arXiv:2603.02610v1 [quant-ph] 3 Mar 2026

全光子与配备量子存储器的量子交换机中的速率-保真度权衡

Panagiotis Promponas、Leonardo Bacciottini、Paul Polakos、
Gayane Vardoyan、Don Towsley 和 Leandros Tassiulas

摘要——**量子纠缠交换机是近期量子网络的关键构成模块，而一个核心设计问题是：近期设备应仅使用光学组件和光子量子比特，还是也应集成量子存储器。我们比较两种架构：一种全光子纠缠生成交换机（EGS），在不存储量子比特的情况下反复尝试贝尔态测量；另一种配备量子存储器的量子交换机，对纠缠进行缓冲，并仅在预示连接可用时触发测量（先通知后交换控制）。这两种设计各有取舍：无存储器的简单运行方式可避免退相干和存储器引起的延迟；基于预告的控制则通过缓冲纠缠，更高效地利用贝尔态测量。我们在统一的硬件抽象下对两种模型进行形式化，并刻画其可实现的速率–保真度区域，进而提出一种基准测试方法，将硬件和协议参数转化为网络级性能。数值评估量化了两种模型的速率–保真度权衡，确定了各架构占优的工作区域，并展示了如何调整硬件和协议参数，以满足特定应用目标。**

I. 引言

在安全通信、分布式传感和分布式量子计算等应用中，能够在相距遥远的用户之间分发高保真度纠缠的量子网络是其基础 [1], [2]。其中一个核心挑战在于，在有损光学信道中，衰减会导致每次纠缠生成尝试的成功概率随距离呈指数下降。这一局限性使得网络需要中间节点，例如采用纠缠交换来在非相邻客户端之间建立纠缠的量子中继器 [3] 和交换机 [4], 。

从量子中继器发展到多用户量子交换机，引入了新一层网络级复杂性：除了底层的物理层误差过程外，交换机还必须在多个用户对之间存在竞争的情况下分配共享资源并调度纠缠交换。量子交换机是连接三个或更多其他节点的网络节点。它作为双侧排队系统运行 [5]：一侧接收端到端纠缠请求，另一侧则由概率性的链路级纠缠生成（LLEG）产生贝尔对 [6] ，这些贝尔对必须通过纠缠交换来满足服务器请求。

大量研究探讨了多种量子交换模型 [4], [7]。一种有用的分类法基于硬件简易性与控制智能之间的主要架构权衡，其根源在于量子存储器的使用以及由此产生的运行逻辑。这一分类为把握当前研究格局提供了框架，并促使本文选择所分析的模型。

第一大类交换机架构由全光子硬件 [8] 构成，这类硬件直接对到达的光子执行纠缠交换，而无需在交换机处将量子比特存储于量子存储器中。在这些通常称为纠缠生成交换机（EGS） [7], [9], [10], 的架构中，由于没有量子存储器，交换机无法在执行贝尔态测量（BSM）之前预示成功的链路级纠缠（LLE）尝试次数。此外，贝尔态测量由一组贝尔态分析器（BSA）执行，每个分析器均可分配给任意用户对。缺乏先兆迫使交换机对预先指定的用户对“盲目”触发贝尔态测量。因此，当两个客户端中至多只有一个成功实现链路级纠缠（即一个纠缠光子到达相应的贝尔态分析器）时，贝尔态分析器往往仍会被触发，从而无法产生端到端纠缠对。近期工作探讨了针对EGS的资源分配方案，包括在资源不可用时丢弃请求的按需阻塞模型 [9], ，以及轮询节点以获取请求的负载均衡策略 [10]。

第二类架构引入量子存储器，从而实现更复杂的动态控制逻辑。这类交换机采用“先预示、后交换”的运行模式。在这种模式下，首先尝试在每个客户端与交换机之间建立LLEG。每次尝试是否成功，都会通过 “ 预示 ” 消息告知交换机的控制器。交换机在连接获取阶段结束后作出调度决策。此时，控制器已知可用贝尔对的集合，因此可以进行最优匹配，将已占用存储器与纠缠交换操作配对。这种方法能更有效地利用存储的贝尔对，因为仅对已预示的链路进行贝尔态测量尝试。配备存储器的交换机已得到广泛研究，相关工作刻画了其吞吐量区域、提出了调度策略 [11], [12], [13], [14], ，并分析了存储器退相干的影响 [15], [16], [5]。

这两种架构提出了一个具体的设计问题：对于给定的硬件栈和目标应用，应部署全光子交换机还是配备存储器的交换机？原则上，量子存储器可以提高端到端速率和

本工作得到美国陆军研究办公室 MURI 项目（项目编号 W911NF2110325）、QuantumCT 项目以及美国国家科学基金会项目（项目编号 CNS 2402862）的支持。P. Promponas（panagiotis.promponas@yale.edu）和 L. Tassiulas 任职于美国康涅狄格州纽黑文耶鲁大学电气与计算机工程系。L. Bacciottini、G. Vardoyan 和 D. Towsley 任职于马萨诸塞大学阿默斯特分校（美国马萨诸塞州阿默斯特）。P. Polakos 任职于 Einblick LLC（美国新泽西州马尔伯勒）。

保真度，但如果相干时间较短或先兆延迟较大，基于量子存储器的设计可能会失去这一优势，甚至不如更简单的全光子交换机。因此，架构优势取决于具体运行条件，而非绝对存在。本文通过将两种模型纳入统一的抽象框架，量化它们在现实硬件假设下可达到的速率—保真度权衡，从而明确这一选择。我们的主要贡献如下：

•我们为全光子交换机和配备量子存储器的交换机分别建立了数学上精确的模型，明确规定其控制策略、物理资源假设和运行逻辑。

• 我们推导了每种模型下可实现吞吐量和保真度的闭式表征，并明确量化了硬件参数对端到端纠缠性能的影响。在此基础上，我们提出了一种基准测试方法，为比较这两种架构提供统一基础，将底层硬件参数与网络级性能联系起来，并确定每种架构占优的运行区间。

• 数值评估量化了两种模型在速率与保真度之间的权衡，并展示了如何调节硬件和协议调节参数，以满足特定应用目标。

II. 预备知识

A. 系统描述

我们考虑一个星型拓扑量子网络，其中中央交换机节点连接至 N 个客户端节点。令 $\mathcal{N}$ 为节点集，令 $\mathcal{F} \triangleq \{(i,j): i<j,\ i,j\in\mathcal{N}\}$。我们将使用 $\mathcal{F}$ 以避免重复计算节点对，并将每个 $(i,j)\in\mathcal{F}$ 称为一个端到端流。因此，流的总数为 $F \triangleq \binom{N}{2} = |\mathcal{F}|$。交换机在客户端之间生成并分发纠缠量子比特对（贝尔对）。每个节点都配备量子存储器，用于存储量子比特，直至建立端到端纠缠。由于硬件限制（例如，它配备了 B个贝尔态分析器），交换机最多可并行执行 B 次贝尔态测量，以生成端到端贝尔对。我们假设采用时隙化运行。LLEG尝试、贝尔态测量和先兆的具体时序取决于模型，并将在 EGS 模型和配备存储器的交换机模型中分别说明。

B. 保真度与量子噪声模型

纠缠保真度量化所生成的态与目标贝尔态的接近程度，不同应用对保真度有不同要求。我们将含噪声的纠缠态建模为维尔纳态，这是一类单参数混合态族，通过维尔纳参数$w\in[0,1]$刻画不完美性。一对量子比特的维尔纳态表示为：

$$\rho_w = w\left|\Phi^+\right\rangle\left\langle\Phi^+\right| + (1-w)\frac{\mathbb{I}}{4},$$

其中，$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ 是其中一个参考贝尔态，$\mathbb{I}/4$ 是完全混合态，$w\in[0,1]$ 是维尔纳参数。这种态的保真度为 $F_w = \frac{3w+1}{4}$。当两个维尔纳参数分别为 w_1 和 w_2 的维尔纳态经过完美操作交换后，所得态的维尔纳参数就是它们的乘积 $w_1\cdot w_2$。即使实际的纠缠态并非维尔纳态，在适当的假设条件下，这种近似也能为保真度分析提供下界 [17]。

参数为 $q\in[0,1]$ 的退极化信道作用于维尔纳态时，只是将其维尔纳参数 w 替换为 $w\cdot q$。我们用参数为 q_{BSM} 的退极化信道来建模交换后的纠缠对的保真度下降。独立地，贝尔态测量尝试成功的概率由 p_{BSA} （针对 EGS 模型中的贝尔态分析器）或 p_{swap} （针对存储器辅助模型）刻画；这些参数影响端到端生成速率，而不影响输出态的保真度。我们还将存储器中的量子退相干建模为参数为 $e^{-\tau/T}$的退极化信道，其中 T 为量子存储器相干时间（假定对所有存储器相同），τ 为在存储器中停留的时间。

C. 链路级纠缠生成模型

我们将客户端与交换机之间生成并分发贝尔对的协议称为链路级纠缠生成（LLEG）；由此产生的贝尔对称为 LLE。尽管具体细节取决于硬件平台，但大多数实现都具有共同的结构：它们使用光源（例如，自发参量下转换（SPDC） [18]或基于量子点、色心或囚禁离子的发射体 [19] ）沿客户端–交换机链路发送纠缠光子。LLEG 本质上是概率性的，因此需要先兆。为区分这种先兆与在相距遥远的节点之间建立端到端纠缠所需的先兆，我们将其称为链路级预告。与硬件无关的抽象可由一组通用参数来描述。对于 i 与交换机之间的 LLEG，定义如下：

成功概率 p_i：一次 LLEG 尝试成功在 i 与交换机之间生成贝尔对的概率。

先兆开销 τ_{hrld}：成功建立 LLE 后（例如，节点中的一个存储器与交换机共享一对纠缠态），为在交换机中预示链路级纠缠（使其能够获知连接关系）所需的开销时间。节点侧可并行确认 LLE，且不会影响交换机的运行，因为节点最终会通过端到端纠缠先兆获知成功或失败。

维尔纳参数 w_i：与节点 i 成功建立的 LLE 相关联的维尔纳参数。因此，w_i 也涵盖了交换机等待链路级预告信号期间，参与此 LLE 的量子比特所经历的退相干。

调节参数 β：源的一个可控属性（例如，自发参量下转换的泵浦功率，或发射体 [17] 的亮度），会影响 p_i 和 w_i。提高 β 会提升 p_i，但由于多对发射，会降低 w_i。因此，p_i 和 w_i 是 β 的函数，分别记为 $p_i(\beta)$ 和 $w_i(\beta)$。

尝试周期 Δt**:** 给定链路上两次连续LLEG尝试开始之间的最小时间间隔。例如，对于每个光脉冲仅进行一次光子对生成尝试的SPDC 光源，$\Delta t = 1/f_{\mathrm{pulse}}$，其中 f_{pulse} 为源脉冲重复率（赫兹）。

复用度 S_i：可在客户端 i与交换机之间执行的并行LLEG尝试的数量。它涵盖空间复用或频分复用，例如使用多个光源或频分复用的SPDC 光源 [20]。我们将每一次此类并行尝试称为复用帧。

符合这一抽象的知名 LLEG 协议包括，例如，[1] 基于 SPDC 光源的三种方案（见 [22], ）、使用发射体的 Barret-Kok 协议 [19] ，以及文献 [17] 中描述的单击方案。后续章节将进一步讨论其中一些协议，因为我们可以根据所用的具体 LLEG 协议，进一步细化交换机模型。最后需要指出，并非所有这些 LLEG 协议都适用于下文描述的所有交换机模型。例如，某些协议（如 [22]中采用基于区块的尝试的中点源协议）可能要求交换机配备存储器，这与第 III 节中的全光子交换机模型不兼容。

III. 逻辑模型 1：全光子交换机

这一纠缠生成交换机（*EGS*）模型描述了一种全光子交换机，它不配备量子存储器，具有N 个客户端节点、多个纠缠源，以及一个由贝尔态分析器组成的交换机内部池 B，这些分析器可以重新绑定到端口对。贝尔态测量具有概率性，其结果会以预示的方式返回端点，但交换机在尝试进行贝尔态测量之前不会执行任何链路级预示。调度器的核心决策是每个贝尔态分析器监听哪个端口对，从而将传入光子路由至同一个贝尔态分析器。由于客户端节点可能拥有多个光源（因此具有复用度S_i），多个贝尔态分析器可以同时服务于同一客户端。类似的交换机模型见文献 [7], [23], [9], [10]。

A. 系统建模

交换机与客户端之间的 LLEG 协议 i 依赖于 S_i 进入交换机入口端口的独立光路（例如光纤模式）。时间被划分为持续时间为Δt的时隙。硬件校准将时隙分组为每个包含 T_{freeze} 个时隙的周期。每个这样的区间称为冻结周期，在此期间控制决策保持固定，仅在周期之间更新 [9]。我们假设周期边界处的校准时间为常数 T_{freeze}且可忽略；若该开销不可忽略，则后续推导的所有每秒速率应按相应的有效占空比进行缩放（即按处于冻结周期的时间比例缩放）。

我们将时隙持续时间设为 LLEG 尝试周期，并对先兆作出两项不同的建模假设。首先，交换机不会等待链路级 先兆后才尝试进行贝尔态测量：贝尔态分析器会在一个周期内分配给各节点对，并对入射光子盲目地进行贝尔态测量。其次，在端到端 层面，我们假设用户节点拥有足够多的本地存储器，可以缓存所有暂定的贝尔对半体，跨越多个时隙，直至经典贝尔态测量结果得到预示。本文仅关注交换机的资源约束。在此假设下，下一个时隙的开始不会因端到端先兆而延迟，且时隙持续时间不受经典通信时间的限制。[2]

控制决策： 我们假设交换机中的贝尔态分析器、每节点源以及物理链路均为同质。因此，调度器在冻结周期 e 内的决策完全由站点分配 变量 $x_{ij} \in \mathbb{Z}_{\geq 0}$ $(i,\ j) \in \mathcal{F}$，所描述，该变量表示在周期 e 内分配多少个贝尔态分析器来服务客户端对 $(i,\ j) \in \mathcal{F}$。可行性条件为：

$$\sum_{(i,j)\in\mathcal{F}} x_{ij} \leq B, \qquad \sum_{j<i} x_{ji} + \sum_{j>i} x_{ij} \leq S_i \quad (\forall i). \tag{1}$$

即，全局贝尔态分析器预算和每节点源预算。我们将由不等式 (1) 定义的凸集记为 $\mathcal{X}$。

逐链路LLEG成功： 根据第 II-C 节，存在多种 LLEG 协议。为说明 $p_i(\beta)$ 的计算，我们考虑一种中点源架构（位于节点与交换机之间），每个时隙进行一次单光子对尝试。令 $p_{\text{pair}}(\beta)$ 表示源成功产生两个纠缠光子的概率。我们预期，$p_{\text{pair}}(\cdot)$ 是关于 β 严格递增的函数（见第 II-C 节）。

当光子锁存到节点存储器且其伴随光子在贝尔态测量的接受范围内到达贝尔态分析器时，LLEG尝试才算成功。因此，LLEG尝试成功的概率为 $p_i(\beta) = p_{\text{pair}}(\beta)\left(\eta_{\text{m}} g_{\text{m}}\right)\left(\eta_{\text{sw}} g_{\text{sw}}\right)$. 其中，$\eta_{\text{m}}$表示半链路和通往节点存储器的耦合损耗，$g_{\text{m}}$ 表示存储器写入/门效率；η_{sw} 表示半链路损耗以及通往任意站点 b的路由损耗，g_{sw} 表示贝尔态测量输入端的时间/频谱/模式接受度。

端到端纠缠生成成功： 当 b两侧各恰有一次 LLE 时，若且唯若两个光源均成功且光子被捕获，时隙内才会发生一次符合尝试。因此，对于在周期 e为（i、j）提供服务的贝尔态分析器 b，

$$p_{ij}^{\text{e2e}}(\beta) \;=\; \left(p_i(\beta)\, p_j(\beta)\right) \xi^2\, p_{\text{BSA}}. \tag{2}$$

其中，ξ 是每个站点 b的探测器效率，p_{BSA}是光学贝尔态分析器的成功概率。下面我们总结此类交换机的运行方式：

时隙操作： ***(i)*** 周期分配：在一个周期开始时，选择满足不等式 (1) 的 $x_{ij}(e)$，并配置任意 $x_{ij}(e)$ 个站点，在该周期内为 $(i,\ j)$ 提供服务。***(ii)*** 尝试：每个节点 i 最多执行 S_i 次 LLE尝试，且成功概率相同 $p_i(\beta)$。***(iii)*** 贝尔态测量：对于每一对 $(i,\ j)$，贝尔态测量的成功概率（在同质性假设下）为$p^{\text{e2e}}(\beta) \triangleq p_{ij}^{\text{e2e}}(\beta)$。***(iv)*** 预示：成功后，贝尔态分析器会将经典结果发送给 $i,\ j$。我们假设下一个时隙在贝尔态测量（步骤 (iii)）之后立即开始。这意味着预示同时发生，与此同时，每个客户端节点会将一个贝尔对半体存储在一个

[1]离散尝试抽象对于脉冲源而言是精确的 [21]。对于连续波长源 [18], ，其光子对生成间隔时间近似服从速率为 λ的泊松分布（β），p_i（β）$= 1 - e^{-\lambda(\beta)\Delta t} \approx \lambda$（$\beta$）$\Delta t$ 对于 λ（β）$\Delta t \ll 1$。我们将窗口内的任何成功视为在时隙边界登记的伯努利试验，以适配离散设定。

[2]如果贝尔态分析器的速度慢于源，或者由于节点存储器稀缺，必须等待端到端预告后才能重复使用，那么只需将时隙持续时间设为这些时间中较长者即可。这只会按一个常数因子缩放有效交换频率，从而缩放所有每秒速率。关于在不采用和采用节点存储器充裕假设的情况下比较EGS架构的效能，将留待未来研究。

专用量子存储器中。关于 EGS 及其时隙操作的图示，请分别参见图 1(a) 和 1(c)。

B. 吞吐量区域

对于每一对 i，j，给定 x_{ij} 时，端到端纠缠对的期望数量为 $p^{\mathrm{e2e}}(\beta)\,x_{ij}$。我们关注的是可实现的长期每时隙服务速率 λ_{ij}，即利用这些可能的调度方案能够为每一对i，j 实现的速率。将这些速率收集为向量 $\lambda=(\lambda_{ij})_{i<j}$，并将吞吐量区域$\Lambda_{\mathrm{EGS}}$ 定义为所有此类可实现向量的集合。

回顾一下，由不等式 (1) 定义的凸集记为 $\mathcal{X}$。用图论术语来解释 $\mathcal{X}$ 会很方便。考虑完全图 $G=(\mathcal{N},\ \mathcal{F})$，其节点集为 $\mathcal{N}$，边集为 $\mathcal{F}=\{(i,\ j):\ i<j\}$。对于任意整数向量 $x=\{x_{ij}\}_{(i,j)\in\mathcal{F}}\in\mathbb{Z}^F_{\geq 0}$，其分量 x_{ij} 指定了分配给边（等价地，分配给点对）$(i,\ j)$ 的容量单位（或等价地，贝尔态分析器）的数量。约束 (1) 随即保证：(i) 此类单位的总数不超过预算 B；(ii) 对于每个节点 i，与节点 i 相连的单位总数不超过 S_i。在组合优化术语中，任何此类 $x\in\mathcal{X}$ 都是完全图上的一个带容量的 b-匹配，其顶点容量为 $b_i=S_i$，并附加一个大小为 B 个单位的基数约束。因此，凸包 $\mathrm{co}(\mathcal{X})$ 可视为一个带容量的 b-匹配多面体，并附加了这一全局预算约束（例如，[24]）。

令 $co(\cdot)$ 表示集合 $\mathcal{X}$ 的凸包，则集合Λ_{EGS} 由 $\Lambda_{\mathrm{EGS}}\equiv co(\mathcal{X})\cdot p^{\mathrm{e2e}}(\beta)/\Delta t$给出。因此，吞吐量区域由下式给出

$$\Lambda_{\mathrm{EGS}}=\left\{\frac{p^{\mathrm{e2e}}(\beta)}{\Delta t}\cdot\lambda:\lambda=\sum_{x\in\mathcal{X}}\delta_x x,\ \sum_{x\in\mathcal{X}}\delta_x=1,\ \delta_x\geq 0,\quad \text{for all } x\in\mathcal{X}\right\}. \tag{3}$$

集合 Λ_{EGS} 包含一个EGS所能实现的所有可能的端到端纠缠生成速率向量。为便于记法，我们去掉了对 β 的依赖。

最大允许总吞吐量： Λ_{EGS} 是由可行站点分配 $\mathcal{X}$（约束 (1)）所诱导的 $\mathbb{R}^F$ 中的高维多面体。为从总吞吐量的角度量化Λ_{EGS} 的“大小”，我们考虑每秒的最大端到端总速率：

$$R^{\mathrm{sec}}_{\mathrm{EGS}}\ \triangleq\ \max_{\lambda\in\Lambda_{\mathrm{EGS}}}\sum_{(i,j)\in\mathcal{F}}\lambda_{ij}\ =\ \max_{\lambda\,\in\,\mathrm{co}(\mathcal{X})}\frac{p_{\mathrm{e2e}}(\beta)}{\Delta t}\sum_{(i,j)\in\mathcal{F}}\lambda_{ij}.$$

换言之，$R^{\mathrm{sec}}_{\mathrm{EGS}}$ 表示吞吐量区域中L1范数最大的向量。显然，该指标的局限在于它无法给出该区域的整体图景。本节的主要定理如下，并为 $R^{\mathrm{sec}}_{\mathrm{EGS}}$提供了一个闭式公式。

定理 1（最大总吞吐量的闭式）。对于采用 B 贝尔态分析器且节点复用度为 $\{S_i\}$ 的 *EGS* 模型，最大期望端到端总成功率为

$$R^{\mathrm{sec}}_{\mathrm{EGS}}\ =\ \frac{p_{\mathrm{e2e}}(\beta)}{\Delta t}\,E^{\max}_{\mathrm{EGS}},\quad \textit{where}$$

$$E^{\max}_{\mathrm{EGS}}\ =\ \min\left\{B,\ \left\lfloor\tfrac{1}{2}\sum_i S_i\right\rfloor,\ \sum_i S_i\ -\ S_{\max}\right\}, \tag{4}$$

和 $S_{\max}\triangleq$ 最大值$_i\,S_i$.

我们现在给出一个引理，为 $E^{\max}_{\mathrm{EGS}}$ 提供直观理解，并且它对于定理 1 的证明以及后续分析也很有用。

引理 1。设 $\mathcal{X}$ 为满足不等式 (1) 的积分多面体，其中 $x=\{x_{ij}\}_{(i,j)\in\mathcal{F}}$。那么，可行分配的最大基数为 $E^{\max}_{\mathrm{EGS}}$，如(4)所示：

$$\max_{x\,\in\,\mathcal{X}}\sum_{(i,j)\in\mathcal{F}}x_{ij}\ =\ E^{\max}_{\mathrm{EGS}}.$$

引理 1 和定理 1 的证明均见附录（第 VII-A 节和第 VII-B 节）。

C. 端到端纠缠保真度

除了吞吐量分析之外，我们现在还将推导任意一对 i，之间按照第 II-C 节所述共享的端到端纠缠态的保真度闭式表达式。

回顾一下，给定一个 LLEG 协议，$w_i(\beta)$ 表示节点i 成功实现一次 LLE 时关联的维尔纳参数（LLEG 和链路级先兆中的交换机相关噪声）。因此，根据第 II-B 节，节点 i、j 之间端到端纠缠的维尔纳参数为

$$w^{\mathrm{e2e}}_{\mathrm{EGS}}(\beta)\ \triangleq\ w_i(\beta)\,w_j(\beta)\,q_{\mathrm{BSM}}\ e^{-2L/(v_fT)},$$

其中，L表示链路长度，v_f表示链路中的光速。最后一项 $e^{-2L/(v_fT)}$ 对应于两个端节点的纠缠量子比特的存储器退相干所关联的维尔纳参数，其时间范围从交换机获知 LLEG 是否成功开始，到端到端纠缠发出预示为止。该表达式假设客户端仅在端到端纠缠发出预示后才使用端到端纠缠对。也可以采用其他模型（例如带后选择的制备与测量），此时相关的计时／退相干时间窗口将有所不同。

作为示例，我们进一步通过假设采用一种 LLEG 协议来具体化这一表达式：每条链路上都设有中点源，它们生成初始 Werner 参数为 $w_0(\beta)$ 的成对粒子（该参数是 β 的递减函数）。回想一下，τ_{hrld}表示与链路级先兆相关的经典通知延迟，即交换机获知客户端–交换机链路尝试是否成功所需的时间。在 EGS 模型中，交换机不会等待收到该信息后再尝试进行贝尔态测量；相反，贝尔态测量会在不知情的情况下触发，随后收到的链路级预示信号则用于判断该次贝尔态测量尝试是有效的（两条链路均成功），还是白费的（至多一条链路成功）。由于我们将源置于链路中点，因此链路尝试成功后，节点会向交换机发送一个经典比特

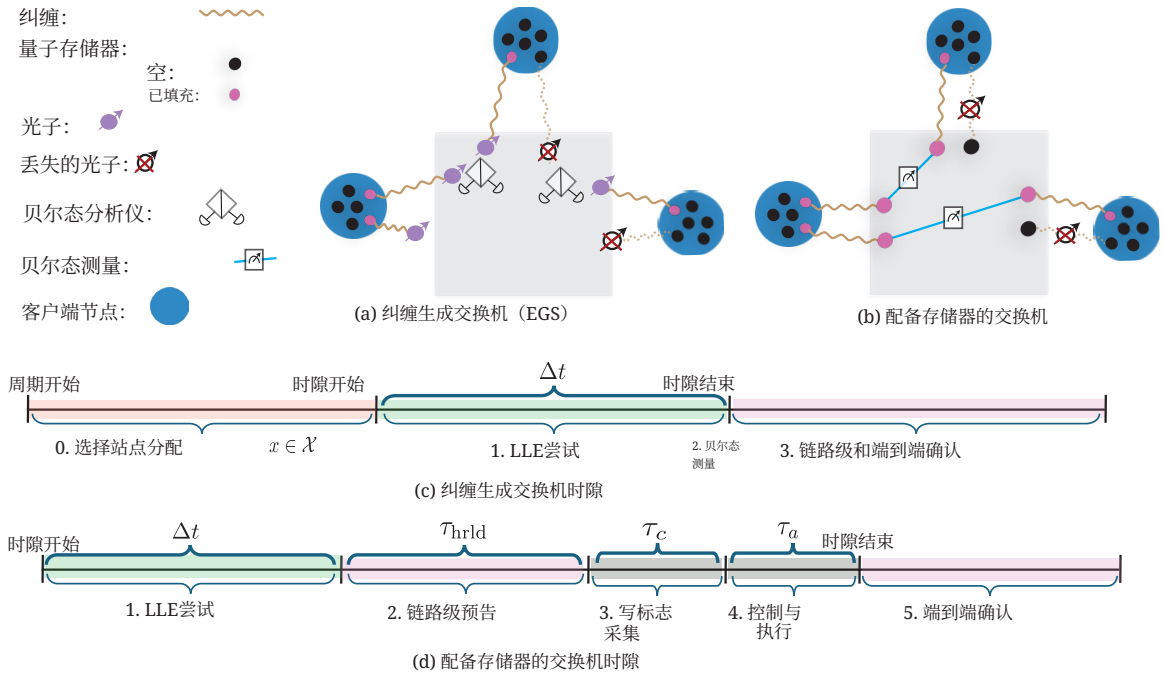


图 1. 图示：(a) EGS，(b) 配备存储器的交换机，(c) EGS 的时隙操作，以及 (d) 配备存储器的交换机的时隙操作。(a) 和 (b) 中的两种交换机对每个 i 和 $B = 2$ 均使用 $S_i = 2$。对于 (b) 中配备存储器的模型，每个节点 i 均使用 $M_i = 2$。

以报告链路级预示信号。因此，对于每一对 $\tau_{\text{hrld}} = L/v_f$ 和 $w_i(\beta) = w_j(\beta) = w_0(\beta)e^{-\tau_{\text{hrld}}/T}$ i, j，因为在这种架构中，只有节点中的状态存储在存储器中，而我们会对 LLE 进行预示（见第 III-A 节）。

交换后的纠缠对所对应的端到端保真度为 $F_{\text{e2eEGS}}(\beta) = \frac{1+3w^{\text{e2e}}{}_{\text{EGS}}(\beta)}{4}$。~~注意，~~ 端到端保真度 $F^{\text{e2e}}_{\text{EGS}}(\beta)$ 随着 β 按 $w_0(\beta)$ 的平方变化。

IV. 逻辑模型 2：带内部存储器的量子交换机（先宣告后交换控制）

本节对近期的带存储器交换机进行建模和分析，并量化最少存储的益处：仅将光子量子比特存入存储器，直到交换机获知时隙的连接关系并作出明智的交换决策。类似模型可见于 [11], [12], [15]–[25]。这种基于连接关系的控制，与 EGS 相比减少了浪费的贝尔态测量尝试，并避免了 EGS 中概率性光学贝尔态分析器造成的瓶颈；但由于额外的存储时间，它会加剧退相干并延长时隙。

A. 系统建模

该交换机每个时隙可执行 B 次贝尔态测量，并配有 $M \geq 2B$ 个量子存储器，可将单量子比特态存储一个时隙。交换机为每个客户端 i 专门分配 $M_i = S_i$ 个量子存储器（因此 $\sum_i M_i = M$）：LLEG 协议的每个复用帧在其入口信道上都有一个专用的交换机存储器。[3] 这简化了阐述，并使该模型可视为 EGS 的直接推广：每个 LLE 都可以在交换机处短暂缓存，直到确定连接关系。这与 EGS架构模型类似，我们假设用户节点有足够多的存储器，可缓存所有暂存的贝尔对半体，直至经典贝尔态测量结果得到预示（即在端到端纠缠先导期间）。在此假设下，贝尔态测量完成后，下一个时隙可以立即开始。

时序与时隙级操作。令 Δt_{mem} 表示一个时隙的持续时间。为说明这一持续时间，我们假设每个复用帧对应一个位于链路中点的源，该源尝试在每个脉冲中发射一对光子，而每个节点存储器则尝试将其锁存。Δt_{mem}（以及相应的量子比特存储时间）由四个串行时延决定：(i) 一次 LLEG 尝试的持续时间，Δt；[4] (ii) 链路级预告时间的经典开销，τ_{hrld}；(iii) 写标志+链路级预告采集，τ_c；以及 (iv) 控制与执行，τ_a：

$$\Delta t_{\text{mem}} = \Delta t + \tau_{\text{hrld}} + \tau_c + \tau_a.$$

在这种情况下，$\tau_{\text{hrld}} = \frac{L}{v_f}$，其中 L 表示节点—交换机光纤长度，v_f 表示物理链路中的光速。此外，$\Delta t = 1/f_{\text{pulse}}$。基于区块尝试的 LLEG 协议的明确扩展将在第 V-A 节中给出。

$\tau_a > 0$ 意味着交换机在获知连接关系后需要时间来优化调度方案。因此，τ_a 取决于调度算法的复杂度。在 EGS 中，我们

[3]若 $M_i \neq S_i$，则每个时隙最多只能实现 $\min\{M_i, S_i\}$ 个链路级纠缠；等价地，在我们的分析中，可将 M_i 替换为 $\min\{M_i, S_i\}$（瓶颈）。

[4]这一 Δt 可以短于一次完整LLEG尝试的持续时间，因为发射可以跨时隙流水线化。具体而言，在时隙中交换机忙于经典处理/决策开销的那一部分，光源不必空闲：它们最多可提前一个单向传播时间发射，从而使相应光子在下一时隙开始时到达。

隐式地假设 τ_a 可以忽略不计，因为我们对 $T_{\text{freeze}}\Delta t$采用固定的控制决策，无需在每个时隙都更改它。虽然在该模型中原则上可以这样做，但这需要为每种连通性模式预先计算不同的控制动作。由于配置数量会随着客户端数量呈指数级增长，因此这种做法无法扩展。时隙 t 内的操作流程如下

（*i*）*LLEG* 尝试：每个 LLEG 协议最多执行S_i 次尝试（*ii*）连通性获取：交换机接收先导消息，并将获取到的连通性存储在内部数据结构中。（*iii*）匹配：控制器计算并实施存储器到 BSM 的映射。（*iv*）*BSM*：在时隙结束前，最多对B 个选定的存储器对进行测量。请注意，我们假设节点中有无限多个存储器，因此这些 BSM 结果的经典先导可以在之后进行，并且可与下一时隙重叠。如果节点存储器有限，则可以通过增加有效时隙持续时间来吸收端到端纠缠先导所产生的延迟。配备存储器的交换机示意图及其时隙操作分别见图 1(b) 和图 1(d)。

LLEG 成功：在链路中点光源、同质存储器的假设下，LLEG 成功概率为$p_i(\beta) = p_{\text{pair}}(\beta)\,\eta_m^2\,g_m^2$. 此处推导针对上文所述的链路中点单次尝试 LLEG； [22] 的基于块的尝试 LLEG 会得到不同的表达式，推导见第 V-A 节。给定 $p_i(\beta)$，时隙 t 中的连通性 $C_i(t)$满足：

$$C_i(t) \sim \text{Binomial}(M_i,\, p_i(\beta))\,. \tag{5}$$

对于每个客户端 i，序列 $\{C_i(t)\}_t$ 是独立同分布的。

控制变量：学习到连接关系后，交换机观测 $C(t) = \{C_i(t)\}_{i\in\mathcal{N}}$，并选择一个跨客户端索引的已占用存储器匹配，即$y_{ij}(t) \in \mathbb{Z}_{\geq 0}$，$(i,\ j) \in \mathcal{F}$，同时满足每个节点的供给约束和每站预算，$\sum_{j<i} y_{ji}\ t\ + \sum_{j>i} y_{ij}\ t\ \leq ( ) ( ) ( ) ( ) C_i\ t\ \ \forall i\ ,$ $t\ \leq\ B.$ 以及 $\sum_{i<j} y_{ij}( )$

每次尝试的交换成功率：在为客户端对 $(i,\ j)$ 调度$y_{ij}(t)$次交换的条件下，每次尝试以p_{swap}的概率成功。对于贝尔态分析器，$p_{\text{swap}} = \xi^2\,p_{\text{BSA}}$。然而，由于在该模型中，贝尔态测量是在存储量子比特之间执行的，并不一定通过线性光学实现，因此交换可以是确定性的（即 $p_{\text{swap}} = 1$），这是存储器辅助交换机的一项优势。因此，在时隙 t 中为客户端对 $(i,\ j)$ 产生的端到端贝尔态数量为$\mu_{ij}(t) \sim$ 二项分布$(y_{ij}(t),\ p_{\text{swap}})$。

B. 吞吐量区域

设 $\mathcal{N}$ 为节点集，$\mathcal{F} \triangleq \{(i,\ j)\colon\ i<j,\ \ i,\ j\in\mathcal{N}\}$。连通状态空间为 $\mathcal{C} \triangleq \prod_{i\in\mathcal{N}}\{0,1,\ldots,M_i\}$。向量 $c\in\mathcal{C}$ 是一个连通状态。随机变量 $C = C(t) = \{C_i(t)\}_{i\in\mathcal{N}}$ 按照式（5）对连通状态进行采样。由于对于每个客户端 i，序列 $\{C_i(t)\}_t$ 都是独立同分布的，因此当时隙无关紧要时，可以省略 t。概率质量函数可分解为

$$\pi(c) \triangleq \mathbb{P}\{C=c\} = \prod_{i\in\mathcal{N}} \binom{M_i}{c_i} p_i(\beta)^{\,c_i}\big(1-p_i(\beta)\big)^{M_i-c_i}.$$

给定状态 c，交换机获知已成功与节点i 建立链路级纠缠的存储器的可用供应 c_i，并可调度至多 B 次贝尔态测量。令

$$\mathcal{Y}(c) \triangleq \Big\{ y\in\mathbb{Z}_{\geq 0}^{F} \ :\sum_{j<i} y_{ji} + \sum_{j>i} y_{ij} \leq c_i\ ,\forall i,\ \sum_{(i,j)\in\mathcal{F}} y_{ij} \leq B\Big\}$$

为状态 c 下满足全局预算约束（预算为 B 个单位，见第 III-B 节）的容量约束 b-匹配集合。向量 $y\in\mathcal{Y}(c)$ 是在给定连接关系 c时贝尔态测量的一种有效调度方案。其凸包 $\text{co}\,\mathcal{Y}(c) \subset \mathbb{R}_{\geq 0}^{F}$，涵盖了以 c 为条件的所有随机化时分调度方案。

平稳因果调度规则（可以是随机化的）是一个映射 $v:\ \mathcal{C}\ \to\ \text{co}\,\mathcal{Y}(c)$；对于每个观测到的状态 c，该规则都会选取可行匹配的一个凸组合。等价地，$v(c)$ 是该时隙内在对 $\mathcal{Y}(c)$ 进行随机化时的期望配对向量。

我们关注所有可实现的长期端到端纠缠速率向量 λ，并将其汇集为吞吐量区域 Λ_{mem}（单位为端到端对/秒）：

$$\Lambda_{\text{mem}} = \Big\{\lambda\in\mathbb{R}_{\geq 0}^{F}:\ \exists\, v:\mathcal{C}\to\text{co}\,\mathcal{Y}(\cdot) \quad \text{s.t. } \lambda\ =\ \frac{p_{\text{swap}}}{\Delta t_{\text{mem}}}\sum_{c\in\mathcal{C}}\pi(c)\,v(c)\Big\}. \tag{6}$$

最大允许总吞吐量：令 $E_{\text{mem}}^{\max}(c)$ 表示状态 c 下交换的最大数量，即

$$E_{\text{mem}}^{\max}(c) \triangleq \max_{y\,\in\,\mathcal{Y}(c)} \sum_{(i,j)\in\mathcal{F}} y_{ij}.$$

为求出 $E_{\text{mem}}^{\max}(c)$的闭式，我们通过对每个节点 i将 S_i 替换为 c_i，来使用引理 1，

$$E_{\text{mem}}^{\max}(c) = \min\Big\{\ B,\ \ \Big\lfloor \tfrac{1}{2}\sum_i c_i\Big\rfloor,\ \sum_i c_i - \max_i c_i\Big\}.$$

因此，最大期望总端到端速率为

$$\begin{aligned} R_{\text{mem}}^{\text{sec}} &\triangleq \max_{\lambda\in\Lambda_{\text{mem}}} \sum_{(i,j)\in\mathcal{F}} \lambda_{ij} \\ &= \frac{p_{\text{swap}}}{\Delta t_{\text{mem}}}\ \max_{v(\cdot)} \sum_{c\in\mathcal{C}}\pi(c) \sum_{(i,j)\in\mathcal{F}} v_{ij}(c) \\ &\text{s.t.}\quad v(c)\in\text{co}\,\mathcal{Y}(c)\quad \forall c\in\mathcal{C}. \end{aligned}$$

由于最大化过程可按 c解耦，因此

$$R_{\text{mem}}^{\text{sec}} = \frac{p_{\text{swap}}}{\Delta t_{\text{mem}}}\ \sum_{c\in\mathcal{C}}\pi(c)\ \max_{y\in\text{co}\,\mathcal{Y}(c)}\ \sum_{(i,j)\in\mathcal{F}} y_{ij}.$$

由于目标函数是线性的，在$\text{co}\,\mathcal{Y}(c)$上，最大值在某个极点处取得，因此

$$\max_{y\in\text{co}\,\mathcal{Y}(c)} \sum_{(i,j)\in\mathcal{F}} y_{ij} = \max_{y\in\mathcal{Y}(c)} \sum_{(i,j)\in\mathcal{F}} y_{ij} \ \triangleq\ E_{\text{mem}}^{\max}(c),$$

因此

$$R_{\text{mem}}^{\text{sec}} = \frac{p_{\text{swap}}}{\Delta t_{\text{mem}}}\ \sum_{c\in\mathcal{C}}\pi(c)\,E_{\text{mem}}^{\max}(c).$$

求得该架构中最大期望总速率的闭式表达式，要比 EGS（参见定理 1）困难得多

不过，由于 $R_{\mathrm{mem}}^{\mathrm{sec}}$ 是一个期望值，我们可以通过对连通性 c 进行采样，利用蒙特卡洛方法估算它。

C. 端到端纠缠保真度

回顾一下，给定一个 LLEG 协议，$w_i(\beta)$ 表示节点 i 成功执行 LLEG 时对应的维尔纳参数（ LLEG 以及交换机相关噪声下的链路级预告）。因此，节点 i,j 之间端到端纠缠的维尔纳参数为

$$w_{\mathrm{mem}}^{\mathrm{e2e}}(\beta) \triangleq w_i(\beta)\, w_j(\beta)\, q_{\mathrm{BSM}}\; e^{-4(\tau_c+\tau_a)/T}\; e^{-2L/(v_f T)}.$$

项 $e^{-4(\tau_c+\tau_a)/T}$ 对应于链路级预告之后、贝尔态测量之前，四个相关量子比特所经历的存储器退相干所对应的维尔纳参数；而 $e^{-2L/(v_f T)}$ 则对应于从贝尔态测量到端到端纠缠被预示期间，两个端节点量子比特的退相干。因此，$F_{\mathrm{meme2e}}(\beta) = \frac{1+3\, w^{\mathrm{e2e}}{}_{\mathrm{mem}}(\beta)}{4}$.

作为演示，我们再次假设链路中点光源每个光脉冲尝试产生一对光子，且每个节点存储器都尝试锁存最先到达的光子。回顾一下，τ_{hrld} 表示成功建立 LLEG 后，在交换机中预示该过程所需的额外时间。由于我们将光源放置在链路中点，LLE 建立后，需要从节点向交换机传输一个经典比特，以便在交换机中进行链路级预告。因此，对于每一对 i,j ，$\tau_{\mathrm{hrld}} = L/v_f$ 以及 $w_i(\beta) = w_j(\beta) = w_0(\beta)\, e^{-2\tau_{\mathrm{hrld}}/T}$ ，因为现在节点和交换机中都配备了存储器。

V. 数值结果

我们考虑一个反映近期硬件参数的默认场景。每当我们偏离这组参数时，都会明确说明发生变化的参数。

基线场景：我们考虑 $N = 6$个客户端节点，复用度为 $S_i = 3$ ，并允许每个时隙最多执行 $B = 8$ 次并发贝尔态测量。对于 EGS 模型中的贝尔态分析器，探测器的效率为 $\xi = 0.90$ 和 $p_{\mathrm{BSA}} = 0.50$。对于存储器辅助模型，我们取 $p_{\mathrm{swap}} = 1$。光纤衰减为每/千米 $\alpha = 0.2$ dB，节点至交换机距离为$L_{\mathrm{node}\to\mathrm{sw}} = 1$ 千米，因此每条半链路（0.5 千米）的透射率为 $\eta_{\mathrm{fiber},1/2} = 10^{-\alpha(0.5)/10} = 0.98$，其中 $g_{sw} = g_m = 0.85$。对于 LLEG 协议，我们使用链路中点的 SPDC 光源：采用 [26], 中的 SPDC 示例，其中产生 n 对符合事件的概率为 $p(n) = (n+1)\,\beta^n\,(\beta+1)^{n+2}$。在此模型下，光源发射光子的概率为 ~~$\Pr(n \geq 1) = \beta(\beta+2)\,(\beta+1)^2$~~，由此得到的初始保真度为 $F_0(\beta) = \frac{3w_0(\beta+1)}{4} + \frac{\Pr(n=1)}{4}\Pr(n\geq 1) + \frac{(\beta+1)(\beta+2)}{2}$。我们设置 $\beta = 0.03$，~~以获得 $F_0(\beta) = 0.96$。假设~~ $v_f = 2.0\times 10^8$ 米 / 秒，得到$\tau_{\mathrm{hrld}} = 5$ μ秒，其中 $\tau_c = 2$ μ 秒、 $\tau_a = 3$ μ 秒，源脉冲速率为 $f_{\mathrm{pulse}} = 10$ 兆赫兹。使用这些 SPDC 光源时，我们假设 EGS架构的 LLEG 协议采用第 III 节所述的单次尝试协议，而存储器辅助架构采用第 V-A 节详细描述的最优基于区块的 LLEG 协议。对于与保真度相关的参数，我们采用 $T = 500$ μ 秒和 $q_{\mathrm{BSM}} = 0.97$。我们使用包含 10^5 个样本的蒙特卡洛方法近似随机变量。

A. 配备存储器的模型中的基于区块的 *LLEG* 协议

本节中，我们研究基于存储器的模型中端到端纠缠率、端到端保真度与用户效用之间的权衡。我们利用所提出的框架，展示硬件设计者如何在基于区块尝试的 LLEG 协议中优化区块大小 K。

基于区块尝试的 LLEG 协议：当 LLEG 协议所用的光子源支持较高的脉冲速率 $f_{\mathrm{pulse}} \gg 1/\Delta t_{\mathrm{mem}}$ 时，在一个时隙开始时连续进行 K 次 LLEG尝试，并且仅将第一个成功的光子载入交换机的量子存储器 [22], 会很方便。通过经典先导，在区块结束时协调载入的是哪个光子；要取得成功，双方必须匹配。将此机制纳入我们的模型无需进行结构性改动，只需根据区块大小调整量子比特存储时间，并提高 LLEG尝试成功的概率 p_i（β）。

假设采用链路中点光源，时隙时长为

$$\Delta t_{\mathrm{mem}} = (K-1)/f_{\mathrm{pulse}} + \tau_{\mathrm{herald}} + \tau_c + \tau_a. \quad (7)$$

K 越大，延迟越高，但 LLEG 成功概率也会提高。当且仅当该块中第一个非空时隙单元传输了两个光子时，才会成功。设每个时隙单元的结果为 $q_2 = p_{\mathrm{pair}}(\eta g_m)^2$, $q_1 = p_{\mathrm{pair}}\eta g_m(1-\eta g_m)$ $q_0 = 1 - p_{\mathrm{pair}}(2\eta g_m - (\eta g_m)^2)$（分别表示两个光子均成功、恰有一个光子成功以及两个光子均未成功）。对于 K 个相互独立的时隙单元，第一个非空结果传输两个光子的概率（即 LLEG 成功的概率）为

$$p(\beta,K) = \sum_{t=1}^{K} q_0^{t-1}\, q_2 = q_2\, \frac{1-q_0^K}{1-q_0} = \frac{\eta g_m}{2-\eta g_m}\left[1 - \left(1 - p_{\mathrm{pair}}(\beta)\cdot(2\eta g_m - (\eta g_m)^2)\right)^K\right].$$

端到端纠缠保真度也会发生变化（参见 IV-C 节）。具体而言，随 K 变化的是成功的量子比特在交换机收到链路级预示之前于存储器中停留的时间，进而影响退相干程度。设 $T_{\mathrm{succ}} \in \{1,\ldots,K\}$ 表示该块中第一个结果为“非空且良好”的时隙单元的索引（即该时隙单元产生了最终用于构成链路级贝尔对的两个光子）。与上文计算成功概率时一样，第一个非空时隙单元结果为 t且传输两个光子的事件，其概率为 $\mathbb{P}\{T_{\mathrm{succ}} = t,\ \text{成功}\} = q_0^{t-1}\, q_2, \quad t = 1,\ \ldots$ $\ldots K$，总成功概率为 $p(\beta,\ K) = \sum_{=}^{K} q^{t-1} q_2$. 因此，

$$\mathbb{P}\{T_{\mathrm{succ}} = t \mid \text{success}\} = \frac{q_0^{t-1} q_2}{p(\beta,K)}, \qquad t = 1,\ldots,K.$$

在中点光源假设下，链路 i 上一次成功的 LLEG 对应于对可能成功时隙单元 t 的混合，每个时隙单元都会导致不同的链路级前存储时间。因此，基于分块的协议中，链路 i 上一次成功的 LLEG 所对应的维尔纳参数为

$$w_i(\beta,K) = \frac{w_0(\beta)}{p(\beta,K)} \sum_{t=1}^{K} q_0^{t-1} q_2\, \exp\!\left(-\frac{2}{T}\left(\tau_{\mathrm{hrld}} + (K-t)\Delta t\right)\right).$$

将其代入端到端表达式（见第 IV-C 节）后，基于分块的 LLEG 协议下节点 i与 j 之间共享对的维尔纳参数变为 $w_{\text{mem block}}^{\text{e2e}}(\beta, K) \triangleq w_i(\beta, K)\, w_j(\beta, K)\, q_{\text{BSM}}\, e^{-4(\tau_c+\tau_a)/T} e^{-2L/(v_f T)}$.

最大聚合吞吐量与 K 和 L： 图 2 展示了 $\hat{R}_{\text{mem}}^{\text{sec}} \triangleq R_{\text{mem}}^{\text{sec}}/F$ 随 K 和距离 L 的变化等高线。在面板 (a) 中，采用默认值 $\beta = 0.03$ 和 $f_{\text{pulse}} = 10$ 兆赫时，较小的 L 有利于使最佳尝试次数达到 $K = 30$（红色虚线矩形）。随着 L 增大，链路成功概率下降，因此需要更大的 K 才能维持较高的端到端速率；然而，进一步增大 K 会使端到端速率变得次优，因为根据式 (7)，时隙持续时间会增加，从而降低尝试速率。随着 L 持续增大，速率会下降，而改变 K 带来的收益也会逐渐减少。在面板 (b) 中，将 β 增大到 0.15 会使图中颜色变亮，并将最优值移向较低的 K：更高的单光子对成功率使我们能够降低 K，从而缩短时隙并提高速率。在面板 (c) 中，降低光源速度（$f_{\text{pulse}} = 1$ 兆赫）也会降低最佳 K，因为每增加一次尝试，时隙持续时间的增幅都会更大（见式 (7)）。

保真度热图与 K 和光源速率： 我们考察端到端保真度随 K 和 f_{pulse}变化的情况。图 3 中的热图显示，对于高速光源，保真度对 K不敏感；因此，可以选择 K ，仅用于优化速率。相比之下，对于较低速率的光源，增加 K 会降低保真度。原因在于，每次额外尝试都会延长时隙（参见公式 (7)），从而增加存储器存储时间并导致退相干。一般而言，提高 K 并不会改善保真度，因此设计问题是速率–保真度权衡：仅当速率增益足以弥补保真度损失时，才增加 K。

通过效用函数实现端到端速率/保真度权衡： 为了研究端到端纠缠速率R 与保真度 F之间的联合权衡，我们采用如下形式的比例公平效用函数：$U = \log\big(R\,Q(F)\big)$，遵循 [17]。我们考虑针对应用的三种 $Q(F)$选择，因此也得到 U：(i) 可蒸馏纠缠，U_{DE}，其中 $Q(F) = D_H(F)$，即哈希产率（渐近单向可蒸馏纠缠）；(ii)密钥效用，U_{SKF}，其中 $Q(F) = S_{\text{BB84}}(F)$，即 BB84 秘钥比例；以及 (iii)负性效用，U_{NGT}，其中$Q(F) = \max\{F - \frac{1}{2}, 0\}$，即超过可分阈值的经缩放纠缠负性。这些效用函数提供了一个兼顾吞吐量与质量的单一目标，并支持在公平基础上比较不同的硬件/模型选择。由于U_{DE} 和 U_{SKF} 仅在高保真度（$F \gtrsim 0.85$）下才具有相关性，而默认场景产生 $F_{\text{mem}}^{\text{e2e}} \in [0.3, 0.8]$（图 3），因此下文采用 $L = 0.1$千米。

图 4 绘制了 U_{DE}，U_{SKF}，U_{NGT} 随区块大小 K 的变化。为便于比较纵轴数值，我们将效用值归一化到 0 至 1。三者均呈现单峰、近似凹形的依赖关系，因此存在最佳区块大小 $K^\star$。$K^\star$ 会随效用（即目标应用）和硬件参数而变化，从而能够针对量子交换机的参数进行面向应用的调优。在当前设定下，三种效用均保持最优的 $K^\star \in [4, 10]$。

B. EGS 与存储器交换的比较

本节比较 EGS 与配备量子存储器的交换机。我们的目标是确定随着参数区间变化，哪种架构占优，并为硬件与模型协同设计提供指导。尽管结果针对所选参数区间，但我们的框架具有普适性，也可用于比较其他感兴趣的参数。我们采用两种互补视角：（i）通过扫描调优参数 β 得到的（R, F）保真度–速率前沿；（ii）面向应用的负性效用 $U_{\text{NGT}} = \log\big(R\max\{F - \frac{1}{2}, 0\}\big)$，它将每个工作点压缩为一个与纠缠实用性相符的单一评分。使用内存模型时，我们会针对每种设定选取最优区块大小 $K^\star$ —— $K^\star(\beta) \in \arg\max_K U_{\text{NGT}}(K, \beta)$。因此，$K^\star$ 取决于具体情况。

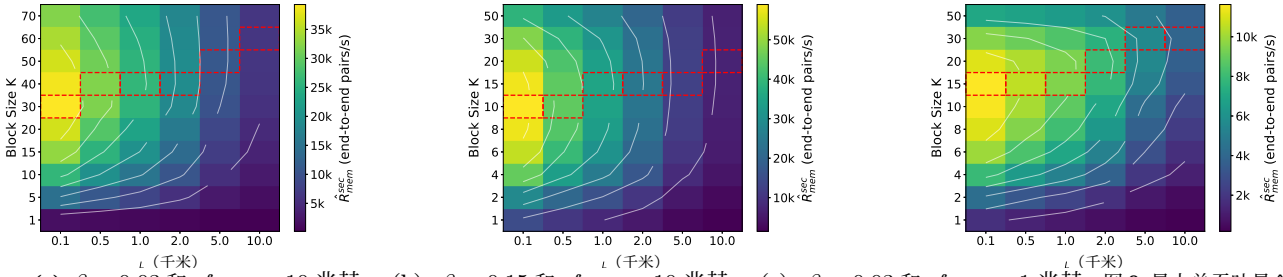
1) 保真度–速率前沿与负性对比 β：我们扫描参数 $\beta \in [0, 0.15]$，并计算 EGS 和配备量子存储器的交换机可实现的归一化端到端速率 $\hat{R}_{\cdot}^{\text{sec}}$及保真度 $F_{\cdot}^{\text{e2e}}$。

从图 5(a) 可见，曲线描绘了 (R, F) 随 β变化的集合。在较大的 β范围内，EGS 的速率和保真度均更高。然而，在中间的高保真度区间（$0.73 \lesssim F \lesssim 0.9$）内，存储器模型更具优势：对于相同的目标保真度，它能实现更高的速率。对于极高的保真度（$F \gtrsim 0.9$），由于交换机存储器中的退相干限制了该参数设置下可达到的保真度，存储器模型无法达到 EGS 曲线。实际应用中，可以根据所需保真度画一条水平线，并从图 5(a) 中读取可实现速率。图 5(b) 中的负性效用表明，对于较大的 β（对应图 5(a) 中较低的保真度），EGS 可获得更高的 U_{NGT}；而对于较小的 β ，在高保真度范围内，经过效用优化区块大小的存储器模型可达到更高的 U_{NGT} 。综合这两幅图可知，模型的优劣取决于所处区间。

2) 区间敏感性：我们改变关键参数，观察 EGS 与量子存储器交换机之间的优劣顺序如何变化。为使每个工作点都对应一个统一且可比较的评分，我们绘制跨越 β 的负性效用差值$\Delta U_{\text{NGT}} \triangleq U_{\text{NGT}}^{\text{mem}} - U^{\text{EGSNGT}}$，并根据相应的（$R$, F）图（如图 5(a) 所示），标出各情况下配备存储器的架构占优的保真度区间。为节省篇幅，省略（R, F）图。

光源速率（图 6(a)）。相较于基线，较慢的光源会使 ΔU_{NGT} 上升。配备存储器的交换机占优，因为时隙不再构成瓶颈，而且连接性学习在保持高保真度的同时提升了速率；对于 $f_{\text{pulse}} = 10$kHz，这意味着其在 $0.65 \lesssim F \lesssim 0.92$以上占优。使用快速光源（例如 1GHz）时，所有 β 条件下的 $\Delta U_{\text{NGT}} < 0$（EGS占优）；内存模型会浪费许多纠缠对，使其一直等待先兆。不过，这种优势假设节点能够在等待先兆期间存储所有状态，而在高 f_{pulse}条件下这可能不切实际。

长度（图 6(b)）。增大 L有利于配备存储器的交换机，并且其占优的 β区间也会扩大。更短的先兆路径可减少退相干并缩短有效时隙



(a) $\beta = 0.03$ 和 $f_{\text{pulse}} = 10$ 兆赫。（b） $\beta = 0.15$ 和 $f_{\text{pulse}} = 10$ 兆赫。（c） $\beta = 0.03$ 和 $f_{\text{pulse}} = 1$ 兆赫。图 2. 最大总吞吐量 $\hat{R}^{\text{sec}}_{\text{mem}}$ 随区块大小 K 和链路长度 L的变化；红色虚线矩形标示了每个 L对应的最优 K 。

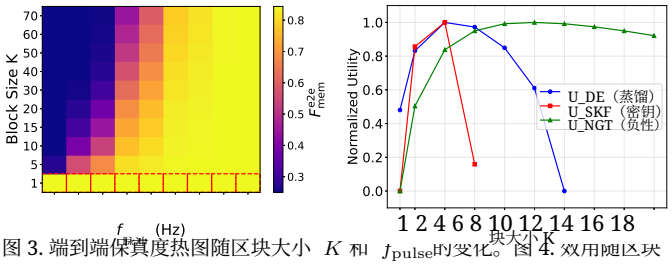
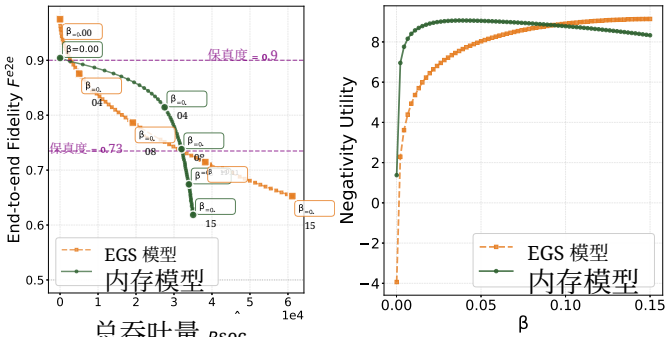


图 3. 端到端保真度热图随区块大小 K 和 f_{pulse}的变化。图 4. 效用随区块大小 K的变化。



(a) 相对于 β的保真度–速率前沿。(b) β对应的负性效用。图 5. EGS 与配备存储器的交换的基线比较。

持续时间，从而提升存储器方案的速率和保真度；在 $L = \{0.1, 0.01\}km$ 中，当 $0.66 \lesssim F \lesssim 0.93$时，其性能优于 EGS。

交换机规模（图 6(c)）。对于更大规模的交换机，曲线总体上与基线保持一致，支持了结论的稳健性。

VI. 结论

本研究提供了一个统一框架，用于刻画全光子量子交换机与配备存储器的量子交换机之间的速率-保真度权衡。我们证明，尽管内部存储器能够实现高效的先宣告后交换控制，但其优势取决于具体运行区间，并受相干时间、宣告延迟和源速率之间相互作用的影响。未来的工作应将这一分析从近期交换机拓展至包括能够支持纠缠蒸馏和纠错的高稳定性存储器，并进一步探索其他交换机架构和 LLEG 协议。

VII. 附录

A. 引理 1 的证明

证明。 写出 $E^\star \triangleq \max_{x\in\mathcal{X}} \sum_{(i,j)\in\mathcal{F}} x_{ij}$ 和 $T \triangleq \sum_i S_i$。我们分两步证明该引理：首先证明可行分配的最大基数不超过 $E^{\max}_{\text{EGS}}$，最后证明它可以达到该值。

(I) 上界： 每个可行分配在每条选定的边上使用一个站点，因此 $E^\star \le B$. 每条边都会消耗两个端点各一个单位的容量，因此

$$2 \sum_{(i,j)\in\mathcal{F}} x_{ij} \le \sum_i S_i = T \Rightarrow E^\star \le \left\lfloor \frac{T}{2} \right\rfloor. \tag{8}$$

最后，至多有 $S_{\max}$ 条边可以使用容量为 $S_{\max}$ 的节点，其余 $T - S_{\max}$ 个单位的容量则分布在其他节点上。即使将所有这些剩余容量都与该节点配对，我们也无法实现超过 $T - S_{\max}$ 条边。因此 $E^\star \le T - S_{\max}$，进而有 $E^\star \le E^{\max}_{\text{EGS}}$.

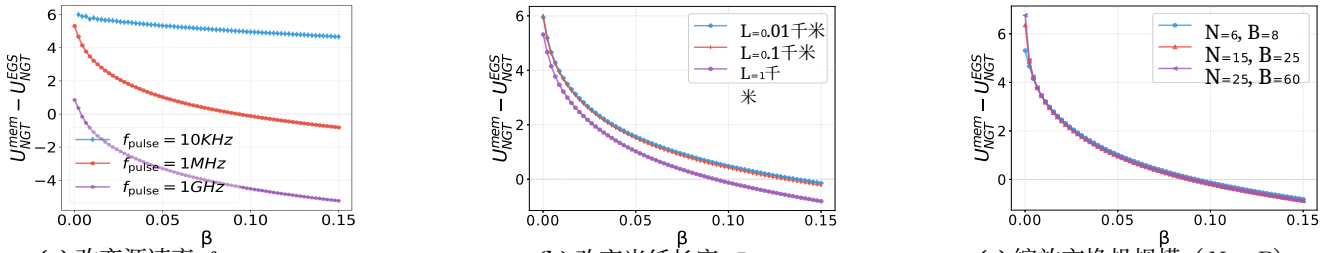
(II) 可实现性： 我们希望构造一个达到 $E^{\max}_{\text{EGS}}$ 的整数向量 $x \in \mathcal{X}$ 。暂且忽略 B ，首先构造该向量，之后会在可实现性的证明中重新考虑它。考虑一个图，对于每个节点 i ，我们创建 S_i 个顶点。可行分配对应于将属于不同节点的顶点配成无序对；每一对 $(i,\ j)$ 都会使 x_{ij} 增加一（同一 $(i,\ j)$ 之间允许存在多对），并消耗 i 和 j 各一个单位的剩余容量。

我们证明，不考虑站点预算约束$\sum_{(i,\ j)\in\mathcal{F}} x_{ij} \le B$ 时，此类配对的最大数量等于

$$\widetilde{E} = \min\left\{ \left\lfloor \frac{T}{2} \right\rfloor,\ T - S_{\max} \right\}. \tag{9}$$

情况 *1*： $T - S_{\max} \le \lfloor T/2 \rfloor$。选取一个满足 $S_{i^\star} = S_{\max}$ 的节点 $i^\star$。将其他节点上的每个容量单位都与 $i^\star$配对。这样共使用 $T - S_{\max}$ 对，并满足每个节点的容量限制：$i^\star$ 贡献 $T - S_{\max} \le S_{\max}$ 对，每个其他节点 j最多贡献 S_j 对，且不会出现自配对。因此 $E \ge T - S_{\max}$，结合上界可知等号成立。

情况 *2*： $T - S_{\max} > \lfloor T/2 \rfloor$。等价地， $S_{\max} < \lceil T/2 \rceil$。我们证明 $\lfloor T/2 \rfloor$ 对总能通过以下贪心规则实现： 每个节点 i 的初始剩余（即未分配）容量为 S_i。每一步，选择剩余容量最大的两个不同节点$u \ne v$ （若并列则任意选择），将 x_{uv} 加 *1*，并将两者的剩余容量都减去 1。当剩余容量为正的节点少于两个时停止。



(a) 改变源速率 f_{pulse}。 (b) 改变光纤长度 L。 (c) 缩放交换机规模（N， B）。

图 6. 使用 $\Delta U_{\text{NGT}}(\beta)$ 时优势性的敏感度。正值表示内存模型更优；负值表示 EGS 更优。

经过 k 步后的剩余容量为 $r_1^{(k)} \geq r_2^{(k)} \geq \cdots \geq r_n^{(k)} \geq 0$，并令 $T^{(k)} = \sum_i r_i^{(k)} = T - 2k$。定义不平衡量 $\Delta^{(k)} \triangleq r_1^{(k)} - \sum_{i\geq 2} r_i^{(k)}$. 由于 $S_{\max} < \lceil T/2 \rceil$，$\Delta^{(0)} = S_{\max} - T - S_{\max} < T - 2\lfloor T/2 \rfloor < 0.$ ( ) 每个贪心步骤都会使 $r_1^{(k)}$ 和 $\sum_{i\geq 2} r_i^{(k)}$ 恰好减少 1，因此在过程运行期间，对所有 k，均有 $\Delta^{(k+1)} = \Delta^{(k)}$。

因此，$\Delta^{(k)} < 0$ 始终成立。若在某个阶段 $T^{(k)} \geq 2$ 但 $r_2^{(k)} = 0$，则只有 $\sum_{i\geq 2} r_i^{(k)} =$，0 $r_1^{(k)} > 0$ $\Delta^{(k)} > 0$ 一个节点为正，因此 和，这意味着，矛盾。因此，只要 $T^{(k)} \geq 2$，就必有 $r_2^{(k)} \geq 1$，且贪心步骤始终可行。因此，该算法恰好执行 $\lfloor T/2 \rfloor$ 步（直到 $T^{(k)} \in \{0, 1\}$）。每一步都会生成一条连接不同节点的有效边，因此我们得到 $\lfloor T/2 \rfloor$ 对。因此，$E \geq \lfloor T/2 \rfloor$，结合上界 (8)，可得 $E = \lfloor T/2 \rfloor$。因此，(9)成立。

现在，为了证明证明中的可实现性步骤，剩下要做的是重新引入 B，因为在式 (9) 中我们假设不存在贝尔态测量预算约束。令 $E_0 \triangleq \widetilde{E}$ 为在没有贝尔态测量约束的情况下构造的配对数。如果 $B \geq E_0$，则证明完毕。如果 $B < E_0$，只需在构造出 B 个配对后截断即可。截断不会破坏可行性。因此，我们明确构造出了具有 $\sum_{(i,\ j)\in\mathcal{F}} x_{ij} = \min\left\{B, \widetilde{E}\right\} = E_{\text{EGS}}^{\max}$. 的 $x \in \mathcal{X}$。结合上界，即可证明 $E^\star = E_{\text{EGS}}^{\max}$。□

B. 定理 *1* 的证明

证明。 目标函数 $\sum_{(i,\ j)\in\mathcal{F}} \lambda_{ij}$ 在多面体 $\text{co}(\mathcal{X})$ 上是线性的，因此最大值在某个顶点处取得，而该顶点对应一个整数分配 $x^\star \in \mathcal{X}$。因此，

$$R_{\text{EGS}}^{\text{sec}} = \frac{p_{\text{e2e}}(\beta)}{\Delta t} \max_{x\in\mathcal{X}} \sum_{(i,j)\in\mathcal{F}} x_{ij} = \frac{p_{\text{e2e}}(\beta)}{\Delta t} E_{\text{EGS}}^{\max},$$

其中，最后一个等式由引理 1 得出。 □

参考文献

[1] R. Van Meter，《量子网络》。John Wiley & Sons，2014年。[2] M. Caleffi 等，“分布式量子计算：综述”，*Computer Networks*，第254卷，第110672页，2024年。[3] H.-J. Briegel、W. D¨ur、J. I. Cirac 和 P. Zoller，“量子中继器：不完美局部操作在量子通信中的作用”，*Physical Review Letters*，第81卷，第26期，第5932页，1998年。[4] G. Vardoyan 等，“量子纠缠分发交换机的随机分析”，*IEEE Transactions onQuantum Engineering*，第2卷，第1–16页，2021年。[5] S. Bhambay、T. Vasantam 和 N. Walton，“量子交换机中的最优调度”，*arXiv*预印本 *arXiv:2501.05380*，2025年。

[6] A. Einstein、B. Podolsky 和 N. Rosen，“量子力学对物理实在的描述是否可以被认为是完备的？”Physical Review，第47卷，第10期，第777页，1935年。[7] S. Gauthier、G. Vardoyan 和 S. Wehner，“量子网络中纠缠生成交换机控制架构”，IEEE Transactions on Quantum Engineering，第4卷，第1–17页，2023年。[8] K. Azuma、K. Tamaki 和 H.-K. Lo，“全光子量子中继器”，Nature Communications，第6卷，第1期，第6787页，2015年。[9] S. Gauthier、T. Vasantam 和 G. Vardoyan，“量子网络枢纽的按需资源分配算法及其性能分析”，载于2024年 IEEE QCE，第1卷。IEEE，2024年，第1748–1759页。[10] G. X. Yau、T. Vasantam 和 G. Vardoyan，“量子纠缠交换中的 d 选项最长队列优先策略”，载于2025年国际量子通信、网络与计算会议（QCNC）。IEEE，2025年，第1–8页。[11] T. Vasantam 和 D. Towsley，“量子交换机的吞吐量最优调度策略”，载于 Quantum Computing, Communication, and Simulation II，第12015卷。SPIE，2022年，第14–23页。[12] I. Tillman 等，“量子交换机容量区域的计算”，载于2024年 IEEE QCE，第1卷。IEEE，2024年，第1868–1878页。[13] W. Dai、A. Rinaldi 和 D. Towsley，“量子交换机中的纠缠交换：协议设计与稳定性分析”，arXiv预印本 arXiv:2110.04116，2021年。[14] P. Promponas 等，“通过高效内存管理最大化灵活量子交换机中的纠缠速率”，IEEE Journal on Selected Areas in Communications，第42卷，第7期，第1749–1762页，2024年。[15] V. Valls、P. Promponas 和 L. Tassiulas，“有纠缠退相干和无纠缠退相干时量子交换机的容量”，IEEE Communications Letters，第27卷，第9期，第2388–2392页，2023年。[16] N. K. Panigrahy 等，“带纠缠纯化的量子交换机容量区域”，载于 IEEE INFOCOM 2023。IEEE，2023年，第1–10页。[17] G. Vardoyan 和 S. Wehner，“量子网络效用最大化”，载于2023年 IEEE QCE，第1卷。IEEE，2023年，第1238–1248页。[18] M. Alshowkan 等，“通过已部署光纤构建可重构量子局域网”，PRX Quantum，第2卷，第4期，2021年10月。[19] S. D. Barrett 和 P. Kok，“利用物质量子比特和线性光学实现高效高保真量子计算”，Physical Review A，第71卷，第6期。[20] K. C. Chen 等，“面向地面和空间量子网络的零附加损耗纠缠光子复用”，Physical Review Applied，第19卷，第5期，2023年5月。[21] P. Kok 和 S. L. Braunstein，“使用参量下转换的后选择与非后选择量子隐形传态”，Physical Review A，第61卷，第4期，2000年3月。[22] C. Jones 等，“量子中继器网络通信协议的设计与分析”，New Journal of Physics，第18卷，第8期，第083015页，2016年。[23] S. Gauthier、G. Vardoyan 和 S. Wehner，“量子网络中纠缠生成交换机的控制架构”，载于《第一届量子网络与分布式量子计算研讨会论文集》，2023年，第38–44页。[24] A. Schrijver 等，《组合优化：多面体与效率》。Springer，第24卷，第2期。[25] P. Promponas、V. Valls 和 L. Tassiulas，“充分利用量子纠缠交换中的有限内存”，载于2023 IFIP Networking。IEEE，2023年，第1–9页。[26] P. Dhara 等，“利用自发参量下转换产生双轨纠缠光子对的预示式复用高效率级联光源”，Physical Review Applied，第17卷，第3期，第034071页，2022年。