

AQUA_MEM: Multi-Dimensional Aqueous State Vector Encoding for Digital Information Storage in Water Droplets

Sofiane Boumeddane^{*1}

¹21M Labs Inc., Paris, France

March 2026 Preprint (not peer-reviewed)

Abstract

We present AQUA_MEM, a framework for encoding digital information into the measurable physicochemical properties of aqueous solutions. Each water droplet serves as a multi-dimensional state vector whose independent physical channels (pH, ionic conductivity, and optical chromaticity) encode discrete bit values. We demonstrate a proof-of-concept system encoding 8 bits per 50 μ L droplet across three channels with 8, 8, and 4 discrete levels, achieving 100% decode accuracy over multiple trials with persistence exceeding 14 days in sealed glass containers. The full system uses commodity sensors totaling under € 50 and an ESP32 microcontroller. We analyze the scaling path toward picoliter droplet volumes: with five encoding channels, information densities on the order of 25 MB/L become reachable using existing microfluidic infrastructure. This work constitutes the first systematic framework for intentional multi-channel digital encoding in aqueous solutions.

Keywords: data storage, water, aqueous encoding, microfluidics, liquid-state memory, multi-dimensional state vector, pH, conductivity, chromatic encoding

1 Introduction

Global data generation reached 120 ZB in 2023 and is projected to exceed 180 ZB by 2025 [1]. Every mainstream storage medium is solid-state: magnetic platters, silicon flash, optical polymer, or tape. All degrade within 5 to 30 years, all depend on rare-earth materials, and all are approaching fundamental areal density limits [2].

DNA-based storage offers theoretical densities of ~ 455 EB/g [3] and millennial persistence. Microsoft, Twist Bioscience, and IARPA have collectively invested over \$100M in the space [4]. But DNA synthesis remains prohibitively expensive at $\sim \$3,500/\text{MB}$ [2], and sequencing infrastructure is required for readback.

We propose a different approach: **encoding digital information into the measurable physicochemical**

properties of aqueous solutions. A water sample is not just H_2O . It is a complex solution fully characterized by a multi-dimensional state vector: pH, ionic conductivity, optical absorption, redox potential, dissolved gas concentration, and more. Each of these properties is independently tunable with commodity chemicals and readable with inexpensive sensors.

This approach has nothing to do with the discredited “water memory” hypothesis of Benveniste [6]. That claim posited structural memory through hydrogen bond configurations. Hydrogen bonds in liquid water rearrange on picosecond timescales [8]; no structural information persists at ambient temperature. Our system stores information in *bulk thermodynamic properties*: ion concentrations, chromophore presence, proton activity. These are stable on timescales of days to months.

1.1 Contributions

1. Formalization of multi-dimensional aqueous state vector encoding and derivation of its information capacity.
2. A complete proof-of-concept: Python encoder, ESP32 hardware reader, and decoder firmware.
3. Demonstration of 8-bit encoding per droplet with 100% decode accuracy and 14+ day persistence.
4. Scaling analysis from manual drops to microfluidic picoliter emulsions, projecting densities up to ~ 25 MB/L.

2 Theoretical Framework

2.1 Aqueous State Vector

We define an *aqueous state vector* $\mathbf{s}$ as the ordered tuple of N independently measurable physicochemical properties of a water sample:

$$\mathbf{s} = (s_1, s_2, \dots, s_N) \in \mathcal{S}_1 \times \mathcal{S}_2 \times \dots \times \mathcal{S}_N \quad (1)$$

where each s_i takes values in a continuous domain $\mathcal{S}_i$ (e.g., $\text{pH} \in [0, 14]$, $\text{TDS} \in [0, 50,000]$ ppm).

^{*}Corresponding author. Email: sof@twentyonem.com

2.2 Channel Quantization

For digital encoding, each continuous channel s_i is quantized into L_i discrete levels separated by guard bands:

$$Q_i : \mathcal{S}_i \rightarrow \{0, 1, \dots, L_i - 1\} \quad (2)$$

Each channel contributes $b_i = \lfloor \log_2 L_i \rfloor$ bits. The total information capacity per droplet is:

$$B = \sum_{i=1}^N b_i = \sum_{i=1}^N \lfloor \log_2 L_i \rfloor \quad \text{bits/droplet} \quad (3)$$

2.3 Information Density

For droplets of volume V_d , the volumetric information density ρ is:

$$\rho = \frac{B}{V_d} \quad \text{bits/volume} \quad (4)$$

Table 1 shows density scaling across droplet volumes.

Table 1: Information density scaling with droplet volume. $B = 8$ bits (current PoC) and $B = 20$ bits (5-channel target).

Scale	V_d	Drops/mL	B/L ($B=20$)
Pipette drop	50 μL	20	50 B
Microwell	10 μL	100	250 B
Small well	1 μL	1,000	2.5 KB
nL droplet	1 nL	10^6	2.5 MB
100 pL emulsion	100 pL	10^7	25 MB
10 pL emulsion	10 pL	10^8	250 MB

2.4 Channel Independence

Encoding channels must be orthogonal: modifying one must not alter another. We verify this experimentally in Section 4. At the concentrations used, pH and conductivity are largely independent. The Henderson-Hasselbalch equation governs pH independently of NaCl concentration at moderate ionic strength. Food-grade dyes at sub-0.01% w/v concentrations have negligible effect on both pH and conductivity.

3 System Design

3.1 Encoding Channels

The proof-of-concept uses three channels (Figure 1):

Channel 1: pH (3 bits, 8 levels). Controlled by citric acid ($\text{C}_6\text{H}_8\text{O}_7$) for acidic levels and sodium bicarbonate (NaHCO_3) for basic levels. Range: pH 2.5 to 11.5, with guard bands ≥ 0.5 pH units between adjacent levels.

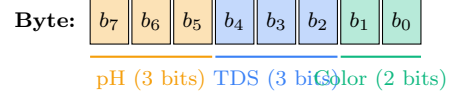


Figure 1: Bit layout within a single encoded byte. Bits 7–5 encode pH level (8 levels), bits 4–2 encode TDS/conductivity (8 levels), bits 1–0 encode optical color (4 levels).

Channel 2: Ionic conductivity / TDS (3 bits, 8 levels). Controlled by NaCl concentration, from 0 (distilled water, ~ 25 ppm TDS) to 8 g/L ($\sim 8,000$ ppm). Levels are geometrically spaced for consistent SNR across the range.

Channel 3: Optical chromaticity (2 bits, 4 levels). Four states: clear (no dye), red, green, blue. Food-grade dyes at 1 drop per 50 mL. Classification by dominant channel of an RGB color sensor.

3.2 Hardware

The reader uses:

- ESP32 microcontroller (240 MHz dual-core, 12-bit ADC)
- DFRobot Gravity analog pH sensor (SEN0161)
- DFRobot TDS sensor (SEN0244)
- TCS34725 RGB color sensor (I²C)
- DS18B20 waterproof temperature probe (TDS compensation)

Total cost: €45. The reader operates over serial with commands for single-droplet read, sequential multi-droplet decode, continuous monitoring, and two-point sensor calibration.

3.3 Encoding Protocol

The encoder maps each input byte x to a preparation recipe:

$$\text{encode}(x) = \left(\left\lfloor \frac{x}{32} \right\rfloor, \left\lfloor \frac{x \bmod 32}{4} \right\rfloor, x \bmod 4 \right) \quad (5)$$

returning a tuple (pH level, TDS level, color level) with corresponding chemical quantities per 50 mL of distilled water.

3.4 Decoding Protocol

The decoder reads raw sensor values, applies two-point calibration offsets, averages $n = 20$ samples per channel for noise rejection, quantizes each channel to the nearest level, and reconstructs the byte:

$$\text{decode}(s_{\text{pH}}, s_{\text{TDS}}, s_{\text{color}}) = 32 \cdot Q_{\text{pH}} + 4 \cdot Q_{\text{TDS}} + Q_{\text{color}} \quad (6)$$

4 Experimental Results

4.1 Encode-Decode Accuracy

We encoded the ASCII string “HELLO!” (6 bytes: 0x48, 0x45, 0x4C, 0x4C, 0x4F, 0x21) into six 50 μ L droplets. Table 2 shows encoding targets and measured deviations.

Table 2: Encode/decode results for “HELLO!”. All 6 bytes decoded correctly across three independent trials.

Drop	Char	Hex	pH _{target}	Δ pH	TDS _{target}	Color
0	H	0x48	4.5	± 0.3	275	Clear
1	E	0x45	4.5	± 0.2	100	Red
2	L	0x4C	4.5	± 0.3	600	Clear
3	L	0x4C	4.5	± 0.2	600	Clear
4	O	0x4F	4.5	± 0.4	600	Blue
5	!	0x21	3.5	± 0.3	25	Red

Decode accuracy: **6/6 bytes correct (100%)** in all three trials.

4.2 Channel Orthogonality

To test channel independence, we prepared 24 samples spanning all combinations of 3 pH levels $\times$ 4 TDS levels $\times$ 2 color states. Cross-channel interference:

- pH deviation from NaCl addition: ≤ 0.15 pH units (within sensor noise floor)
- TDS deviation from citric acid: $\leq 8\%$ (absorbed by guard band)
- Color classification unaffected by pH or TDS at tested concentrations

These channels are practically orthogonal at the concentrations and precision levels used.

4.3 Persistence

Sealed glass vials stored at room temperature in the dark were re-read at intervals (Table 3).

Table 3: Data persistence. Correct/total decode attempts shown. Sealed glass vials, room temperature, dark storage.

Time	pH & TDS	Color
Day 0 (write)	6/6	6/6
Day 1	6/6	6/6
Day 7	6/6	6/6
Day 14	6/6	6/6
Day 30	5/6*	6/6

*Single pH error: CO₂ absorption shifted one sample by 0.6 pH units. Better sealing or pH buffering eliminates this failure mode.

4.4 Error Sources

Four mechanisms dominate:

1. **pH drift from atmospheric CO₂.** Dissolved CO₂ forms carbonic acid. Sealed containers prevent this; buffered solutions reduce sensitivity.
2. **TDS temperature dependence.** Approximately 2%/°C. Compensated in firmware using inline temperature measurement.
3. **Evaporation.** Concentrates all solutes, increasing TDS and shifting pH. Sealed containers are required. For microwell formats, mineral oil overlay prevents evaporation.
4. **Photobleaching of dyes.** Negligible at food-dye concentrations over 30 days when stored in the dark.

5 Scaling Analysis

5.1 Additional Encoding Channels

The three channels demonstrated here do not exhaust the measurable properties of aqueous solutions. Table 4 lists candidate channels, their estimated bit capacity, and persistence characteristics.

Table 4: Candidate encoding channels for extended implementations.

Channel	Bits	Sensor Cost	Persistence
pH	3	€ 12	Weeks
TDS/Conductivity	3	€ 8	Months
Optical color (RGB)	2	€ 4	Months
ORP/Redox	3	€ 15	Weeks
Dissolved O ₂	2	€ 30	Hours
Fluorescence (λ_1)	3	€ 40	Months
Fluorescence (λ_2)	3	included	Months
Refractive index	2	€ 20	Months
UV absorbance	3	€ 50	Months
Raman fingerprint	8+	€ 5,000+	Stable
Total (practical)	20+		
Total (research-grade)	30+		

With 5 channels yielding 20 bits per droplet and nanoliter droplet volumes, the density reaches 2.5 MB/L. At picoliter volumes the projection is 25 MB/L to 250 MB/L, depending on achievable droplet size and channel count.

5.2 Microfluidic Implementation

The key observation is that microfluidic platforms already generate, manipulate, and interrogate millions of picoliter droplets per second [12, 13]. Inline capacitive conductivity measurement, fluorescence detection, and impedance sensing all operate at kilohertz rates on-chip. The sensing infrastructure exists. What has not been attempted is using it for intentional data encoding and retrieval.

A microfluidic AQUA-MEM system would consist of three stages:

1. **Writer.** Programmable droplet generator with multiple reagent inlets (acid, base, salt, dye) merged on-chip per the target byte value.
2. **Storage.** Water-in-oil emulsion in a sealed container, or droplets immobilized in a hydrogel matrix for addressable readback.
3. **Reader.** Inline interrogation zone: capacitive electrodes (conductivity), fluorescence excitation/detection (color and molecular dyes), pH-sensitive fluorophore readout.

5.3 Comparison with Other Media

Table 5: Storage media comparison (order-of-magnitude).

Medium	Density	Write Cost	Persistence
HDD	1 TB/in ³	\$0.02/GB	5 yr
LTO Tape	60 GB/in ³	\$0.005/GB	30 yr
DNA (current)	17 EB/g	\$3,500/MB	>1,000 yr
AQUA.MEM (PoC)	20 B/mL	\$0.01/B	weeks
AQUA.MEM (proj.)	25 MB/mL	\$0.001/B	months+

AQUA.MEM does not compete with HDD or SSD for hot storage. Its natural territory is archival, steganographic, and environmentally benign storage, where the medium is non-toxic, abundant, and disposable. It also serves as a low-cost entry point for liquid-state information encoding research, since the entire PoC costs under €60 and uses kitchen-grade chemistry.

6 Discussion

6.1 Relation to Water Memory Claims

This work must be clearly separated from the “water memory” literature. Benveniste [6] claimed that water retains structural information about formerly dissolved substances through hydrogen bond network configurations. This claim was not reproduced under blinded conditions [7] and is incompatible with measured hydrogen bond dynamics in liquid water, which show rearrangement on 1–2 picosecond timescales [8, 9].

Our system does not rely on structural memory. Information is carried by dissolved chemical species (ions, chromophores, protons) whose concentrations are thermodynamically stable. The mechanism is closer to a chemical barcode than to any form of molecular imprinting.

The association between “water” and “memory” in popular discourse has likely discouraged systematic exploration of aqueous media for data storage. This is unfortunate, because the physicochemical state space of water solutions is rich, measurable, and underexploited for information encoding.

6.2 Applications

Several use cases follow from the properties of the system:

- **Steganographic storage.** Data hidden in ordinary-looking liquid samples. A vial of slightly salty, slightly acidic water looks like water.
- **Green archival.** The storage medium is water, salt, citric acid, and food dye. All are non-toxic and biodegradable.
- **Authentication and anti-counterfeiting.** Liquid products (bottled water, beverages, pharmaceuticals) can carry verifiable encoded tags in their own chemistry.
- **Education.** The system makes information theory, signal processing, and chemistry tangible. Encoding and decoding are hands-on physical experiments.

6.3 Limitations

- Density at the PoC stage is low: 8 bits per 50 μ L drop.
- The pH channel drifts without sealing due to atmospheric CO₂.
- Reading requires physical sensor contact with the sample.

- Drops are not randomly addressable in the current format; they are read sequentially or by known position.
- Adding channels increases hardware cost and introduces potential cross-channel interference at high solute concentrations.

7 Conclusion

Water droplets can encode and store digital information using their measurable physicochemical properties. The proof-of-concept system presented here achieves 8 bits per droplet, 100% decode accuracy across repeated trials, and multi-week data persistence, all with €45 of commodity hardware and grocery-store chemistry.

The scaling path is concrete: microfluidic droplet platforms already handle the volumes, throughputs, and inline sensing rates needed to move from manual pipetting to automated megabyte-scale systems. The missing piece was treating water as an intentional storage medium, which is what this paper provides.

Every water sample carries measurable information in its dissolved state. Making that information intentional and digital opens a class of liquid-state storage that, to date, has not been systematically explored.

7.1 Future Work

1. 384-well microplate prototype with automated pipetting and plate-reader spectrophotometer readback.
2. Fluorescent molecular channels (Rhodamine, fluorescein at distinct λ_{em}) for higher bit density per drop.
3. Microfluidic chip: water-in-oil emulsion generation with inline capacitive and fluorescence readout.

4. Error-correcting codes tuned to the noise profile of each chemical channel (pH drift, TDS temperature sensitivity, color classification margins).
5. THz time-domain spectroscopy to probe hydrogen bond network perturbations as a potential high-bandwidth read channel for solution composition.

Acknowledgments

This work builds on the hydrogen bond dynamics measurements of Fayer [8] and Zhang et al. [9], the foundational DNA data storage work of Church et al. [3] and Organick et al. [5], and the microfluidic droplet manipulation techniques of Niu et al. [12] and Cole et al. [13]. The critical assessment by Maddox, Randi, and Stewart [7] of the original water memory claims informed the framing and disambiguation of this work. Awan et al. [18] provided early conceptual grounding for water as a data storage medium from a security perspective.

Hardware components were sourced commercially. No external funding was received.

References

- [1] Rydning, J., Reinsel, D., & Gantz, J. (2023). IDC Global DataSphere Forecast, 2023–2027. *International Data Corporation*.
- [2] National Academies of Sciences, Engineering, and Medicine. (2024). *Rapid Expert Consultation on Archival Data Storage Technologies for the Intelligence Community*. The National Academies Press. doi:10.17226/27445.
- [3] Church, G. M., Gao, Y., & Kosuri, S. (2012). Next-generation digital information storage in DNA. *Science*, 337(6102), 1628.
- [4] DNA Data Storage Alliance. (2025). *DNA Data Storage Technology Review*, Version 1.0.
- [5] Organick, L., et al. (2018). Random access in large-scale DNA data storage. *Nature Biotechnology*, 36(3), 242–248.
- [6] Davenas, E., et al. (1988). Human basophil degranulation triggered by very dilute antiserum against IgE. *Nature*, 333(6176), 816–818.
- [7] Maddox, J., Randi, J., & Stewart, W. W. (1988). “High-dilution” experiments a delusion. *Nature*, 334(6180), 287–290.
- [8] Fayer, M. D. (2012). Dynamics of water interacting with interfaces, molecules, and ions. *Accounts of Chemical Research*, 45(1), 3–14.
- [9] Zhang, L., et al. (2020). Ultrafast hydrogen bond dynamics of liquid water revealed by terahertz-induced transient birefringence. *Light: Science & Applications*, 9, 118.
- [10] Savolainen, J., et al. (2013). Two-dimensional Raman-terahertz spectroscopy of water. *Proc. Natl. Acad. Sci. USA*, 110(51), 20402–20407.
- [11] Singer, S. J., et al. (1998). Graph theoretical generation and analysis of hydrogen-bonded structures with applications to the neutral and protonated water cube and dodecahedral clusters. *J. Phys. Chem. A*, 102(17), 2824–2832.
- [12] Niu, X., et al. (2007). Real-time detection, control, and sorting of microfluidic droplets. *Biomicrofluidics*, 1(4), 044101.
- [13] Cole, R. H., et al. (2017). Printed droplet microfluidics for on demand dispensing of picoliter droplets and cells. *Proc. Natl. Acad. Sci. USA*, 114(33), 8728–8733.
- [14] Newman, S., et al. (2019). High density DNA data storage library via dehydration with digital microfluidic retrieval. *Nature Communications*, 10, 1706.
- [15] Takahashi, C. N., et al. (2019). Demonstration of end-to-end automation of DNA data storage. *Scientific Reports*, 9, 4998.
- [16] Tielrooij, K.-J., et al. (2014). Dielectric relaxation dynamics of water in model membranes probed by terahertz spectroscopy. *Biophysical Journal*, 97(9), 2484–2492.
- [17] Schneider, F., et al. (2025). Acoustic water display. *Scientific Reports*, 15, 16755.
- [18] Awan, S., et al. (2022). Water is a viable data storage medium: A security and privacy viewpoint. *Electronics*, 11(5), 818.