同步輻射蛋白質結晶學設施計算及網路系統之架構

Computing and Network Architecture of SPXF Core Facility

趙俊雄、許嘉妮、簡玉成

國家同步輻射研究中心

X 光生物結構小組

中華民國九十二年十二月十日

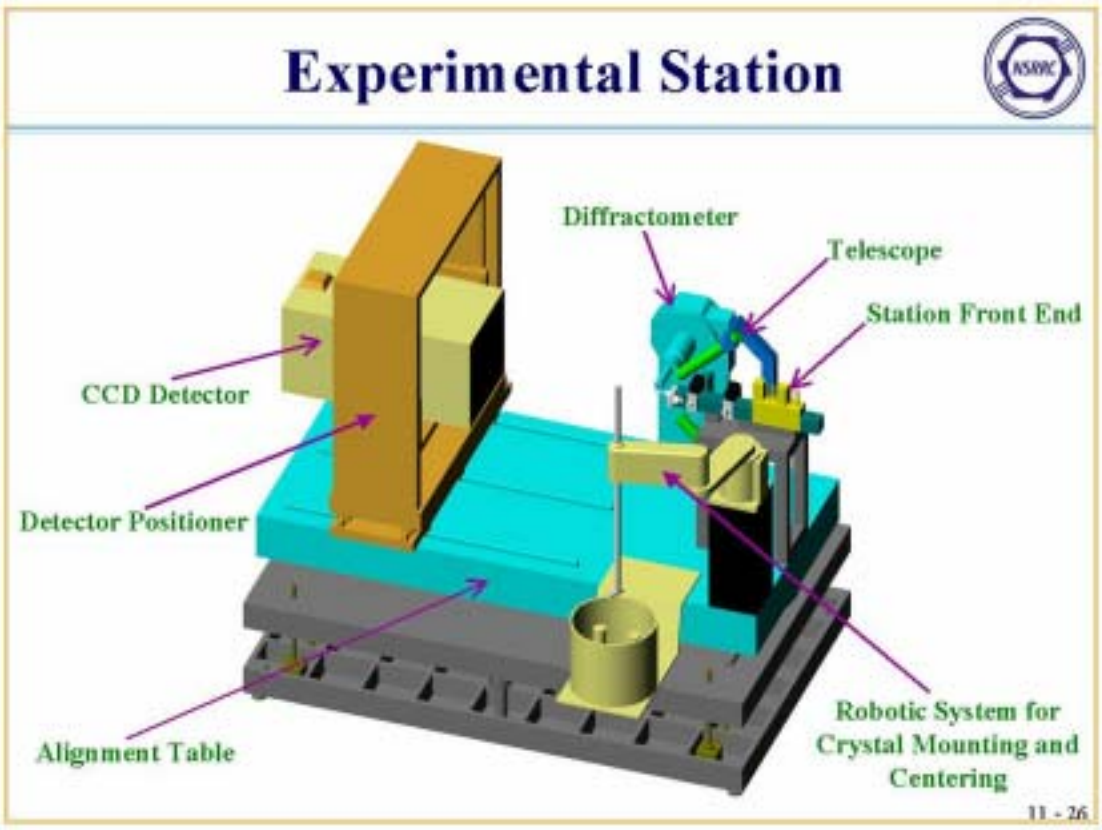
1. 前言

為配合基因體醫學國家型計畫，本中心將興建兩條蛋白質結晶學專屬的高效能光束線與實驗站。為了達成『高效能』這個目的，除了光強度夠強之外，還要有硬體 (如更快的 X 光偵測器，機器人等等) 的配合。而高效能的結果，則是產生更大量的實驗數據，因此需要更快的電腦，高頻寬且穩定的網路架構，以及高容量的儲存設備。

2. SPXF 光束線

根據基因體醫學國家型科技計畫於 2001 年九月的要求，國家同步輻射研究中心將興建兩條同步輻射蛋白質結晶學設施 (Synchrotron Protein Crystallography Facility，SPXF) 光束線，編號為 BL13B 以及 BL13C。

光束線及實驗站外觀如下圖所示：



BL13B 光束線選用平面雙晶體單色光器，反射面為 Si(111)，能量範圍為 6.5 ~ 19 keV，BL13C 光束線則選用曲面單色光器，反射面為 Si(111)。BL13B 預計的樣品處光通量 (0.2 mm 針孔，能量 12.65 keV) 為 3.5×10^{11} 光子/秒，與世界其他頂尖的蛋白質結晶學光束線相當。

在這樣的光強度之下，每一張影像預計只需要 10 秒的曝光時間，一個小時內即可取到一組完整的數據。

3. X 光面積偵測器

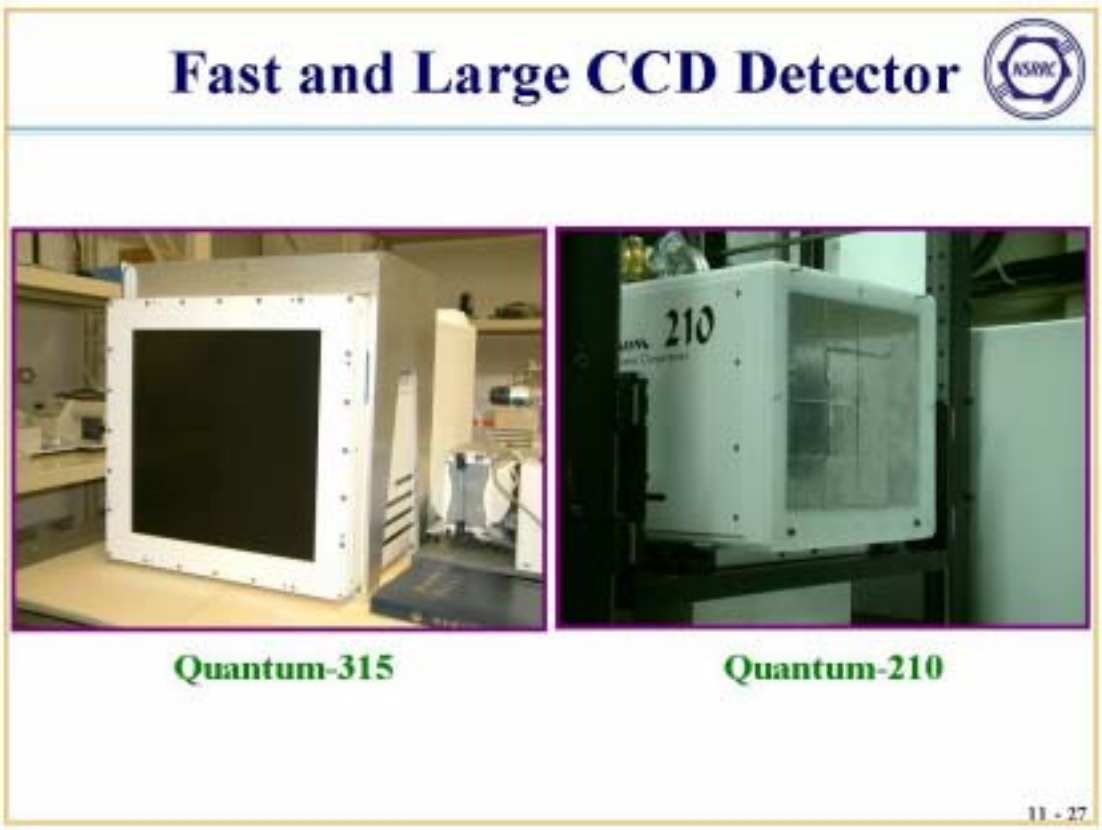
SPXF 光束線使用 ADSC 公司所出品的 Quantum 315 (尺寸為 315 mm × 315 mm) 以及 Quantum 210 (尺寸為 210 mm × 210 mm) CCD detector。

Quantum 315 產生的影像檔案大小為 72 MB，讀取時間小於 2 秒，預計曝光時間為 10 秒/影像，因此每 8 個小時產生數據大小為 202 GB。

Quantum 210 產生的影像檔案大小為 32 MB，讀取時間小於 2 秒，預計曝光時間為 10 秒/影像，因此每 8 個小時產生數據大小為 100 GB。

假設兩條光束線同時運作，每秒鐘在網路上面傳輸的資料大小為

83.2 MB，8 個小時產生的總資料量為 300 GB。



4.網路架構

4.1 資料傳輸

在實驗期間，CCD detector 會持續產生數據，如果因為網路頻寬不足造成塞車，除了資料會損失之外，亦可能造成整個 CCD 系統運作不正常。為了應付每秒 83.2 MB 的傳輸量，我們必須使用超高速乙太網路 (gigabit ethernet) 或是光纖網路 (fibre channel)。

超高速乙太網路傳輸速率可達每秒十億個位元 (1 gigabit/s)，使用的網路傳輸協定為 TCP/IP。 然而 TCP/IP 的制定目標是讓不同平台的電腦間交換資料和傳遞訊息，用來傳遞大型檔案時，由於會佔用 CPU 資源，加上 TCP/IP 本身的限制，不僅效能會降低，同時失敗率也會增加。

光纖網路傳輸速率可達每秒 1.12 gigabits/s 以上，可支援現有以及未來的所有協定 (http://www.fibrechannel.org/OVERVIEW/software.html)。在眾多的光纖網路協定中，FCP (Fibre Channel Protocol) 將 SCSI 上層協定對應到光纖網路傳輸層中，因此應用程式可以透過光纖網路使用 SCSI

來存取資料。 SCSI 協定的用途是快速存取儲存設備 (如光碟，硬碟等) 上的資料，即使資料量很大的時候，也不會增加電腦主機的負擔。

不論就應用上或是將來的擴充性而言，光纖網路都是優於超高速乙太網路的。

4.2 資料儲存

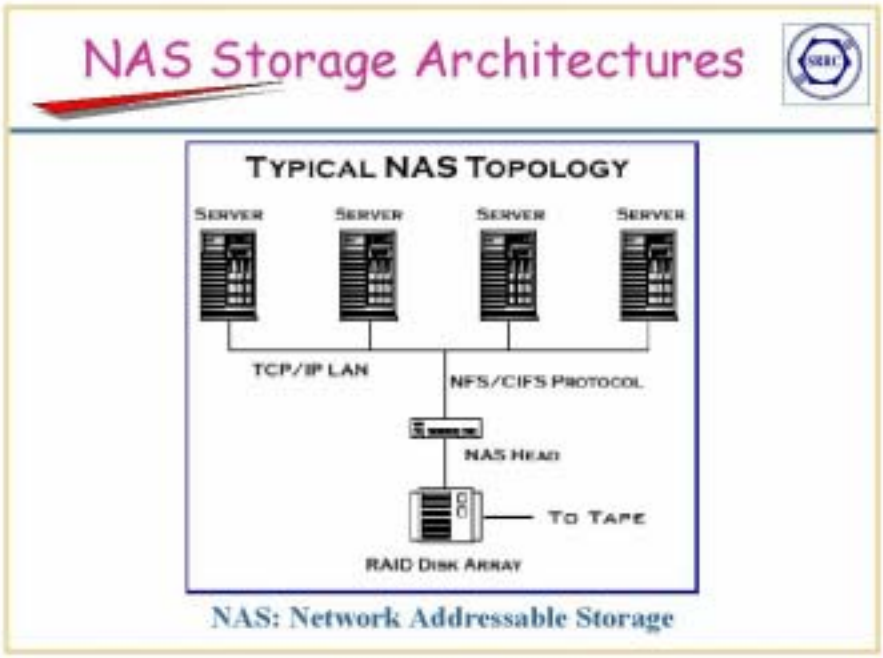
最簡單的儲存方式為直接連接儲存 (Direct Attached Storage，DAS)，例如 PC 上的內接或外接硬碟。這種方式的好處是成本較低，傳輸品質良好，缺點是儲存的資料很難共享，且較容易有安全上的顧慮。

另外一種方式是透過網路來存取資料。目前較常見的方式有 NAS 以及 SAN 兩種。

4.2.1 NAS

網路附加存放區 (Network Attached Storage，NAS) 使用連接到網路的特殊裝置來儲存資料。這些裝置都有指定的網路協定位址 (Internet Protocol Address)， 用戶端可以透過伺服器或直接由這些裝置來存取資料。

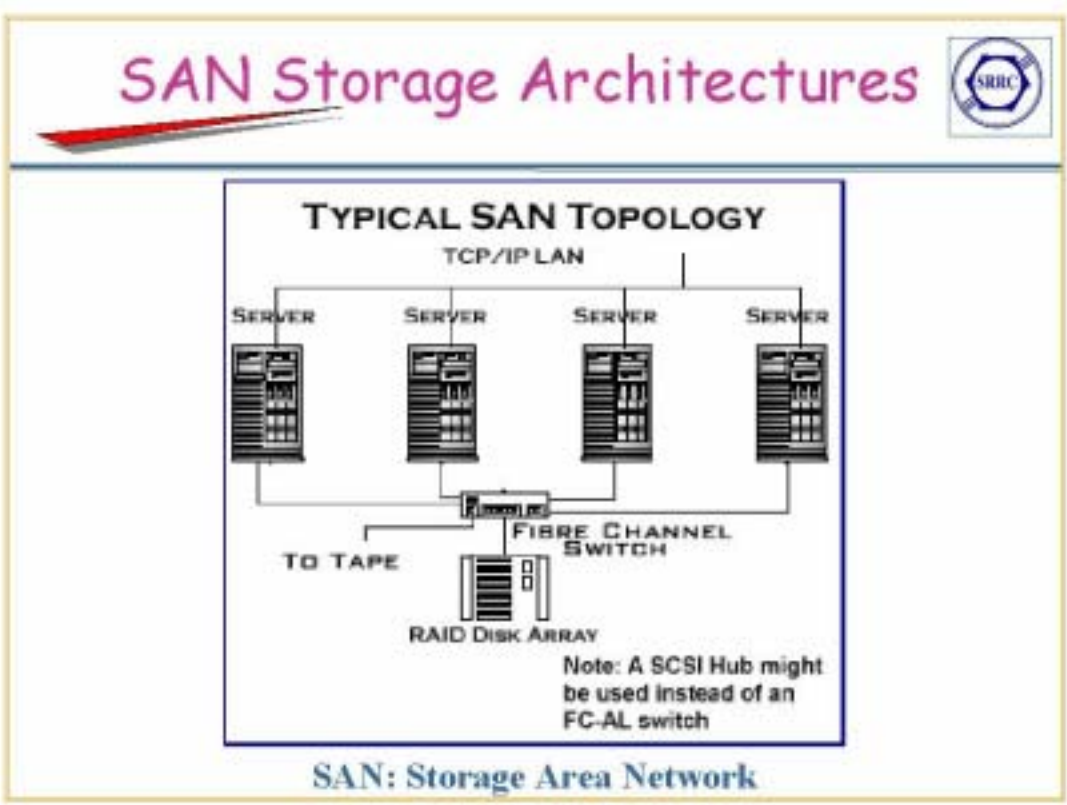
NAS 只能使用 TCP/IP 以及乙太網路架構，因此對於網路以及 CPU 負載比較難以掌控，且效能容易受到網路流量影響，並不適合用來存取大型的檔案。



4.2.2 SAN

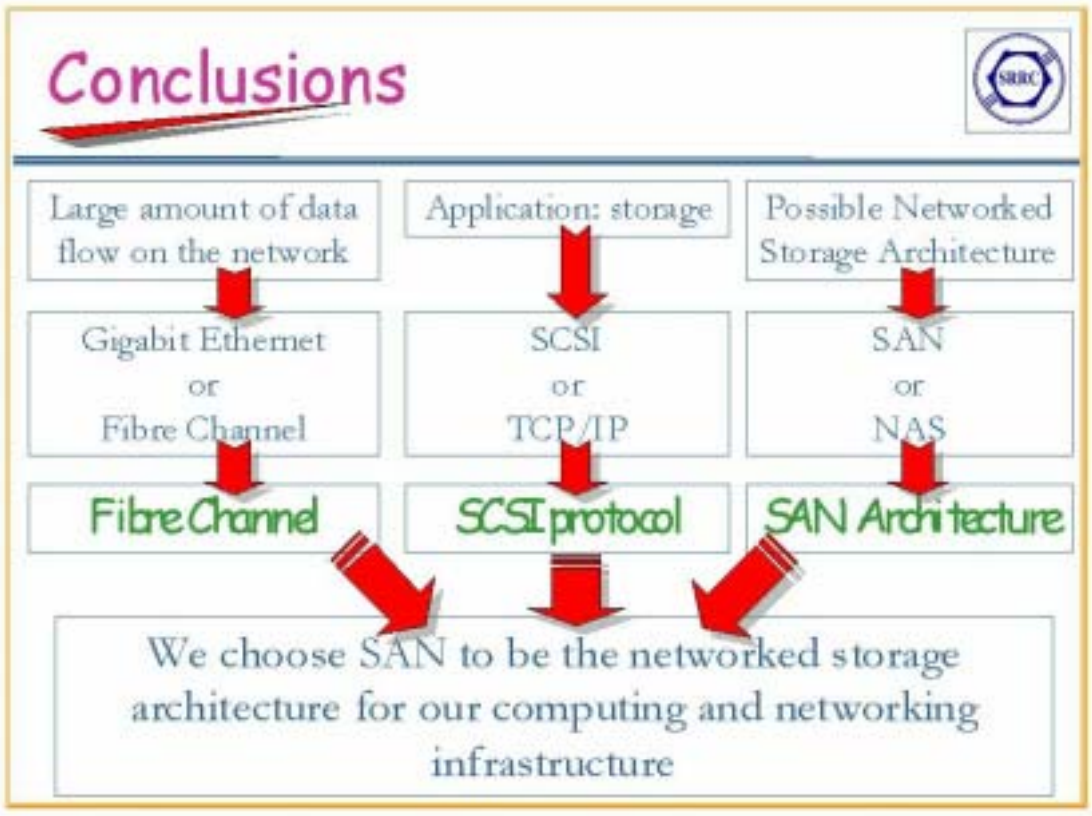
儲存區域網路 (Storage Area Network，SAN) 是由數個互相連接的儲存裝置，與一個伺服器或伺服器叢集 (cluster) 所形成的網路，這個網路也可以連接到其他的網路上面。 SAN 使用一個特別的交換器 (類似於乙太網路交換器) 來連接各個裝置，其他所有伺服器都透過這個交換器來存取資料。

SAN 存取資料的網路與其他網路是分開的，因此不會受到其他網路流量的影響，同時若使用 SCSI 作為上層協定，大量資料存取也不會造成電腦主機的負擔，從應用的觀點而言，SAN 是優於 NAS 的。



4.3 結論

由於資料流量大，我們選擇使用光纖網路作為資料傳輸的媒介；為了快速存取儲存設備上的資料，我們選擇使用 SCSI 作為上層傳輸協定；儲存資料的方式則使用 SAN 架構。



SAN 所使用的檔案系統非常多，每一家電腦公司都有自己獨特的檔案系統。由於使用軟體的限制，我們使用 SGI 電腦公司的 SAN 產品，使用的檔案系統稱為 CXFS (Clustered XFS)。

在一般的情況下，雖然 SAN 架構提供一群電腦到儲存設備的高速連線，但並未消除 DAS 以及 NAS 的瓶頸。即使在 SAN 架構下，儲存設備仍然是指定給特定的系統，因此在分享資料時，我們仍然使用傳統的檔案共享方式— NFS，FTP 以及 CIFS/Samba 等等。這些方法都會降低系統的效能和可用度，並增加系統的複雜度。

為了解決這個問題，SGI 公司根據 XFS 技術發產出 SAN 專用的 Infinite Storage Shared File System CXFS 檔案系統。 XFS 檔案系統是一個真實的 64 位元檔案系統，從 1990 年開始到現在一直都是 SGI IRIX 作業系統的核心，不論是擴充性 (可支援 9 million TB 大小的檔案，以及 18 million TB 大小的檔案系統) 、效能 (7 GB/sec) 以及可靠度都是無庸置疑的。

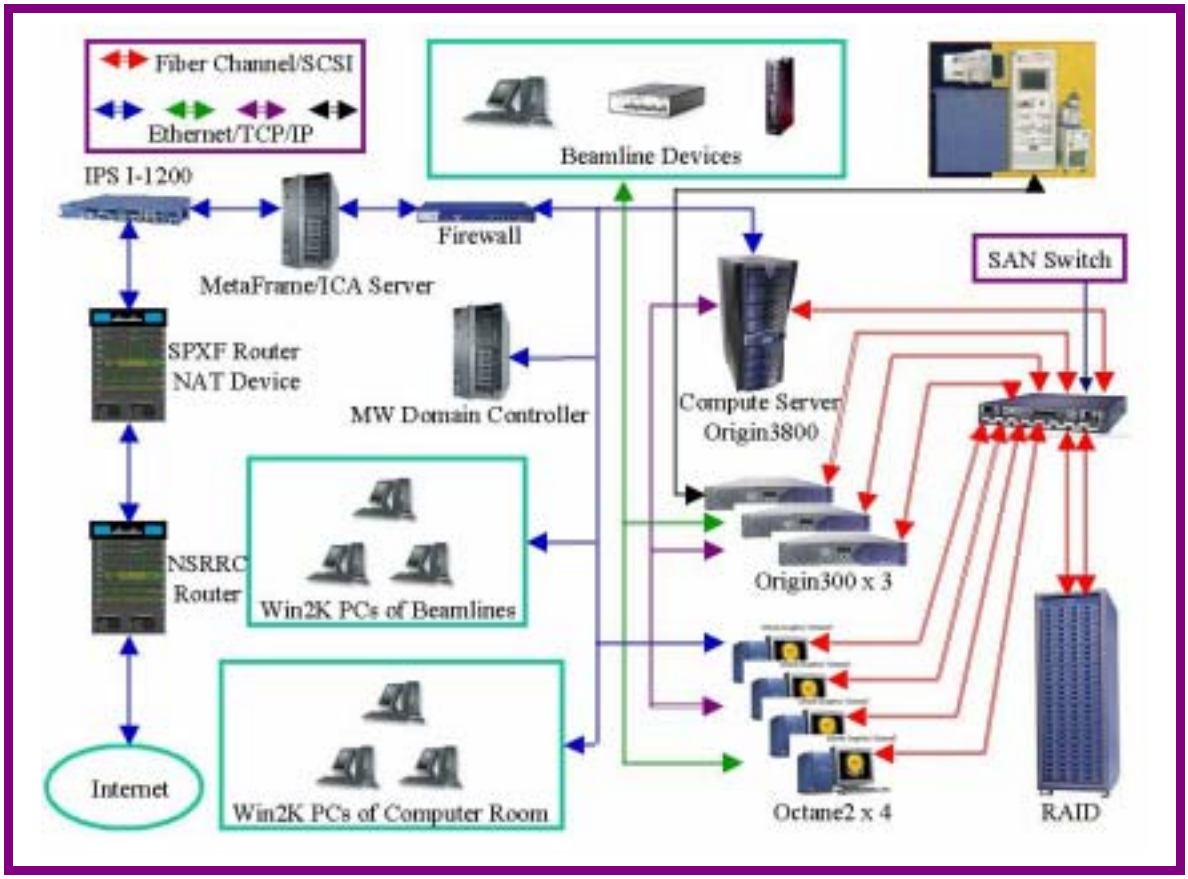
CXFS 檔案系統除了保有 XFS 的特性外，還具備以下特性：

◆ CXFS 檔案系統是掛載 (mounted) 在整個叢集 (cluster) 之上的，因此叢集中所有平台的主機都可以存取檔案系統中所有的資料；

- 所有檔案的解釋資料 (metadata，關於檔案的資訊，例如檔案名稱，大小，位置，存取權限等等) 由解釋資料伺服器 (metadata server) 管理。每一個 CXFS 檔案系統只有一個有效的解釋資料伺服器；若叢集中同時存在多個 CXFS 檔案系統，則這個叢集可以同時擁有多個有效的解釋資料伺服器；
- 叢集中所有主機都可以同時讀寫同一個檔案，而不會造成資料毀損；
- 當解釋資料伺服器發生異常 (如當機) 時，可由叢集中另一個節點接管伺服器的工作，且不會影響到其他正在進行的工作 (如資料寫到硬碟的上動作)；
- 支援 SGI IRIX、Solaris 8、Solaris 9、Windows NT 4.0、Windows 2000、64-bit Linux for SGI、Red Hat 7.3，以及 IBM AIX 5L 等系統。

5.電腦網路架構

根據上述結論，整個 SPXF 高效能計算及高速資料網路系統的架構如下圖所示：



圖中紅線部份代表以光纖為傳輸媒體的 SCSI Data Network，傳輸速率為每秒 100MByte。除了多工模式下能達到又快又穩且容忍錯誤外，最特殊的是使用一叢集式檔案系統 CXFS，SPXF 的 11 台 SGI/IRIX 電腦可同時存取共用此一檔案系統，大幅降低用戶使用上的負擔。

紫線部份代表以同軸電纜為傳輸媒體的 TCP/IP CXFS Metadata Network，傳輸速率為每秒 100MBit。此網路是 SCSI CXFS Data Network 用來交換 Metadata 的專用網路，不可以有不屬 CXFS 系統的電腦存在，以免 CXFS 受其他電腦間的訊息/資料傳輸的干擾而混亂失效。

黑線部份代表以光纖為傳輸媒體的 TCP/IP Detector Network，傳輸速率為每秒 1000MBit。此網路上只有 Detector Frame Grabbers 及 Data Collection Computer，以確保 Detector 持續後傳的數據不會受到網路上其他電腦間的訊息/資料傳輸的干擾而中斷。

綠線部份代表以同軸電纜為傳輸媒體的 TCP/IP BluIce Network，傳輸速率為每秒 1000MBit。此網路包含所有實驗相關的硬體控制器及控制電腦，透過此網路可控制 Beamline 及 Station 上的所有設備。

藍線部份代表以同軸電纜為傳輸媒體的 TCP/IP Remote Access Network，傳輸速率為每秒 100MBit。是 SPXF 內部電腦存取 Internet 的網路。為了保護昂貴的硬體設備，此網路有防火牆、防入侵偵測系統、及 ICA MetaFrame Server 來強化存取上的安全。