



靜宜大學資訊工程學系

畢業專題成果報告書

Department of Computer Science and Information Engineering

Providence University

Degree Dissertation

融合 GNN 與 Transformer 技術的智能網路優化

學生： 資工四 A 411134468 陳世汶

資工四 B 411154646 林庭緯

指導教授： 劉建興 教授

西元 2025 年 12 月

融合 GNN 與 Transformer 技術的智能網路優化

學生：陳世汶 林庭緯

指導教授：劉建興

靜宜大學資訊工程學系

摘要

本專題研製之為滿足大規模超可靠低延遲通訊（mURLLC）對超低延遲、高可靠性與可擴展性的需求，我們提出結合 Transformer 的 G4U 與 PG4U 框架。原有 G4U 透過無線聚合減少訊號與計算開銷，PG4U 則以並行聚合提升穩健性，但兩者皆受限於局部訊息傳遞，難以捕捉長距離依賴關係。為克服此問題，新架構在嵌入更新階段整合 Transformer 編碼器，使模型能全域關注所有節點，實現跨節點依賴建模且無需額外訊號交換。

誌謝

在本次專題研究的完成過程中，我衷心感謝所有曾協助與支持我的人。特別感謝我的指導老師，劉建興老師。研究期間提供耐心的指導與寶貴的建議，使我在專題的規劃與實作過程中獲得重要的成長與啟發。同時，也感謝組員在討論與合作中的投入與努力，讓我們得以順利完成本次專題。本專題的成果凝聚了所有參與者的心力，在此再次致上最誠摯的感謝。

目	錄	
摘要	1
誌謝	1
目錄	2
第一章、	引言	3
1.1	貢獻	3
第二章、	系統模型與問題定式化.....	4
第三章、	GNN 混合變壓器於分散式波束成形與功率控制.....	6
3.1	現有 GNN 框架.....	7
3.2	Transform 工作.....	9
3.3	G4U with Transformer 框架.....	10
第四章、	PG4U 框架與訓練損失函數.....	12
4.1	PG4U with Transformer 框架.....	13
第五章、	模擬結果.....	15
第六章、	結論.....	18
仿真設定系統		
參數表	19
參考文獻	20

第一章、引言

第六代（6G）無線網路的出現，加劇了對能夠實現超高可靠性、毫秒級延遲以及能源高效大規模運作的通信系統的需求。在大規模超可靠低延遲通信（mURLLC）場景中，網路必須同時滿足嚴苛的 1 毫秒級延遲、近乎零封包遺失的極高可靠性，以及支援密集部署（如工業物聯網、協作式自動駕駛與關鍵任務控制系統）的廣域連接能力。這些嚴峻的需求對分散式架構提出了重大挑戰，因為干擾快速變化、拓撲高度動態，而集中式最佳化因訊號開銷與延遲過大而不可行。圖神經網路（GNNs）提供了一種自然的建模方式，能將直接通信鏈路映射為節點，干擾路徑映射為邊，並透過訊息傳遞實現分散式最佳化。在此基礎上，G4U 框架利用空中（OTA）聚合，每幀僅進行一次嵌入更新，以降低訊號交換與計算成本。然而，它依賴於本地鄰居資訊，限制了捕捉長距依賴的能力。PG4U 變體則透過引入平行聚合通道提升穩健性，但仍受限於局部化的訊息傳遞。為了突破這一限制，我們將 Transformer 編碼器整合進上述框架，透過自注意力機制在圖嵌入中建模全域依賴。此整合不僅保留 OTA 聚合的低延遲與低開銷特性，更能有效協調整個網路，從而在複雜干擾與異質通道條件下顯著提升波束成形與功率控制的效能。

1.1 貢獻

本文基於 G4U 框架及其平行變體 PG4U [1]，並做出以下主要貢獻：

- 1) 提出 G4U 和 PG4U with Transformer 增強架構，該架構在嵌入更新階段插入 Transformer 編碼器，能在維持低信令開銷與低計算複雜度的同時，有效捕捉網路範圍內的全域依賴關係。

- 2) 設計 Transformer 編碼器，使其能夠處理由 G4U 和 PG4U 的局部更新函數產生的逐幀節點嵌入，無需額外的消息傳遞步驟，並保持原有的延遲特性。
- 3) 透過在多種網路拓撲與通道條件下的大量仿真，證明所提出方法相較於基線模型能顯著降低訓練損失與 QoS 中斷機率，突顯將全域注意力機制融入大規模 mURLLC 系統之分散式波束成形與功率控制的效益。

第二章、系統模型與問題定式化

我們考慮一個由 K 條通訊鏈路 L_1, L_2, \dots, L_K 組成的下行 mURLLC 網絡，其中每條鏈路連接一個基地台 (BS) 及其關聯的用戶設備 (UE)，所有鏈路均工作在相同的頻寬 B 上[2]。每個基地台都配備有系統包含 N_t 根天線，每個 UE 只有一個天線，所有 BS 同步傳輸。在持續時間為 T_f 的每個訊框中，每個 BS 都會向其 UE 傳送一個 b 位元的封包 [2]。幀長度小於通道相干時間，因此通道係數在幀內保持不變，並在連續幀之間表現出時間相關性，如圖 1 所示。在幀 t 中，鏈路 L_i 的直接通道向量為用 $\mathbf{h}_{i,i} \in \mathcal{C}^{N_t}$ 表示其中， $\mathbf{h}_{i,j} \in \mathcal{C}^{N_t}$ 表示基地台 i 到用戶設備 j 的干擾通道，其中 $i \neq j$ 。全域通道張量 $\mathbf{H}(t) \in \mathcal{C}^{K \times K \times N_t}$ 表示為

$$\mathbf{H}(t) = \mathbf{L} \circ \tilde{\mathbf{G}}(t), \quad (1)$$

其中， $\mathbf{L}(t) \in \mathcal{R}^{K \times K \times N_t}$ 包含大尺度通道增益， $\tilde{\mathbf{G}}(t) \in \mathcal{C}^{K \times K \times N_t}$ 是小尺度衰退係數， \circ 表示逐元素乘法[2]。小尺度衰退演化如下：

$$\tilde{\mathbf{G}}(t+1) = \rho \tilde{\mathbf{G}}(t) + \tilde{\mathbf{A}}(t), \quad \tilde{\mathbf{A}}(t) \sim \mathcal{CN}^{K \times K \times N_t}(0, 1 - \rho^2), \quad (2)$$

相關係數 $\rho \in [0,1]$ 。設備用戶 i 在幀 t 接收到的訊號為

$$y_i(t) = \mathbf{h}_{i,i}^H(t) \mathbf{v}_i(t) s_i + \sum_{j=1, j \neq i}^K \mathbf{h}_{j,i}^H(t) \mathbf{v}_j(t) s_j + n_i(t), \quad (3)$$

其中， s_i 為發射符號， $n_i(t) \sim CN(0, \sigma^2)$ 為加性高斯白噪聲 (AWGN)， $v_i(t) \in C^{N_t}$ 為波束成形和功率控制向量 [3]。連結 L_i 的瞬時訊號與干擾加雜訊比 (SINR) 為：

$$\xi_i(H(t), V(t)) = \frac{|h_{i,i}^H(t)v_i(t)|^2}{\sum_{j=1, j \neq i}^K |h_{j,i}^H(t)v_j(t)s_j|^2 + \sigma^2}, \quad (4)$$

其中 $V(t) = [v_1(t), \dots, v_K(t)]^T \in C^{K \times N_t}$ [3]。令 T_o 表示策略推理期間產生的總開銷，包括訊號延遲和計算延遲。

剩餘的資料傳輸時間為 $T_d = T_f - T_o$ 。鏈路 i 的誤塊機率 (BLER) 由 [4]、[5] 紿出。

$$\varepsilon_i(V(t), H(t)) = Q\left(\frac{-bln2 + B(T_f + T_o)\ln[1 + \xi_i(t)]}{\sqrt{B(T_f - T_o)V_i(t)}}\right), \quad (5)$$

其中 $V_i(t) = 1 - 1/[1 + \xi_i(t)]^{-2}$ 是通道色散[4]，且 $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ 為高斯 Q 函數。我們將最差鏈路 BLER 定義為 (6)，為了提高訓練穩定性 [6] 和 [7]，我們採用基於對數的效用函數 (7)。

$$\varepsilon_{max}(t) = \max_{i=1, \dots, K} \varepsilon_i(V(t), H(t)), \quad (6)$$

$$U(t) := \log_{10}(\varepsilon_{max}t + 10^{-\beta}) + \beta, \beta \geq 0, \quad (7)$$

基於式 (7) 中的效用函數和式 (6) 中的最差連結 BLER 定義，最佳化問題可以表述為：

$$\begin{aligned} & \arg \min_{v_1(t), \dots, v_K(t)} E[U(t)], \\ \text{s.t. } & Eq. (7) \text{ and (6)}, 0 \leq |v_i(t)|^2 \leq P_{max}, \forall i. \end{aligned} \quad (8)$$

其中期望值是基於固定的 K ，在不同的網路拓撲與通道實現下取平均 [3]。此公式定義了大規模 mURLLC 網路中分散式波束成形和功率控制的最佳化目標和限制條件。在第三章中，我們提出了一個基於 GNN 的混合 Transformer 框架，該框架透過捕捉局部和全局依賴關係，同時保持低延遲和低開銷的特性，解決了上述問題。

第三章、GNN 混合變壓器於分散式波束成形與功率控制

基於第二章的系統模型，我們將 mURLLC 網路表示為一個圖，其中每個通訊鏈路 L_i 被建模為一個節點，干擾關係以邊表示。直接通道增益 $h_{(i,i)}(t)$ 被視為節點特徵，而交叉通道增益 $h_{(i,j)}(t)$ 則被視為邊特徵。這些特徵被組織成前面定義的 CSI 張量 $H(t)$ 。這種圖抽象使得空間訊息傳遞機制能夠以分散式的方式捕捉局部干擾模式。如圖 1 所示，實體網路可以對應到圖 2 所示的對應圖拓撲結構。

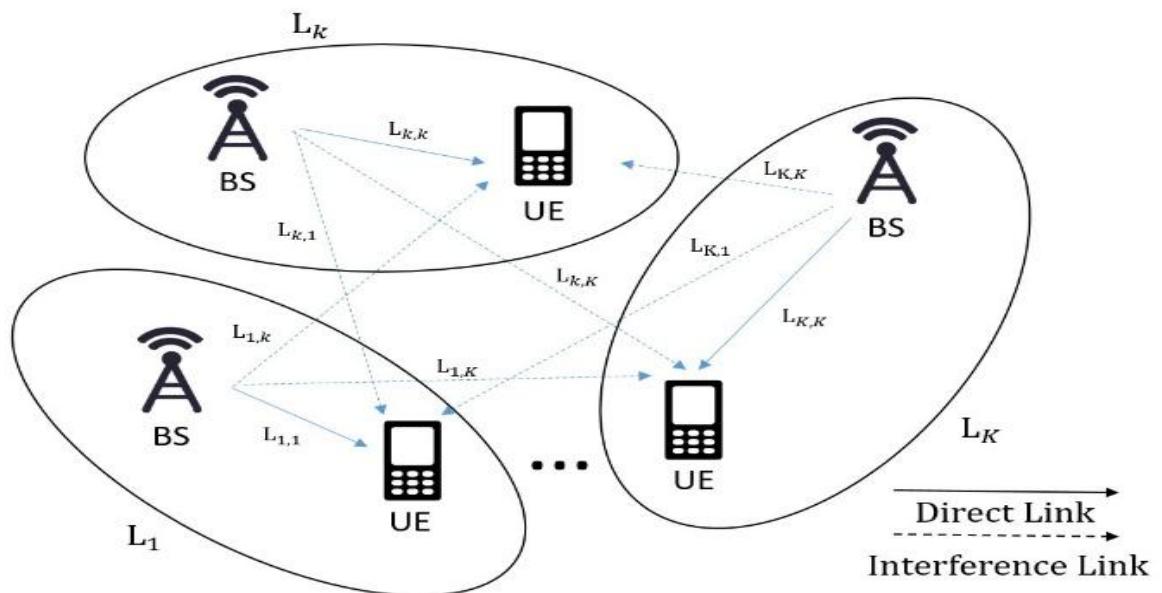


圖 1. 基於 [1] 的具有 K 個連結的所考慮的 mURLLC 網路的系統模型。

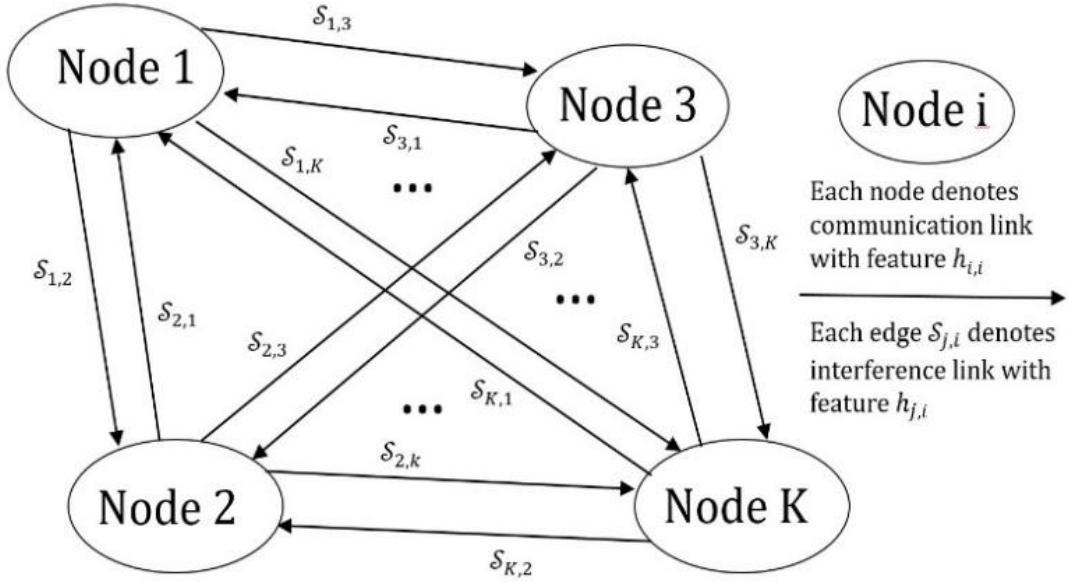


圖 2. 依照 [1] 中的公式， K -link mURLLC 網路的圖模型。

3.1 現有 GNN 框架

我們首先回顧了 G4U 擴展的標準空間 GNN 內核，該內核使用第二章中的系統定義並遵循 [3] 中的消息傳遞公式來表示。在第 t 帀的第 n 層 GNN 中，節點 i 持有嵌入 $e_i^{(n)}(t)$ 。

訊息的產生、聚合和更新步驟定義如下：

$$m_{j \rightarrow i}^{(n)}(t) = \Phi \left(h_{j,j}(t), h_{j,i}(t), e_j^{(n)}(t); \theta \right), \quad (9)$$

$$a_i^{(n)}(t) = agg \left(m_{j,i}^{(n)}(t), j \in \mathcal{N}(i) \right), \quad (10)$$

$$e_i^{(n+1)}(t) = U \left(e_i^{(n)}(t), a_i^{(n)}(t), h_{i,i}(t); \omega \right), \quad (11)$$

$$\nu_i(t) = \Omega \left(e_i^{(N)}(t); \varphi \right). \quad (12)$$

這裡 $agg(\cdot)$ 是一個置換不變的聚合函數（例如，求和、平均值或最大值）。映射 Φ 、 U 和 FNN 的參數在所有節點上共享。經過 N 層訊息傳遞後，每個基地台輸出其波束成形和功率控制向量 $v_i(t)$ 。根據第二節，SINR 和解碼錯誤機率透過式 (4)-(5) 計算，訓練目標由式 (6) 中的效用函數定義。設 $\mathcal{GNN}(\cdot; \Theta)$ 表示參數為 $\Theta = \{\theta, \omega, \varphi\}$ 的 GNN 模型。因此，式 (7) 中的最佳化可以重寫為

$$\min_{\Theta} E[U(t)] \text{s.t. Eq. (7) and (6),}$$

$$V(t) = \mathcal{GNN}(H(t); \Theta), 0 \leq \|v_i(t)\|^2 \leq P_{max}, \forall i. \quad (13)$$

透過輸出歸一化來強制執行功率約束：

$$\hat{v}_i = \frac{v_i}{\max(\|v_i\|, 1)}, \forall i. \quad (14)$$

以分散式方式部署時，傳統的 GNN 需要逐幀獲取直接通道狀態資訊 (CSI) 和干擾通道狀態資訊 (CSI) 以形成式 (8)，然後進行多次訊息交換以實現式 (9)。對於 K 條連結和 N_t 根天線，通道估計的複雜度為 $n_{ce} = K^2 N_t$ ，訊息傳遞的複雜度為 $n_{mp} = NK(K - 1)$ 。因此，信令開銷和計算延遲分別為

$$T_s = K^2 N_t t_{ce} + NK(K - 1) t_{mp}, \quad (15)$$

$$T_c = n_c t_c = (2N + 1) t_c. \quad (16)$$

由於每個節點每層除了最終輸出 FNN 外，還要執行一個訊息 FNN 和一個更新 FNN [3]，這種開銷會減少公式 (5)-(6) 中的可用資料傳輸時間 $T_d = T_f - T_o$ ，因此需要重新設計 mURLLC 的核心。5GNR 相容的幀實現如圖 3 所示，並在 [3] 中有更詳細的討論。

3.2 Transformer 工作

為了克服上一小節中指出的局部性限制，我們引入了一個 Transformer 編碼器，該編碼器作用於每一幀內的節點嵌入集合。雖然空間 GNN 可以透過鄰域訊息傳遞實現局部協調，但其感受野仍然受限於層數，這限制了它們捕捉長程依賴關係的能力，尤其是在由式(4)中的 SINR 表達式引起的干擾結構下。

令 $x_i(t)$ 表示幀 t 中與節點 i 關聯的標記。此標記由中間 GNN 嵌入 $e_i^{(n)}(t)$ 或式(20)中定義的更新後的 G4U 嵌入 $e_i(t)$ 構成，取決於編碼器的插入點。為了編碼結構訊息，每個標記都添加了一個位置訊號 p_i ，該訊號源自局部可觀測的圖特徵，例如歸一化度或路徑損耗加權干擾摘要。這些訊號被設計為置換不變的，並且不需要除第二節系統模型中已有的 CSI 之外的額外 CSI。編碼器對標記集 $\{x_i(t) + p_i\}_i$ 應用多頭自注意力機制。對於每個頭，產生查詢矩陣、鍵矩陣和值矩陣，並用於計算注意力輸出，計算方法如下：

$$Attn(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V. \quad (17)$$

其中 $Q, K, V \in R^{K \times d_k}$ 是投影後的標記矩陣， d_k 是鍵維度。所有頭部輸出連接並進行線性投影，然後進行殘差連接、層歸一化，最後連接一個位置前饋網絡，遵循標準的 Transformer 架構。為了維持 mURLLC 所需的分散式可行性和低延遲特性，我們在 3.3 節中採用了基於核的線性近似式 (17)。這個近似式使得每個節點僅使用局部查詢和透過單次 OTA 廣播獲得的全局聚合統計資訊來計算其注意力輸出。由此產生的細化步驟在 (20a)-(20d) 中正式定義，將 (17) 中的完整 softmax 注意力機制替換為與 G4U 訊號模型相容的計算。因此，該編碼器作為

G4U 核的輕量級、拓撲感知增強，允許每個節點在不違反單輪推理結構或引入額外信令開銷的情況下整合全局上下文資訊。其與 G4U 流程的整合詳見 3.3 節。

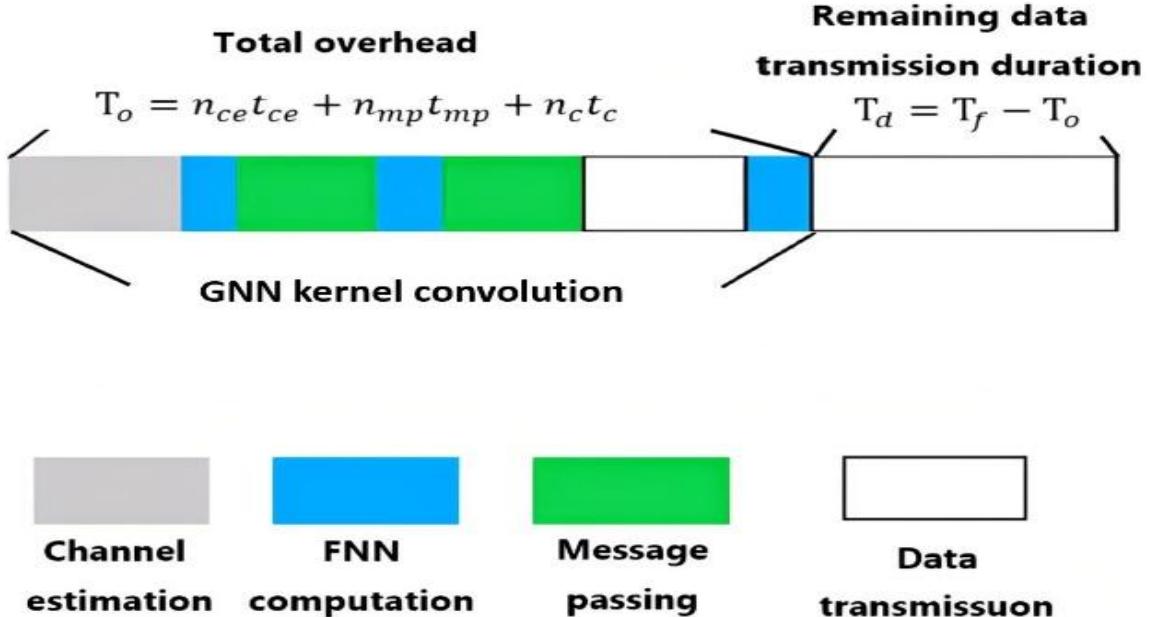


圖 3. 在 5G NR 系統中應用基於 GNN 的策略的幀結構設計。

3.3 G4U with Transformer 框架

基於前文所述的 G4U 內核，我們在局部嵌入更新和輸出頭之間引入了全域注意力機制，如圖 4 所示。這種改進使得每個連結在更新自身之前都能關注到其他所有連結。透過推斷其傳輸策略，從而改善整個網路的協調性。此改進採用基於核函數的線性注意力近似實現，它保留了 G4U 的單輪 OTA 訊號結構，並保持與 mURLLC 延遲限制的兼容性[3]、[8]、[9]。在每個幀 t 的開始，每個節點使用正交導頻序列估計其直接鏈路 $h_{i,i}(t)$ 的通道狀態資訊 (CSI)。基於前一個嵌入 $e_i(t-1)$ 和估計的 CSI，透過前饋神經網路計算導頻波束形成器。

$$v_i^{pilot}(t) = \Phi(e_i(t-1), h_{i,i}; \theta), \quad (18)$$

所有節點隨後使用各自的波束形成器發射導頻訊號。每個接收器透過 OTA 聚合干擾併計算鄰域摘要。

$$\tilde{a}_i(t) = \text{agg} \left(|h_{j,i}^H(t) v_i^{pilot}(t)|^2, j \in \mathcal{N}(i) \right), \quad (19)$$

利用直接通道 $h_{i,i}(t)$ 、聚合干擾 $\tilde{a}_i(t)$ 和先前的嵌入 $e_i(t-1)$ ，每個節點更新其局部圖嵌入。

$$e_i(t) = U(e_i(t-1), \tilde{a}_i(t), h_{i,i}(t); \omega), \quad (20)$$

為了融入全域上下文，我們使用 Transformer 編碼器對更新後的嵌入 $e_i(t)$ 進行細化。我們沒有計算公式(17)中定義的完整 softmax 注意力，而是採用了一種基於核函數的線性近似，從而可以透過 OTA 聚合實現分散式執行。每個節點首先建立其 token，並將其投影到查詢空間、鍵空間和值空間。

$$\begin{aligned} x_i(t) &= e_i(t) + p_i, q_i^h(t) = x_i(t)W_Q^h, k_i^h(t) = x_i(t)W_K^h, \\ v_i^h(t) &= x_i(t)W_V^h. \end{aligned} \quad (20.a)$$

其中 p_i 編碼了局部拓樸感知的位置資訊（例如，度數或路徑損失摘要），而 W_Q^h, W_K^h, W_V^h 是特定於頭部的投影矩陣。然後透過 OTA 聚合全域統計資料。

$$\begin{aligned} S_h^{(0)}(t) &= \sum_{j=1}^K \phi(k_j^h(t)), \\ S_h^{(1)}(t) &= \sum_{j=1}^K \phi(k_j^h(t)) \otimes v_j^h(t). \end{aligned} \quad (20.b)$$

其中 $\phi(\cdot)$ 是核特徵圖（例如，隨機傅立葉特徵），它將式 (17) 中的 softmax 注意力線性化，從而允許透過單次 OTA 廣播計算求和。然後，每個節點在本地重建其注意力輸出。

$$z_i^h(t) = \frac{\phi\left(q_i^h(t)\right)^T S_h^{(1)}(t)}{\phi\left(q_i^h(t)\right)^T S_h^{(0)}(t)},$$

$$\tilde{e}_l(t) = LN\left(x_i(t) + \text{concat}_h\left(z_i^h(t)\right) W_o\right). \quad (20.c)$$

其中 $LN(\cdot)$ 表示層歸一化， W_o 是輸出投影矩陣。然後應用帶有殘差歸一化的逐位置前饋細化。

$$\tilde{e}_l(t) \leftarrow LN\left(\tilde{e}_l(t) + FFN(\tilde{e}_l(t))\right). \quad (20.d)$$

其中 $FFN(\cdot)$ 表示位置前饋網路。然後，將精細化的嵌入 $\tilde{e}_l(t)$ 傳遞給輸出頭，以產生最終的傳輸向量 $v_i(t)$ ，隨後進行如式(13)所示的功率歸一化。

$$v_i(t) = \Omega(\tilde{e}_l(t); \varphi), \quad (21)$$

由於 Transformer 編碼器完全基於本地維護的嵌入和 OTA 聚合的統計資料運行，因此它不會引入額外的導頻傳輸或訊息傳遞輪次。訊號配置與原始 G4U 設計[8]和[9]保持一致，僅略微增加了本地計算量。這確保了與 5G-NR 同步的兼容性，並保持了 URLLC 所需的單次聚合速率。

第四章、PG4U 框架與訓練損失函數

基於第三章所述的基於 Transformer 的 G4U 框架，我們將其擴展到 PG4U 變體，以進一步降低 mURLLC 網路的延遲。PG4U 利用連續幀間 CSI 的時間相關性，允許波束成形和功率控制策略推斷利用前一幀的 CSI。這使得前饋網路能夠

在資料傳輸階段執行，每幀僅需一次導頻傳輸，從而降低信令開銷和運算延遲。基於此原理，我們開發了基於 Transformer 的 PG4U，其中 Transformer 模組經過改進，能夠更有效地對 CSI 的時間依賴性進行建模，從而實現更準確、更有效率的策略生成，以便與 5GNR 系統無縫整合。

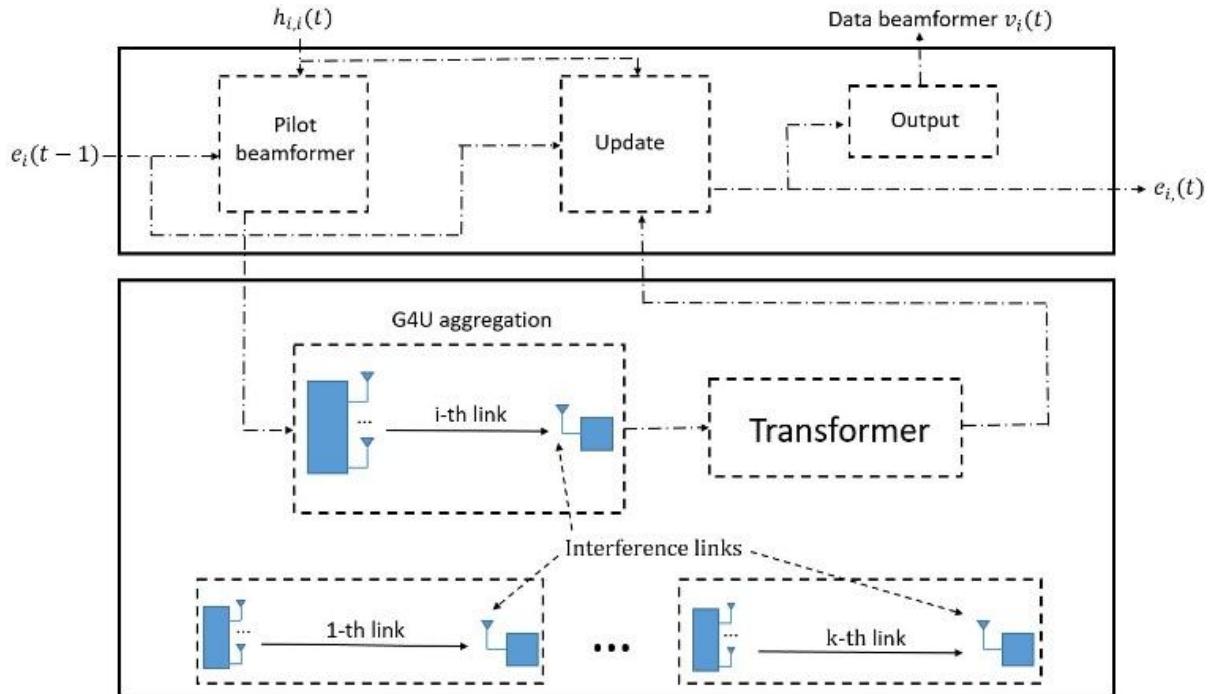


圖 4. 採用 Transformer 框架的所提出的 G4U 分散式實現示意圖。

4.1 PG4U with Transformer 框架

基於第三節中介紹的基於 Transformer 的 G4U 框架，PG4U 變體引入了並行聚合，以利用連續幀間通道狀態資訊 (CSI) 的時間相關性，從而降低 mURLLC 網路的延遲。PG4U 的核心操作包括：G4U 聚合、嵌入更新和輸出波束成形。

$$\tilde{a}_l(t) = \text{agg} \left(|h_{j,i}^H(t)\tilde{v}_j(t)|^2, j \in N(i) \right), \quad (22)$$

$$e_i(t) = U(e_i(t-1), \tilde{a}_l(t), h_{i,i}(t); \phi), \quad (23)$$

$$V_i(t-1) = \Omega(e_i(t)). \quad (24)$$

在每個訊框的開始，所有節點使用正交導頻序列廣播導頻訊號，導頻波束形成器在前一幀中確定。每個節點從接收到的疊加導頻中獲得聚合訊息 $\tilde{a}_l(t)$ ，並根據命題1和命題2[1]估計其直接鏈路通道狀態資訊 $h_{i,i}(t)$ ：

$$h_{i,i} = \text{diag}\left(\frac{1}{\tilde{\nu}_l^*}\right) \tilde{S}_l \tilde{y}_l^H. \quad (25)$$

在資料傳輸階段，基地台利用前一幀的資料波束形成器，同時基於 $\tilde{a}_l(t)$ 和 $h_{i,i}(t)$ ，透過更新型 FNN 更新局部嵌入 $e_l(t)$ 。更新後的嵌入被重塑成一個序列，並由式 (17) 中的 Transformer 編碼器進行處理，從而能夠捕捉長程依賴關係和干擾模式。變換後的嵌入隨後被傳遞給輸出型 FNN，以確定下一幀的波束形成器，之後進行功率歸一化，如式 (17) 所示。

$$P_{norm} = \frac{P}{\max(1, \|P\|_2)}. \quad (26)$$

由於通道估計、PG4U 聚合和 Transformer 優化都在資料傳輸階段執行，因此訊號開銷仍然很小。

$$T_s = KN_t t_{ce}. \quad (27)$$

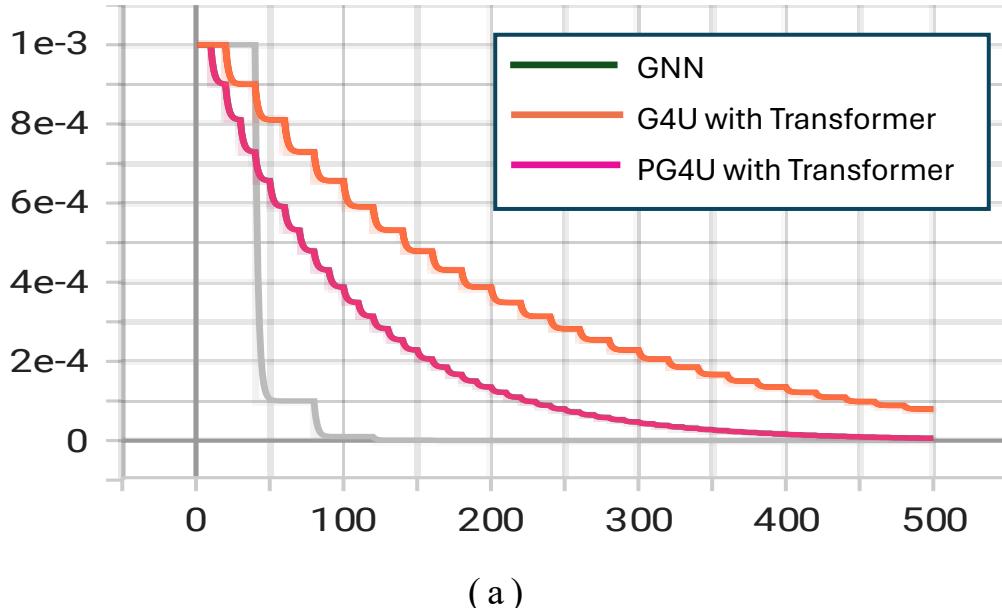
類似地，當 FNN 和 Transformer 處理在資料傳輸階段完成時，計算延遲 T_c 可以忽略不計。與原始 G4U 和 PG4U 相比，PG4U with Transformer 框架在保持低開銷、單導頻設計的同時，改進了對高相關通道中複雜干擾相關性的建模。在每個基地台 (BS) 處，所有 FNN 和 Transformer 編碼器都會執行，每個用戶設備 (UE) 傳回聚合訊息及其本機頻道狀態資訊 (CSI)。這使得基地台能夠使用前一幀的波束形成器進行傳輸，同時使用 Transformer 增強模型更新嵌入並計算下一幀的波束形成器。以上是對所提出的 PG4U with Transformer 的描述，它擴展了 G4U with Transformer 的原理，利用時間 CSI 相關性進一步降低大規模多通道

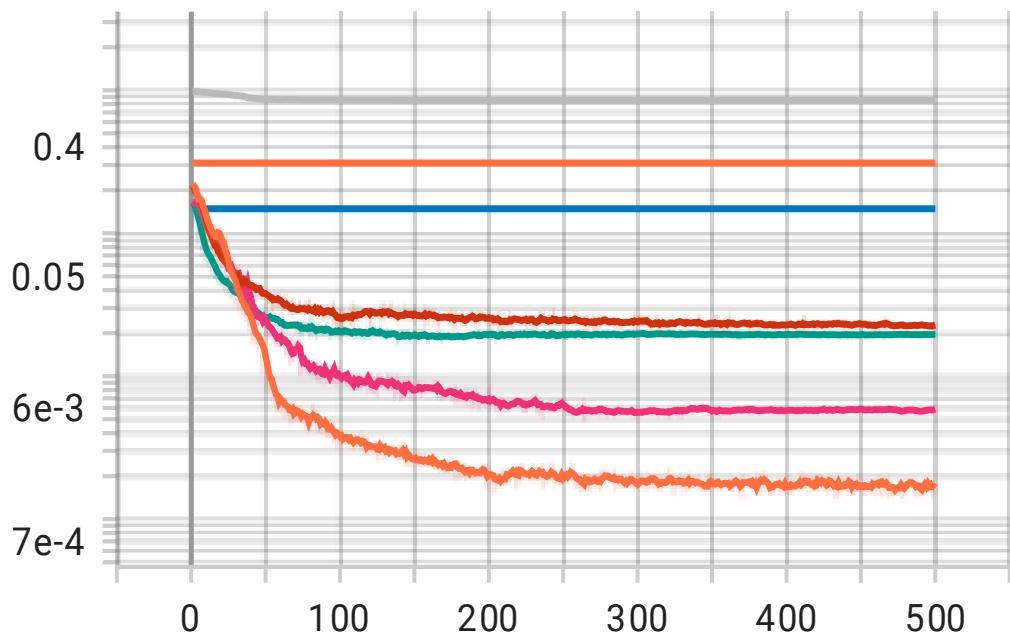
快速響應低延遲通訊（mURLLC）網路中的延遲。下一節將介紹用於模型最佳化的訓練損失函數。

第五章、模擬結果

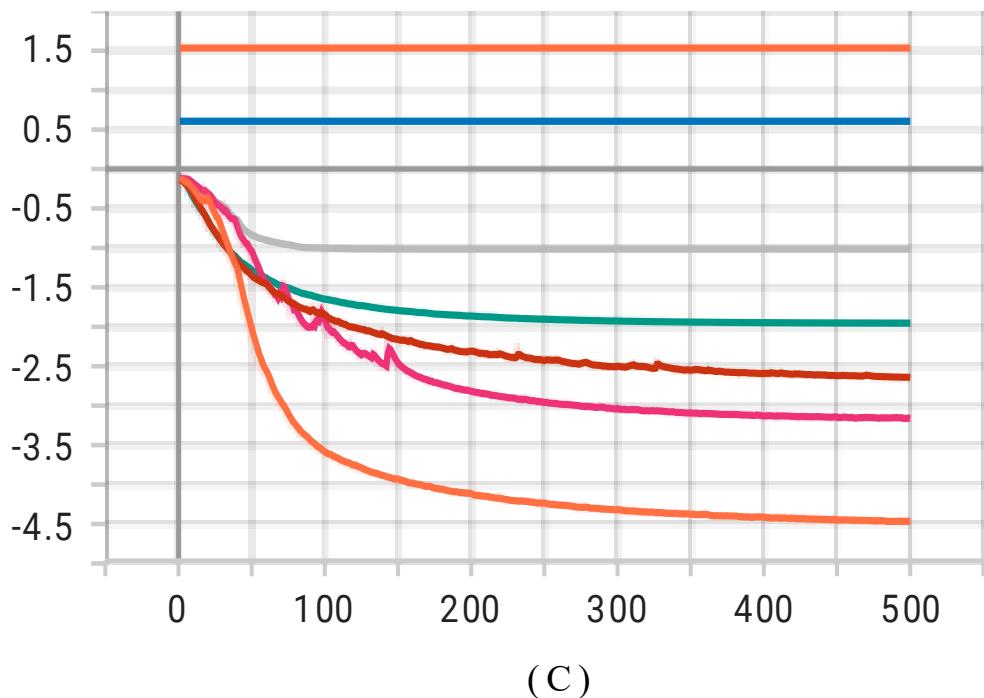
本研究的模擬設定在通道建模、天線部署與訓練配置方面皆遵循文獻 [1]，唯訓練輪數延長至 500 回合。評估的方法包含基準的 G4U 與 PG4U，及其加入 Transformer 的版本；後者透過全域注意力（global attention）以捕捉大規模 mURLLC 網路中的長距依賴關係。作為比較，我們亦考慮三種額外策略：不具自適應協調的等功率配置（EPA）、可考量 (15)-(16) 式中訊號交換開銷 T_s 與計算延遲 T_s 的分散式 GNN、以及在未建模延遲與開銷情形下採行十次迭代的集中式 WMMSE 演算法。資料集中包含 2×10^4 組隨機網路佈局用於訓練，以及 5×10^4 組佈局用於測試，每組佈局皆位於 $500\text{ m} \times 500\text{ m}$ 區域，並設有 $K = 20$ 條 mURLLC 連結。CSI 於 10 個連續訊框上生成，通道模型採 (2) 式並取 $\rho = 0.99$ ，以代表行人移動場景 [2]；大尺度衰落遵循文獻 [10]，並加入 3 dB 對數常態遮蔽。每個訊框長度為 1 ms、頻寬為 5 MHz。每個 FNN 的處理時間為 $100\text{ }\mu\text{s}$ ，通道估測與訊息傳遞開銷為 $0.2\text{ }\mu\text{s}$ 。除非另有說明，所有模擬參數皆彙整於表一。在表一的模擬配置下，圖 5(a)-(c) 的趨勢展示了各方法的相對表現。圖 5 繪製了以 (6) 式定義、並取 $\beta = 5$ 的平均效用，針對 GNN、G4U、PG4U 及其 Transformer 加強版本，於相同驗證資料集上隨訓練回合 的變化情形。G4U-Transformer 的收斂速度比基準 G4U 快超過 50%，並在訓練結束時達到約低 94% 的損失。PG4U-Transformer 也較 PG4U 有明顯損失降幅，但因 PG4U 本身基準較低，相對改善幅度較小。WMMSE 的損失與 G4U-Transformer 相近，但因集中式迭代運算而需更多回合；EPA 則保持不變，而分散式 GNN 收斂較慢，並停留於較高的損失。在圖 5(b) 的 QoS 中斷率（outage probability）比較中，加入

Transformer 的修正可使 G4U 的中斷率降低逾 84%，PG4U 則降低約 52%。G4U-Transformer 的中斷率最低，即使是完全分散式運作，也勝過集中式的 WMMSE•EPA 與 GNN 則呈現最高的中斷率，顯示其干擾協調能力不足。圖 5(a) 的吞吐量比較顯示，在理想假設下 WMMSE 仍達到最高平均總速率，而 G4U-Transformer 則在保持分散式運作下，表現僅比 WMMSE 低約 5%；PG4U-Transformer 亦優於 PG4U。相對之下，GNN 和 EPA 因有效功率配置與波束成形能力較弱，其吞吐量明顯偏低。圖 5(c) 顯示端到端執行延遲。PG4U 及其 Transformer 版本因能進行平行聚合，在自適應方案中達到最低延遲；G4U-Transformer 稍慢，但仍完全符合 mURLLC 的延遲要求，並呈現出可靠度、吞吐量與延遲之間的良好平衡。WMMSE 則因集中式迭代運算成本，延遲最高。綜上所述，將 Transformer 的全域注意力融入分散式 GNN 最佳化架構，特別是在 G4U 框架中，可提升收斂速度、降低中斷率並提高頻譜效率，同時仍能將延遲維持在大規模 mURLLC 網路所需的限制之內。

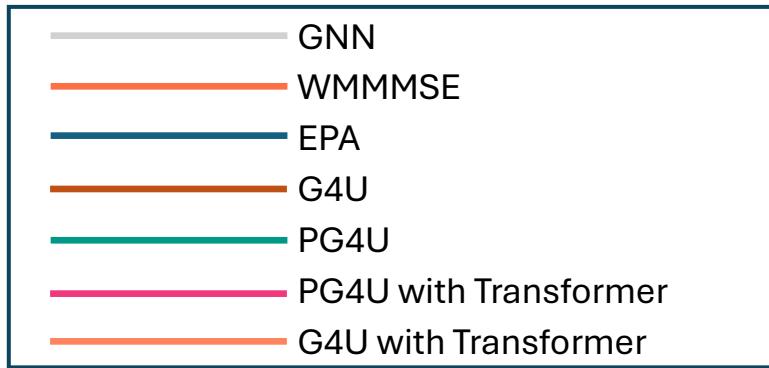




(b)



(C)



(d)

圖 5. 使用不同損失函數對訓練輪數的平均效用進行比較：(a) 學習率 (b) QoS 中斷率和 (c) 訓練損失 (d) b 和 c 的圖例。

第六章、結論

我們在兩種分散式最佳化架構 G4U 與 PG4U，其加入 Transformer 的增強版本，用於大規模 mURLLC 網路中的聯合波束成形與功率控制。此方法的核心創新在於將 Transformer 編碼器嵌入至訊息傳遞過程，使自注意力 (self-attention) 能透過捕捉傳統局部聚合無法建模的長距相依關係來精煉圖嵌入 (graph embeddings)。這項增強使框架在僅需單輪、低延遲訊息交換的前提下，仍能獲得干擾形態的全域視角。模擬結果顯示，整合 Transformer 可顯著加速收斂、提升可靠度並提高頻譜效率，同時不會超出 mURLLC 的延遲限制。特別是，G4U-Transformer 在中度相關通道中表現出色，利用注意力機制有效平衡遠距節點間的干擾；而 PG4U-Transformer 則在高度相關或短訊框情境下提供了良好的「穩健性—延遲」權衡。這些結果證明，基於注意力的精煉是一種能有效補強 GNN 聚合的強大方法，使分散式最佳化更接近集中式效能。未來工作將探討多小區協同、動態拓撲調適，以及用於快速時變環境的學習式通道預測。

參數	數值	參數	數值
Number of antennas N_t	4	Channel correlation coefficient	0.99
Number of frames	10	Carrier frequency	2.4 GHz
Number of links	20	Number of training layouts	2×10^4
Noise power	-174 dBm/Hz	Number of testing layouts	5×10^4
Antenna height	2 m	Channel estimation overhead t_{ce}	0.2 μs
Maximum transmit power	40 dBm	Computation delay t_c	100 μs
Bandwidth	5 MHz	Number of bits	128
Message passing overhead t_{mp}	0.2 μs	Log-normal shadowing	3 dB

表一、仿真設定係統參數

參考論文

- [1] Y. Gu, C. She, S. Bi, Z. Quan and B. Vucetic, "Graph Neural Network for Distributed Beamforming and Power Control in Massive URLLC Networks," in IEEE Transactions on Wireless Communications, vol. 23, no. 8, pp. 9099–9112, Aug. 2024.
- [2] A. Doshi, S. Yerramalli, L. Ferrari, T. Yoo, and J. G. Andrews, "A deep reinforcement learning framework for contention-based spectrum sharing," IEEE J. Sel. Areas Commun., vol. 39, no. 8, pp. 2526–2540, Aug. 2021.
- [3] Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "Graph neural networks for scalable radio resource management: Architecture design and theoretical analysis," IEEE J. Sel. Areas Commun., vol. 39, no. 1, pp. 101–115, Jan. 2021.

- [4] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, “Quasi-static multipleantenna fading channels at finite blocklength,” *IEEE Trans. Inf.Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.
- [5] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [6] Z. Gu et al., “Knowledge-assisted deep reinforcement learning in 5G scheduler design: From theoretical framework to implementation,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2014–2028, Jul. 2021.
- [7] X. Liu, C. She, Y. Li, and B. Vucetic, “Edge-wise gated graph neural network for user association in massive URLLC,” in Proc. IEEE Globecom Workshops (GC Wkshps), Dec. 2021, pp. 1–6.
- [8] Y. Gu, C. She, Z. Quan, C. Qiu, and X. Xu, “Graph neural networks for distributed power allocation in wireless networks: Aggregation over-theair,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7551–7564, Nov. 2023.
- [9] H. Shi, A. Ajaz, and N. Jiang, “Evaluating the performance of over-theair time synchronization for 5G and TSN integration,” in Proc. IEEE Int. Black Sea Conf. Commun. Netw. (BlackSeaCom), May 2021, pp. 1–6.
- [10] Propagation Data and Prediction Methods for the Planning of Short Range Outdoor Radiocommunication Systems and Radio Local Area Networks in the Frequency Range 300 MHz to 100 GHz, document I. R. P.1411-11, 2021.