design and implementation of a novel switched rectifier based voltage multiplexer for multilevel inverters. *Electric Power Systems Research*, 247, Article 111760.
- Hataş, H. (2025c). A novel asymmetrical cascaded H-bridge concept using switched rectifier. *International Journal of Circuit Theory and Applications*, 53(8), 4754–4768.
- Hataş, H. (2026). Deep learning-based fault diagnosis for asymmetric cascaded H-bridge multilevel inverters under dynamic motor loads. *International Journal of Circuit Theory and Applications*.

- Hataş, H., Gürçam, K., & Özel, E. (2026). High-frequency-link assisted 11-level packed U-cell inverter with dynamic switched rectifier. In *Proceedings of the 2026 4th Cognitive Models and Artificial Intelligence Conference (AICCONF)* (pp. 1–5).
- Hinago, Y., & Koizumi, H. (2012). A switched-capacitor inverter using series/parallel conversion with inductive load. *IEEE Transactions on Industrial Electronics*, 59(2), 878–887.
- Holmes, D. G., & Lipo, T. A. (2003). *Pulse width modulation for power converters: Principles and practice*. IEEE Press/Wiley.
- Institute of Electrical and Electronics Engineers. (2018). *IEEE standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces* (IEEE Std 1547-2018).
- International Electrotechnical Commission. (2010). *Safety of power converters for use in photovoltaic power systems – Part 1: General requirements* (IEC 62109-1:2010).
- International Electrotechnical Commission. (2011). *Safety of power converters for use in photovoltaic power systems – Part 2: Particular requirements for inverters* (IEC 62109-2:2011).
- International Electrotechnical Commission. (2014). *Utility-interconnected photovoltaic inverters – Test procedure of islanding prevention measures* (IEC 62116:2014).
- Kerekes, T., Teodorescu, R., Rodríguez, P., Vázquez, G., & Aldabas, E. (2011). A new high-efficiency single-phase transformerless PV inverter topology. *IEEE Transactions on Industrial Electronics*, 58(1), 184–191.
- Kjaer, S. B., Pedersen, J. K., & Blaabjerg, F. (2005). A review of single-phase grid-connected inverters for photovoltaic modules. *IEEE Transactions on Industry Applications*, 41(5), 1292–1306.
- Kouro, S., Malinowski, M., Gopakumar, K., Pou, J., Franquelo, L. G., Wu, B., Rodriguez, J., Pérez, M. A., & Leon, J. I. (2010). Recent advances and industrial applications of multilevel converters. *IEEE Transactions on Industrial Electronics*, 57(8), 2553–2580.
- Lezana, P., Pou, J., Meynard, T. A., Rodriguez, J., Ceballos, S., & Richardeau, F. (2010). Survey on fault operation on multilevel inverters. *IEEE Transactions on Industrial Electronics*, 57(7), 2207–2218.
- Malinowski, M., Gopakumar, K., Rodriguez, J., & Pérez, M. A. (2010). A survey on cascaded multilevel inverters. *IEEE Transactions on Industrial Electronics*, 57(7), 2197–2206.

- Meynard, T. A., & Foch, H. (1992). Multi-level conversion: High voltage choppers and voltage-source inverters. In *Proceedings of the IEEE Power Electronics Specialists Conference (PESC)* (pp. 397–403).
- Millán, J., Godignon, P., Perpiñà, X., Pérez-Tomás, A., & Rebollo, J. (2014). A survey of wide bandgap power semiconductor devices. *IEEE Transactions on Power Electronics*, 29(5), 2155–2163.
- Nabae, A., Takahashi, I., & Akagi, H. (1981). A new neutral-point-clamped PWM inverter. *IEEE Transactions on Industry Applications*, IA-17(5), 518–523.
- Rodriguez, J., Bernet, S., Wu, B., Pontt, J. O., & Kouro, S. (2007). Multilevel voltage-source-converter topologies for industrial medium-voltage drives. *IEEE Transactions on Industrial Electronics*, 54(6), 2930–2945.
- Rodriguez, J., Lai, J.-S., & Peng, F. Z. (2002). Multilevel inverters: A survey of topologies, controls, and applications. *IEEE Transactions on Industrial Electronics*, 49(4), 724–738. <https://doi.org/10.1109/TIE.2002.801052>
- Schweizer, M., & Kolar, J. W. (2013). Design and implementation of a highly efficient three-level T-type converter for low-voltage applications. *IEEE Transactions on Power Electronics*, 28(2), 899–907.
- Shelar, S., Bankar, D., & Bakre, S. (2024). Review of revisions of IEEE 519 standard on power system harmonics (1981 to 2022). In *Proceedings of the 21st International Conference on Harmonics and Quality of Power (ICHQP)* (pp. 415–420).
- TaiyangNews. (n.d.). *Ember: Türkiye installed 4.6 GW solar PV capacity in 2025*. Retrieved October 17, 2026, from <https://taiyangnews.info/markets/ember-turkiye-installed-46-gw-solar-pv-2025>
- Vahedi, H., & Al-Haddad, K. (2016). Real-time implementation of a seven-level packed U-cell inverter with a low-switching-frequency voltage regulator. *IEEE Transactions on Power Electronics*, 31(8), 5967–5973.
- Vazquez, S., Leon, J. I., Franquelo, L. G., Rodriguez, J., Young, H. A., Marquez, A., & Zanchetta, P. (2014). Model predictive control: A review of its applications in power electronics. *IEEE Industrial Electronics Magazine*, 8(1), 16–31.
- VDE Verlag. (2013). *Automatic disconnection device between a generator and the public low-voltage grid* (DIN VDE V 0126-1-1:2013-08).
- Venkataramanaiah, J., Suresh, Y., & Panda, A. K. (2017). A review on symmetric, asymmetric, hybrid and single DC sources based multilevel inverter topologies. *Renewable and Sustainable Energy Reviews*, 76, 788–812.

Wu, Y.-K., Lin, J.-H., & Lin, H.-J. (2017). Standards and guidelines for grid-connected photovoltaic generation systems: A review and comparison. *IEEE Transactions on Industry Applications*, 53(4), 3205–3216.

CHAPTER 3

A-IGZO TFTS FOR LARGE-AREA ELECTRONICS: MATERIALS PHYSICS, DEVICE ARCHITECTURES, AND FABRICATION PROCESSES

*Mehmet İzzeddin GÜLER¹ , Tamila ANUTGAN² ,
Mustafa ANUTGAN³*

¹ Res. Asst. Mehmet İzzeddin GÜLER, Thin Film Materials and Devices Laboratory, Materials Research and Development Center, Karabuk University, Karabuk, 78050, Turkey, ORCID: 0009-0001-2059-3670

² Prof. Dr. Tamila ANUTGAN, Thin Film Materials and Devices Laboratory, Materials Research and Development Center, Karabuk University, Karabuk, 78050, Turkey, ORCID: 0000-0002-0287-265X

³ Prof. Dr. Mustafa ANUTGAN, Thin Film Materials and Devices Laboratory, Materials Research and Development Center, Karabuk University, Karabuk, 78050, Turkey, ORCID: 0000-0003-2535-2646

1. Introduction

Thin-film transistors (TFTs) are fundamental building blocks of electronic systems such as flat-panel displays and detectors; the TFTs that can be fabricated at low temperatures allow flexible large-area electronic devices, that is, electronic circuits spread over large surfaces on flexible substrates (Nomura et al., 2004). In active-matrix flat-panel displays and X-ray detectors, pixels are addressed through a backplane consisting of TFT arrays (Mativenga et al., 2012). The choice of semiconductor used in the TFT backplane is decisive for the performance and manufacturing requirements of these applications; mobility, off-state current, uniformity, process temperature, and stability must be considered together (Kamiya et al., 2010).

These criteria have also guided the evolution of TFT backplane technologies. Hydrogenated amorphous silicon (a-Si:H) TFT, which has been widely used for a long time, has low mobility and can exhibit threshold voltage (V_{th}) shifts under prolonged gate bias (Powell, 1989). These characteristics make it difficult to meet the requirements of next-generation displays that demand high current-driving capability and fast operation (Kamiya et al., 2010). Low-temperature polycrystalline silicon (LTPS) TFT offers high mobility but has limitations in achieving device uniformity over large areas and in reducing manufacturing cost (Kamiya et al., 2010). Organic TFTs, on the other hand, offer printability and mechanical flexibility, and their field-effect mobilities have risen to more than $1\text{--}10\text{ cm}^2\text{V}^{-1}\text{s}^{-1}$, exceeding those of benchmark a-Si:H TFTs ($0.5\text{--}1\text{ cm}^2\text{V}^{-1}\text{s}^{-1}$). However, the uniformity and operational stability of organic TFTs are still widely regarded to be insufficient for display TFT backplanes comparable to those based on oxide or polysilicon TFTs (Sirringhaus, 2014); their instability arises from structural and energetic disorder in the semiconductor and from extrinsic factors such as oxidation and moisture (Jiang et al., 2019). The low-temperature polycrystalline oxide (LTPO) architecture, which combines the high drive current of LTPS with the low off-state current of oxide TFTs in the same pixel circuit, shows that these technologies can be complementary as well as competing (Fan et al., 2022).

Among these options, amorphous indium–gallium–zinc oxide (a-IGZO), a representative of amorphous oxide semiconductors (AOS) (Kamiya et al., 2010), has come to the fore remarkably quickly. A working hypothesis for finding wide-band-gap, electrically conducting amorphous oxides was proposed in 1996, identifying amorphous double oxides containing heavy-metal cations with an $(n-1)d^{10}ns^0$ electronic configuration as promising candidates (Hosono, Kikuchi, et al., 1996; Hosono, Yasukawa, et al., 1996). A field-effect mobility of approximately $80\text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ was demonstrated in 2003 for TFTs with a single-

crystalline $\text{InGaO}_3(\text{ZnO})_5$ channel (Nomura et al., 2003), and the first TFTs with an a-IGZO channel were reported in 2004 as transparent devices fabricated at room temperature on flexible substrates (Nomura et al., 2004). By early 2010, a-IGZO TFT-based prototype displays had reached 37 inches for liquid-crystal displays (LCDs) and 19 inches for organic light-emitting diode (OLED) displays (Kamiya et al., 2010), and around 2012, LCDs using a-IGZO TFT backplanes for smartphones and tablet computers appeared (Hosono, 2018). Nomura et al. (2004) obtained saturation mobilities of $6\text{--}9\text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ in devices fabricated on flexible polyethylene terephthalate (PET) substrates, demonstrating that low-temperature processability and relatively high mobility can be achieved simultaneously. In addition to these experimental results, low off-state current, large-area uniformity, and relatively simple fabrication processes have increased the importance of a-IGZO for TFT backplane technologies (Kamiya et al., 2010). Hall measurements of carrier transport have shown that mobility in a-IGZO increases with carrier concentration and that this behavior can be explained by percolation conduction through distributed potential barriers near the conduction-band edge (Takagi et al., 2005).

However, the multicomponent oxide nature of this material makes TFT behavior sensitive to deposition (sputtering) and annealing conditions. While oxygen incorporation during sputtering affects the electrical properties of the a-IGZO film and the TFT behavior (Kwon et al., 2011), hydrogen diffusion from the gate insulator into the active layer, i.e. a-IGZO, can either improve or degrade device characteristics depending on the annealing conditions (Nam et al., 2018). First-principles calculations further reveal that the formation energy of oxygen vacancies is generally low in local environments surrounded by In atoms and tends to increase as the number of neighboring Ga atoms increases (Noh et al., 2011). Experimentally, a comparison of two sputtering-target compositions in which the In ratio was kept constant while the Ga/Zn ratio was increased ($\text{In:Ga:Zn} = 2:1:2$ and $2:2:1$) showed that devices with higher Ga content exhibited lower mobility but better bias stability (Huh et al., 2011). These findings require cation composition of the sputtering target, a-IGZO film defect structure, and fabrication conditions to be considered together.

The translation of a-IGZO material properties into TFT device performance also depends on the contacts and TFT channel geometry. Jeong et al. (2008) showed that high contact resistance can reduce the field-effect mobility extracted from short-channel a-IGZO TFTs. Similarly, in structures where the active layer extends beyond the source/drain electrodes along the TFT channel-width direction, current flowing along the channel edges increases the effective channel width beyond its mask value and can lead to an overestimation of the field-effect mobility (Mativenga et al., 2012). Therefore, when interpreting measured TFT device

parameters, it is important to evaluate carrier transport in the TFT channel together with the contributions of the contacts and the device geometry.

In this chapter, a-IGZO TFTs are discussed at the material and device levels. First, silicon-based TFT backplane technologies and their limitations are summarized (Section 2); then the electronic structure that explains the high mobility and low off-state current of a-IGZO TFT, together with the roles of cations and subgap defect states, are examined (Section 3). Afterwards, TFT device architectures, active-layer fabrication processes, extraction of basic device parameters, and the effects of TFT channel geometry are addressed (Section 4). Finally, the open problems and research gaps in these areas are discussed (Section 5).

2. Silicon-Based TFT Backplane Technologies and Their Limitations

In a-Si:H, the main TFT backplane material of large-area electronics, the directional sp^3 hybrid orbitals that form the conduction paths are significantly distorted by the disorder of the amorphous structure (Kamiya et al., 2010). In amorphous materials, structural disorder manifests itself most clearly in the distribution of bond angles (Hosono, 2006). This disorder converts the states at the conduction band edge into tail states, which lie just below the mobility edge and in which electrons cannot move freely; the states deep in the band gap, on the other hand, arise from defects in the network, largely dangling bonds (Powell, 1989). At room temperature, the current is carried by a small fraction of electrons above the mobility edge; because electrons are repeatedly trapped in and released from the tail states (multiple trapping and release), the effective mobility remains low (Powell, 1989). In a-Si:H TFTs, therefore, the threshold voltage is determined by the deep states and the field-effect mobility by the tail states; typical mobility values are on the order of $1 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ or below (Powell, 1989). Under prolonged gate bias, the V_{th} shifts through charge trapping in the silicon nitride gate insulator and the creation of metastable dangling-bond states in the amorphous silicon (Powell, 1989).

LTPS is generally obtained by laser crystallization of amorphous silicon films (Matsuo et al., 2019). Owing to the ordered arrangement of atoms over large regions, LTPS offers a carrier mobility approximately 200 times that of a-Si:H, which allows the driving circuits to be fabricated on the panel as well and thus allows the display bezel to be narrowed (Geng et al., 2023). However, because polysilicon grains are irregular, the uniformity between TFTs decreases; the narrow process window of the laser process and the high manufacturing cost also hinder scalability and yield in large panels (Geng et al., 2023). In the comparison by Geng et al. (2023), the main application area of LTPS is given as high-end mobile active-matrix LCDs (AMLCDs) and active-matrix OLED (AMOLED) displays. Matsuo et al. (2019), for their part,

list high film quality, large grain size, and control of the locations where grains form as the basic requirements for LTPS in applications where driver and control circuits are integrated on the panel.

Unlike in display applications, radiation hardness is also a decisive criterion in X-ray detectors (Street, 2022). In the study by Park et al. (2021), no V_{th} shift was observed in a-IGZO TFTs exposed to X-rays up to a total dose of 2000 Gy, whereas LTPS TFTs showed a pronounced negative V_{th} shift accompanied by significant degradation of mobility, subthreshold swing (SS), and off-state current. The authors attribute this difference to the strong ionic bonds in amorphous oxides being more stable against X-rays than the covalent bonds in silicon (Park et al., 2021). The properties discussed in this section are summarized in Table 1. As the table shows, a-IGZO lies between the two silicon technologies in mobility and bias stability, while outperforming both in off-state current and TFT uniformity (Geng et al., 2023).

Table 1
Comparison of a-Si:H, LTPS, and a-IGZO TFT Backplane Technologies

Property	a-Si:H	LTPS	AOS (a-IGZO)
Typical mobility ($\text{cm}^2\text{V}^{-1}\text{s}^{-1}$)	~0.5	~100	~10
Off-state current	Low	Relatively high	Very low
TFT uniformity	Good	Moderate	Very good
Circuit integration	n-type only	CMOS (n- and p-type)	n-type only
Bias stress stability	Poor	Good	Moderate
Process temperature ($^{\circ}\text{C}$)	250–350	350–550	30–350
Number of masks	4–5	7–9	4–8
Largest substrate size achieved (as of 2019)	Gen 11	Gen 6	Gen 8.5
X-ray hardness (up to 2000 Gy)	Negative V_{th} shift and leakage increase above ~100 Gy	Significant degradation	No V_{th} shift
Main limitation	Low mobility; V_{th} shift under bias	Non-uniformity; narrow laser process window; high cost	Moderate bias stability; sensitivity to light and process conditions

Note. Values and assessments are typical and may vary with device architecture and process conditions. The rows for mobility, off-state current, TFT uniformity, circuit integration, bias stress stability, process temperature, number of masks, and substrate size are based on the table that Geng et al. (2023, Table 1) adapted from Jiang et al. (2019); the values in the AOS (a-IGZO) column are those given in that table for amorphous oxide semiconductors. Because that table gives off-state currents without specifying device size and measurement conditions, and bias stability without specifying stress conditions, these two properties are shown qualitatively. TFT uniformity was assessed on the basis of the standard deviation (σ) values in the same table (a-Si:H $\sigma < 0.3$; LTPS $\sigma < 0.5$; AOS $\sigma < 0.2$). The X-ray hardness row is taken from Park et al. (2021) for LTPS and a-IGZO and from Boudry and Antonuk (1996) for a-Si:H; the latter study used ^{60}Co γ -irradiation, found the effects to be small below approximately 100 Gy, and showed that annealing at 200 $^{\circ}\text{C}$ for 2 h restored the preirradiation characteristics. The main limitation row is based on Geng et al. (2023), Powell (1989), Jeong (2013), and Kamiya et al. (2022). The substrate sizes and numbers of masks are the typical values given in that table and reflect the state reported in 2019; the source does not specify the parameter or unit to which the σ values refer. Post-deposition

annealing at temperatures up to approximately 400 °C is discussed in Section 4.2. The 2,940 mm × 3,370 mm substrate listed as Gen. 11 in the source table is also referred to as Gen. 10.5 (Kim et al., 2024).

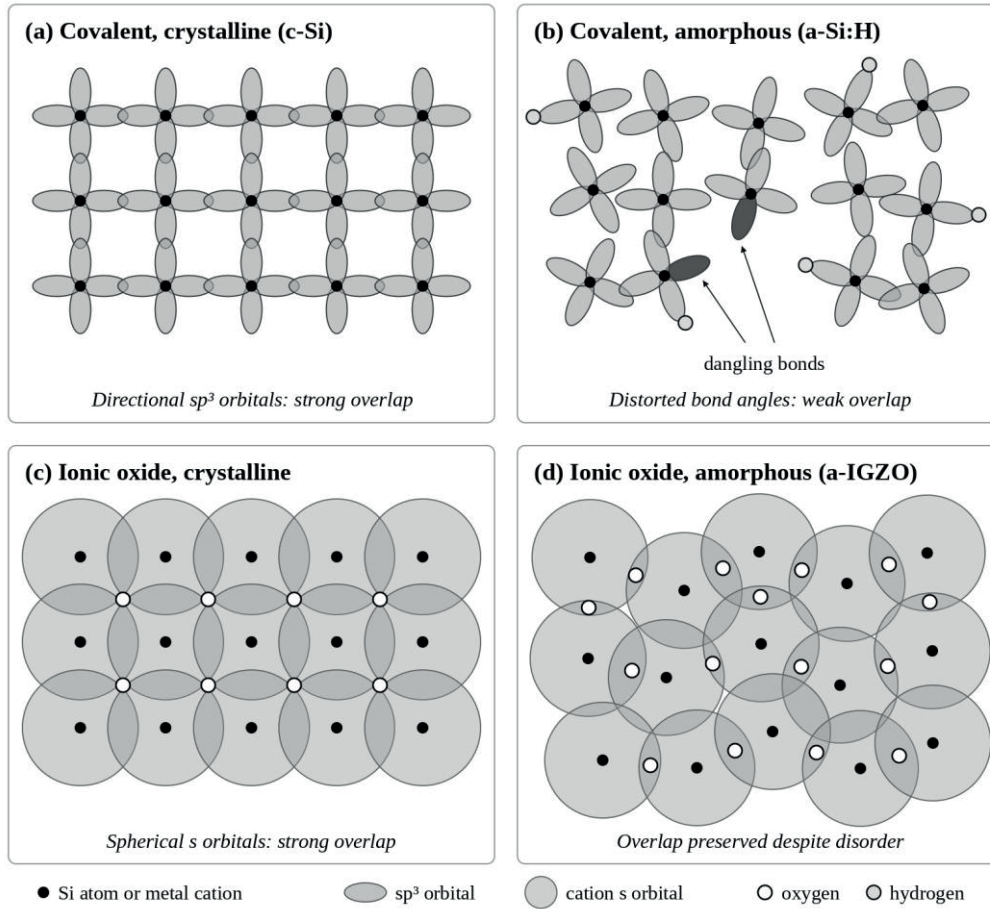
3. Materials Physics of a-IGZO

3.1. Ionic Bonding and ns Orbital Overlap

In ionic AOS the valence band maximum is formed largely by oxygen 2p orbitals and the conduction band minimum by the vacant s orbitals of heavy metal cations (Hosono, 2006); in a-IGZO, the largest contribution to the conduction band minimum comes from the In s orbital (Noh et al., 2011). When the spatial spread of the isotropic s orbitals is larger than the intercation distance, the overlap of neighboring orbitals remains insensitive to the bond angle distribution (Hosono, 2006). Therefore, unlike the directional sp^3 orbitals in a-Si:H mentioned in Section 2, a continuous conduction path is preserved in IGZO even in the amorphous structure (Hosono, 2006; Nomura et al., 2004; Figure 1). Experimentally, a-IGZO still has an exponential tail of localized states below the conduction band edge, with a characteristic energy of approximately 27 meV, and trapped carriers outnumber free carriers at all energy levels examined, so that conduction remains trap-limited (S. Lee & Nathan, 2016). The advantage of a-IGZO therefore lies not in the absence of tail states but in their low density (Section 3.3) and in the s-orbital character of the conduction band edge, which first-principles (DFT) calculations show to be relatively insensitive to disorder (Noh et al., 2011). Owing to this property of the s orbitals, Hall mobilities above $10 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ have been obtained in a-IGZO films deposited at room temperature, and saturation mobilities of 6–9 $\text{cm}^2\text{V}^{-1}\text{s}^{-1}$ in transparent TFTs fabricated on PET substrates (Nomura et al., 2004). These values are approximately one order of magnitude above the field-effect mobility of a-Si:H TFTs, which is on the order of $1 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ (Powell, 1989). Consistent with this s-orbital-based conduction path, the electron effective mass in a-IGZO ($\approx 0.34 m_0$, where m_0 is the free-electron mass) is almost the same as that in crystalline IGZO ($\approx 0.32 m_0$), which implies that the electronic structure and carrier transport of a-IGZO are similar to those of the crystalline phase (Takagi et al., 2005).

Figure 1

Overlap of the Orbitals Forming the Conduction Path in Crystalline and Amorphous Covalent and Ionic Semiconductors



Note. Panel (a) is shown as a two-dimensional square projection of the tetrahedral silicon network. The orbitals are schematic; dark orbitals represent dangling bonds, and small gray circles represent hydrogen atoms terminating dangling bonds (Si-H bonds). The dangling-bond density in (b) is exaggerated for visualization. Drawn by the authors based on Hosono (2006) and Nomura et al. (2004).

3.2. The Role of the Cations

The cation composition directly determines the electrical properties of a-IGZO. As the In fraction increases, the carrier density and mobility rise, whereas Ga, which forms strong bonds with oxygen, suppresses free-carrier generation and shifts the TFT turn-on voltage in the positive direction. The cation ratios thus allow the background carrier density to be controlled over a wide range (Fortunato et al., 2012). DFT calculations also reveal the defect-level counterpart of this experimental trend: the formation energy of an oxygen vacancy increases with the number of neighboring Ga atoms and generally remains low in environments surrounded by In atoms (Noh et al., 2011). Consequently, In, which increases mobility, also facilitates the formation of defects that are a source of instability (Noh et al., 2011). Geng et al.

(2023) likewise note that methods aimed at improving stability can also affect mobility, and that a balance between mobility and stability must therefore be maintained.

The reasons for this relationship have been examined in detail by Shiah et al. (2021). The study showed that the sensitivity of amorphous oxide semiconductors to extrinsic impurities and defects is determined by the position of the conduction band minimum and the resulting dopability. High-mobility oxides with a deep conduction band minimum tend to accept electrons from external sources, and under negative bias stress applied at room temperature, electrons donated by carbon monoxide-related impurities shift the Fermi level. These impurities were found to affect the stability of high-mobility indium-tin-zinc oxide (ITZO) TFTs but not that of low-mobility IGZO TFTs; building on these findings, stable ITZO TFTs with a mobility of approximately $70 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ were fabricated (Shiah et al., 2021).

3.3. Accumulation-Mode TFT Operation and the Density of States

In crystalline silicon metal-oxide-semiconductor field-effect transistors (MOSFETs), conduction begins with the inversion channel formed when the surface potential (ψ_s) reaches twice the bulk Fermi potential (ϕ_F), that is, when $\psi_s \approx 2\phi_F$. In transistors operating in accumulation mode, such as oxide TFTs, by contrast, the conduction threshold remained ambiguous for a long time because of the complex density of states (DOS) distribution in the mobility gap (S. Lee & Nathan, 2016). Because of the deep and tail states created in the band gap by structural disorder, conduction by either free or trapped carriers can dominate, depending on the position of the Fermi level (S. Lee & Nathan, 2016). S. Lee and Nathan (2016) defined the conduction threshold in a-IGZO TFTs as the transition of the Fermi level from the deep states to the tail states, and showed that this point corresponds to where the energy dependence of the accumulated carrier density changes from linear to exponential and that the V_{th} can be extracted directly from the second derivative of the drain current (I_D) with respect to the gate-source voltage (V_{GS}) (Figure 2).

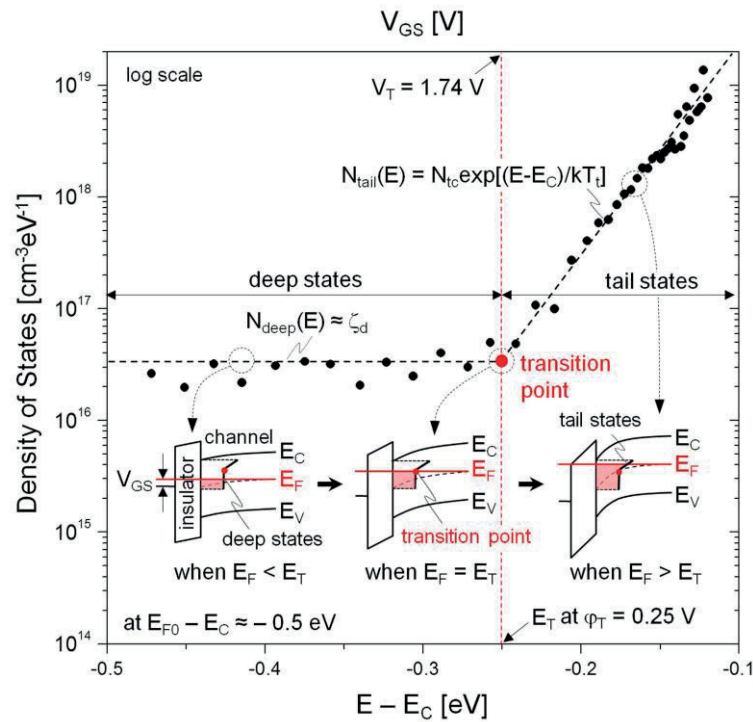
Device simulations show that the density of conduction-band tail states at the band edge in a-IGZO is two to three orders of magnitude lower than that in typical a-Si:H (10^{21} – $10^{22} \text{ cm}^{-3}\text{eV}^{-1}$) and that the low densities of tail and deep gap states lead to small SS values and high mobilities (Hsieh et al., 2008). Accordingly, a-IGZO TFTs exhibit SS of approximately 100 mV/decade, comparable to that of polysilicon TFTs, and the devices can be operated at voltages below 5 V (Kamiya et al., 2010).

Because of the disordered structure, potential barriers of different heights form above the conduction band minimum in a-IGZO, and depending on temperature, electrons travel either

over these barriers or along paths through lower barriers (Kamiya et al., 2010). As the carrier density increases, the effective barrier height that electrons must overcome decreases and mobility increases; this behavior is the opposite of that in single-crystal semiconductors, where mobility generally decreases owing to ionized impurity scattering (Kamiya et al., 2010). When a negative V_{GS} is applied to a-IGZO TFTs, no hole conduction channel forms, and the off-state current remains low and flat; Kamiya et al. (2010) suggest a high density of deep subgap states above the valence band maximum among the possible explanations for this behavior.

Figure 2

Subgap Density of States in an a-IGZO TFT and the Gate-Voltage-Driven Transition of the Fermi Level from Deep States to Tail States (S. Lee & Nathan, 2016)



Note. N_{deep} and N_{tail} denote the densities of the deep and tail states, E_T the transition energy, and ϕ_T the surface potential corresponding to the transition. The band diagrams at the bottom show the cases in which the Fermi level (E_F) is below, above, and exactly at the transition energy. The upper V_{GS} axis is illustrative and only marks the V_{th} corresponding to the transition ($V_T = 1.74$ V). Reproduced without modification from S. Lee and Nathan (2016), Scientific Reports, 6, 22567, under the Creative Commons Attribution 4.0 International (CC BY 4.0) license (<https://creativecommons.org/licenses/by/4.0/>).

3.4. Subgap Defect States

The main defects that determine the stability of a-IGZO are associated with oxygen and hydrogen (Ide et al., 2019; Kamiya et al., 2022). According to DFT calculations, most oxygen vacancies (V_o) are deep donors whose energy levels lie deep in the band gap; some, however, unlike in crystalline oxides and depending on the type and arrangement of the atoms surrounding the vacancy, can behave like shallow donors whose levels lie within the conduction band and which spontaneously release their electrons into the conduction band (Noh et al.,

2011). Under illumination, neutral V_O states can be ionized to V_O^{2+} ; because oxygen vacancies exhibit negative-U behavior, the singly ionized V_O^+ state is unstable (Noh et al., 2011). Because the metal atoms surrounding the vacancy relax outward during ionization, an energy barrier forms between the ionized and neutral states (Noh et al., 2011). V_O^{2+} states can therefore remain stable at room temperature without spontaneously returning to the neutral state (Ryu et al., 2010); the return to the neutral state by electron capture occurs at room temperature for some defects, whereas for others overcoming this barrier requires thermal treatment at up to 400 °C, and some ionized vacancies that create a large void around the vacancy site do not return to the neutral state even at 500–600 °C, becoming shallow donors instead (Noh et al., 2011). Such light-induced ionization of oxygen vacancies is one of the explanations for the instability observed in devices subjected to negative bias illumination stress (NBIS) (Ryu et al., 2010).

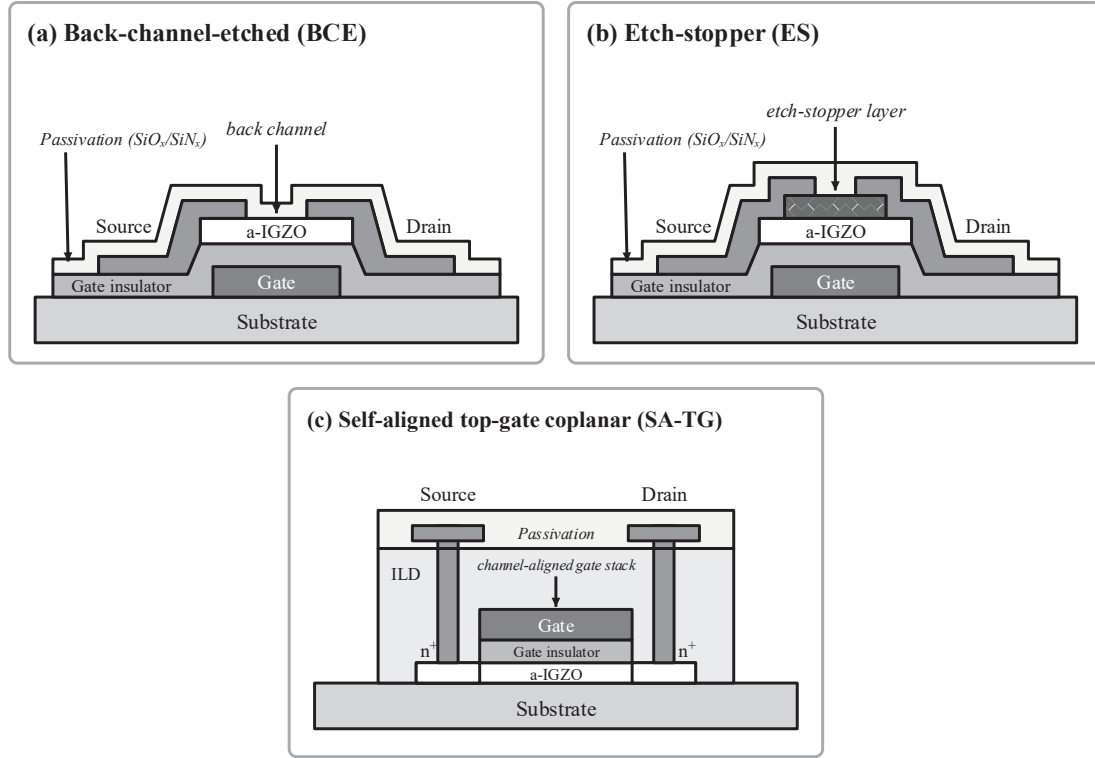
Hydrogen, on the other hand, behaves in two ways. Hydrogen that ionizes as H^+ in the form of $-OH$ donates one free electron. Hydrogen occupying the place of an O^{2-} ion, that is, an oxygen site (Ho ; V_O-H complex), ionizes as H^- , but because it carries one less negative charge than the O^{2-} ion it replaces, it acts as a net donor and again releases one free electron (Kamiya et al., 2022). Some hydrogen, on the other hand, passivates electron traps and thereby improves TFT performance (Hanyu et al., 2013); photoresponse spectroscopy places these traps approximately 0.3 eV below the conduction band edge (Hanyu et al., 2014), and at high concentrations hydrogen causes additional instability (Kamiya et al., 2022). H^- species bonded to metal cations ($M-H$) can be present at levels of approximately 10^{20} cm^{-3} ; they form subgap states above the valence band maximum and have been associated with NBIS instability (Bang et al., 2017). Weakly bonded oxygen incorporated into the a-IGZO active layer under excessively oxidizing conditions during deposition forms electron traps; the fact that these weakly bonded oxygen and hydrogen species can leave the layer even well below 200 °C can cause temperature-induced instability (Kamiya et al., 2022). Another type of defect associated with excess oxygen is the peroxide ($O-O$) structure. Choi et al. (2022) showed that in commercial SA-TG (self-aligned top-gate coplanar) a-IGZO TFTs fabricated with a high oxygen flow rate to keep the V_{th} positive, bias and illumination instabilities accelerate as the channel oxygen content increases, meaning that the conventional V_O model is inadequate for these devices, and they explained these instabilities with an excess-oxygen peroxide model.

4. Device Architectures and Fabrication Processes

4.1. Device Architectures

Figure 3

Schematic Cross-Sections of a-IGZO TFT Architectures: (a) Back-Channel-Etched (BCE), (b) Etch-Stopper (ES), (c) Self-Aligned Top-Gate Coplanar (SA-TG)



Note. Layer thicknesses are not to scale. ILD: interlayer dielectric; n⁺: S/D regions made conductive. The back channel in (a) is exposed during etching of the S/D electrodes and is subsequently covered by the passivation layer. Drawn by the authors based on L. Zhang et al. (2022) and Son et al. (2025).

TFTs are classified according to whether the gate and the source/drain (S/D) contacts are located on opposite sides (staggered) or on the same side (coplanar) of the channel, and according to the position of the gate electrode (bottom-gate or top-gate) (Fortunato et al., 2012). In oxide TFTs, conventional bottom-gate structures are implemented in two forms: back-channel-etched (BCE) and etch-stopper (ES) (L. Zhang et al., 2022; Figure 3). In the BCE structure, which according to L. Zhang et al. (2022) is used in most commercial applications such as mobile phones and televisions because of its simple photolithography process and low cost, the oxide semiconductor is exposed to back-channel damage during etching of the S/D electrodes. In the ES structure, by contrast, an etch-stopper layer formed on the active layer protects the channel during this step, but it requires an additional process step and increases cost (L. Zhang et al., 2022). Moreover, because the S/D electrodes in the ES structure must overlap the edges of the etch-stopper layer, the overlap between the gate and the S/D electrodes

becomes larger (Figure 3b). More generally, the additional layers of the ES process enlarge the inter-layer overlap area, which raises the parasitic capacitance and lowers the aperture ratio (Chung et al., 2018). Chung et al. (2018) showed that in a clean-interface ES process developed on Gen 8.5 (2200 mm × 2500 mm) glass substrates, the V_{th} shift under negative bias temperature illumination stress (NBTIS) and positive bias temperature stress (PBTS), with ± 30 V bias applied for 3600 s at 60 °C, was markedly reduced compared with the BCE structure (-0.51 V and $+1.94$ V versus -3.88 V and $+5.58$ V, respectively).

Top-gate and coplanar structures have also been developed to reduce parasitic components (L. Zhang et al., 2022). Among these, the SA-TG structure has been widely adopted in AMOLED backplanes because it provides low parasitic capacitance and good process control (Son et al., 2025). In this structure, it has been proposed that hydrogen diffusing into the channel during fabrication from the S/D extensions, which are made conductive (n^+), passivates defects in short-channel devices (Son et al., 2025). In the SA-TG structure, the gate insulator and the gate electrode are defined over the channel with the same pattern, and the regions of the active layer outside this stack are made conductive to form the S/D regions; the overlap between the gate and the S/D, and the associated parasitic capacitance, are thus minimized (Figure 3c; L. Zhang et al., 2022; Son et al., 2025).

4.2. Active Layer Deposition, Annealing, and Hydrogen

The a-IGZO active layer is mostly deposited by sputtering (Geng et al., 2023). For example, Kwon et al. (2011) and Anutgan et al. (2026) deposited films by RF magnetron sputtering in an Ar/O₂ gas mixture. The oxygen fraction in the gas mixture, and hence the oxygen partial pressure, is the main process variable that determines the amount of oxygen incorporated into the film and thereby the carrier density (Kwon et al., 2011). Kwon et al. (2011) showed that the film resistivity can be varied by approximately seven orders of magnitude by changing the oxygen partial pressure, that increasing the oxygen partial pressure raises the V_{th} , and that the best transfer characteristics are obtained when the ratio of the flux of O₂ molecules reaching the film during growth to the IGZO flux is approximately 1. Anutgan et al. (2026) also reported that the deposition rate decreases and the resistivity of the deposited films increases as the O₂ flow rate increases. The oxygen partial pressure must be carefully optimized to suppress both oxygen deficiency and the incorporation of excess or weakly bonded oxygen (Kamiya et al., 2022). This parameter also alters the subgap DOS distribution and thereby affects bias and illumination instabilities (Choi et al., 2022; Yang et al., 2022).

Post-deposition annealing rearranges the defect balance, but the direction of the effect depends on the ambient atmosphere and on the passivation layer (W. Zhang et al., 2021). X-ray photoelectron spectroscopy of a-IGZO films annealed in different atmospheres showed that annealing in O₂ markedly reduced the vacancy-related O 1s component, whereas annealing in air left it almost unchanged; the hydrated-oxide-related component increased in all atmospheres. In TFTs with a-SiO_x passivation, annealing in air and O₂ lowered the mobility (from 3.07 to 2.08 and 1.37 cm²V⁻¹s⁻¹) and raised the V_{th} (to 5.1 and 8.0 V), whereas annealing in N₂ or vacuum slightly increased the mobility (to 3.78 and 4.03 cm²V⁻¹s⁻¹) while also raising the V_{th} (from 3.6 V to 4.7 and 5.4 V). All annealed devices showed better positive bias stability than the unannealed device: the V_{th} shift after stress decreased from 3.67 V to 1.88, 2.50, 1.13, and 1.38 V for annealing in air, O₂, N₂, and vacuum, respectively (W. Zhang et al., 2021).

Hydrogen diffusing from the gate insulator into the active layer can also markedly affect device characteristics: Nam et al. (2018) reported that hydrogen from an Al₂O₃ gate insulator grown by atomic layer deposition (ALD) passivates defects and improves the transfer parameters and reliability; in vacuum post-annealing at 200–400 °C, the devices degraded at low temperatures (200–250 °C), which was attributed to the de-passivation of hydrogen in the a-IGZO layer through the SiO₂ passivation layer, and improved at higher temperatures as hydrogen diffused from the gate insulator into the active layer. On the other hand, the hydrogen that passivates electron traps desorbs at approximately 300–430 °C, and TFTs annealed at 400 °C or above show degraded characteristics, which can be recovered by reintroducing hydrogen through annealing in wet O₂ (Hanyu et al., 2013). As a simplified picture, these findings suggest distinguishing weakly bonded species, which can leave the layer well below 200 °C (Kamiya et al., 2022), from trap-passivating hydrogen, which desorbs at approximately 300–430 °C (Hanyu et al., 2013); however, as the de-passivation already observed at 200–250 °C by Nam et al. (2018) shows, the temperatures at which hydrogen is lost or supplied also depend on the hydrogen source and the passivation structure. The importance of process history also manifests itself in the thermal response; it has been reported that even a-IGZO films with similar resistivity can exhibit markedly different thermal transient behaviors depending on their fabrication history (Anutgan et al., 2026).

4.3. Basic Device Parameters and Their Extraction

The comparability of mobility, V_{th}, and stability values reported in different studies depends on how these parameters are extracted. In a-IGZO TFTs operating in accumulation mode, the I_D is in most cases described with sufficient accuracy by the gradual channel

approximation developed for crystalline MOSFETs (Fortunato et al., 2012; Kamiya et al., 2010). The relations used are, in the linear region ($V_{DS} < V_{GS} - V_{th}$),

$$I_D = \mu_{FE} C_i \frac{W}{L} \left[(V_{GS} - V_{th}) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (1)$$

and in the saturation region ($V_{DS} \geq V_{GS} - V_{th}$),

$$I_D = \frac{WC_i \mu_{sat}}{2L} (V_{GS} - V_{th})^2. \quad (2)$$

Here, C_i is the gate insulator capacitance per unit area, and W and L are the TFT channel width and length, respectively. The field-effect mobility (μ_{FE}) is extracted from the transconductance in the linear region at small V_{DS} ($V_{DS} \ll V_{GS} - V_{th}$), where the quadratic term in Equation (1) can be neglected ($g_m = \partial I_D / \partial V_{GS}$), and the saturation mobility (μ_{sat}) from the slope of the $\sqrt{I_D}$ – V_{GS} curve; V_{th} is determined from the point at which the linear part of this curve intersects the voltage axis (Figure 4; Fortunato et al., 2012; Kamiya et al., 2010). Because the V_{th} can also be determined as the V_{GS} corresponding to a given I_D (Fortunato et al., 2012) or from the second derivative of the I_D with respect to the V_{GS} (S. Lee & Nathan, 2016), the extraction method must be stated when comparing reported V_{th} values. The V_{GS} at which the current begins to rise from the off-state level is defined as the turn-on voltage (V_{on}) (Fortunato et al., 2012).

The SS is the change in V_{GS} required to increase the I_D by one decade in the subthreshold region (Fortunato et al., 2012):

$$SS = \left(\frac{\partial \log_{10} I_D}{\partial V_{GS}} \right)^{-1} \quad (3)$$

Because SS is sensitive to the trap density in the TFT channel and at the interface, the maximum total trap density (N_t , $\text{cm}^{-2}\text{eV}^{-1}$) is estimated using the relation

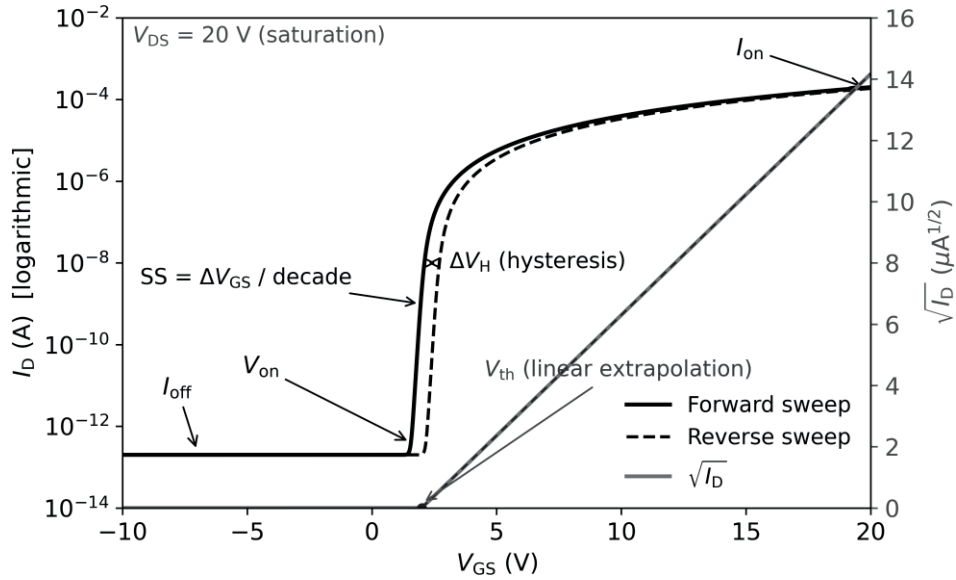
$$N_t \approx \left[\frac{SS \log_{10} e}{kT/q} - 1 \right] \frac{C_i}{q} \quad (4)$$

given by Rolland et al. (1993); because C_i/q has the unit $\text{cm}^{-2}\text{V}^{-1}$, N_t is numerically equal to its value in $\text{cm}^{-2}\text{eV}^{-1}$ when energy is expressed in eV. Since $kT/q \approx 25.9$ mV at room temperature ($T = 300$ K), the theoretical lower limit of SS is approximately 60 mV/decade. In addition, the on/off current ratio (I_{on}/I_{off}) is reported, and hysteresis (ΔV_H), defined as the voltage difference between the forward and reverse sweeps, is used as a quick indicator of charge trapping (Fortunato et al., 2012). Because contact resistance and channel-edge effects can distort the

extraction of these parameters, as discussed in Section 4.4, it is important that mobility values be reported together with the device geometry and the extraction method.

Figure 4

Definition of the Basic Device Parameters on the Transfer Characteristic of an a-IGZO TFT



Note. The curves are schematic, generated with a simple analytical model to illustrate the parameter definitions; they are not measured data. Left axis: I_D on a logarithmic scale (forward and reverse sweeps); right axis: $\sqrt{I_D}$ ($\mu A^{1/2}$), with $V_{DS} = 20$ V. Drawn by the authors based on Fortunato et al. (2012) and S. Lee and Nathan (2016).

4.4. Channel Geometry and Scaling Effects

In real TFT devices, W and L give rise to parasitic effects that, beyond the scaling of the current with the W/L ratio, affect both parameter extraction and stability. As the TFT channel becomes shorter, the channel resistance decreases while the series contact resistance remains approximately constant, so its share of the total resistance increases and the extracted field-effect mobility falls below the true channel mobility (Jeong et al., 2008). In bottom-gate a-IGZO TFTs, it has been shown that a high contact resistance ($R_C W \approx 200 \Omega \cdot \text{cm}$) lowers the normalized field-effect mobility by approximately 22% in devices with a channel length of 10 μm , whereas contact-limited behavior is almost absent in devices with low contact resistance ($R_C W \approx 23 \Omega \cdot \text{cm}$) (Jeong et al., 2008).

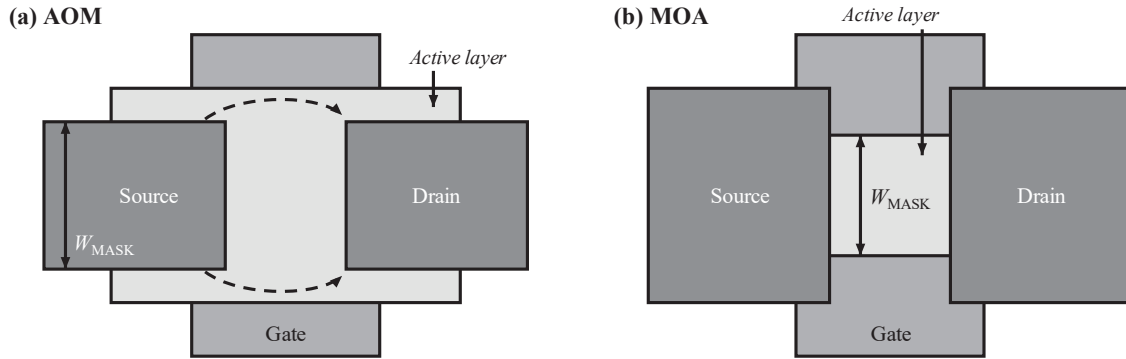
A similar parameter-extraction problem exists along the channel-width direction. In bottom-gate inverted staggered TFTs designed with the active layer extending beyond the S/D electrodes along the W direction (active-over-metal, AOM; Figure 5a), current also flows along the channel edges through the active-layer regions lying outside the S/D electrodes. These edge currents do not form a separate parasitic channel but widen the main channel; consequently, the effective channel width (W_{EFF}) exceeds the mask value (W_{MASK}) and can be expressed as W_{EFF}

$= W_{\text{MASK}} + \Delta W$ (Mativenga et al., 2012). Mativenga et al. (2012) determined the extension at each edge to be approximately $1.35 \mu\text{m}$ for a-Si:H TFTs and approximately $2.55 \mu\text{m}$ for a-IGZO TFTs. Because this extension is relatively large in narrow devices, the field-effect mobility calculated with W_{MASK} is increasingly overestimated as the channel narrows, whereas it becomes independent of the W when W_{EFF} is used (Mativenga et al., 2012). This design-related behavior differs from a fabrication-related edge effect reported in earlier studies. When the edges of the active layer are poorly etched, the gate insulator becomes thinner and the electric flux density higher at the channel corners, so that more electrons are trapped at the edges than in the main channel under gate bias stress. The edges then act as parasitic transistors in parallel with the main channel, so that one physical device behaves as more than one transistor, and a hump appears in the subthreshold region of the transfer characteristics (Mativenga et al., 2012).

Figure 5

Edge Effects Along the Channel-Width Direction in Bottom-Gate Inverted Staggered TFTs:

Top Views of (a) Active-Over-Metal (AOM) and (b) Metal-Over-Active (MOA) Designs



Note. Top view; dimensions are not to scale. In (a), the dashed arrows indicate the current flowing along the channel edges through the active-layer regions lying outside the S/D electrodes; consequently, the effective channel width exceeds W_{MASK} , which here corresponds to the S/D width. In (b), W_{MASK} corresponds to the width of the active layer and is approximately equal to the effective channel width. AOM and MOA denote the in-plane overhang of the active layer or of the S/D electrodes along W , not the vertical layer order (Tsormpatzoglou et al., 2012). Drawn by the authors based on Mativenga et al. (2012).

To minimize this effect, a design in which the S/D electrodes extend over the active area along the W direction (metal-over-active, MOA; Figure 5b) is recommended; in this structure, the width of the active layer is a good approximation of W_{EFF} (Mativenga et al., 2012). In the MOA structure, however, the a-IGZO edges remain within the channel, so, according to Tsormpatzoglou et al. (2012), defects that may be induced at the edges by the plasma process

used to define the active layer are reflected in the device behavior. Indeed, the near-interface gate-insulator trap density extracted from low-frequency noise measurements was found to be approximately $2 \times 10^{12} \text{ cm}^{-2} \text{ eV}^{-1}$ in MOA devices and approximately $5 \times 10^{11} \text{ cm}^{-2} \text{ eV}^{-1}$ in AOM devices, and hysteresis in the transfer characteristics was observed only in the MOA devices (Tsormpatzoglou et al., 2012). Since a low trap density is important for applications such as medical imaging, where noise must be kept as low as possible (Mativenga et al., 2012), channel-edge design involves a trade-off between the accuracy of parameter extraction and interface quality. Gate–S/D overlap capacitances, in turn, are a significant contribution that must be separated out in DOS extraction based on capacitance–voltage measurements (Bae et al., 2013).

The effect of geometry on stability, meanwhile, appears in different directions depending on the device architecture and the type of stress. In bottom-gate staggered a-IGZO TFTs, within the investigated range ($W = 3\text{--}50 \text{ }\mu\text{m}$, $L = 3\text{--}10 \text{ }\mu\text{m}$), increasing W and L markedly improved reliability under +30 V positive bias stress, with the most stable behavior obtained at $W/L = 50/10 \text{ }\mu\text{m}$ (Cho et al., 2024). In SA-TG structures, by contrast, short-channel devices have been reported to be more stable under positive bias temperature stress, which was associated with a reduction in both the subgap DOS in the channel and the near-interface trap density in the gate insulator (Son et al., 2025). Also in SA-TG structures, degradation under self-heating stress has been shown to be non-uniform along the channel length and to become more pronounced in wide-channel devices, mainly because of increased electron trapping in the gate insulator (D.-H. Lee et al., 2023).

5. Open Problems and Research Gaps in the Literature

The studies discussed in the preceding sections show that, although a-IGZO TFTs are widely used in commercial displays, some fundamental problems at the material and device levels still await solution. These problems can be grouped under three headings.

The first is achieving high mobility and high stability together in the same device. Although high-mobility, stable devices have been demonstrated by reducing impurities, this result was obtained in TFTs with ITZO channels (Shiah et al., 2021). Because the impurity sensitivity addressed in that study is specific to high-mobility compositions with a deep conduction band minimum and was not observed in low-mobility IGZO (Shiah et al., 2021), the open question is whether the same impurity control can be achieved in other high-mobility compositions and under mass-production conditions.

The second problem is the reliable identification of defect types. In X-ray photoelectron spectroscopy (XPS), the component of the O 1s spectrum in the 531–532 eV region is commonly assigned to oxygen vacancies. However, many contributions, such as hydroxyl groups, surface-bound water, and organic oxygen species, can fall within this region (Easton & Morgan, 2025). Easton and Morgan (2025) note that, although it has been shown that oxygen vacancies cannot be identified in this way by ex situ XPS, this assignment remains widespread in many fields, catalysis in particular. Spectroscopic findings should therefore be interpreted together with electrical DOS analyses. For this purpose, optical capacitance–voltage methods have been developed that extract the subgap DOS directly by separating out the parasitic capacitance contribution (Bae et al., 2013). The fact that the same instability can be explained by V_O ionization (Yang et al., 2022) or peroxide (Choi et al., 2022) models under different process conditions also shows that defect types must be verified in the context of the device and the process.

The third problem is that the interactions between process history, channel geometry, and stability have not been examined in an integrated manner. The effects of oxygen flow rate (Yang et al., 2022), annealing atmosphere (W. Zhang et al., 2021), and channel dimensions (Cho et al., 2024) on stability have mostly been addressed in separate studies. Nor do the existing findings point to a single trend: in bottom-gate devices, stability under positive bias stress improves as W and L increase (Cho et al., 2024), whereas in SA-TG devices short channels have been reported to be more stable under positive bias temperature stress (Son et al., 2025) and wide channels to degrade more under self-heating stress (D.-H. Lee et al., 2023). As a possible reason for the greater stability of short-channel devices in SA-TG structures, passivation of defects by hydrogen diffusing into the channel from the S/D extensions has been proposed (Son et al., 2025). The fact that a significant portion of the recent and detailed findings on channel-length-dependent stability mechanisms comes from SA-TG structures with such S/D extensions raises the question of the extent to which these mechanisms can be transferred to bottom-gate structures, which have no S/D extension regions. In this context, it becomes important to evaluate device performance not only in terms of peak mobility but in a multi-criteria manner, together with V_{th} , SS, I_{on}/I_{off} , hysteresis, and post-stress parameter shifts.

6. Conclusion

In this chapter, a-IGZO TFTs have been examined in terms of materials physics, device architecture, and fabrication processes. The advantage of a-IGZO over silicon-based TFTs rests on the combination of high mobility, preserved even in the amorphous structure owing to the spherical symmetry of the s orbitals that form the conduction band minimum, and an extremely low off-state current. Although the cation composition allows these properties to be tuned over a wide range, the fact that In, which increases mobility, also facilitates the formation of oxygen vacancies shows that the trade-off between mobility and stability begins at the material level.

On the other hand, the behavior of a-IGZO TFTs depends strongly on subgap states associated with oxygen and hydrogen, on fabrication history, and on parasitic effects arising from geometry. The fact that the same instability is explained by different defect models under different process conditions in the literature, and that the effect of TFT geometry on stability appears in different directions depending on the architecture and the type of stress, shows that these factors cannot be treated independently of one another.

Accordingly, the priority research areas can be summarized as follows: achieving high mobility and high stability together in the same device, reliably identifying defect types with complementary electrical and spectroscopic methods, and examining process history, geometry, and stability together on the same experimental platform.

References

- Anutgan, T., Güler, M. İ., & Anutgan, M. (2026). On the thermal transient behaviors of a-IGZO TFTs: Roles of oxygen flow rate during sputtering and post-deposition air annealing. *Materials Science in Semiconductor Processing*, 210, 110703. <https://doi.org/10.1016/j.mssp.2026.110703>
- Bae, H., Choi, H., Oh, S., Kim, D. H., Bae, J., Kim, J., Kim, Y. H., & Kim, D. M. (2013). Extraction technique for intrinsic subgap DOS in a-IGZO TFTs by de-embedding the parasitic capacitance through the photonic C–V measurement. *IEEE Electron Device Letters*, 34(1), 57–59. <https://doi.org/10.1109/LED.2012.2222014>
- Bang, J., Matsuishi, S., & Hosono, H. (2017). Hydrogen anion and subgap states in amorphous In–Ga–Zn–O thin films for TFT applications. *Applied Physics Letters*, 110(23), 232105. <https://doi.org/10.1063/1.4985627>
- Boudry, J. M., & Antonuk, L. E. (1996). Radiation damage of amorphous silicon, thin-film, field-effect transistors. *Medical Physics*, 23(5), 743–754. <https://doi.org/10.1118/1.597668>
- Cho, J., Choi, J., Nguyen, M. P., Trinh, T. T., Song, J.-K., Kim, Y.-S., Pham, D. P., & Yi, J. (2024). Gate bias stress reliability of a-InGaZnO TFTs under various channel dimension. *Microelectronics Reliability*, 153, 115308. <https://doi.org/10.1016/j.microrel.2023.115308>
- Choi, S., Park, J., Hwang, S.-H., Kim, C., Kim, Y.-S., Oh, S., Baeck, J. H., Bae, J. U., Noh, J., Lee, S.-W., Park, K.-S., Kim, J.-J., Yoon, S. Y., Kwon, H.-I., & Kim, D. H. (2022). Excessive oxygen peroxide model-based analysis of positive-bias-stress and negative-bias-illumination-stress instabilities in self-aligned top-gate coplanar In-Ga-Zn-O thin-film transistors. *Advanced Electronic Materials*, 8(5), 2101062. <https://doi.org/10.1002/aelm.202101062>
- Chung, J.-M., Zhang, X., Shang, F., Kim, J.-H., Wang, X.-L., Liu, S., Yang, B., & Xiang, Y. (2018). Enhancement of a-IGZO TFT device performance using a clean interface process via etch-stopper nano-layers. *Nanoscale Research Letters*, 13, 164. <https://doi.org/10.1186/s11671-018-2571-9>
- Easton, C. D., & Morgan, D. J. (2025). Critical examination of the use of X-ray photoelectron spectroscopy (XPS) O 1s to characterize oxygen vacancies in catalytic materials and beyond. *Journal of Vacuum Science & Technology A*, 43(5), 053205. <https://doi.org/10.1116/6.0004686>

- Fan, C.-L., Chen, C.-Y., Liu, S.-Y., & Lin, W.-Y. (2022). AMOLED pixel circuit using LTPO technology supporting variable frame rate from 1 to 120 Hz for portable displays. *Micromachines*, 13(9), 1505. <https://doi.org/10.3390/mi13091505>
- Fortunato, E., Barquinha, P., & Martins, R. (2012). Oxide semiconductor thin-film transistors: A review of recent advances. *Advanced Materials*, 24(22), 2945–2986. <https://doi.org/10.1002/adma.201103228>
- Geng, D., Wang, K., Li, L., Myny, K., Nathan, A., Jang, J., Kuo, Y., & Liu, M. (2023). Thin-film transistors for large-area electronics. *Nature Electronics*, 6(12), 963–972. <https://doi.org/10.1038/s41928-023-01095-8>
- Hanyu, Y., Abe, K., Domen, K., Nomura, K., Hiramatsu, H., Kumomi, H., Hosono, H., & Kamiya, T. (2014). Effects of high-temperature annealing on operation characteristics of a-In-Ga-Zn-O TFTs. *Journal of Display Technology*, 10(11), 979–983. <https://doi.org/10.1109/JDT.2014.2352860>
- Hanyu, Y., Domen, K., Nomura, K., Hiramatsu, H., Kumomi, H., Hosono, H., & Kamiya, T. (2013). Hydrogen passivation of electron trap in amorphous In-Ga-Zn-O thin-film transistors. *Applied Physics Letters*, 103(20), 202114. <https://doi.org/10.1063/1.4832076>
- Hosono, H. (2006). Ionic amorphous oxide semiconductors: Material design, carrier transport, and device application. *Journal of Non-Crystalline Solids*, 352(9–20), 851–858. <https://doi.org/10.1016/j.jnoncrysol.2006.01.073>
- Hosono, H. (2018). How we made the IGZO transistor. *Nature Electronics*, 1(7), 428. <https://doi.org/10.1038/s41928-018-0106-0>
- Hosono, H., Kikuchi, N., Ueda, N., & Kawazoe, H. (1996). Working hypothesis to explore novel wide band gap electrically conducting amorphous oxides and examples. *Journal of Non-Crystalline Solids*, 198–200, 165–169. [https://doi.org/10.1016/0022-3093\(96\)80019-6](https://doi.org/10.1016/0022-3093(96)80019-6)
- Hosono, H., Yasukawa, M., & Kawazoe, H. (1996). Novel oxide amorphous semiconductors: Transparent conducting amorphous oxides. *Journal of Non-Crystalline Solids*, 203, 334–344. [https://doi.org/10.1016/0022-3093\(96\)00367-5](https://doi.org/10.1016/0022-3093(96)00367-5)

- Hsieh, H.-H., Kamiya, T., Nomura, K., Hosono, H., & Wu, C.-C. (2008). Modeling of amorphous InGaZnO₄ thin film transistors and their subgap density of states. *Applied Physics Letters*, 92(13), 133503. <https://doi.org/10.1063/1.2857463>
- Huh, J.-Y., Jeon, J.-H., Choe, H.-H., Lee, K.-W., Seo, J.-H., Ryu, M.-K., Park, S.-H. K., Hwang, C.-S., & Cheong, W.-S. (2011). Effects of the composition of sputtering target on the stability of InGaZnO thin film transistor. *Thin Solid Films*, 519(20), 6868–6871. <https://doi.org/10.1016/j.tsf.2011.01.400>
- Ide, K., Nomura, K., Hosono, H., & Kamiya, T. (2019). Electronic defects in amorphous oxide semiconductors: A review. *Physica Status Solidi (a)*, 216(5), 1800372. <https://doi.org/10.1002/pssa.201800372>
- Jeong, J. K. (2013). Photo-bias instability of metal oxide thin film transistors for advanced active matrix displays. *Journal of Materials Research*, 28(16), 2071–2084. <https://doi.org/10.1557/jmr.2013.214>
- Jeong, J. K., Chung, H.-J., Mo, Y.-G., & Kim, H. D. (2008). Comprehensive study on the transport mechanism of amorphous indium-gallium-zinc oxide transistors. *Journal of The Electrochemical Society*, 155(11), H873–H877. <https://doi.org/10.1149/1.2972031>
- Jiang, C., Cheng, X., & Nathan, A. (2019). Flexible ultralow-power sensor interfaces for e-skin. *Proceedings of the IEEE*, 107(10), 2084–2105. <https://doi.org/10.1109/JPROC.2019.2936105>
- Kamiya, T., Nomura, K., & Hosono, H. (2010). Present status of amorphous In–Ga–Zn–O thin-film transistors. *Science and Technology of Advanced Materials*, 11(4), 044305. <https://doi.org/10.1088/1468-6996/11/4/044305>
- Kamiya, T., Nomura, K., Ide, K., & Hosono, H. (2022). Defects and relevant properties. In H. Hosono & H. Kumomi (Eds.), *Amorphous oxide semiconductors: IGZO and related materials for display and memory* (pp. 93–103). Wiley. <https://doi.org/10.1002/9781119715641.ch4>
- Kim, S. H., Kim, M., Lee, J. H., Kim, K., Park, J. S., Lim, J. H., & Oh, S. (2024). Optimizing length scalability of InGaZnO thin-film transistors through lateral carrier profile engineering and negative ΔL extension structure. *Advanced Electronic Materials*, 10, 2400012. <https://doi.org/10.1002/aelm.202400012>

- Kwon, S., Noh, J. H., Noh, J., & Rack, P. D. (2011). Quantitative calculation of oxygen incorporation in sputtered IGZO films and the impact on transistor properties. *Journal of The Electrochemical Society*, 158(3), H289–H293. <https://doi.org/10.1149/1.3530779>
- Lee, D.-H., Jeong, H.-S., Kim, Y.-G., Kim, M.-H., Son, K. S., Lim, J. H., Song, S.-H., & Kwon, H.-I. (2023). Quantitative analysis of channel width effects on electrical performance degradation of top-gate self-aligned coplanar IGZO thin-film transistors under self-heating stresses. *Journal of Semiconductor Technology and Science*, 23(1), 79–87. <https://doi.org/10.5573/JSTS.2023.23.1.79>
- Lee, S., & Nathan, A. (2016). Conduction threshold in accumulation-mode InGaZnO thin film transistors. *Scientific Reports*, 6, 22567. <https://doi.org/10.1038/srep22567>
- Mativenga, M., Um, J. K., Kang, D. H., Mruthyunjaya, R. K., Chang, J. H., Heiler, G. N., Tredwell, T. J., & Jang, J. (2012). Edge effects in bottom-gate inverted staggered thin-film transistors. *IEEE Transactions on Electron Devices*, 59(9), 2501–2506. <https://doi.org/10.1109/TED.2012.2205258>
- Matsuo, N., Heya, A., & Hamada, H. (2019). Review—Technology trends of poly-Si TFTs from the viewpoints of crystallization and device performance. *ECS Journal of Solid State Science and Technology*, 8(4), P239–P252. <https://doi.org/10.1149/2.0211903jss>
- Nam, Y., Kim, H.-O., Cho, S. H., & Park, S.-H. K. (2018). Effect of hydrogen diffusion in an In–Ga–Zn–O thin film transistor with an aluminum oxide gate insulator on its electrical properties. *RSC Advances*, 8(10), 5622–5628. <https://doi.org/10.1039/C7RA12841J>
- Noh, H.-K., Chang, K. J., Ryu, B., & Lee, W.-J. (2011). Electronic structure of oxygen-vacancy defects in amorphous In-Ga-Zn-O semiconductors. *Physical Review B*, 84(11), 115205. <https://doi.org/10.1103/PhysRevB.84.115205>
- Nomura, K., Ohta, H., Takagi, A., Kamiya, T., Hirano, M., & Hosono, H. (2004). Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors. *Nature*, 432(7016), 488–492. <https://doi.org/10.1038/nature03090>
- Nomura, K., Ohta, H., Ueda, K., Kamiya, T., Hirano, M., & Hosono, H. (2003). Thin-film transistor fabricated in single-crystalline transparent oxide semiconductor. *Science*, 300(5623), 1269–1272. <https://doi.org/10.1126/science.1083212>

- Park, S., Song, M.-K., Sung, T., & Kwon, J.-Y. (2021). Effect of X-ray irradiation on a-IGZO and LTPS thin-film transistors for radiography applications. *Applied Surface Science*, 550, 149237. <https://doi.org/10.1016/j.apsusc.2021.149237>
- Powell, M. J. (1989). The physics of amorphous-silicon thin-film transistors. *IEEE Transactions on Electron Devices*, 36(12), 2753–2763. <https://doi.org/10.1109/16.40933>
- Rolland, A., Richard, J., Kleider, J.-P., & Mencaraglia, D. (1993). Electrical properties of amorphous silicon transistors and MIS-devices: Comparative study of top nitride and bottom nitride configurations. *Journal of The Electrochemical Society*, 140(12), 3679–3683. <https://doi.org/10.1149/1.2221149>
- Ryu, B., Noh, H.-K., Choi, E.-A., & Chang, K. J. (2010). O-vacancy as the origin of negative bias illumination stress instability in amorphous In–Ga–Zn–O thin film transistors. *Applied Physics Letters*, 97(2), 022108. <https://doi.org/10.1063/1.3464964>
- Shiah, Y.-S., Sim, K., Shi, Y., Abe, K., Ueda, S., Sasase, M., Kim, J., & Hosono, H. (2021). Mobility–stability trade-off in oxide thin-film transistors. *Nature Electronics*, 4(11), 800–807. <https://doi.org/10.1038/s41928-021-00671-0>
- Sirringhaus, H. (2014). 25th anniversary article: Organic field-effect transistors: The path beyond amorphous silicon. *Advanced Materials*, 26(9), 1319–1335. <https://doi.org/10.1002/adma.201304346>
- Son, D.-H., Oh, C.-E., Lee, H.-W., Jeong, C.-Y., Jang, J.-M., Ahn, B.-D., Bae, J.-U., & Kwon, H.-I. (2025). Analysis of physical mechanisms for channel-length-dependent PBTS reliability in SA TG coplanar IGZO TFTs. *Scientific Reports*, 15, 43556. <https://doi.org/10.1038/s41598-025-27549-x>
- Street, R. A. (2022). Oxide TFTs and their application to X-ray imaging. In H. Hosono & H. Kumomi (Eds.), *Amorphous oxide semiconductors: IGZO and related materials for display and memory* (pp. 503–517). Wiley. <https://doi.org/10.1002/9781119715641.ch23>
- Takagi, A., Nomura, K., Ohta, H., Yanagi, H., Kamiya, T., Hirano, M., & Hosono, H. (2005). Carrier transport and electronic structure in amorphous oxide semiconductor, a-InGaZnO₄. *Thin Solid Films*, 486(1–2), 38–41. <https://doi.org/10.1016/j.tsf.2004.11.223>

- Tsormpatzoglou, A., Hastas, N. A., Khan, S., Hatalis, M., & Dimitriadis, C. A. (2012). Comparative study of active-over-metal and metal-over-active amorphous IGZO thin-film transistors with low-frequency noise measurements. *IEEE Electron Device Letters*, 33(4), 555–557. <https://doi.org/10.1109/LED.2012.2185677>
- Yang, G. W., Park, J., Choi, S., Kim, C., Kim, D. M., Choi, S.-J., Bae, J.-H., Cho, I. H., & Kim, D. H. (2022). Total subgap range density of states-based analysis of the effect of oxygen flow rate on the bias stress instabilities in a-IGZO TFTs. *IEEE Transactions on Electron Devices*, 69(1), 166–173. <https://doi.org/10.1109/TED.2021.3130219>
- Zhang, L., Yu, H., Xiao, W., Liu, C., Chen, J., Guo, M., Gao, H., Liu, B., & Wu, W. (2022). Strategies for applications of oxide-based thin film transistors. *Electronics*, 11(6), 960. <https://doi.org/10.3390/electronics11060960>
- Zhang, W., Fan, Z., Shen, A., & Dong, C. (2021). Atmosphere effect in post-annealing treatments for amorphous InGaZnO thin-film transistors with SiO_x passivation layers. *Micromachines*, 12(12), 1551. <https://doi.org/10.3390/mi12121551>