

WHAT A QUANTUM COMPUTER SEES WHEN IT READS A BOOK THE CASE FOR QUANTUM AGENTS

CHRISTOPHE JURCZAK

APRIL 2026

[arXiv:2605.14188](https://arxiv.org/abs/2605.14188)

The barrier to entry to quantum hardware has quietly collapsed. Cloud-accessible processors, a mature software stack, and agentic LLM tooling now let a single investigator carry an application end to end — without first becoming a quantum physicist. I wanted to test that claim at its limit, in a field about as far from quantum computing as one can travel: the structural analysis of literary texts.

The bridge is the Maximum Independent Set, a problem neutral-atom processors solve in their own physics. Each unit of a book — page, canto, chapter — becomes an atom; semantic similarity becomes a physical blockade constraint; the machine's lowest-energy configuration is the book's structural backbone. From it I introduce *rigidity*, a metric for how unique that backbone is. I then invert the pipeline: choose a graph the hardware encodes natively and write a book to match it, extending OuLiPo's constrained writing to graph topology. Twenty-nine engineered texts and eight historical works run on Pasqal's FRESNEL at up to one hundred atoms. What is reported is an application layer, not a speedup.

The next step takes the same agentic approach to physics: working with Pasqal's team to carry theory papers smoothly onto Pasqal QPUs, from paper to protocol elaboration, hardware run and data analysis.

QUANTONATION

QOuLiPo: What a quantum computer sees when it reads a book

Christophe Jurczak¹ 

Quantonation, Dallas, TX, USA

May 15, 2026

Abstract. What does a book look like to a quantum computer? This paper takes eight classical works of the Renaissance and its late-antique inheritance — from Augustine through Dante, Giambullari, and Galileo — and runs each one through a neutral-atom quantum processor. The bridge is the mathematics of graphs. Encoding texts as graphs is by now standard in the digital humanities; what is new is that the graph is not a representation the computer manipulates but the physical configuration the hardware itself realises. Each textual unit (page, canto, chapter, paragraph, depending on the work) becomes an atom; for engineered exact unit-disk designs, graph edges *are* physical blockade constraints, while for natural texts the processor receives the best 2D unit-disk approximation to the semantic graph. A question literary scholars actually ask is then read directly off the atoms: which subset of a book’s units is mutually so distinct that no additional unit can be added without violating their mutual independence?

Three contributions follow. First, we introduce *rigidity* ρ , a new metric that measures how unique a book’s structural backbone is — discriminating Marguerite de Navarre’s *Heptaméron* (rigid, twelve-nouvelle hard core) from Boethius’s *De Consolatione Philosophiae* (fully fungible, every chapter substitutable). Second, we invert the pipeline: rather than extracting a graph from existing prose, we pick a target graph the hardware encodes natively, and *write* a book whose structure matches it. The twenty-nine texts written this way, collected under the name *QOuLiPo*, extend the OuLiPo tradition to graph-topological constraints and, together with the eight natural texts of the corpus, form a benchmark distribution against which neutral-atom hardware can be tracked as it scales. Third, we run both natural and engineered texts on Pasqal’s FRESNEL processor at sizes up to one hundred atoms: the hardware behaves faithfully across the corpus and, on engineered texts purpose-built to match its native geometry, reaches high approximation ratios and, on the cleanest instances, returns the exact backbone; on natural texts the semantic graph must be embedded approximately into the planar register, exposing that embedding step as the present bottleneck and better embedders as the path to fuller recovery. The framework already opens to a new mode the hardware natively supports — three-dimensional registers that escape the planar embedding bottleneck.

A cloud-accessible quantum machine plus an agentic coding environment now lets a single investigator run this pipeline end-to-end. What is reported is an application layer, not a speedup — humanistic instances ready to load onto neutral-atom processors as they scale, already complementing classical text analysis. The Digital Humanities community has a stake in building familiarity with this hardware now: the engineered-corpus design choices made today fix the benchmark distribution future hardware will be measured against.

1 Introduction

The Maximum Independent Set (MIS) problem asks for the largest subset of mutually non-adjacent vertices in a graph. In physics, MIS occupies a privileged position on *unit-disk graphs*: graphs whose vertices can be drawn as points in the plane such that two vertices are connected if and only if they lie within a fixed distance. An array of cold atoms (typically rubidium, caesium, or ytterbium) held in optical tweezers, where no two atoms within a blockade radius can simultaneously occupy an excited state, can be driven by an adiabatic ramp of the laser detuning into a configuration that solves MIS on the unit-disk graph defined by the array, as its lowest-energy arrangement under the Rydberg Hamiltonian. No software compilation sits between the problem and the hardware. The physical setting is the *Rydberg Hamiltonian* on a neutral-atom register [1, 2]; quantum MIS on this platform has been proposed [3] and demonstrated [4, 5]. The hardware that realises it is the neutral-atom quantum processors built by several companies.

This mapping is not a metaphor. The Rydberg blockade *is* the independence constraint, enforced

¹christophe@quantonation.com

by the physics: any two atoms within R_b cannot both be driven to the Rydberg state, exactly as the Independent Set constraint forbids any two adjacent vertices from both being selected. Laser detuning plays the role of a per-vertex inclusion reward, and the ground state prepared by the adiabatic protocol is the MIS of the blockade-adjacency graph [1, 2, 6, 7]. The hardware, the register geometry, and a concrete MIS read-out are shown in Figure 1.

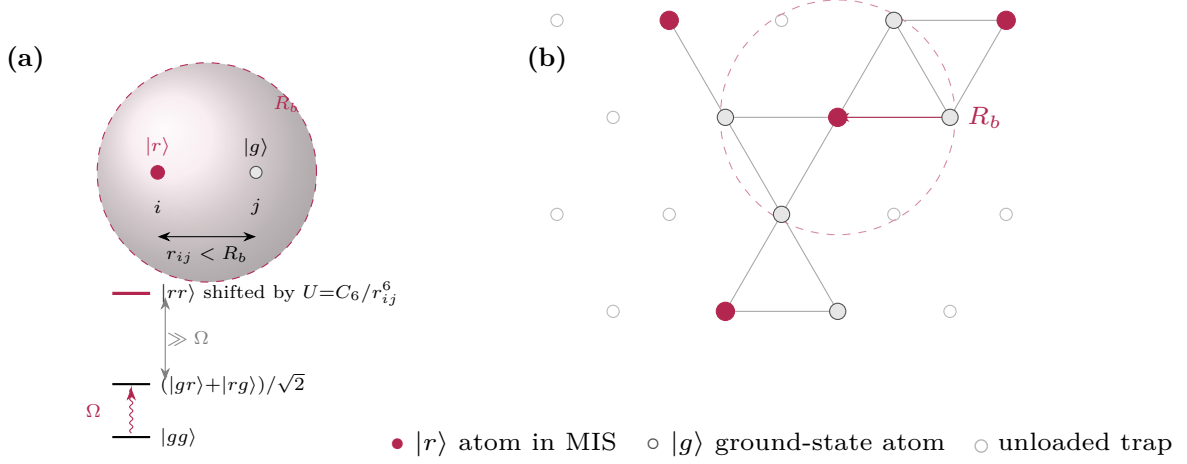


Figure 1: Neutral-atom MIS in one picture. (a) *Rydberg blockade*. Two atoms held within a blockade radius R_b cannot both be driven to the excited Rydberg state $|r\rangle$: the doubly-excited level $|rr\rangle$ is shifted up by an interaction energy $U = C_6/r_{ij}^6$ much larger than the Rabi coupling Ω that the laser would need to populate it, so only the symmetric singly-excited combination $(|gr\rangle + |rg\rangle)/\sqrt{2}$ is reachable. The blockade radius R_b is the distance at which $U(R_b) = \hbar\Omega$; for typical $\Omega/2\pi \sim 1\text{--}5$ MHz on rubidium-87, $R_b \approx 5\text{--}10\ \mu\text{m}$. (b) *Atom register = text graph*. The faint open circles are an optical-tweezer lattice (FRESNEL uses a triangular trap layout); atoms are loaded into a chosen subset, and that loaded subset is the register that runs the computation. Each loaded atom represents one textual unit (a page, a canto, a chapter); the gray edges connect every loaded pair within blockade range R_b and are therefore the edges of the underlying graph. The register's ground state under the Rydberg drive is a Maximum Independent Set of that graph: a largest subset of atoms (red, in $|r\rangle$) with no two within R_b of each other. The dashed circles around the red atoms mark their blockade exclusion zones — no second atom can be excited inside them, which is exactly the independence constraint. The register is not a representation of the graph: it *is* the graph, encoded in matter. After [2, 4].

Here we bring MIS to an unexpected domain: the structural analysis of literary texts, specifically the books that the Renaissance recovered and printed. Any book can be decomposed into units (nouvelles, cantos, chapters, printed pages or paragraphs), each embedded as a vector in a high-dimensional semantic space. Connecting units whose embeddings are sufficiently similar produces a graph, and the MIS of that graph selects what we call the *structural backbone* of the book: the largest set of pairwise-dissimilar units, with every omitted unit adjacent to at least one selected one.

A graph can have many distinct optimal MIS solutions, all of the same size. We introduce *rigidity* $\rho \in [0, 1]$ as the fraction of backbone nodes common to every optimal solution: $\rho = 1$ means the backbone is unique (every selected unit is irreplaceable); $\rho = 0$ means the backbone is maximally degenerate (any element can be substituted by another structurally equivalent one). Rigidity is the central new quantity of this paper: it characterises the space of possible structures a text permits.

Contributions. (i) We establish MIS as a structural analysis tool for historical texts in their original languages *under a fixed multilingual sentence-transformer embedding* (**multilingual-e5-large-instruct**), validating it across eight corpora spanning four centuries and three languages, and showing that rigidity ρ , a property of the (text, embedding) pair, discriminates between qualitatively different kinds of textual organisation. The results are striking: Dante's *Inferno* is moderately rigid ($\rho = 0.444$); Marguerite de Navarre's *Heptaméron* at $k = 16$ has a 12-nouvelle hard core in every optimum ($\rho = 0.857$); Boethius's *De Consolatione* is fully fungible ($\rho = 0$, hundreds of optima whose intersection is empty). Backbone *size* (MIS/N) varies moderately (0.10–0.30); rigidity varies dramatically and reveals structural properties invisible to word-frequency methods [8] and topic models [9]. (ii) We invert the pipeline: rather than extract a graph from given prose, we pick a target graph with prescribed properties (unique

MIS by double-domination, 3^{17} equivalent backbones by disjoint cliques, exact unit-disk realisations on king lattices, a dodecahedron by 30-motif edge-assignment, and more) and *write* prose so the page-level adjacencies match the target. This extends OuLiPo (the *Ouvroir de Littérature Potentielle*, founded 1960) [10–13] to graph-topological constraints; we engineer twenty-nine such texts at register sizes (16–100 atoms) matched to the hardware. The designed graphs are ground truth by construction, so the engineered corpus is also a calibration instrument for the embedder, the classical Integer Linear Programming (ILP) solver, and the Quantum Processing Unit (QPU) pipeline. (iii) We run both natural and engineered texts on Pasqal’s FRESNEL² neutral-atom QPU across register sizes $N = 25$ –100. On every row the approximation ratio against the canonical adjacency the QPU solves sits in the published 0.84–1.00 band of neutral-atom MIS benchmarks [4, 5] — exact-UDG engineered rows and simulated-annealing (SA) embedded k -NN rows alike. The recovery rate of the underlying *text*-MIS is bounded above by SA recall on the natural-text rows: the upstream 2D embedding step is what loses information, not the processor.

The Renaissance was itself a technology revolution applied to texts: humanist editors collated manuscripts, the Aldine press stabilised them in pocket octavos, and a body of classical and contemporary works that had been scattered across monastic libraries became reproducible, comparable, and portable. Applying today’s most advanced computational technology, neutral-atom quantum processors, to those same objects is therefore not an arbitrary pairing but a continuation: each generation reads the inherited corpus through the most analytic tools it has. We read these same texts (Ausonius, Dante, Marguerite de Navarre, Giambullari, Augustine, Lactantius, Boethius, Galileo) and ask which admit a unique structural backbone, which admit many, and whether a quantum processor can recover that backbone by physical means.

Methodology: a single-investigator agentic loop. Three developments now make such bridging possible at the scale of a single investigator. Cloud access exposes the machines; large-language-model (LLM) agents now translate a domain question (“what are the structural chapters of this book?”) through the full pipeline (text $\rightarrow$ embedding $\rightarrow k$ -NN graph $\rightarrow$ atom register $\rightarrow$ laser pulse), submit it, and explain the result back in the vocabulary of the originating field. Shiraishi *et al.* [14] demonstrated an LLM-agent interface to quantum emulators built on the Model Context Protocol (MCP), Anthropic’s open standard for connecting language-model agents to external tools and services; this paper is the Pasqal/FRESNEL instantiation. Concretely, every QPU and EMU_MPS (Pasqal Cloud’s matrix-product-state classical emulator of the Rydberg Hamiltonian) submission reported here was authored through an agentic loop with Anthropic’s Claude Opus (versions 4.6 and 4.7, 1-million-token context for the larger campaigns): the agent reads the text-graph artefacts, calls Pulser (Pasqal’s open-source Python library for designing atom registers and pulse sequences [15]) to construct the register and pulse, submits to the Pasqal Cloud, and reports approximation ratio, validity, and Hamming-weight statistics. The loop runs end-to-end through standard Bash, Python, and file-system tool calls; an MCP-style native interface to the Pasqal Cloud, analogous to [14], would streamline this further but is not a prerequisite. Operationally, the cloud QPU was stable enough to support an iterative single-investigator workflow on a fixed timeline: FRESNEL_CAN1, hosted at the DistriQ quantum innovation zone in Sherbrooke, Québec, was available from the user’s perspective every day except for planned maintenance, with no submission lost to an unscheduled outage, and no result in this paper depends on privileged hardware access. Day-to-day reliability of this kind is the precondition without which a single-investigator agentic loop is not viable in practice; it is what lets a project plan around quantum hardware in the same way it would plan around any other piece of cloud infrastructure. Taken together, these three pieces (agentic interface, mature SDK, reliable cloud hardware) bring the barrier to entry to quantum hardware lower than commonly assumed: with agentic LLM tooling on top of a stable cloud QPU, domain specialists (literary scholars, historians, philologists) can drive a real QPU end-to-end without first becoming quantum physicists.

What this paper is, and is not. This paper is deliberately written as a boundary object between quantum computing and computational literary studies [16–18]. Its contribution is not a new quantum algorithm nor a finished literary reading. It is an application-layer and corpus proposal: a reproducible interface between text structure and hardware-native neutral-atom MIS, plus a benchmark distribution (the QOuLiPo corpus) calibrated against the published (N, d) neutral-atom MIS landscape of Cazals *et al.* [5], where N is register size and d graph density. The pieces a humanist reader needs (a deposited

²The FRESNEL processor is named after Augustin-Jean Fresnel (1788–1827), the French physicist whose wave theory of light, Fresnel equations, and stepped-lens design are foundational to optics and to the laser-based atom-trapping techniques the machine relies on.

corpus, an ILP-verified backbone, a falsifiable rigidity metric) and the pieces a physics reader needs (an exact-UDG benchmark family, a hardware-native pipeline, a regime where classical simulation hits its ceiling) sit on the same object. Static 2D-UDG MIS admits a polynomial-time approximation scheme (PTAS) and is not a quantum-utility target at our sizes (MIS on general graphs remains NP-hard); what this paper establishes, on the way to the harder regimes addressed in §6, is that the hardware behaves faithfully on the corpus. The paper should be read as a manifesto for a category that did not exist before, populated densely enough to be tested.

This contribution sits at the intersection of two strands of prior work. On the text side, MIS has been applied to extractive text summarisation before [19, 20], but not as a structural-analysis tool for literary corpora, and never as far as we’re aware on quantum hardware. Graph-based representations of texts have a growing track record in computational linguistics: complex-network features correlate with text quality and authorship [21–23]; character-interaction networks reveal narrative structure [24–26]; grammar-aware quantum natural-language-processing (NLP) models have been run on gate-based hardware for sentence classification [27]. Our approach differs: unit-level semantic similarity builds the graph, and MIS extracts a whole-graph backbone rather than ranking individual nodes.

A handful of bridges between quantum computing and the humanities have already been built. Barzen and Leymann [28, 29] have called for a “quantum humanities” research programme; Miranda’s edited volume [30] collects the existing work. Published hardware experiments concentrate in two small clusters: quantum natural-language processing (QNLP) on superconducting and trapped-ion processors, scaling from toy sentence classification to hundred-sentence experiments [27, 31, 32], and quantum-assisted music composition collected in Miranda’s volume. QOuLiPo is distinct from the QNLP lineage in its compute modality (analog Rydberg-blockade MIS on an atom register rather than parameterised quantum circuits) and extends the platform base of the field by adding neutral atoms to the hardware already represented. The pipeline is end-to-end on real hardware, with constrained writing as the inverse output of a designed graph for the engineered half of the corpus.

2 From Text to Graph

Figure 2 shows the end-to-end pipeline on a 1544 book, from page photograph to bitstring histogram. The graph-construction step itself (which is what this section describes) proceeds through four stages. First, the source text is segmented into N units. We adopt a tiered granularity rule: when the author divided the text into discrete structural units (cantos, nouvelles, prosimetric pieces, poems, chapters, paragraphs — per text, see Appendix A), we use those. When no such division exists and page-faithful alignment with the original edition is available, we fall back to the printed page — the smallest division the historical reader actually encountered. For long continuous prose whose authorial division is too coarse to give the pipeline statistical resolution (Galileo’s *Dialogo*, four days of dialogue at $N=4$, being the canonical case), we fall back to fixed analytic windows — 400 words for Galileo, chosen as the median word count of a printed page of the 1632 Landini *editio princeps*, so that the analytic window remains commensurate with what the historical reader encountered. The choice of unit per text is reported explicitly in Appendix A.

Sourcing faithful transcriptions of early-modern texts in their original language is harder than it sounds. Two failure modes appear. First, the text exists online but only as a modernised critical edition: spelling normalised, long-s resolved, ligatures expanded. Ausonius is representative: his poetry is freely available through the Loeb Classical Library [34], The Latin Library [35], and DigiLiBLT [36], but no open transcription preserves the orthography of the Aldine 1517 *editio princeps* we actually photographed. Second, no open-access transcription exists at all. Giambullari’s 1544 *De’l sito, forma, & misura dello Inferno di Dante* (Florence, Neri Dortelata) is the Accademia Fiorentina’s first systematic reconstruction of Dante’s infernal topography — building on Antonio Manetti’s earlier geometric attempt (c. 1480s, printed posthumously by Giunti in 1506) and Landino’s 1481 commentary, and feeding directly into the lineage that culminates in Galileo’s 1587 Florentine Academy lectures on the geometry of Dante’s *Inferno* [37, 38] — yet no digital edition of the *editio princeps* is publicly available: not in Bibliotheca Italiana, not in the Accademia della Crusca, not on Wikisource, not on Google Books.

Both gaps are closed by a single per-edition low-rank-adaptation (LoRA) fine-tune of a vision-language model on a handful of manually-corrected reference pages, run on photographs taken with a consumer smartphone rather than a flatbed scanner. The Giambullari case is documented elsewhere [33]; the Ausonius case is a 211-page transcription of the Aldine 1517 *Opera*. The remaining six corpus texts are sourced from validated original-language transcriptions in open repositories (see Appendix A); smartphone photography plus fine-tuned VLM-OCR reduces the infrastructure barrier to a level accessible to

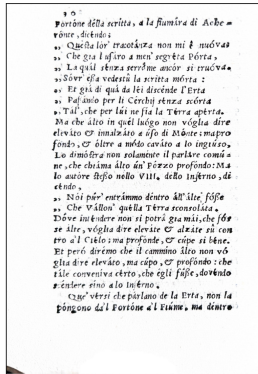
private collectors who own period editions but lack institutional digitisation facilities.

Second, each unit is encoded as a 1024-dimensional vector using `intfloat/multilingual-e5-large-instruct` [39], the top-ranked multilingual model on the MMTEB semantic-similarity benchmark [40]. It handles Latin, Italian, and French without language-specific tuning. Third, pairwise cosine similarities are computed. Fourth, an edge is placed between units i and j if j is among the k nearest neighbours of i (or vice versa) and their cosine similarity exceeds a threshold t . We call the resulting object G_{text} .

Three distinct graphs appear throughout this paper, and we keep them notationally separate. G_{text} is the semantic k -NN graph just defined: one node per textual unit, edges from cosine adjacency. G_{reg} is the physical *register graph* realised on the atom array after embedding (atom positions placed in 2D, bilayer, or 3D by the simulated-annealing procedure of §6; edges = pairs within the Rydberg blockade radius R_b). G_{design} is the engineered *target* graph for the QOuLiPo texts of §4, specified by construction, of which the prose is a literary realisation. For natural texts G_{text} is the canonical graph and G_{reg} approximates it lossily (Table 3). For engineered exact-UDG and exact unit-ball-graph (UBG; the bilayer and 3D analogues of UDG) texts (*Le venticinque stanze*, *Il castello*, *Pascal/Ménil*, *The Friar's Notebook*, plus the Archimedean and king-grid families of Table 7) $G_{\text{design}} = G_{\text{reg}}$ by construction, and the MIS benchmark is G_{design} itself. For engineered k -NN texts (*Nithard's Wager*, *Le Pari de Nithard II/III*, etc.) G_{design} is the pre-prose target and is the canonical graph on which the QPU is benchmarked.

But if the pipeline starts with a neural embedder, isn't every structural claim hostage to that embedder? We address this in Appendix B, running four natural texts (Dante, Boethius, Galileo, Giambullari) through three competitive multilingual transformers (`e5-large-instruct`, `bge-m3`, `arctic-embed-l-v2`) and highlight two findings. The backbone ratio MIS/ N is stable across all three: backbone *size* is an embedder-robust property of the text. The rigidity ρ is not: the three models disagree on which corpus is most rigid, so rigidity should be read as a property of the (text, embedding) pair. The engineered

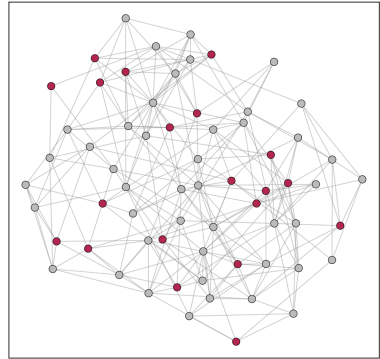
1. 1544 page photograph (Neri Dortelata, Firenze)



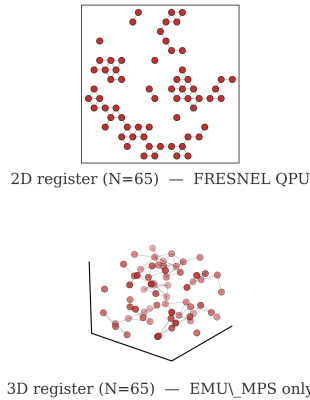
2. LightOn OCR

Fortóne délla scritta, a la fumàra di Ache-
vónte, dicéndo:
„Quèsta lór' tracotànzà non mi è nuóva
„Che già l'usáro a men' segreta pórtà,
„La quál sènza serróme ancór si truóva
„Sóvr' éssa vedestú la scritta mórtà:
„Et già di quá da léi discénde l'Érta
„Pašándo per li Cérchij sènza scórtà
„Tàl' che per lui ne fía la Tèrra apértà.
Ma che álto in quel luógo non vóglia dire
eleváto & innalzáto a úso di Mónte; ma pro
fóndo, & óltre a módo caváto a lo ínqúso.
Lo dímostra non soláménte il paríre comú-
ne, che chiáma álto un' Pózzo profóndo; Ma
lo autóre stéso nello VIII. dello Injérno, di
céndó,
„Nói púr' entrámmo déntro ál' álte, fósse
„Che Vállon' quèlla Tèrra sconsoláta.
Dóve inténdere non si pótrà già mái, che fós
se álte, vóglia dire eleváto & alzáte sú con-
tro a'l Cíelo; ma profónde, & cúpe si béne.
Et però díremo che il cammíno álto non vó
glia dire eleváto, ma cúpo, & profóndo: che
tále conveníva cértó, che égli fússe, dovéndó
scéndere síno a lo Injérno.
Que' vèrsi che párlano de la Érta, non la
póngono dà'l Fortóne a'l Fúme, ma déntro

3. k-NN semantic graph N=65, k=8, e5-large-instruct



4. Rydberg register (2D: QPU / 3D: emulator)



5. Results histograms

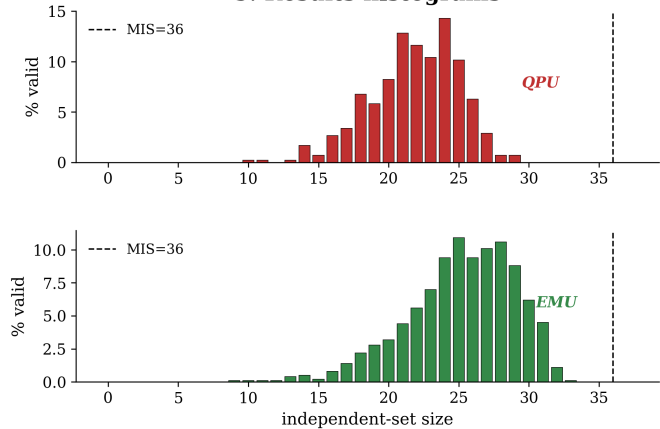


Figure 2: End-to-end pipeline on Giambullari's 1544 treatise: (1) original page photograph, (2) fine-tuned vision-language-model OCR (VLM-OCR) at 2.48% character error rate (CER) [33], (3) k -NN text graph, (4) 2D QPU and 3D emulator registers, (5) bitstring histograms.

texts of §4, whose graph structure is fixed by construction, serve as the calibration anchor that makes the rigidity metric interpretable across the corpus.

The pipeline has three free parameters: N (the number of units; our natural-text corpus spans $N = 27$ for Ausonius’s *Epigrammata* up to $N = 151$ for Giambullari’s full treatise), k (the k -NN parameter controlling graph density; default 8 or 16), and a cosine similarity threshold $t = 0.78$ (almost never binding because k dominates).

In stage 5 of Figure 2 the 2D histogram is the QPU output and the 3D histogram is produced by a classical *emulator*, since no current processor accepts 3D atom layouts. More generally, the emulator is the workhorse of every step of the pipeline that precedes a QPU shot. The emulator is a GPU-based numerical simulation that integrates the same Rydberg Hamiltonian the physical processor would implement, producing the bitstring distribution the real device would output — either as ideal noiseless dynamics or with calibrated hardware-noise models (laser dephasing, SPAM, finite-temperature broadening) when closer-to-device predictions are needed. Our runs use Pasqal Cloud’s tensor-network backend **EMU_MPS** [41], which represents the quantum state as a matrix product state [42] and is efficient as long as the transient entanglement stays below a bond-dimension ceiling the GPU can hold in memory. Most runs were performed on Pasqal Cloud’s A100 back-end, with a subset of larger instances run on a separate cloud platform with H100/H200 GPUs. The practical scaling limit is set by bond-dimension growth and depends on N , pulse duration T , bond dimension χ , and graph density; for the adiabatic pulses used in this paper, Vovrosh *et al.* [43] locate the exponential-cost crossover in the $N \approx 100$ –150 range, with engineered low-entanglement UDG instances reaching further before the back-end stops converging. This is the regime where direct classical simulation of the quantum dynamics, rather than classical solution of MIS itself, becomes resource-limited. In practice we use the emulator to develop and debug the workflow on each instance, then run a subset of the debugged and optimised cases on the QPU.

Throughout the rest of this paper we use *QPU* as shorthand for FRESNEL_CAN1 and *EMU* for EMU_MPS. Marker shapes and colours in the figures follow the same convention.

3 The Structural Backbone

Two natural questions arise about the MIS backbone of a text: how big it is, and how unique it is. The first is uninformative — backbone size MIS/N stays in a narrow 0.10–0.30 band across the corpus. The second is what discriminates between texts, and we call it *rigidity*.

Rigidity ρ is the fraction of MIS nodes that appear in every optimal solution. $\rho = 1.0$: unique backbone, every selected unit irreplaceable. $\rho = 0.0$: maximally degenerate, every selected unit substitutable. The

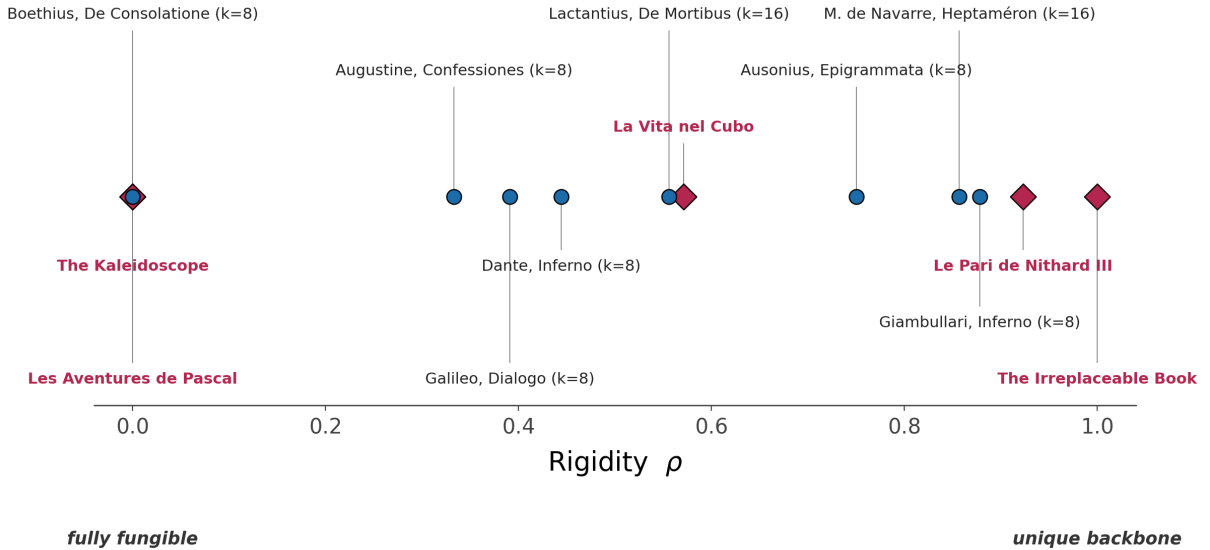


Figure 3: The corpus laid out along the rigidity axis $\rho \in [0, 1]$, between the *fully fungible* limit ($\rho = 0$, every selected unit has a structural substitute) and the *unique backbone* limit ($\rho = 1$, every selected unit is irreplaceable). *Blue circles*: natural texts (at the k_{nn} each text is most discriminating at; see Table 1). *Red diamonds*: engineered QOuLiPo texts written for this paper.

underlying object — the intersection of all optimal solutions, which we call the *persistent core* (the standard *persistence set* of integer programming [44–46]) — is standard in combinatorial optimisation, and in the Ising/Rydberg picture of §1 it is precisely the degeneracy of the MIS ground state: $1 - \rho$ counts how many backbone slots the ground-state manifold can swap between optima. What is new is applying it to literary texts as a structural metric. Figure 3 plots the spectrum across the corpus.

We believe the rigidity spectrum exposes structural properties that word-frequency methods such as Voyant Tools [8] and topic models do not directly access. Dante’s *Inferno* sits at a moderate ρ because the *terzina* structure creates local substitutability while the cosmological progression anchors the overall arc; Galileo’s *Dialogo* sits low because each day’s discussion is relatively self-contained, so modular reshuffling preserves the argumentative structure; the *Heptaméron* at $k = 16$ rises because of a hard core of *nouvelles* distributed across the frame days that no optimal backbone can omit. Boethius’s *De Consolatione* at the opposite extreme raises a substantive question (whether the structural fungibility reflects the philosophical argument’s internal coherence or simply the text’s topical homogeneity at the embedding level) which we leave to Boethius specialists. The numerical values and the table-format readings follow in Table 1.

3.1 What does a MIS backbone actually look like? An example

It is worth showing the pipeline once on a text small enough to print in full. We wrote a twelve-paragraph English mini-text, *Twelve Hours of a Coastal Town*, with a four-cluster structure: paragraphs 1–3 are the dawn departure, 4–6 the mid-morning market, 7–9 the silent noon, 10–12 the late-afternoon return. Each paragraph is embedded by `e5-large-instruct` and connected to its $k = 3$ nearest neighbours. The MIS is computed by ILP. The resulting graph has $N = 12$ vertices, $E = 26$ edges, and density $d = 2E/(N(N - 1)) = 0.39$. The MIS has size 4, and one of the nine optimal solutions is $\{3, 4, 7, 10\}$, exactly one from each cluster. The full text is in Appendix C; here are the four selected paragraphs:

(¶3) *From the cliff above the beach an old fisherman watches the two crews vanish. He no longer goes to sea. He keeps a small notebook of weather signs and writes one line each morning, always about wind direction and the colour of the horizon.*

(¶4) *By mid-morning the market square is loud with bargaining. Fishwives shout the prices of mackerel and sardines, butchers slap the boards with their cleavers, the children of the town run between the stalls stealing peaches when no one is looking.*

(¶7) *After lunch the town falls silent. The shutters close one by one, the dogs sleep in the shadow of the church wall, and the only sound is the rope of a small boat creaking against its bollard down at the harbour. Even the gulls have stopped screaming.*

(¶10) *By late afternoon the tide has come in and is licking the keels of the boats hauled up that morning. The same boy who carried the bait bucket at dawn now sits on the gunwale of his father’s boat, mending a torn net with a wooden needle, his fingers already as quick as a man’s.*

Those four paragraphs are a coherent micro-summary: a single scene from each phase of the day, no scene repeating, every theme represented. The backbone in the sense used here is the maximum independent set, which in this deliberately clustered example selects one representative from each thematic phase. The omissions are instructive: the morning departure (¶1, ¶2) collapses into ¶3, the cheese-seller and priests (¶5, ¶6) into ¶4, the cat and beekeeper (¶8, ¶9) into ¶7, the second-crew return (¶11, ¶12) into ¶10. The independence constraint forces one representative per cluster rather than the most vivid or quotable.

Compare this to popularity-weighted summarisation (TextRank, LexRank, centroid-based scoring). Those methods rank units by centrality and return the top-scoring ones. The result: they systematically pick units that *cluster together*. The two most central market paragraphs both get selected while the quietest noon paragraph never makes the cut. MIS does the opposite: the independence constraint *forbids* picking two similar paragraphs, forcing the backbone to spread across the thematic space. This is what we want from a structural backbone of a literary text.

The persistent core consists of the two paragraphs in every optimal solution: $\{\text{¶3}, \text{¶10}\}$, the cliff-watcher at dawn and the boy mending nets at evening — giving $\rho = 0.5$. Even when the algorithm can swap the noon and market choices, those two anchors are irreplaceable. $\rho = 0.5$ means two of the four backbone paragraphs are structural invariants; the other two have equally good substitutes.

The same operation, applied to Dante or Giambullari, performs this task at scales where the reader can no longer verify by inspection.

3.2 Rigidity readings across the corpus

Rigidity is computed classically with off-the-shelf ILP, fast at our corpus sizes ($N \leq 151$): finding one optimal MIS takes well under a second on every natural text, and enumerating the optimal-MIS manifold up to a 500-solution cap takes between a fraction of a second (*Heptaméron* at $k=16$, 6 distinct optima) and $\sim$ two minutes (Boethius, where enumeration hits the cap); per-instance ILP timings are listed in Appendix D, Table 6. For the rows where enumeration hits the 500-solution cap, we lift ρ from a sampled estimate to an exact certified value via per-vertex ILP-exclusion (testing, for each backbone page v , whether MIS shrinks when v is forced out), which removes any dependence on the cap.

Table 1 collects rigidity across the corpora with a one-sentence literary reading. “# opt.” is the number of distinct optimal MIS solutions (exact below the 500-solution cap, lower-bounded otherwise). Each row uses the k_{nn} at which rigidity is maximally discriminating for that text. The literary readings in Table 1 and in the rest of this section are sanity checks that the rigidity metric produces human-legible distinctions, not closed interpretations: ρ is a property of the (text, embedding) pair — a situated metric, in the sense of recent work on computational hermeneutics [47] — and a different embedding may reorder the rigidity ranking, as Appendix B documents directly.

We fix **e5-large-instruct** throughout for reproducibility; under that fixed embedder, MIS/ N is stable across alternatives but ρ is not (Appendix B).

3.3 Is the backbone real? Three checks

With the pipeline fixed, three further worries remain before reading the values of Table 1 as properties of the texts themselves rather than of the apparatus: that the same numbers would come out of a graph with the right edge count alone; that they depend on which particular optimal backbone the solver happens to return; and that simpler off-the-shelf summarisation would do the same job. We address each in turn.

First, is it the semantics or just the edge count? If we keep the number of units and the number of edges fixed, but reshuffle which unit connects to which, the rigidity disappears. Giambullari’s canonical graph ($N = 151$, $E = 866$, MIS = 41) sits far below random ensembles built on the same (N, E) : both Erdős–Rényi random graphs (each pair of units connected with the same fixed probability; 100 trials) and configuration-model graphs (edges reshuffled while keeping each unit’s number of neighbours fixed;

Corpus	k_{nn}	ρ	# opt.	Literary reading
M. de Navarre	16	0.857	6	Frame narrative: a 12-nouvelle hard core spread across several frame days
Lactantius	16	0.556	7	Chronological narrative of the deaths of the persecuting emperors; semi-rigid spine
Dante	8	0.444	17	The macro-descent through the circles anchors the arc; canto-to-canto transitions within a circle remain locally substitutable
Giambullari	8	0.878	24	Highly rigid topographical reconstruction; 36 of 41 backbone pages in every optimum
Galileo	8	0.391	≥ 500	Four dialogue-days, each topically self-contained; high modularity within each day’s discussion
Ausonius	8	0.750	4	27 short occasional poems (satirical, epigraphic, descriptive); 6 of the 8 backbone poems irreplaceable, the MIS forcing strongly distinct backbone selections in a dense graph ($d = 0.464$)
Augustine	8	0.333	5	Meditation on the opening of Genesis; semi-rigid, with three chapters present in every optimal backbone
Boethius	8	0.000	≥ 500	Fully fungible: no single section is structurally privileged by the MIS criterion

Table 1: Per-text rigidity values across the natural-text corpus, with a one-line literary reading for each. k_{nn} = number of nearest neighbours each node is connected to in the cosine-similarity graph; ρ = rigidity (fraction of MIS nodes present in every optimal solution); # opt. = number of distinct optimal MIS solutions (lower-bounded at the 500-solution enumeration cap).

65 trials) yield larger MIS than the real graph, at the strongest empirical bound the trial counts allow ($p \leq 1/(n+1) \approx 0.01$ for ER, ≈ 0.015 for the configuration model; no randomised graph in either ensemble matched the real text’s rigidity). Rigidity discriminates even more sharply: the real Giambullari graph reaches certified $\rho = 0.878$, while randomised graphs of the same size do not. The identity of the edges — the semantics of the text — is what compresses the backbone, not the edge count. The reading would fail this check if rigidity rankings on real texts were statistically indistinguishable from those on degree-matched random graphs; none of our corpus shows that pattern.

Second, do different optimal backbones tell the same story? The MIS algorithm typically returns one of several equally-good backbones. If the literary reading depended on which one the solver happened to find, the metric would be unstable. We sampled three of the 24 optimal backbones of Giambullari (each 41 pages) and generated a plain-text summary of each with Claude Sonnet 4.5, plus a fourth summary of the full 151-page text. Re-embedding the four summaries gives pairwise cosine similarity above 0.95 throughout: distinct optimal backbones generate the same narrative. This is a qualitative robustness check rather than an independent validation — the same model family writes the QOuLiPo prose and these summaries, so we read the null-model and per-vertex ILP certifications above as carrying the structural load. The check would fail if summaries from distinct optimal backbones diverged substantially.

Third, would simpler off-the-shelf methods give the same answer? We tried two standard summarisation methods that, like MIS, pick a small subset of units (k -center clustering [48, 49] and greedy facility-location [50, 51]) at the same subset size as our MIS, on thirteen rows of the natural-text corpus. Both overlap with our backbone only modestly (median around 30% and 20% respectively) and both routinely pick adjacent, semantically near-duplicate units. The reason is structural: neither method forbids picking two similar units, so neither produces a non-redundant backbone of mutually-dissimilar units. MIS does, and that is what makes the difference.

4 QOuLiPo: Engineering the Text Graph

QOuLiPo (pronounced “koolipo”) is a portmanteau of *quantum* and *OuLiPo*, and is introduced here for the first time. It is a constrained-writing project that treats a book’s semantic graph as an object to be *designed* rather than discovered, and designs it under constraints tailored to quantum hardware (with a focus on neutral-atom architectures in this paper). The result is a family of readable literary texts whose

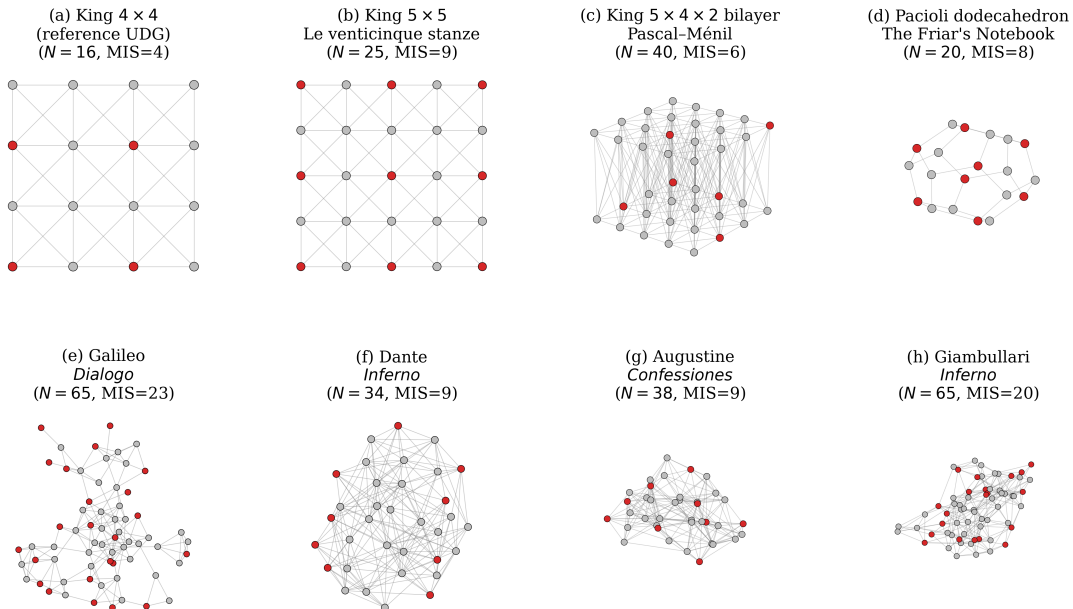


Figure 4: Register targets, with the classical MIS highlighted on each graph. *Top row* — *QOuLiPo (engineered) texts*: graphs designed by construction as exact unit-disk or unit-ball graphs. *Bottom row* — *natural texts*: k -nearest-neighbour graphs in multilingual sentence-embedding space. Red nodes are members of an exact MIS computed by ILP.

graph-theoretic invariants are controlled by construction. Figure 4 previews the visual flavour of these register targets, alongside the natural-text graphs against which they are contrasted.

4.1 Ancestors: from Ausonius to Perec

OuLiPo (the *Ouvroir de littérature potentielle*, founded in 1960 by Raymond Queneau and François Le Lionnais [52]) produces literary works under explicit formal constraints. Queneau’s *Cent mille milliards de poèmes* [10] combines fourteen sonnets by permutation; Perec’s *La Disparition* [11] excludes the letter *e*; Roubaud’s $\in$ [53] treats poems as set-theoretic elements.

Where classical OuLiPo constrains letters (lipograms), verses (cento), or chapter order (knight’s tours), QOuLiPo extends the tradition to the design of the graph itself: the constraint is the canonical graph G_{design} , and the prose is its literary realisation. The notation $(G_{\text{text}}, G_{\text{reg}}, G_{\text{design}})$ and the broad split between exact-UDG and k -NN-engineered texts were introduced in §2 alongside the natural-text pipeline; a finer decomposition appears later in this section, and the full inventory in Table 7. What QOuLiPo adds beyond classical OuLiPo is that the constraint is now an object the QPU can execute on: each text is simultaneously a readable literary work and a quantum-benchmark instance whose graph structure is known by construction.

The synthesis of constrained literature with combinatorial mathematics is older than OuLiPo. In the preface to his *Cento Nuptialis* (c. 370 AD), Ausonius of Bordeaux (late-Latin poet, tutor to Emperor Gratian) compares his cento technique (a wedding poem composed of recombined Virgilian half-lines) to the Stomachion, the geometric puzzle attributed to Archimedes whose 14 pieces admit 17,152 configurations [54]. Ausonius writes: “*like that game which the Greeks call ostomachion . . . from various joinings, a thousand species are simulated*” [55]. The text we OCR’d for this work (the Aldine 1517 *editio princeps*) thus contains, in its own preface, the constraint-literature $\leftrightarrow$ combinatorial-mathematics parallel this paper operationalises. Following [56], we treat Ausonius as an ancient Oulipian. The Stomachion preface is the strongest ancient-source evidence for the claim.

A particularly striking twentieth-century precedent for us is Perec’s later and more ambitious novel *La Vie mode d’emploi* [12], whose 99 chapters are ordered by an open knight’s tour on a 10×10 grid representing the elevation of a Parisian apartment building. The chapter order is a Hamiltonian path on the 10×10 knight’s graph; each of the 99 chapters is further overdetermined by 42 thematic elements distributed via 21 Graeco-Latin bi-squares.

What Perec’s scholarship has to our knowledge not observed is that this constraint graph appears not to admit a 2D unit-disk realisation (we have not found one in the literature nor constructed one ourselves). Perec chose his constraint for literary reasons — because it forced the narrator to skip around the building unexpectedly, and because the deliberately omitted 66th chapter is the structural counterpart of the novel’s protagonist Bartlebooth’s failure to complete his life-project — but the constraint he chose is a non-planar, bounded-degree graph of exactly the kind discussed in §6. Half a century before the neutral-atom hardware community began wrestling with non-unit-disk encodings, the most famous OuLiPo novel had already built itself on exactly such a graph.

Italo Calvino, the first Italian OuLiPo member and Perec’s nearest peer in the combinatorial tradition, provides a parallel lineage of graph-structured fiction: *Il castello dei destini incrociati* [57] constructs narratives from tarot-card adjacencies; *Se una notte d’inverno un viaggiatore* [58] nests ten interrupted novels inside a frame that is itself a combinatorial traversal of narrative beginnings; and the *Lezioni americane* [59] (in their fifth memo on *Multiplicity*) theorise the OuLiPo programme explicitly as a literature of combinatorial networks. Outside OuLiPo proper, Jorge Luis Borges (*La biblioteca de Babel*, *El jardín de senderos que se bifurcan*, both 1941 [60]) extends the same combinatorial imagination into Spanish-language fiction. Our engineered texts continue this combinatorial lineage; and where Perec’s and Calvino’s graph structures were chosen for literary reasons (with no hardware available against which to test them), we design ours explicitly with the properties of neutral-atom processors in mind.

4.2 The QOuLiPo design space

We generated twenty-nine constrained texts (full inventory in Appendix E, Table 7). The corpus is organised along three orthogonal axes (design tier, geometric class, rigidity) and Figure 5 shows where every text lands in the resulting three-dimensional space.

The generation mechanism produces the prescribed canonical graph G_{design} and steers the embedder’s G_{text} toward it. Each page is assigned m threads from a pool of T thematic threads; pages sharing at least τ threads are the edges of G_{design} . The prose is then written with thread-specific keywords so that the embedder reads each prescribed pair as high cosine similarity, and G_{text} approximates G_{design} .

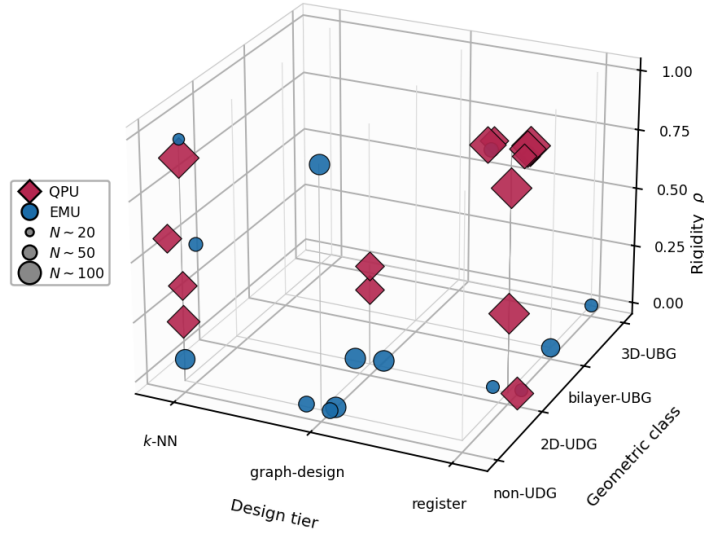


Figure 5: The full QOuLiPo corpus (twenty-nine engineered texts) in the three-dimensional design space (design tier $\times$ geometric class $\times$ rigidity). Each marker is one engineered text; markers are slightly offset within each (tier, geometric-class, ρ) bucket so all 29 are visible. Red diamonds: corpus members run on the QPU (FRESNEL_CAN1). Blue circles: corpus members run only on the emulator. Marker size scales with N .

The k -NN parameter then sets the realised density of G_{text} through the rule $d \approx k/N$; for high-density targeting ($d \geq 0.30$) we set $k \geq 0.30N$.

The prose is produced as constrained generation in the OuLiPo sense: the human supplies the constraint (target graph, thread / register / poetic-form assignments, narrative voice, rigidity target, language) and Anthropic’s Claude (Opus 4.6 / 4.7) generates page-level prose that satisfies it. Validation is mechanical (ILP MIS and k -NN edge recovery, in place of the syllable counts or vowel-avoidance of classical OuLiPo); the author validates each page against the deposited graph and iterates where the constraints are not met, with several texts going through five or more revision cycles. Two distinct quantities are worth separating here. For register-led exact-UDG and exact-UBG texts, register fidelity is 1.000 by construction — the geometric identity $G_{\text{design}} = G_{\text{reg}}$ holds before any prose is written. Separately, the *semantic-realisation fidelity* — the edge recall between the embedder-derived G_{text} and the prescribed G_{design} once the prose exists — is a property of the writing, not the geometry. The canonical files (`source/page_NNN.txt` in the Zenodo deposit accompanying this paper [61]; full description in §8) are the final versions. The corpus varies in literary ambition: some texts are pursued as serious literary objects, others kept as benchmark instances. As a first attempt, the literary pieces invite contributions from writers, poets, and OuLiPians to turn these graph-bound exercises into genuinely artistically relevant works.

The engineered texts arrange themselves along three increasingly strict design tiers. In the k -NN-led regime, a target graph guides the prose generation, but the canonical graph is the embedder’s k -NN of the final prose (*Nithard’s Wager*, *Le Pari de Nithard II/III*, *La Vita nel Cubo*, *Sonetti dal Tesseracto*, *La Partition du Texte*); the constraint is on (N, d) and on the rough thread structure, not on the precise edge list. In the *graph-design-led* regime the canonical graph is fixed exactly — as a $K_{25,25}$ bipartite (*The Trial of Nithard*), a Sierpinski hierarchy (*The Fractal Book*), 17 disjoint K_3 (*The Kaleidoscope*), a hub-and-spoke trap for greedy heuristics (*The Reader’s Trap*), a planar 5×10 grid (*The Map of the Text*), or a unique-MIS double-domination construction (*The Irreplaceable Book*) — but the text has no register coordinates: pages are graph nodes, and the prose realises the prescribed adjacencies. In the *register-led* regime pages are placed at explicit register coordinates so that $G_{\text{design}} = G_{\text{reg}}$ holds exactly — the king-grid families (*Le venticinque stanze*, *Il castello*, *The Book of Eighty-One Squares*, *Eighty-One Faces of a Pomegranate*, the Pascal/Ménil pair), the centred-hex texts (*Treize sur trente-sept*, *Hours of the Day*), the Archimedean tilings (*Kagome of Late Summer*, *The Postman’s Songbook*), *The Incarnate Graph* (random UDG placement), and the bilayer and three-dimensional extensions (*Pascal-Ménil bilayer*, *The Friar’s Notebook*).

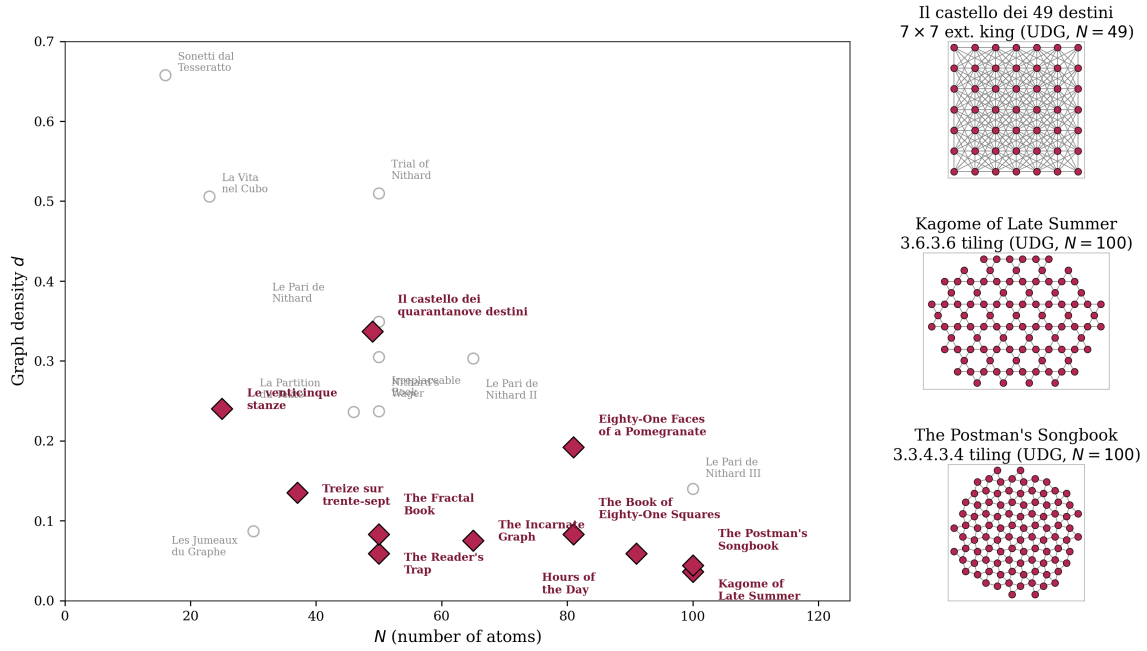


Figure 6: Engineered QOuLiPo corpus on the (N, d) plane. *Filled red diamonds*: 2D-UDG texts run on the QPU. *Open circles*: non-UDG engineered texts (k -NN-led and graph-design-led). Bilayer / 3D extensions and EMU-only 2D-UDG texts are not shown; see Table 7 for the full inventory.

A second axis cuts across this tier ladder: the geometric class of the canonical graph. The k -NN-led family is non-UDG by construction, since the canonical graph is a 1024-dimensional cosine k -NN, not a unit-disk graph. The register-led family is 2D-UDG, or bilayer/3D-UBG for the bilayer and dodecahedron texts, by construction. The graph-design-led family is mixed: most members are 2D-UDG-realizable (*The Map of the Text* as a planar grid, *The Reader's Trap* as a hub-spoke star, *The Kaleidoscope* as 17 well-separated triangles, *The Fractal Book* as a Sierpinski hierarchy at moderate density), but three appear non-UDG: the $K_{25,25}$ bipartite of *The Trial of Nithard* (which contains $K_{4,4}$ and is therefore not unit-disk by a standard kissing-number argument), the double-domination construction of *The Irreplaceable Book*, and *Les Jumeaux du Graphe* (C_{30} plus eight same-parity chords) all defeated our embedding attempts; we have not produced formal non-embeddability proofs for the latter two.

The third dimension of Figure 5 is rigidity ρ . The corpus is designed to cover the full $[0, 1]$ interval, with rigid backbones, maximally degenerate constructions, and intermediate values all represented.

Figure 6 positions the engineered QOuLiPo corpus on the (N, d) design plane, where each graph occupies a point defined by the register size N (one atom per vertex) and the graph-theoretic density $d = 2|E|/(N(N-1))$. These two quantities decide which constraints the graph stresses on the QPU. N sets the hardware budget: FRESNEL_CAN1 supports up to 100 atoms, so the texts that exercise the QPU most are those that hit the upper end of that budget. Density d controls how constrained the MIS problem is: sparse graphs have many degenerate independent-set configurations and are trivially solved by any heuristic, while dense graphs have small MIS, larger branching factors, and empirically slower classical runtimes; Cazals *et al.* [5] document this trend systematically on UDG instances at full triangular-lattice filling and $N \geq 50$. Engineering the QOuLiPo corpus to span the (N, d) plane was therefore an explicit design goal. QPU outcomes are reported in §5.

4.3 Beyond 2D: bilayer and 3D text designs

The QOuLiPo corpus also includes two texts whose graph structure requires *more* than a planar register. *The Friar's Notebook* ($N = 20$) is twenty short pieces, each in a different literary register (confession, ledger, sonnet, letter to Leonardo, recipe, dream, sermon, dialogue, prayer, epitaph, ...), apocryphally attributed to the Renaissance mathematician Luca Pacioli (1447–1517), whose *De divina proportione* was illustrated by Leonardo, and arranged at the twenty vertices of a regular dodecahedron — a 3D unit-ball graph that is provably not realisable as a 2D UDG or even a bilayer graph. It is the simplest QOuLiPo text that *requires* full 3D.

The second is a two-layer engineered text written specifically for a bilayer register geometry. *Les Aventures de Pascal* is a twenty-page apocryphal French novel about Blaise Pascal (the philosopher and mathematician), organised on a 5×4 grid whose axes are the five eras of Pascal’s life (Clermont childhood, Paris salons, vacuum experiments, Port-Royal, posthumous fortunes) and four cardinal threads (reason, wager, vacuum, mystic). Each page sits at one (era, thread) cell and carries the corresponding thematic markers so that king-adjacent pages (the eight grid neighbours — orthogonal and diagonal, after the chess king’s move) share vocabulary. *Scholia de M  nil*, a parallel second text, is 20 pages of fictional 19th-century scholarly annotation by “Adolphe M  nil”, archivist at the Biblioth  que Nationale in 1878, laid out on an identical 5×4 grid whose columns are four editorial concerns (philology, biography, science-history, theology). Together the two texts form a 40-page bilayer whose graph structure — a $5 \times 4 \times 2$ king lattice — is fixed by construction. The self-reference is deliberate: Pascal is being read on Pasqal.

The physical realisation of both the dodecahedron and the bilayer is taken up in §6, where they serve as engineered anchors for the natural-text 3D and bilayer experiments.

5 Quantum MIS on FRESNEL QPU

The text graphs of §3 and the engineered constructions of §4 are now loaded onto the QPU — one atom per textual unit (page, canto, novella, paragraph, chapter — whatever the author’s segmentation gave us in §2), with one Rydberg blockade per edge of the realised register graph G_{reg} — exact for the engineered-UDG rows where $G_{\text{design}} = G_{\text{reg}}$ by construction, lossy for the natural-text rows where G_{reg} approximates G_{text} through the SA embedder. The MIS that ILP computes on G_{reg} is the same MIS that the array’s quantum ground state encodes in matter; the rest of this section reports what the hardware actually returns, on both halves of the corpus.

5.1 Setup: register and pulse

We implement the text-graph MIS pipeline on Pasqal’s FRESNEL_CAN1 (100-qubit, 2D rubidium register, hosted at the DistriQ quantum innovation zone in Sherbrooke, Qu  bec) in *analog* mode: the physical interactions *are* the computation, no circuit is compiled, no gates are applied — Feynman’s 1982 vision [62] of computing with matter, made concrete by the Rydberg blockade introduced in §1. All

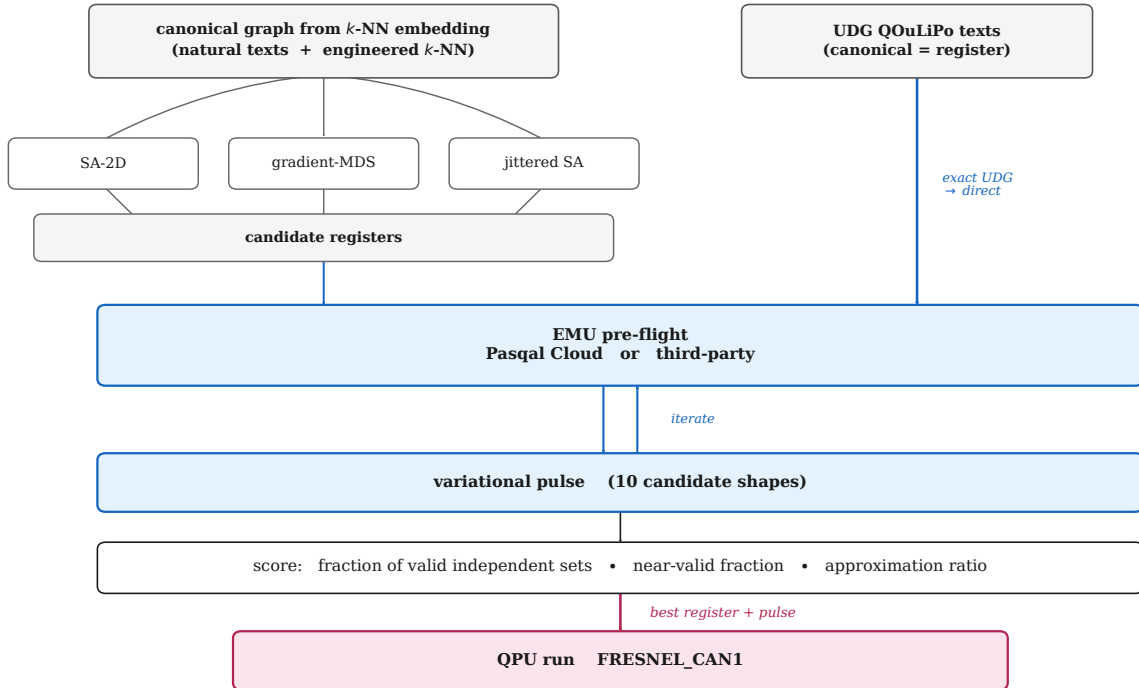


Figure 7: Optimisation toolkit: the EMU $\leftrightarrow$ QPU loop. The canonical graph (text k -NN or designed) feeds candidate (register, pulse) pairs through EMU pre-flight; the best pair is submitted to the QPU.

natural texts fit the QPU’s register at their natural granularity (Dante *Inferno* $N=34$, Augustine $N=38$, Lactantius $N=52$, Giambullari $N=65$ at the central-chapter abridgement of the $N=151$ full treatise, Galileo $N=65$; see Appendix A); for Dante we report QPU results on the *Inferno* at $N=34$ (canto level) and a separate full-*Divina Commedia* run at $N=100$ cantos, both in Table 2. QOuLiPo texts use the full 100-qubit capacity. Field of view ($\leq 46\mu\text{m}$ radius) and minimum spacing ($\geq 5\mu\text{m}$) impose geometric constraints, all satisfied by the graphs of Table 2.

Figure 7 summarises the EMU $\leftrightarrow$ QPU optimisation loop used throughout this section. EMU is a faithful proxy of the QPU in its convergent regime; where it does not converge (several large- N runs in this campaign hit the bond-dimension ceiling described in §2), we transfer design rules from small- N topology-matched controls to their large- N targets, the *Treize sur trente-sept* / *Hours of the Day* pair being one such transfer.

Register engineering (the simulated-annealing fit of G_{text} to FRESNEL’s triangular lattice, maximising *edge recall* — the fraction of canonical G_{text} edges that survive within R_b on the embedded register) and pulse choice (the standard baseline ramp $\Omega = 3.3\text{ rad}/\mu\text{s}$, $T = 4\mu\text{s}$, extended to $T = 6\mu\text{s}$ at $N \geq 81$) are documented in detail in Appendix F. Two facts matter for what follows. *For natural texts*, the SA edge-recall ceiling caps below 0.5 at $N \geq 30$ in 2D (quantified in §6, Table 3), and the dense graphs one would most want the QPU to recover are precisely those that lose the most information at the embedding stage. *For engineered exact-UDG texts*, the register is constructed so that edge recall is 1.0 exactly — $G_{\text{design}} = G_{\text{reg}}$ by construction — and any deviation from ratio 1.0 on QPU is attributable to hardware noise alone, with neither embedder mismatch nor register-fidelity loss in the way.

Where an SA-embedded or exact-UDG instance nevertheless fails on hardware (zero valid IS), the practical remedy is a tighter blockade margin r_{nnn}/R_b between the smallest non-edge distance and the blockade radius, achieved by re-running SA with all non-edges pushed comfortably above R_b ; this single lever lifts several rows of Table 2 from 0/1000 valid IS to ratios in the 0.84–1.00 band, and is the largest hardware-side margin documented in this paper. Pulse engineering, by comparison, is a second-order refinement: EMU-selected variational pulse shapes improve **valid IS%** by +3–6 percentage points over the baseline ramp on the engineered exact-UDG rows of Table 2, with the approximation ratio pinned at 1.000 (full sweep in Appendix F).

5.2 QPU results

Table 2 reports data for 19 QPU submissions, all at 1000 shots per row except *The Incarnate Graph* at 402 shots, organised in three blocks: natural k -NN texts (SA-embedded), engineered k -NN texts (also SA-embedded), and exact-UDG / designed-graph engineered rows. The best-shot ratio is $r = \max_{s \text{ valid}} |s| / \alpha(G)$, where the maximum runs over the 1000 shots forming a valid IS on the benchmark graph, $|s|$ is the IS size, and $\alpha(G)$ is the ILP-exact MIS; rows with no valid shot are reported as $r = 0$. Near-valid% on SA-embedded rows is the fraction of shots with $|w - \text{MIS}| \leq 1$ (near-MIS occupation indicator); on exact-UDG rows it is the fraction of shots violating at most two canonical edges (soft-blockade tail indicator); the two are not interchangeable.

On natural-text rows the pipeline cascades through two upstream stages before any QPU shot: a 1024-dimensional cosine k -NN graph built on the embedded text units, then a simulated-annealing fit of that adjacency onto FRESNEL’s triangular lattice. Table 2 reports for each row the approximation ratio against the realised register MIS (the standard convention in the neutral-atom MIS literature [4, 5]). On every bigger natural-text row we ran — Augustine, Lactantius, Galileo, Marguerite de Navarre, Boethius, and Dante ($N = 38$ to 100) — this register-MIS ratio is ≥ 0.92 at 14–60% strict valid-IS rates; on the engineered k -NN rows with EMU-selected variational pulses it is 1.000 at $\geq 60\%$ valid IS. The QPU delivers on the graph it is sent.

What the literary user receives is one step removed. The relevant quantity is the canonical text-MIS ratio — the fraction of the actual backbone the pipeline returns end-to-end — bounded above by the SA recall of the realised register. In practice, with the simulated-annealing embedder used throughout this paper, we observe that recall caps below 0.5 at $N \geq 30$; Galileo’s *Dialogo* reaches that ceiling exactly and accordingly delivers the highest canonical recovery on hardware: 14 of the 23 backbone pages, ratio 0.61. The near-valid Hamming-weight column in Table 2 carries a complementary signal: on Galileo, 31.5% of shots have Hamming weight within one of the canonical MIS size, and on Ausonius 36.2% — the QPU output clusters tightly around the right occupation regime even where individual shots fail strict independence, giving a post-processable structural signal where the strict-valid rate alone would say nothing. The upstream embedding step is the bottleneck on canonical-text MIS recovery, not the processor; closing that gap with better embedders — gradient-MDS, SDP relaxations, learned register

placement, or the 3D escape of §6 — is a natural follow-up.

On the exact-UDG and designed-graph engineered rows the canonical graph is by construction realised by the register, and the QPU does what the physics promises: on the cleanest instance, *Le venticinque stanze* ($N = 25$, an exact 5×5 king-UDG graph; the same row shown in Figure 8), the best QPU shot reaches the classical 9-page MIS at ratio 1.000, with 53.1% valid independent sets across 1000 shots. The exact-MIS shot fraction is 2.4%; the remaining valid shots cover sub-MIS independent sets across the structural backbone. On these rows the Rydberg blockade *is* the independence constraint, with no compilation in between. The hardware exposes per-atom occupation, near-valid Hamming-weight statistics, and the full shot histogram natively in microseconds, complementing rather than duplicating the single optimum a classical ILP solver returns.

Two exact-UDG constructions in our campaign sit at FRESNEL’s parameter edge and are reported here as hardware-boundary diagnostics rather than as part of the headline 0.84–1.00 band of Table 2. *Il castello* (extended-king on a 7×7 grid) would require $R_b > 14.14 \mu\text{m}$ at the standard spacing, which corresponds to $\Omega < 0.11 \text{ rad}/\mu\text{s}$ — below FRESNEL’s amplitude floor $\Omega \geq \pi/10 \approx 0.314 \text{ rad}/\mu\text{s}$; the row in Table 2 is the king-only subgraph (Chebyshev ≤ 1 , $R_b = 8 \mu\text{m}$, MIS = 16) at ratio 0.938. *Eighty-One Faces of a Pomegranate* ($N=81$, multi-distance $\sqrt{5}$ king) goes the other way: at $R_b = 11.5 \mu\text{m}$ the $\sqrt{5}$ -distance pairs sit at $U/\Omega \approx 1.4$ — the energy gap between the blockaded $|rr\rangle$ state and the singly-excited level is only marginally above the Rabi drive, a physics constraint that no pulse shape fixes — and the QPU returns only ratio 0.615. The natural remedy here is not dimensional escape (§6) but *per-atom local detuning* — on Pasqal’s hardware roadmap — or ancilla-augmented embedding [63]. A wider Rabi-frequency window on the device, together with local-control extensions, would together extend the class of canonical graphs realisable as exact UDG; on this axis, hardware flexibility buys more useful instances than additional qubits.

Where the QPU comes closest to literally reading the book is on Galileo’s *Dialogo*: the SA register loses about half the canonical edges, yet the QPU still recovers 14 of the 23 pages of Galileo’s structural backbone — a non-trivial subset of the four-day dialogue, anchored by the most distinctive astronomical demonstrations on each day. Figure 8 shows another case, *Le venticinque stanze*: the QPU returns the

Text	N	k	d	Best-shot ratio	Valid IS%	Near-valid%
<i>Natural k-NN texts (SA-embedded)</i>						
Ausonius <i>Epigrammata</i>	27	8	0.464	1.000	10.0%	36.2%
Augustine <i>Confessiones</i> (paragraph)	38	8	0.293	1.000	13.5%	4.9%
Lactantius <i>De mortibus persecutorum</i>	52	8	0.232	1.000	61.8%	33.0%
Galileo <i>Dialogo</i>	65	8	0.079	0.966	55.9%	31.5%
Boethius <i>Consolatio</i>	72	8	0.146	0.968	46.4%	2.4%
M. de Navarre <i>Heptaméron</i>	72	8	0.165	0.969	51.9%	16.4%
Dante <i>Divina Commedia</i>	100	8	0.121	0.917	43.7%	1.7%
<i>Engineered k-NN texts (SA-embedded, variational pulse)</i>						
Nithard’s Wager	50	16	0.305	1.000	59.5%	—
Le Pari de Nithard	50	16	0.349	1.000	60.6%	—
<i>Exact-UDG and designed-graph engineered texts</i>						
Le venticinque stanze	25	—	0.240	1.000	53.1%	85.3%
Treize sur trente-sept	37	—	0.135	0.923	0.7%	8.6%
Il castello dei quarantanove destini [‡]	49	—	0.133	0.938	24.1%	62.1%
The Reader’s Trap	50	—	0.059	1.000	65.4%	93.9%
The Fractal Book	50	—	0.083	0.947	60.5%	92.9%
The Incarnate Graph	65	—	0.075	0.955	3.2%	27.1%
The Book of Eighty-One Squares [§]	81	—	0.083	0.840	2.9%	18.2%
Hours of the Day [§]	91	—	0.059	0.839	23.9%	62.4%
Kagome of Late Summer [§]	100	—	0.036	0.923	18.4%	62.3%
The Postman’s Songbook [§]	100	—	0.044	0.861	30.3%	73.5%

Table 2: QPU results across the corpus. **Best-shot ratio** r is the largest valid-IS shot size divided by the ILP-exact MIS of the benchmark graph (the realised register graph G_{reg} for natural-text and k -NN engineered rows; $G_{\text{design}} = G_{\text{reg}}$ by construction for exact-UDG and design-led engineered rows). **Valid IS%** is the fraction of shots forming a valid IS. **Near-valid%** is the regime-dependent occupancy diagnostic defined in §5.2. [§] Pulse duration $T = 6 \mu\text{s}$; other rows $T = 4 \mu\text{s}$. [‡] King-only QPU subgraph of the extended-king construction; see §5.2.

classical 9-page MIS exactly, a symmetric set of nine stanze spanning all three floors of the building (*piano terra*, *secondo piano*, *sottotetto*) and three positions across the building’s depth (*nord*, *cortile*, *sud*). At register fidelity 1.0 — which holds here by construction — the atom array’s blockade pattern reproduces the text graph edge-for-edge, and the adiabatic protocol drives the system into the MIS of *this* text.

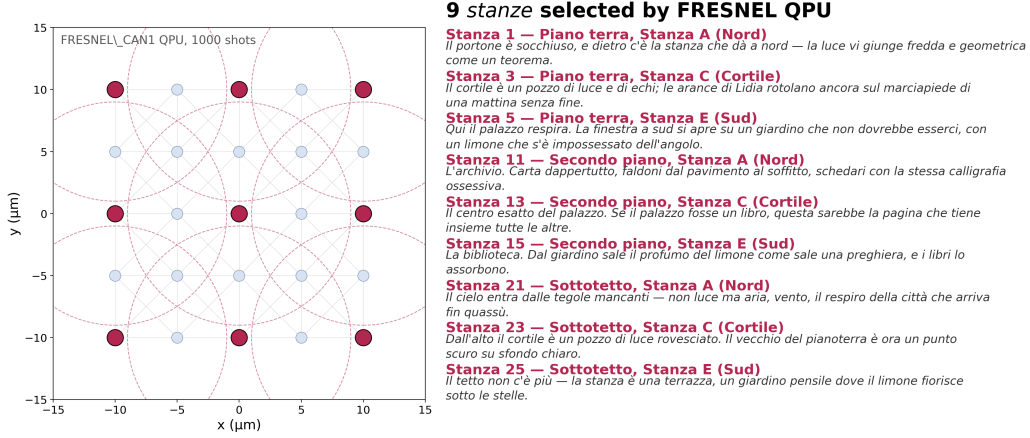


Figure 8: The quantum backbone of *Le venticinque stanze*. Left: register; the 9 atoms of the best valid independent set are in red. Right: the 9 selected *stanze*, spanning all three floors of the building (*piano terra*, *secondo piano*, *sottotetto*) and three positions across the building’s depth (*nord*, *cortile*, *sud*).

Scaling forward, the natural target for outperformance over classical heuristics is not 2D-UDG — which is classically tractable at any desired accuracy [64] — but bounded-degree non-UDG graphs at high N , where no PTAS is known and classical heuristics empirically slow. A random k -regular construction at $N \approx 100$ –200 and $k \approx 10$ is one potential target in that direction; the 256-qubit neutral-atom processor of Leclerc *et al.* [65] is the hardware reach. Within 2D-UDG, a 16×16 extended-king text ($N = 256$, $d \approx 0.33$) remains a useful scale-up benchmark — it shows the hardware can be loaded at the new size — even if asymptotic outperformance is not in this class.

6 A third dimension: for the atoms and for the texts

The natural texts and the k -NN engineered texts all yield k -NN graphs of sentence embeddings, not unit-disk graphs. The SA recall ceiling cited in §5 is moreover *genre-independent*: across frame-tale (*Decameron*), narrative poem (*Divina Commedia*), encyclopedic miscellany (*Noctes Atticae*), aphorism (*Novum Organum*), and scientific dialogue (Galileo’s *Discorsi*), at $N = 100$ recall sits in the band 0.21–0.27 at density $d \approx 0.10$ –0.14; and a random 12-regular graph at $N=75$ (no community structure) gives SA recall 0.21 in the same band. The ceiling is *geometric* — a property of bounded-degree graphs whose degree exceeds the 2D-UDG kissing number of 6, not a literary failure of natural-text clustering — and no pulse engineering, variational refinement, or noise tuning can recover the edges lost at the register-embedding stage. Dalyac and Henriët [63] prove that any graph of maximum degree $\Delta \leq 6$ admits an exact 3D-array embedding (unlike 2D). Our k -NN text graphs typically exceed this sufficient condition — symmetrised $k=8$ neighbours can produce maximum degrees well above 6 — so the Dalyac–Henriët theorem does not apply directly. Table 3 below should be read as empirical evidence that the third dimension relaxes the 2D recall ceiling, not as a theorem guaranteeing exact recovery.

A distinctive property of neutral-atom quantum processors — unique among the current quantum-computing platforms (superconducting qubits, trapped ions, photonics, semiconductor spins) — is that the register itself can be three-dimensional: optical tweezers position atoms in arbitrary 3D arrangements at the same micron scale that fixes the blockade radius [2]. 3D registers have been demonstrated in the laboratory [66] but are not yet exposed on commercial cloud QPUs — FRESNEL itself is 2D, and the discussion in this section is therefore forward-looking, anchored on hardware roadmaps rather than on present submissions.

Before full 3D register reconfigurability becomes available, a simpler intermediate device could already capture part of the 3D benefit: two parallel atom arrays separated by a distance $h \leq R_b$, allowing inter-layer blockade while requiring only a modest extension of existing single-plane trapping optics. To quantify the gain of bilayer and full 3D over the 2D baseline, we ran simulated annealing in three lattice

modes — 2D, bilayer (two stacked planes at $h = 0.80 R_b$), and 3D — on the natural-text rows where 2D recall sits well below the 0.5 threshold. Table 3 reports the resulting edge recall: the lift is monotonic across the dimensional containment $\text{UDG} \subset \text{unit-ball-in-slab} \subset \text{unit-ball-in-}\mathbb{R}^3$ on every row, with bilayer closing roughly 30–50% of the 2D-to-3D gap.

Text	k	N	d	2D recall	2L recall	3D recall
Ausonius <i>Epigr.</i>	8	27	0.464	0.344	0.442	0.571
Dante <i>Inferno</i>	8	34	0.330	0.368	0.465	0.546
Augustine <i>Confessiones</i>	8	38	0.293	0.364	0.466	0.583
Galileo <i>Dialogo</i>	8	65	0.079	0.497	0.515	0.558
Boethius <i>Consolatio</i>	8	72	0.146	0.339	0.392	0.543

Table 3: Edge-recall ladder for the five natural-text rows where 3D register escape lifts recall above the 0.5 threshold (the level above which the QPU-recovered backbone substantially matches the canonical text-MIS). Discrete-site simulated-annealing embedder [63] in 2D, bilayer 2L ($h = 0.80 R_b$), and 3D modes; spring-layout init, 30,000 iterations $\times$ 15 restarts.

The 3D recovery has been validated end-to-end on EMU for three natural texts of the corpus. With the baseline adiabatic pulse, the 3D-embedded Ausonius register ($N = 27$, 3D recall 0.571) recovers 6 of the 10 canonical text-MIS pages (canonical ratio 0.60); the 3D-embedded Augustine register ($N = 38$, recall 0.583) recovers 6 of the 12 canonical pages (0.50); and the 3D-embedded Boethius register ($N = 72$, recall 0.543) recovers 13 of the 14 canonical pages (ratio 0.929) — the cleanest of the three, with the low valid-IS rate at this larger graph size reflecting the register’s incomplete coverage of canonical edges. The Pascal–Ménil bilayer (Figure 9) realises a $5 \times 4 \times 2$ king lattice whose graph is an *exact* bilayer unit-ball graph by construction; run on EMU with the standard adiabatic pulse (intra-layer spacing $a = 5 \mu\text{m}$, layer separation $h = 3 \mu\text{m}$, $R_b = 8 \mu\text{m}$, 1000 shots; the values a and h here are graph-embedding parameters of a forward-looking bilayer geometry on EMU, not currently exposed FRESNEL loading constraints), it recovers the classical MIS exactly (ratio 1.00, valid-IS fraction 78.6%). Four Pascal pages alternate with

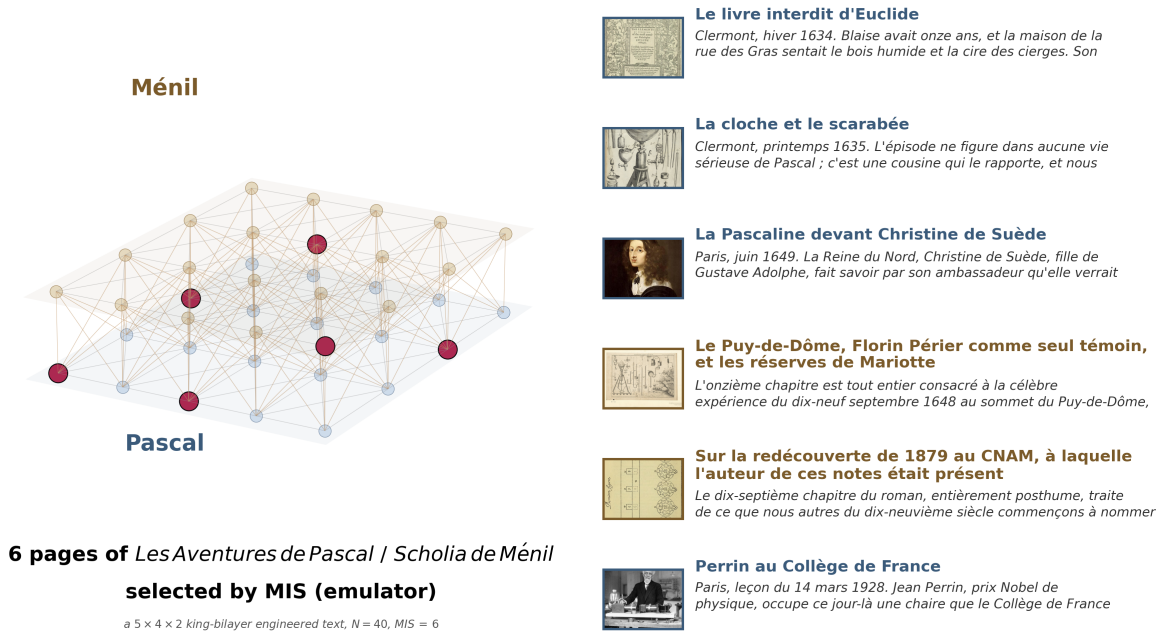


Figure 9: *Les Aventures de Pascal* (lower layer, blue) and *Scholia de Ménil* (upper layer, amber): the $5 \times 4 \times 2$ king-bilayer engineered text, $N=40$, with the six pages of the MIS in red. The right panel lists the MIS pages in deposited reading order. Illustrations, top to bottom: Billingsley’s *Euclid*, 1570; Boyle, *New Experiments*, 1660; S. Bourdon, portrait of Queen Christina of Sweden; Deulland, plate of Pascal’s Puy-de-Dôme experiment from Bélidor’s *Architecture hydraulique*; Belair, *Machine arithmétique*; J. Perrin, 1927 — all public-domain, Wikimedia Commons.

two Ménil scholia, the layers commenting on each other at chapter-matched positions. Figure 9 is the controlled case: unlike the three natural-text rows above, where bilayer/3D improves a lossy embedding, the Pascal–Ménil bilayer is designed directly as an exact bilayer unit-ball text graph, and the canonical MIS is recovered with no embedding loss. Each of these rows crosses the canonical-MIS recovery threshold that 2D embedding could not reach.

3D entanglement grows faster than 2D under the same protocol, so MPS-based emulators that handle ~ 100 atoms efficiently in 2D saturate at substantially smaller scales in 3D: a 100-qubit 3D Rydberg machine would likely already escape classical exact simulation, making 3D-capable hardware the natural next target for literary-MIS benchmarks. Staying in 2D and encoding non-UDG graphs through ancilla-gadget constructions [64, 67] performs poorly at FRESNEL-scale qubit counts, so we favour the native 3D path. End-to-end QPU validation on bilayer or 3D-capable Rydberg hardware is the natural next experimental step.

7 Envoi

The category this paper proposed in the Introduction — a boundary object between neutral-atom MIS and the structural analysis of literary texts — is now populated densely enough to be tested. Across the realised registers, the processor returned valid independent-set distributions and the sampling outputs (shot histograms, Hamming-weight statistics, register diagnostics) that classical ILP does not produce. Where natural-text recovery is partial, the bottleneck sits at the 2D embedding stage, not in the hardware; the forward-looking extension reported in §6 — 3D register geometries that lift canonical-backbone recovery on the emulator, with a bilayer construction recovering its backbone exactly — targets the hardware generation in which the planar embedding ceiling identified here no longer bounds the campaign.

Two openings are offered beyond this paper. Rigidity ρ , the (text, embedding)-pair metric introduced here, is portable: any deposited corpus with a fixed embedder and an ILP-verified MIS can be characterised the same way, and we hope the field tests it against kinds of textual organisation we did not run. The same framework also crosses out of text. A pilot extension in preparation applies the MIS-backbone construction to a visual dossier: artworks as nodes, multimodal visual embeddings layered with the ICON 2.0 ontology [68] and IICONGRAPH constructions [69, 70], MIS as the structural backbone of an iconographic canon — to our knowledge a first. Beyond these openings, we invite the digital-humanities, quantum-hardware, and benchmark communities to extend the corpus to harder canonical instances and to track its behaviour across processor generations as the hardware scales.

A note on scale, for the humanist reader. In the quantum-computing community the central scaling challenge is to grow from today’s ~ 100 -atom machines to thousands and ultimately hundreds of thousands of atoms or qubits, and neutral atoms held in optical tweezers are uniquely positioned to reach that regime: defect-free arrays of 1024 atoms in a cryogenic environment [71] and 6100-atom alkaline-earth tweezer arrays [72] are already experimentally demonstrated, and are expected within the next few years to reach commercial cloud-accessible platforms of the kind used in this paper. As the platform grows, the corpus we deposit grows with it — larger registers admit longer natural texts at finer chunking, denser engineered graphs, and constraints whose canonical sizes were too large to instantiate at $N = 100$.

Each generation reads its inherited corpus through the tools it has; with cloud-exposed neutral-atom processors and agentic LLM interfaces, that toolset has just acquired a new member. Outstanding instruments are now within reach of literary scholars, historians, musicologists, and philologists alike — from here it is for the imagination to take the lead, and we know that imagination has no limits.

8 Data availability

The QOuLiPo corpus is openly deposited on Zenodo [61] (DOI 10.5281/zenodo.20074378, CC-BY-4.0): 29 engineered QOuLiPo folders, 11 natural-text companion corpora, per-folder `metadata.json`, explicit-edge-list graph files, ILP-verified MIS solutions and embedding distances (`mis_analysis_all.csv`), a strict consistency verifier (`verify.py -strict -lint`), and a standalone single-page HTML navigator. The OCR-derived digitisation of the 1544 Giambullari is at [33].

9 Disclosures

The author is a co-founder of Pasqal, the company whose FRESNEL_CAN1 neutral-atom processor and EMU_MPS cloud emulator are used throughout this paper, and a co-founder and managing partner at Quantonation, a deep-tech venture capital firm with positions in quantum hardware companies including Pasqal. Both relationships are disclosed for transparency; all QPU results and the bulk of the EMU runs were obtained on standard Pasqal Cloud accounts available to any external user, with a subset of larger EMU instances run on commercial H100/H200 cloud resources (see §2). The 1544 *editio princeps* of Giambullari’s *Del sito, forma, & misure dello Inferno di Dante* and the 1517 Aldine *Opera* of Ausonius are part of the author’s personal collection of Renaissance and humanist books; these physical witnesses were the originals photographed and OCR’d for the corpus described in §2 and Appendix A.

Use of generative AI. The QOuLiPo prose deposited with this paper was generated by Anthropic’s Claude (Sonnet 4.5 / Opus 4.6 / Opus 4.7) under the author’s design, validation, and editorial guidance, as described in §4. All FRESNEL_CAN1 and EMU_MPS submissions reported in §5 and §6 were authored through an agentic Claude Opus loop calling Pulser and the Pasqal Cloud, as described in §1.

10 Acknowledgments

The author thanks Constantin Dalyac, Loïc Henriët, Alexandre Dauphin, and Mourad Beji at Pasqal for technical discussions on FRESNEL, register embedding, and noise characterisation; Igor Carron at LightOn for discussions on the VLM-OCR pipeline used to digitise the Giambullari and Ausonius sources; and Bernard Maréchal (zazipo.net) for inspiring the constrained-writing framing.

References

- [1] L. Henriët, L. Beguin, A. Signoles, T. Lahaye, A. Browaeys, G.-O. Reymond, and C. Jurczak, “Quantum computing with neutral atoms,” *Quantum*, vol. 4, p. 327, 2020.
- [2] A. Browaeys and T. Lahaye, “Many-body physics with individually controlled Rydberg atoms,” *Nature Physics*, vol. 16, pp. 132–142, 2020.
- [3] H. Pichler, S.-T. Wang, L. Zhou, S. Choi, and M. D. Lukin, “Quantum optimization for maximum independent set using Rydberg atom arrays,” *arXiv preprint arXiv:1808.10816*, 2018.
- [4] S. Ebadi, A. Keesling, M. Cain, T. T. Wang, H. Levine, D. Bluvstein, G. Semeghini, A. Omran, J.-G. Liu, R. Samajdar, X.-Z. Luo, B. Nash, X. Gao, B. Barak, E. Farhi, S. Sachdev, N. Gemelke, L. Zhou, S. Choi, H. Pichler, S.-T. Wang, M. Greiner, V. Vuletić, and M. D. Lukin, “Quantum optimization of maximum independent set using Rydberg atom arrays,” *Science*, vol. 376, no. 6598, pp. 1209–1215, 2022.
- [5] P. Cazals, A. François, L. Henriët, L. Leclerc, M. Marin, Y. Naghmouchi, W. da Silva Coelho, F. Sikora, V. Vitale, R. Watrigant, M. Witt Garzillo, and C. Dalyac, “Identifying hard native instances for the maximum-independent-set problem on neutral-atom quantum processors,” *Physical Review Applied*, vol. 25, no. 3, 2026.
- [6] D. Jaksch, J. I. Cirac, P. Zoller, S. L. Rolston, R. Côté, and M. D. Lukin, “Fast quantum gates for neutral atoms,” *Phys. Rev. Lett.*, vol. 85, no. 10, pp. 2208–2211, 2000.
- [7] E. Urban, T. A. Johnson, T. Henage, L. Isenhower, D. D. Yavuz, T. G. Walker, and M. Saffman, “Observation of Rydberg blockade between two atoms,” *Nature Physics*, vol. 5, pp. 110–114, 2009.
- [8] G. Rockwell and S. Sinclair, *Hermeneutica: Digital Humanities, Interpretation and Technology*. MIT Press, 2016.
- [9] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
- [10] R. Queneau, *Cent Mille Millions de Poèmes*. Gallimard, 1961.
- [11] G. Perec, *La Disparition*. Denoël, 1969.

- [12] G. Perec, *La Vie mode d'emploi*. Paris: Hachette, 1978.
- [13] N. Berkman, *OuLiPo and the Mathematics of Literature*. Peter Lang, 2022.
- [14] M. Shiraishi, I. Hamamura, T. Ishigaki, and T. Kadowaki, “A model context protocol server for quantum execution in hybrid quantum-hpc environments,” *arXiv preprint arXiv:2604.08318*, 2026.
- [15] H. Silvério, S. Grijalva, C. Dalyac, L. Leclerc, P. J. Karalekas, N. Shammah, M. Beji, L.-P. Henry, and L. Henriët, “Pulser: An open-source package for the design of pulse sequences in programmable neutral-atom arrays,” *Quantum*, vol. 6, p. 629, 2022.
- [16] F. Moretti, *Distant Reading*. Verso, 2013.
- [17] M. L. Jockers, *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013.
- [18] T. Underwood, *Distant Horizons: Digital Evidence and Literary Change*. University of Chicago Press, 2019.
- [19] T. Uçkan and A. Karcı, “Extractive multi-document text summarization based on graph independent sets,” *Egyptian Informatics Journal*, vol. 21, no. 3, pp. 145–157, 2020.
- [20] C. Hark, “Using graph-based maximum independent sets with large language models for extractive text summarization,” *Applied Sciences*, vol. 15, no. 12, p. 6395, 2025.
- [21] L. Antigueira, M. d. G. V. Nunes, O. N. Oliveira Jr., and L. d. F. Costa, “Strong correlations between text quality and complex networks features,” *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 12, pp. 2603–2611, 2009.
- [22] D. R. Amancio, O. N. Oliveira Jr., and L. d. F. Costa, “Identification of literary movements using complex networks to represent texts,” *New Journal of Physics*, vol. 15, p. 043048, 2013.
- [23] D. R. Amancio, “Probing the topological properties of complex networks modeling short written texts,” *PLoS ONE*, vol. 10, no. 2, p. e0118394, 2015.
- [24] D. K. Elson, N. Dames, and K. R. McKeown, “Extracting social networks from literary fiction,” in *Proceedings of the 48th Annual Meeting of the ACL*, pp. 138–147, 2010.
- [25] F. Moretti, “Network theory, plot analysis,” *New Left Review*, vol. 68, pp. 80–102, 2011.
- [26] V. Labatut and X. Bost, “Extraction and analysis of fictional character networks: a survey,” *ACM Computing Surveys*, vol. 52, no. 5, p. 89, 2019.
- [27] K. Meichanetzidis, A. Toumi, G. de Felice, and B. Coecke, “Grammar-aware sentence classification on quantum computers,” *Quantum Machine Intelligence*, vol. 5, no. 1, p. 16, 2023. arXiv:2012.03756.
- [28] J. Barzen and F. Leymann, “Quantum humanities: a vision for quantum computing in digital humanities,” *SICS Software-Intensive Cyber-Physical Systems*, vol. 35, pp. 153–158, 2019.
- [29] J. Barzen, “From digital humanities to quantum humanities: Potentials and applications,” in *Quantum Computing in the Arts and Humanities* (E. R. Miranda, ed.), Springer, 2022. arXiv:2103.11825.
- [30] E. R. Miranda, ed., *Quantum Computing in the Arts and Humanities: An Introduction to Core Concepts, Theory and Applications*. Springer, 2022.
- [31] R. Lorenz, A. Pearson, K. Meichanetzidis, D. Kartsaklis, and B. Coecke, “QNLP in practice: Running compositional models of meaning on a quantum computer,” *Journal of Artificial Intelligence Research*, vol. 76, pp. 1305–1342, 2023. arXiv:2102.12846.
- [32] H. Wazni, K. I. Lo, L. McPheat, and M. Sadrzadeh, “Large-scale structure-aware pronoun resolution using quantum natural language processing,” *Quantum Machine Intelligence*, vol. 6, no. 2, p. 60, 2024.
- [33] C. Jurczak, “A 2.48% CER VLM-OCR pipeline for a heavily accented 1544 Florentine treatise,” Apr. 2026. Preprint. <https://zenodo.org/records/19503091>.

- [34] Harvard University Press, “Loeb classical library.” <https://www.loebclassics.com/>. Subscription bilingual Greek/Latin–English series; founded 1911.
- [35] The Latin Library, “The latin library: A free resource for latin texts.” <https://www.thelatinlibrary.com/>. Open-access Latin transcriptions, classical and post-classical.
- [36] DigiLiBLT, “DigiLiBLT: Biblioteca digitale di testi latini tardoantichi.” <https://digiliblt.uniupo.it/>. Università degli Studi del Piemonte Orientale; open-access digital library of Late Antique Latin texts.
- [37] J.-M. Lévy-Leblond, “A mathematical physicist in hell: Galileo on the geometry of Dante’s Inferno,” in *Imagine Math 8: Dreaming Venice* (M. Emmer and M. Abate, eds.), Springer, 2022.
- [38] M. A. Peterson, *Galileo’s Muse: Renaissance Mathematics and the Arts*. Harvard University Press, 2011.
- [39] L. Wang, N. Yang, X. Huang, L. Yang, R. Majumder, and F. Wei, “Multilingual E5 text embeddings: A technical report,” *arXiv preprint arXiv:2402.05672*, 2024.
- [40] K. Enevoldsen *et al.*, “MMTEB: Massive multilingual text embedding benchmark,” in *International Conference on Learning Representations (ICLR)*, 2025.
- [41] Pasqal, “emu-mps: A matrix-product-state emulator for neutral-atom quantum processors.” <https://github.com/pasqal-io/emulators>, 2024. Open-source tensor-network backend for Pasqal Cloud.
- [42] U. Schollwöck, “The density-matrix renormalization group in the age of matrix product states,” *Annals of Physics*, vol. 326, no. 1, pp. 96–192, 2011.
- [43] J. Vovrosh, T. Mendes-Santos, A. Mamann, K. Bidzhiev, M. Hayes, B. Ximenez, L. Béguin, C. Dalyac, and A. Dauphin, “Resource assessment of classical and quantum hardware for post-quench dynamics,” *arXiv preprint arXiv:2511.20388*, 2025.
- [44] G. L. Nemhauser and J. Trotter, Leslie E., “Vertex packings: Structural properties and algorithms,” *Mathematical Programming*, vol. 8, no. 1, pp. 232–248, 1975.
- [45] P. L. Hammer, P. Hansen, and B. Simeone, “Roof duality, complementation and persistency in quadratic 0–1 optimization,” *Mathematical Programming*, vol. 28, no. 2, pp. 121–155, 1984.
- [46] E. Boros and P. L. Hammer, “Pseudo-Boolean optimization,” *Discrete Applied Mathematics*, vol. 123, no. 1–3, pp. 155–225, 2002.
- [47] C. Kommers, R. Ahnert, M. Antoniak, E. Benetos, S. Benford, M. Bunz, B. Caramiaux, S. Concanon, M. Disley, J. Dobson, Y. Du, E. Duéñez Guzmán, K. Francksen, E. Gius, J. Gray, R. Heuser, S. Immel, R. J. So, S. Leigh, D. Livingston, H. Long, M. Martin, G. Meyer, D. Mihai, A. Noel-Hirst, K. Ostherr, D. Parker, Y. Qin, J. Ratcliff, E. Robinson, K. Rodriguez, A. Sobey, T. Underwood, A. Vashistha, M. Wilkens, Y. Wu, Z. Yuan, and D. Hemment, “Computational hermeneutics: Evaluating generative AI as a cultural technology,” *SSRN preprint*, 2025. SSRN abstract ID 5409144, submitted August 2025.
- [48] T. F. Gonzalez, “Clustering to minimize the maximum intercluster distance,” *Theoretical Computer Science*, vol. 38, pp. 293–306, 1985.
- [49] D. S. Hochbaum and D. B. Shmoys, “A best possible heuristic for the k -center problem,” *Mathematics of Operations Research*, vol. 10, no. 2, pp. 180–184, 1985.
- [50] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, “An analysis of approximations for maximizing submodular set functions – I,” *Mathematical Programming*, vol. 14, pp. 265–294, 1978.
- [51] A. Krause and D. Golovin, “Submodular function maximization,” *Tractability: Practical Approaches to Hard Problems*, pp. 71–104, 2014.
- [52] OuLiPo, *Atlas de littérature potentielle*. Paris: Gallimard, 1981.
- [53] J. Roubaud, *∈*. Paris: Gallimard, 1967.

- [54] R. Netz and W. Noel, *The Archimedes Codex: Revealing the Secrets of the World's Greatest Palimpsest*. Da Capo Press, 2011.
- [55] D. M. Ausonius, “Praefatio ad Centonem Nuptialem,” ca. 370. Modern critical edition: R. P. H. Green (ed.), *The Works of Ausonius*, Oxford: Clarendon Press, 1991, pp. 132–134; Loeb edition: H. G. Evelyn-White (trans.), *Ausonius*, vol. 1, Cambridge MA: Harvard University Press, 1919/1985, pp. 370–373. Photographed in this work from the Aldine *editio princeps*, Venice, 1517.
- [56] A. Blossier-Jacquemot, *Les Oulipiens antiques: pour une anthropologie des pratiques d’écriture à contraintes dans l’antiquité*. PhD thesis, Université Paris VII Diderot, 2009.
- [57] I. Calvino, *Il castello dei destini incrociati*. Torino: Einaudi, 1973.
- [58] I. Calvino, *Se una notte d’inverno un viaggiatore*. Torino: Einaudi, 1979.
- [59] I. Calvino, *Lezioni americane: sei proposte per il prossimo millennio*. Milano: Garzanti, 1988.
- [60] J. L. Borges, *El jardín de senderos que se bifurcan*. Buenos Aires: Editorial Sur, 1941. Collection containing *La biblioteca de Babel* and the title story *El jardín de senderos que se bifurcan*; both later included in *Ficciones* (1944).
- [61] C. Jurczak, “QOuLiPo corpus — canonical text-graph data,” 2026. Dataset. <https://doi.org/10.5281/zenodo.20074378>.
- [62] R. P. Feynman, “Simulating physics with computers,” *International Journal of Theoretical Physics*, vol. 21, no. 6–7, pp. 467–488, 1982.
- [63] C. Dalyac and L. Henriët, “Embedding the MIS problem for non-local graphs with bounded degree using 3D arrays of atoms,” *arXiv preprint arXiv:2209.05164*, 2022.
- [64] P. Cazals, A. Sorondo, V. Onofre, C. Dalyac, W. Coelho, and V. Vitale, “Quantum optimization on rydberg atom arrays with arbitrary connectivity: Gadgets limitations and a heuristic approach,” *Physical Review A*, vol. 112, p. 062416, 2025.
- [65] L. Leclerc, S. Julià-Farré, G. Silva Freitas, G. Villaret, B. Albrecht, L. Béguin, *et al.*, “One-to-one quantum simulation of the low-dimensional frustrated quantum magnet TmMgGaO₄ with 256 qubits,” *arXiv preprint arXiv:2603.20372*, 2026.
- [66] D. Barredo, V. Lienhard, S. de Léséleuc, T. Lahaye, and A. Browaeys, “Synthetic three-dimensional atomic structures assembled atom by atom,” *Nature*, vol. 561, no. 7721, pp. 79–82, 2018.
- [67] M.-T. Nguyen, J.-G. Liu, J. Wurtz, M. D. Lukin, S.-T. Wang, and H. Pichler, “Quantum optimization with arbitrary connectivity using Rydberg atom arrays,” *PRX Quantum*, vol. 4, no. 1, p. 010316, 2023.
- [68] B. Sartini, S. Baroncini, M. van Erp, F. Tomasi, and A. Gangemi, “ICON: An ontology for comprehensive artistic interpretations,” *Journal on Computing and Cultural Heritage*, vol. 16, no. 3, pp. 59–76, 2023. ICON ontology v2.0 at <https://w3id.org/icon/ontology>; formalises Panofsky’s tripartite pre-iconographic / iconographic / iconological framework.
- [69] B. Sartini, “IICONGRAPH: improved iconographic and iconological statements in knowledge graphs,” in *The Semantic Web — 21st European Semantic Web Conference, ESWC 2024, Heraklion, Crete, Greece, May 26–30, 2024, Proceedings*, Springer, 2024. arXiv:2402.00048; dataset Zenodo DOI 10.5281/zenodo.10294589, CC-BY 4.0.
- [70] S. Baroncini and F. Mele, “Describing monastic iconography using semantic data: A preliminary investigation,” in *AIUCD 2025: Conference of the Italian Association for Digital Humanities*, 2025.
- [71] D. Lim, H. Mamann, G. Pichard, *et al.*, “Defect-free arrays at the thousand-atom scale in a 4-K cryogenic environment,” *arXiv preprint arXiv:2604.07205*, 2026.
- [72] H. J. Manetsch, G. Nomura, E. Bataille, K. H. Leung, X. Lv, and M. Endres, “A tweezer array with 6,100 highly coherent atomic qubits,” *Nature*, vol. 647, pp. 60–67, 2025.

- [73] J. Chen, S. Xiao, P. Zhang, K. Luo, D. Lian, and Z. Liu, “BGE M3-Embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation,” *arXiv preprint arXiv:2402.03216*, 2024.
- [74] P. Yu, L. Merrick, G. Nuti, and D. Campos, “Arctic-Embed 2.0: Multilingual retrieval without compromise,” 2024.
- [75] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge: Cambridge University Press, 2008.
- [76] C. Dalyac, L. Leclerc, L. Vignoli, M. Djellabi, W. da Silva Coelho, B. Ximenez, A. Dareau, D. Dreon, V. E. Elfving, A. Signoles, L.-P. Henry, and L. Henriët, “Graph algorithms with neutral atom quantum processors,” *European Physical Journal A*, vol. 60, p. 177, 2024.

Supporting Information

The supporting information collects the corpus and methodology references for the main text:

- A. Corpus.** The eight natural texts used throughout, with author, date, language, segmentation, description and source edition.
- B. Embedding validation.** Cross-embedder sensitivity study on four natural texts $\times$ three multilingual transformers.
- C. *Twelve Hours of a Coastal Town*: full text.** Complete English mini-text.
- D. Classical-benchmark results.** Classical MIS solver timings (ILP, greedy, simulated annealing).
- E. QOuLiPo: full corpus inventory.** Per-text catalogue of the twenty-nine constrained QOuLiPo texts.
- F. Register and pulse optimisation.** Register engineering (SA, blockade margin, case studies) and pulse design (baseline, variational sweep).

A Corpus

Text	Author	Date	Lang	N	Unit	What it is	Source / edition
<i>De'l sito, forma, & misure dello Inferno di Dante</i>	P.F. Giambullari	1544	IT	151 / 65	printed page (octavo)	Renaissance Florentine treatise reconstructing the geometry of Dante's Hell from textual clues in the <i>Divina Commedia</i> ; central document of this edition project.	Neri Dortelata, Firenze, 1544; OCR [33].
<i>Dialogo sopra i due massimi sistemi del mondo</i>	Galileo Galilei	1632	IT	65	400-word window	Defence of heliocentrism as a four-day conversation between three speakers; the book that triggered Galileo's trial by the Inquisition.	Landini, Firenze, 1632; Wikisource transcription.
<i>Confessiones</i> (Book XIII)	Augustine of Hippo	c. 400	LAT	38 / 22	paragraph / chapter	Final book of the spiritual autobiography: a meditation on the opening verses of Genesis and on grace in self-knowledge.	J.-P. Migne, <i>PL</i> 32; Latin Library.
<i>De mortibus persecutorum</i>	Lactantius	c. 315	LAT	52 / 29	chapter / grouped	"On the deaths of the persecutors." Pamphlet narrating the violent ends of Roman emperors who had persecuted Christians, Nero to Diocletian.	Creed (ed.), Oxford 1984; Latin Library.
<i>Inferno</i>	Dante Alighieri	c. 1320	IT	34	canto	First part of the <i>Divina Commedia</i> ; Dante's descent through 34 cantos of Hell, guided by Virgil.	Petrocchi critical ed.; Wikisource.
<i>Heptaméron</i>	Marguerite de Navarre	1558–59	FR	72	nouvelle	72 framed nouvelles told by ten travellers stranded in a Pyrenean abbey, each followed by group discussion; French counterpart to the <i>Decameron</i> .	Gruget, Paris, 1559; Wikisource.
<i>De consolazione philosophiae</i>	Boethius	c. 524	LAT	72	prose/verse piece	Philosophical dialogue from prison: Lady Philosophy visits the condemned Boethius and reasons through fortune, providence, and the highest good.	Loeb Classical Library; Latin Library.
<i>Epigrammata</i>	Ausonius of Bordeaux	c. 370	LAT	27	individual poem	27 short occasional epigrams from the late-Latin poet's collected works.	Aldine <i>Opera</i> , Venice, 1517; author's OCR.

Table 4: Primary corpus of natural text-graph instances. Where two N values appear separated by “/” (Giambullari, Augustine, Lactantius), the row’s *Unit* column lists which chunking serves which use.

The texts span four centuries (4th–17th), three languages (Latin, Italian, French), and deliberately heterogeneous genres. The “Unit” column reports the segmentation that determines N following the tiered granularity rule of §2: authorial divisions where they exist, printer’s-page fall-back where they do not, and fixed analytic windows for continuous prose (Galileo). Several texts appear at more than one granularity, each table using the one most informative for its argument: *Giambullari* $N=151$ for rigidity, $N=65$ on the QPU; *Augustine* $N=38$ (paragraph) throughout including QPU, $N=22$ (chapter) in classical benchmarking only; *Lactantius* $N=52$ (chapter) throughout, $N=29$ (grouped) in benchmarking only; *Dante* $N=34$ (canto) for the full *Divina Commedia* in benchmarking.

B Embedding validation

The paper fixes a single embedder for natural-text graphs — `intfloat/multilingual-e5-large-instruct` [39, 40], top-ranked on the MMTEB multilingual benchmark and handling Latin, Italian, and French without language-specific tuning. The natural question is whether a different embedder would change the structural conclusions: are MIS/N and ρ stable across embedders, or are they artefacts of the e5 family?

To answer this we re-ran the full pipeline on four natural texts spanning languages and lengths (Dante’s *Inferno*, $N=68$, Italian; Boethius’ *Consolatio Philosophiae*, $N=72$, Latin; Galileo’s *Dialogo*, $N=130$, Italian; Giambullari, $N=151$, Italian — the cross-embedder check uses a finer chunking than the rest of the paper for *Inferno* (half-canto, $N=68$ instead of the canto-level $N=34$ used elsewhere) and *Dialogo* (200-word window, $N=130$ instead of the 400-word $N=65$ of §5), so that each text has enough nodes to give the rigidity test statistical resolution under all three embedders) with three competitive multilingual embedders: `intfloat/multilingual-e5-large-instruct` [39], `BAAI/bge-m3` [73], and `Snowflake/snowflake-arctic-embed-l-v2.0` [74]. Each embedder feeds the same downstream pipeline — cosine k -NN at $k=8$, exact ILP MIS, and per-vertex ILP-exclusion for ρ ($t_{\text{lim}} = 15$ s per vertex, 32-way parallel) — so only the embedding model varies. Table 5 gives the side-by-side numbers.

Reading the table. Table 5 is designed for *within-row* comparison across the three embedders and is run as a self-contained sensitivity experiment with its own graph-build settings (in particular, a symmetric-union k -NN graph construction that yields denser edge sets than the canonical mutual- k -NN build used in Tables 1 and 4). Absolute MIS, $|E|$, and ρ values in Table 5 therefore differ from the canonical Table 1 numbers even on the same text at the same N and k (e.g. Boethius $N=72$, $k=8$: $\rho = 0.000$ canonical vs $\rho = 0.333$ here, on a denser graph), and the two tables should not be cross-compared cell-by-cell. The single conclusion this appendix supports is the within-row one: across the three embedders, MIS/N is stable, ρ is not.

Text	Embedder	$ E $	MIS	MIS/N	ρ
Dante <i>Inferno</i> ($N=68$)	e5-large-instruct	416	21	0.309	0.476
	bge-m3	405	17	0.250	0.118
	arctic-embed-l-v2	411	18	0.265	0.444
Boethius <i>Consolatio</i> ($N=72$)	e5-large-instruct	399	18	0.250	0.333
	bge-m3	401	18	0.250	0.444
	arctic-embed-l-v2	403	19	0.264	0.474
Galileo <i>Dialogo</i> ($N=130$)	e5-large-instruct	829	40	0.308	0.475
	bge-m3	815	37	0.285	0.297
	arctic-embed-l-v2	765	34	0.262	0.706
Giambullari ($N=151$)	e5-large-instruct	864	40	0.265	0.325
	bge-m3	894	43	0.285	0.558
	arctic-embed-l-v2	923	41	0.272	0.195

Table 5: Cross-embedder sensitivity on four natural texts. Each text is run through three multilingual transformers; downstream pipeline (k -NN at $k=8$, ILP MIS, certified ρ) is held fixed. MIS/N stays in the 0.25–0.31 band across all twelve (text, embedder) pairs; ρ varies from 0.118 to 0.706 and reorders the texts.

Two findings emerge. First, MIS/N is embedder-stable: the spread within each text is at most 0.06, and all four texts cluster in the same 0.25–0.31 band irrespective of which transformer produced the embedding. Backbone *size* is therefore a robust property of the text. Second, ρ is not stable, and the instability is informative rather than noisy: the embedders disagree on which text is the most rigid (Galileo at $\rho = 0.71$ under `arctic-embed-l-v2` drops to $\rho = 0.30$ under `bge-m3`), and the disagreement preserves the qualitative spectrum — semi-rigid versus modular — while permuting the ranking. The interpretation is that ρ is a property of the (*text*, *embedder*) pair, not of the text alone, and any cross-corpus comparison of rigidity should fix the embedder. The paper fixes `e5-large-instruct` throughout for consistency, motivated by its top rank on the MMTEB multilingual benchmark [40].

C Twelve Hours of a Coastal Town: full text

Complete text of the twelve-paragraph mini-text used for the worked example of §3.1. The MIS backbone is $\{\P3, \P4, \P7, \P10\}$, one paragraph from each of the four thematic clusters (dawn departure, mid-morning market, silent noon, late-afternoon return); the persistent core is $\{\P3, \P10\}$ — the cliff-watcher and the boy mending nets, present in every one of the nine optimal backbones.

($\P1$) *At first light the fishermen drag their wooden boats across the wet sand. The hulls leave parallel grooves that the next tide will erase. Nets are loaded, oars are checked, the youngest boy carries the bait bucket without complaint.*

($\P2$) *The harbour wakes by degrees. A second crew of fishermen pushes off from the stone jetty, their boat heavier than the first, their nets longer. Gulls follow them out, screaming for what has already been thrown overboard.*

($\P3$) *From the cliff above the beach an old fisherman watches the two crews vanish. He no longer goes to sea. He keeps a small notebook of weather signs and writes one line each morning, always about wind direction and the colour of the horizon.*

($\P4$) *By mid-morning the market square is loud with bargaining. Fishwives shout the prices of mackerel and sardines, butchers slap the boards with their cleavers, the children of the town run between the stalls stealing peaches when no one is looking.*

($\P5$) *At the centre of the market a cheese-seller has set up a long table covered with white cloths. She arranges the rounds of pecorino in pyramids and offers thin slices on the point of a knife to anyone who looks hungry. The bargaining around her stall is louder than anywhere else.*

($\P6$) *The market smells of fish and lemon and warm bread. Two priests pass through on their way to the harbour chapel, refusing to bargain for anything, blessing the children who try to steal their hats. The square will not empty until the heat of midday makes everyone retreat indoors.*

($\P7$) *After lunch the town falls silent. The shutters close one by one, the dogs sleep in the shadow of the church wall, and the only sound is the rope of a small boat creaking against its bollard down at the harbour. Even the gulls have stopped screaming.*

($\P8$) *A single cat walks the empty length of the main street, stops, looks at nothing in particular, then disappears under a parked cart. The shadow of the bell tower has reached the fountain. No one draws water; the noon hour belongs to no one.*

($\P9$) *In the orchard behind the church a beekeeper turns his hives one quarter turn east, as he does every Tuesday afternoon while the town sleeps. The bees do not protest. He works in silence, bare-handed, and is back inside his cool kitchen before the first shutter reopens.*

($\P10$) *By late afternoon the tide has come in and is licking the keels of the boats hauled up that morning. The same boy who carried the bait bucket at dawn now sits on the gunwale of his father's boat, mending a torn net with a wooden needle, his fingers already as quick as a man's.*

($\P11$) *The second crew returns rowing slowly, their boat low in the water with the day's catch. The harbour wall is lined with women and children waiting to count the baskets. The old fisherman from the cliff has come down at last, and stands at the back of the crowd, saying nothing.*

($\P12$) *At sunset the harbour is loud again, but it is a different loudness from the morning. Lanterns are lit on the jetty, the fish are sorted into baskets by size and not by species, and the cheese-seller from the market square walks down with two empty crates to buy a share of the catch for tomorrow.*

D Classical-benchmark results

We benchmarked three classical MIS solvers on all text-graph instances — exact integer linear programming (ILP), a greedy heuristic (maximum-degree-first), and simulated annealing — and report the per-instance numbers in Table 6.

The hardest single-instance ILP solve in the natural-text corpus is *Boethius* ($k=8$, $N=72$, ≥ 500 distinct optima): the MIS itself is found in 0.06 s, but exhaustive enumeration of the optimal-MIS manifold up to a 500-cap takes ~ 106 s of cumulative ILP time ($\dagger\dagger$). *Heptaméron* at $k=16$ is the single-instance 1.5 s case ($\dagger$, MIS plus 6 distinct optima). The hardest QOuLiPo engineered instance is *Le Pari de Nithard III* ($k=16$, $N=100$); the small- N standout is *Sonetti dal Tesseratto* ($N=16$, $d=0.658$, the densest QOuLiPo graph, where the Q_4 hypercube k -NN saturation is genuinely harder than its size suggests). The exact-UDG QOuLiPo rows (Stanze, Castello, Treize, Squares, Faces, Hours, Kagome, Postman) are omitted from the table as classically trivial: 2D-UDG MIS admits a PTAS, and CBC solves each in < 0.1 s. Erdős–Rényi random graphs at the same density are 20–40 $\times$ harder than the engineered texts: the engineered structure regularises the MIS landscape that ILP exploits.

Graph	N	d	ILP (exact)	t_{ILP}	Greedy	SA
Augustine ($k_{\text{nn}}=8$)	38	0.293	9	0.12 s	7	9
Lactantius ($k_{\text{nn}}=8$)	52	0.232	13	0.07 s	11	13
Heptaméron ($k_{\text{nn}}=16$)	72	0.331	14	1.5 s [†]	—	—
Boethius ($k_{\text{nn}}=8$)	72	0.146	14	106 s ^{††}	—	—
Nithard’s Wager ($k=16$)	50	0.305	14	0.07 s	13	14
Le Pari de Nithard ($k=16$)	50	0.349	10	0.08 s	10	10
Le Pari de Nithard II ($k=16$)	65	0.303	8	1.0 s	8	7
Galileo 65 ($k=16$)	65	0.249	14	0.16 s	13	13
Giambullari 65 ($k=8$)	65	0.120	20	0.06 s	20	19
Giambullari 151 (full)	151	0.077	41	2.3 s	40	40
Le Pari de Nithard III ($k=16$)	100	0.222	20	1.5 s	18	17
Le Pari de Nithard III ($k=25$)	100	0.336	15	1.0 s	13	12
Sonetti dal Tesseratto ($Q_4, k=8$)	16	0.658	4	0.72 s	—	—
<i>ER random</i> ($d \approx 0.3$)	100	0.298	15	33 s	14	12
<i>ER random</i> ($d \approx 0.4$)	100	0.399	11	42 s	11	10
<i>ER random</i> ($d \approx 0.5$)	100	0.495	9	46 s	7	7

Table 6: Classical MIS benchmark (CBC ILP). [†] *Heptaméron*: time includes MIS plus enumeration of 6 distinct optima. ^{††} *Boethius*: MIS in 0.06 s; the reported 106 s is cumulative ILP time for enumerating the optimal-MIS manifold up to the 500-solution cap.

E QOuLiPo: full corpus inventory

Table 7 catalogues the twenty-nine engineered QOuLiPo texts, sorted by N . Rigidity ρ is the fraction of MIS pages common to every optimal solution (entries showing $\geq$ hit the 500-solution enumeration cap). Source data: [61].

Text	Lang	N	$ E $	d	ρ	MIS	MIS/ N	# opt.	UDG	HW	Designed property
Sonetti dal Tesseratto	IT	16	79	0.658	1.000	4	0.250	1		EMU	Q_4 hypercube; 16 sonnets at $k=8$
Les Aventures de Pascal	FR	20	47	0.247	0.000	5	0.250	4	✓	EMU (2L)	King 5×4 , single layer of bilayer pair
Scholia de Ménéil	FR	20	47	0.247	0.000	5	0.250	5	✓	EMU (2L)	King 5×4 , single layer (FR commentary)
The Friar's Notebook	EN	20	30	0.158	0.000	8	0.400	5		EMU (3D)	Pacioli dodecahedron, 30 register-aware motifs (3D UBG)
La Vita nel Cubo	IT	23	128	0.506	0.571	7	0.304	5		EMU	Truncated-cube conceit, $k=8$
Le venticinque stanze	IT	25	72	0.240	1.000	9	0.360	1	✓	QPU	King 5×5
The Twenty-Five Rooms	EN	25	72	0.240	1.000	9	0.360	1	✓	EMU	King 5×5 (EN twin of <i>venticinqu stanze</i>)
Les Jumeaux du Graphe (EN)	EN	30	38	0.087	0.000	14	0.467	21		EMU	C_{30} + 8 same-parity chords
Les Jumeaux du Graphe (FR)	FR	30	38	0.087	0.000	14	0.467	21		EMU	FR twin of EN, shared edge list
Treize sur trente-sept	FR	37	90	0.135	1.000	13	0.351	1	✓	QPU	Centered-hex patch, $R=3$
Pascal-Ménéil bilayer	FR	40	240	0.308	0.000	6	0.150	≥ 500		EMU (2L)	$5 \times 4 \times 2$ king bilayer (UBG)
La Partition du Texte	FR	46	244	0.236	0.091	11	0.239	25		EMU	Five poetic-form communities, $k=8$
Il castello dei quarantanove destini	IT	49	396	0.337	1.000	9	0.184	1	✓	QPU	Extended-king 7×7 (Cheb. ≤ 2); king-only on QPU
Nithard's Wager (EN)	EN	50	374	0.305	0.571	14	0.280	6		QPU	Hard zone, $k=16$
Le Pari de Nithard (FR translation)	FR	50	427	0.349	0.400	10	0.200	19		QPU	Hard zone, $k=16$
The Irreplaceable Book	EN	50	290	0.237	1.000	17	0.340	1		EMU	Unique MIS via double domination; max degree 19 (not 2D-UDG-realizable)
The Map of the Text	EN	50	85	0.069	0.000	25	0.500	2	✓	EMU	Planar grid 5×10
The Trial of Nithard	EN	50	625	0.510	0.000	25	0.500	2		EMU	Bipartite $K_{25,25}$ (provably non-UDG)
The Reader's Trap	EN	50	72	0.059	0.417	24	0.480	≥ 500	✓	QPU	Hub-spoke; greedy-vs-optimum gap maximised
The Fractal Book	EN	50	102	0.083	0.316	19	0.380	≥ 267	✓	QPU	Self-similar (Sierpinski)
The Kaleidoscope	EN	51	51	0.040	0.000	17	0.333	3^{17}	✓	EMU	Maximal degeneracy: 17 disjoint K_3
The Incarnate Graph	EN	65	155	0.075	0.000	17	0.338	≥ 385	✓	QPU	Random UDG placement (insufficient blockade margin)
Le Pari de Nithard II	FR	65	631	0.303	0.250	8	0.123	≥ 255	✓	QPU	Hard zone at $N=65$, $k=16$
The Book of Eighty-One Squares	EN	81	272	0.084	1.000	25	0.309	1	✓	QPU	King 9×9
Eighty-One Faces of a Pomegranate	EN	81	622	0.192	1.000	13	0.160	1	✓	QPU	Extended-king 9×9 ($\sqrt{5}$ reach)
Hours of the Day	EN	91	240	0.059	1.000	31	0.341	1	✓	QPU	Centered-hex patch, $R=5$
Kagome of Late Summer	EN	100	180	0.036	0.846	39	0.390	8	✓	QPU	Kagome (3.6.3.6) Archimedean tiling
The Postman's Songbook	EN	100	220	0.044	0.333	36	0.360	97	✓	QPU	Saeb-square (3.3.4.3.4) Archimedean tiling
Le Pari de Nithard III	FR	100	691	0.140	0.923	26	0.260	4	✓	EMU	100-atom designed k -NN target, $k=8$

Table 7: Full QOuLiPo corpus inventory (twenty-nine constrained texts), sorted by register size N . HW codes: QPU = adiabatic FRESNEL_CAN1 run; EMU = EMU_MPS sweep (Pascal Cloud A100 back-end for most runs, commercial H100/H200 cloud for the larger instances); EMU (2L) and EMU (3D) = bilayer / 3D EMU runs.

F Register and pulse optimisation

This appendix gives the detailed accounting of register engineering and pulse design summarised in §5.1.

F.1 Register engineering

The register embedding proceeds differently for the two halves of the corpus. For natural texts G_{text} is not unit-disk in general, so it must be mapped onto physical atom positions — a non-trivial geometric optimisation that produces G_{reg} as the lossy realisation. We use simulated annealing (SA) on discrete sites of a triangular lattice, which handles the non-convex landscape of blockade constraints robustly without requiring gradient information, and which Dalyac and Henriët [63] also adopt for neutral-atom register embedding. The SA embedder places N atoms on the lattice such that atom pairs within the blockade radius $R_b = 8.0 \mu\text{m}$ (G_{reg} edges) correspond as closely as possible to G_{text} edges; the optimiser runs 16,000–20,000 iterations with 6–10 random restarts (move + swap moves), maximising *edge recall* — the fraction of canonical edges $(u, v) \in G_{\text{text}}$ whose embedded coordinates lie within R_b in G_{reg} , the standard information-retrieval recall metric [75] applied to graph adjacency. For the engineered exact-UDG texts the register embedding is trivially exact: place atoms at integer grid coordinates with spacing d and set $R_b = d\sqrt{2} + \epsilon$. Every king adjacency falls within R_b ; every non-adjacent pair is at distance $\geq 2d$, outside R_b . Edge recall is 1.0 by construction ($G_{\text{design}} = G_{\text{reg}}$), no SA needed.

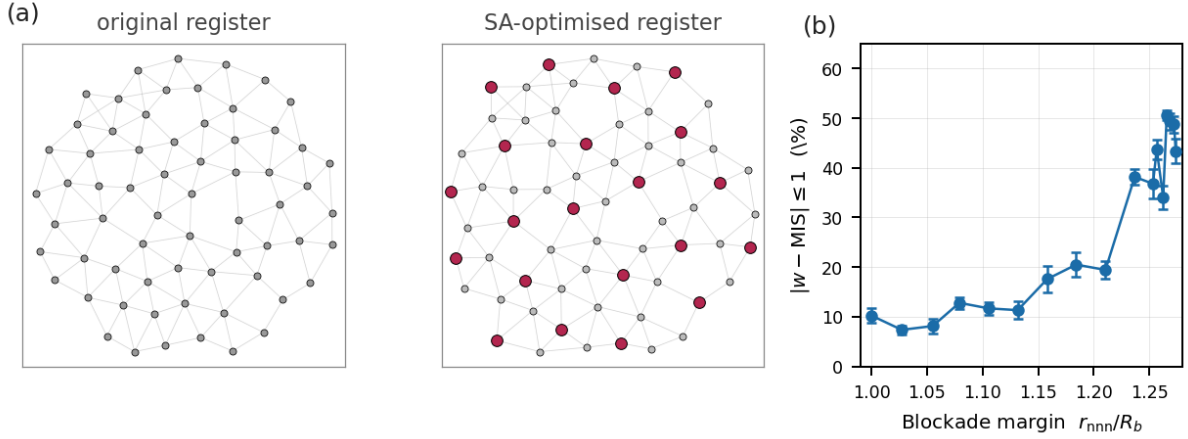


Figure 10: Register engineering on *The Incarnate Graph* ($N = 65$, random UDG). (a) Two QPU runs on the same canonical graph and the same baseline pulse, two different registers. Left: original 2D placement, 0 valid IS out of 1000 shots (ratio 0.000). Right: SA-optimised placement, with the largest valid independent set returned by the QPU shown in red (21/22, ratio 0.955, 402 shots). (b) Blockade-margin curve on the same canonical graph: $|w - \text{MIS}| \leq 1$ fraction — the fraction of shots whose Hamming weight is within one of the exact MIS size — versus achieved blockade margin r_{nnn}/R_b , EMU simulation.

Edge recall 1.0 by construction is necessary but not sufficient for clean MIS recovery on hardware, as *The Incarnate Graph* ($N=65$, random-placement exact 2D UDG) makes plain: 0/1000 valid IS (i.e. shots whose bitstring forms an independent set on the canonical graph), with the Hamming-weight distribution centred near $\text{HW} \approx 18$ (close to $\text{MIS}=22$) — the system reaches the right occupation count but cannot satisfy independence. The reason is geometric: an exact-UDG canonical graph is necessary but not sufficient for clean MIS recovery, because the register — the physical placement of atoms — must also satisfy a sufficient *blockade margin*, which we use as shorthand for the ratio r_{nnn}/R_b between the smallest non-edge (next-nearest-neighbour) distance and the blockade radius. The qualitative requirement $r_{\text{nn}} \ll R_b \ll r_{\text{nnn}}$ for clean Rydberg-blockade embeddings of unit-disk graphs is well-established (Dalyac *et al.* [76], Box 1). The practical implication is that when a register layout fails on hardware (zero valid IS, as on the original *Incarnate* run — see the case studies below), one can often recover it by repositioning the atoms — keeping the canonical graph and the pulse fixed — so that all non-edges sit comfortably above R_b . How far above is achievable is graph-dependent: each canonical graph has its own geometric ceiling on r_{nnn}/R_b , beyond which no register can be found.

Treize sur trente-sept (centred-hex axial-coordinate bug, refit to a proper triangular layout) recovers to ratio 1.000 on EMU and 0.923 on QPU; *The Kaleidoscope* (17 disjoint K_3 , refit to hex-packed cluster

centres) recovers to 1.000 on EMU. On *The Incarnate Graph* (Figure 10a) the original random UDG placement left 17 non-edges in the soft tail at $\leq 1.1 R_b$; an SA register pushing all non-edges beyond $1.25 R_b$ rescues the row on QPU. Re-spacing the *Hours of the Day* lattice ($N=91$ centred-hex) from $a=5\mu\text{m}$ to $a=6\mu\text{m}$ similarly lifts the blockade margin from 1.08 to 1.30, taking the row from 0/1000 valid IS to ratio 0.839 at 23.9% valid IS on 1000 shots. The same approach was used to design two non-lattice 2D-UDG instances directly for the QPU: *The Reader's Trap* (1.000/65.4%) and *The Fractal Book* (0.947/60.5%).

Figure 10b quantifies the rule. We swept the achieved blockade margin r_{nnn}/R_b on the *Incarnate* canonical graph from 1.00 to 1.27, re-engineering the register at each target by SA, reusing the same canonical edges and baseline pulse. The $|w-\text{MIS}| \leq 1$ fraction — a softer cross-platform metric than strict valid IS, well-defined whether the canonical graph is exact-UDG or k -NN — transitions monotonically with the achieved margin, identifies an empirically sufficient threshold for clean recovery on this graph, and exposes the graph's geometric saturation cap beyond which no register can do better.

Register engineering, especially for non-lattice canonical graphs, is a necessary step rather than a cosmetic one. SA on atom positions suffices for every case here; further techniques are surveyed in [76].

F.2 Pulse optimisation

The atoms are driven by a simple baseline pulse, constructed with Pasqal's open-source Pulser library [15]: constant Rabi frequency $\Omega = C_6/R_b^6 = 3.30 \text{ rad}/\mu\text{s}$ and a linear detuning sweep from -3Ω to $+3\Omega$ over $T = 4 \mu\text{s}$, chosen for reproducibility rather than optimality. For the larger instances ($N \geq 81$) we extend the ramp to $T = 6 \mu\text{s}$ — the QPU's adiabatic-pulse ceiling — because the inverse spectral gap of the MIS scales with system size, and the longer ramp is necessary for the system to follow the instantaneous ground state through the gap-closing region into the MIS configuration.

The pulse-engineering axis is in principle fertile, and the EMU loop of Figure 7 is well-suited to exploring it cheaply. We ran a 10-shape variational sweep — baseline, longer T , wider/tighter detuning sweep, Ω -trapezoid envelope, four-knot asymmetric ramps in the spirit of the variational adiabatic shaping of Ebadi *et al.* [4], multi-point cubic detuning interpolation, and jittered-register controls — on EMU for the strong-tier exact-UDG rows. On EMU the variational shapes give substantial lifts on the validity and structural-readout indicators while leaving the ratio at 1.000. On QPU the corresponding lifts are more modest: EMU-selected variational pulses improve **valid IS%** by +3–6 percentage points over baseline, ratio pinned at 1.000. *The Reader's Trap* and *The Fractal Book* (the two non-lattice 2D-UDG instances) ran with an Ω rise-plateau-fall envelope rather than the flat baseline and cleared 60% valid IS at ratio ≥ 0.95 . For *The Postman's Songbook* (snub-square Archimedean tiling) the multi-distance geometry required reducing Ω to $1.63 \text{ rad}/\mu\text{s}$ and proportionally rescaling the detuning span; the standard $\Omega = 3.30$ pulse over-drives the next-nearest pairs of the snub-square lattice. On natural-text rows the trade-off is more nuanced: pulse variants that improve near-valid coverage tend to reduce the strict approximation ratio, and vice versa. Pulse engineering at FRESNEL scale is therefore a second-order refinement on top of a working register, not a substitute for one. Noise-aware pulse optimisation [5] and Floquet / dynamical-decoupling pulse design [76] are natural follow-ups.

QUANTONATION — INVESTING IN THE PHYSICS FRONTIERS SHAPING THE 21ST CENTURY

QUANTONATION VENTURES
SOCIÉTÉ PAR ACTIONS SIMPLIFIÉES – 131 538€ SHARE CAPITAL
REGISTERED UNDER NUMBER 849 813 522 RCS PARIS
58, RUE D'HAUTEVILLE -75010 PARIS
APPROVED BY THE AMF UNDER NUMBER GP-202211

QUANTONATION