А почему просто воткнутые сфп в коммутатор умирают?

Причины могут быть разные, основные: брак электрических компонентов и деградация лазера. Как правило, брак компонентов трансивера проявляется в краткосрочной перспективе - появляются ошибки, либо трансивер вовсе не работает, а вторая причина считается естественным износом оборудования. Естественный износ может быть значительно ускорен при эксплуатации модулей в ненадлежащих условиях

А отечественные трансиверы это прямо льют корпуса в россии, делают платы и напаивают компоненты всё в россии?

Российский трансивер на мой взгляд — это напаивание компонентов и сборка в корпус, думаю, подтвержденные российские производители именно так и делают. Однако у каждого производителя свои тех процессы, которые могут отличаться как в большую степень локализации производства, так и в меньшую. Если интересует правовая точка зрения на данный вопрос – можно изучить ПП № 878, часть 3.

Почему не бывают, или где прячутся wdm qsfp 40г модули?

WDM трансиверы существуют в вариации 100G. Проблема отсутствия 40G таких трансиверов - в нерентабельности. Для подобной передачи необходимо использовать модуляцию PAM4. В отличии от 100G, где подобная технология востребована благодаря подключениям 400G - 4x100G (FR1, DR1 etc.) - для 40г потребности в модуляции PAM4 нет, соответственно нет чипов, которые могли бы модулировать сигнал на 40г в одну лямбду.

В чем разница wdm и bidi модулей?

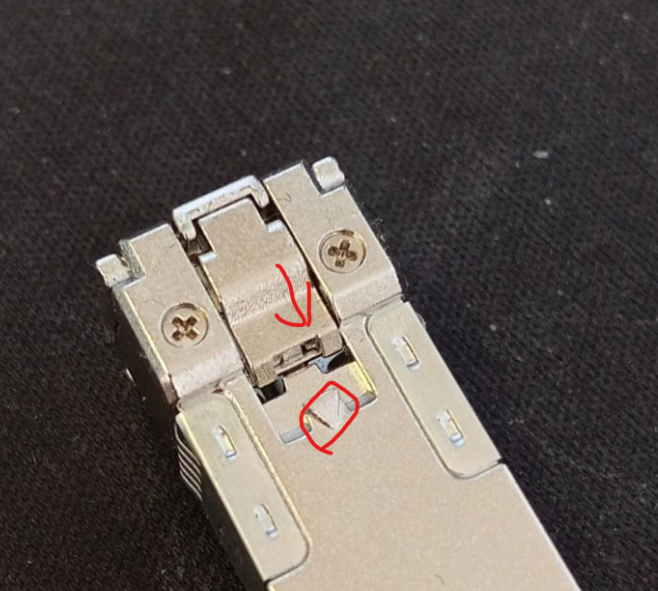
WDM - сокращение от Wavelength Division Multiplexing (Спектральное уплотнение каналов) Bidi - bidirectional (двунаправленный). Это два разных слова обозначающие одну технологию, т.е. разницы нет.

Почему иногда модули застревают в sfp портах, это некачественные модули или разъемы?

Причина может быть как в разъемах, так и в трансиверах. Либо пластина в порту слишком изношена и намертво фиксирует любой модуль, так же механизм на трансивере может быть сделан таким образом, при котором отгибающее давление будет небольшим и на портах с жесткой фиксирующей пластиной достать модуль будет проблематично.

Приложу фотографии для понимания механики процесса. Модуль фиксируется в порту SFP гибкой пластинкой с отверстием, при извлечении специальный механизм двигается по направлению стрелочки на втором фотографии и отгибает пластинку, позволяя фиксатору на SFP вылезти из отверстия.

Небольшой лайфхак по извлечению застрявших трансиверов: если язычок оторван - можно подогнуть эту пластинку снизу модуля, чем то небольшим и плоским, например, плоской отверткой, модуль при этом нужно аккуратно тянуть на себя и немного вверх.  
Главное не прилагать чрезмерных усилий, иначе фиксатор сделает загиб на пластинке (на первой фотке хороший пример, пластинка достаточно сильно погнута) и пользоваться портом будет некомфортно (будет застревать больше модулей)

Есть ли реальная экономия при строительстве оптической сети внутри зданий на мм волокне и модулях по сравнению с см?

Точный ответ, к сожалению, дать не могу, т.к. не занимаюсь проектированием сетей и оценкой стоимости, однако, в связи с растущими требованиями к многомодовым патчкордам (появление стандартов ОМ4 и ОМ5) - стоимость постройки может превысить аналогичную сеть на одномодовых трансиверах. Однако, сами трансиверы стоят дешевле одномодовых аналогов, теоретически, это более выгодно в долгосрочной перспективе на малых расстояниях.

Почему задержка сигналов на медных портах свитчей ниже, чем на оптических, но на медных сфп рж45 модулях выше, чем на оптических?

По порядку  
Почему задержка на медных портах ниже, чем на оптике:  
Медные порты в свитчах часто интегрированы непосредственно в чипсет коммутатора, а так же нет необходимости преобразовывать электрический сигнал в формат, пригодный для SFP.  
Почему оптика быстрее RJ45 SFP:  
Rj45 получает сигнал в формате под SFP, по одной линии, в формате, пригодном для оптики. И после этого формат, пригодный для оптики необходимо преобразовать обратно в формат, для передачи по 4 парам.

Зачем нужно фирмваре в модулях еппром, и почему модули разных производителей не совместимы или не корректно работают с разными коммутаторами и сетевыми картами, если модули — это операционные усилители со светодиодом и фотодиодом?

Модули — это не совсем ОУ со светодиодом и фотоприемником.   
Технологии развиваются, позволяя использовать в одном стандартизированном порту (SFP, QSFP, XFP etc.) модули, требующие различных условий от порта коммутатора. Внутренняя память несет в себе много различных функций, например:  
информация о производителе, артикуле и S/N - позволяет идентифицировать модуль в порту не доставая его из коммутатора физически. Информация с датчиков - температура, напряжение, ток на лазере и т.д.   
С помощью eeprom осуществляются функции управления трансивером: некоторые чипы требуют переключения режима (rate select) с помощью управляющих байтов. В стандартах SFF описано много различных функций, которые могут быть реализованы. Ну и конечно, поставщики оборудования пользуются полями vendor specific для внедрения защит от использования сторонних трансиверов в своем оборудовании.  
Несмотря на общую стандартизированной форм-факторов, начинка модуля может сильно отличаться от завода к заводу. И некоторые реализации, к сожалению, бывают не совместимы с определенным оборудованием.

Почему модули sfp+ не могут работать на гигабите?

Неоднозначный вопрос. На самом деле большое количество SFP+ вполне могут работать в режиме 1G. Все зависит от поддержки скорости чипом, который установлен в SFP модуль, а также поддержки переключения скоростей коммутатором. Не каждый коммутатор даст понизить скорость порта при установленной SFP+ с 10G до 1G. В таком оборудовании, модуль заработает только при установке на него прошивки от гигабитного трансивера.

Какое предельное расстояние для sfp+ модулей 10g (wdm и dual).

Из личного опыта - удавалось запускать без ошибок двухволоконные 1550nm модули на 120км, одноволоконные 49/55 - на 100км.