Acknowledgements

Joint work with:

Benjamin P. Smith  
*University of Toronto*

Andrew Hunt and John Lodge  
*Communications Research Centre, Ottawa*

Arash Farhood  
*Cortina Systems Inc., Sunnyvale/Ottawa*

Thank you to Drs. Nader Alagha and Fernando Kuipers for the invitation!
Physics: Enabling Technologies

1. Low-loss optical fiber (∼54 THz bandwidth)
2. Optical amplifiers
3. Laser transmitters and Mach-Zehnder modulators
Fiber-Optic Communication Systems: Challenges

Reliability

\[ P_e < 10^{-15} \]
Fiber-Optic Communication Systems: Challenges

**Reliability**

\[ P_e < 10^{-15} \]

**Speed**

100 Gb/s per-channel data rates
<table>
<thead>
<tr>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_e &lt; 10^{-15}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Gb/s per-channel data rates</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-Linearity</th>
</tr>
</thead>
<tbody>
<tr>
<td>The fiber-optic channel is non-linear in the input power</td>
</tr>
</tbody>
</table>
# Outline of Talk

## Part I

**Binary Error-Correcting Codes for High-Speed Communications**

- Syndrome-based iterative decoding
- Performance optimization for product-like codes
- Staircase codes
- FPGA-based simulation results
- Analytical error-floor predictions

## Part II

**Spectrally-Efficient Fiber-Optic Communications**

- Memoryless capacity estimates, with and without digital backpropagation
- Pragmatic coded modulation via staircase codes
Part I

Binary Error-Correcting Codes for High-Speed Communications

most reliable

least reliable

Π

Π

Π
This part is about . . .

- FEC for the binary symmetric channel (BSC)
- After optical (and/or) electrical compensation, suitable as forward-error-correction (FEC) for 100Gb/s PD-QPSK systems \textit{without} soft information
This part is about . . .

- FEC for the binary symmetric channel (BSC)
- After optical (and/or) electrical compensation, suitable as forward-error-correction (FEC) for 100Gb/s PD-QPSK systems without soft information
Existing Solutions
Performance Measure

Net Coding Gain

Given a particular BER, we can obtain the corresponding $Q$ via

$$Q = \sqrt{2} \text{erfc}^{-1} \left( 2 \cdot \text{BER} \right).$$

The coding gain (CG) of an error-correcting code is defined as

$$\text{CG (in dB)} = 20 \log_{10} \left( \frac{Q_{\text{out}}}{Q_{\text{in}}} \right),$$

and the Net Coding Gain (NCG) is

$$\text{NCG (in dB)} = \text{CG} + 10 \log_{10}(R),$$

where $R$ is the rate of the code.
Reed-Solomon (255,239) Code

- Coding Symbols are bytes
- Depth-16 interleaving corrects (some) bursts up to 1024 bits
- Coding Gain = 6.1 dB
- Net Coding Gain = 5.8 dB

Constraint: Rate and framing structure fixed for future generations
Concatenated Codes

Product-like codes with algebraic component codes
<table>
<thead>
<tr>
<th>Code</th>
<th>NCG</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.2</td>
<td>8.88 dB @ $10^{-15}$</td>
<td>Outer RS, Inner CSOC</td>
</tr>
<tr>
<td>I.3</td>
<td>8.99 dB @ $10^{-15}$</td>
<td>Outer BCH ($t = 3$), Inner BCH ($t = 10$)</td>
</tr>
<tr>
<td>I.4</td>
<td>8.67 dB @ $10^{-15}$</td>
<td>Outer RS, Inner BCH ($t = 8$)</td>
</tr>
<tr>
<td>I.5</td>
<td>8.5 dB @ $10^{-15}$</td>
<td>Outer RS, Inner Product ($t = 1$)</td>
</tr>
<tr>
<td>I.6</td>
<td>8.02 dB @ $10^{-15}$</td>
<td>LDPC</td>
</tr>
<tr>
<td>I.7</td>
<td>8.09 dB @ $10^{-15}$</td>
<td>Outer BCH ($t = 4$), Inner BCH ($t = 11$)</td>
</tr>
<tr>
<td>I.8</td>
<td>8.00 dB @ $10^{-15}$</td>
<td>RS(2720,2550)</td>
</tr>
<tr>
<td>I.9</td>
<td>8.63 dB @ $7 \cdot 10^{-14}$</td>
<td>~Product BCH ($t = 3$), Erasure Dec?</td>
</tr>
</tbody>
</table>
Objectives

**Increased Net Coding Gain**

NCG ⇒ Shannon Limit of BSC (9.97 dB at $10^{-15}$)

**Error Floor**

- Error floor $\ll 10^{-15}$
- Lower error floor ⇒ ‘Insurance’ in presence of correlated errors

**Block Length**

$n \approx 2 \cdot 10^6$ or less

**Low Implementation Complexity**

Dataflow considerations at 100Gb/s
Implementation Considerations
Hardware Considerations: Product vs. LDPC

**Product Code**
- Algebraic component codes
- Syndrome-based decoding

**LDPC Code**
- SPC component codes
- Belief propagation decoding
The codewords of a linear \((n, k)\) block code \(C\) are the solutions of a homogeneous system of simultaneous linear equations, i.e.,

\[
v \in C \iff vH^\top = 0.
\]

Let \(v \in C\) be sent and let \(r\) be received. Define the \textit{error pattern} \(e\) as \(e = r - v\) so that

\[
r = v + e.
\]

Then the \textit{syndrome} corresponding to \(r\) is

\[
s = rH^\top = (v + e)H^\top = vH^\top + eH^\top = eH^\top.
\]

An error-correcting decoder must infer the most likely error pattern from the syndrome.
Decoding an \((n, k)\) component codeword

- \(n\) received symbols \(\Rightarrow n - k\) symbol syndrome
- \(R = \frac{239}{255}, n \approx 1000, n - k \approx 32\)
  - For high-rate codes, syndromes provide compression

\(~ 3 \) decodings/component
\(~ \leq 96 \) bits/decoding
\(~ \frac{2}{1000} \) components/symbol
\Rightarrow 0.768 \) bits/symbol

Total Dataflow
At 100Gb/s, 76.8 Gb/s internal dataflow
LDPC Belief Propagation Decoding

- ~ 15 iterations
- 2 messages/iteration · edge
- ~ 5 bits/message
- ~ 3 edges/symbol

⇒ 450 bits/symbol

Total Dataflow
At 100Gb/s, 45Tb/s internal dataflow!
Dataflow Comparison

$$76.8 \text{ Gb/s} \ll 45 \text{ Tb/s}$$

2–3 orders of magnitude (huge implementation challenge for soft message-passing LDPC decoders).
Our Solution
Coding with Algebraic Component Codes

Graph Optimization

Degrees of Freedom

- Mixture of component codes (e.g., Hamming, BCH)
- Multi-edge-type structures
Consider a sequence of $m$-by-$m$ matrices $B_i$

$$B_{-1} \quad B_0 \quad B_1 \quad B_2 \quad \cdots$$

and a linear, systematic, $(n = 2m, k = 2m - r)$ component code $C$

Encoding Rule

$$\forall i \geq 0, \text{ all rows of } [B_{i+1}^TB_i] \text{ are codewords in } C$$
Staircase Codes: Construction

\[ B_{-1} \]

\[ B^T_0 \]

\[ B^T_2 \]

\[ B_1 \]

\[ B_3 \]
Staircase Codes: Construction

\[ B_{-1} \]

\[ B_0^T \]

\[ B_1 \]

\[ B_2^T \]

\[ B_3 \]
Staircase Codes: Construction

$B_{-1}$

$B_T^0$

$B_T^2$

$B_1$

$B_3$
Staircase Codes: Construction

\[ B_{-1} \]
\[ B_{0}^{T} \]
\[ B_{1} \]
\[ B_{2}^{T} \]
\[ B_{3} \]
Staircase Codes: Construction

\[ B_{-1} \]
\[ B_0^T \]
\[ B_1 \]
\[ B_2^T \]
\[ B_3 \]
Staircase Codes: Construction

\[ B_{-1} \]

\[ B^T_0 \]

\[ B^T_2 \]

\[ B_1 \]

\[ B_3 \]
Staircase Codes: Construction

\[
B_{-1} \quad B^{T} \quad B_{0}^{T} \quad B_{1} \quad B_{2}^{T} \quad B_{3}
\]
Staircase Codes: Properties

- Hybridization of recursive convolution coding and block coding
  - Recurrent Codes of Wyner-Ash (1963)

\[ R = 1 - \frac{r}{m} \]

- Variable-latency (sliding-window) decoder
The multiplicity of (minimal) stalls of size \((t + 1) \times (t + 1)\) is

\[
K = \binom{m}{t+1} \cdot \sum_{j=1}^{t+1} \binom{m}{j} \binom{m}{t+1-j},
\]

and the corresponding contribution to the error floor, for transmission over a binary symmetric channel with crossover probability \(p\), is

\[
\text{BER}_{\text{floor}} = K \cdot \frac{(t + 1)^2}{m^2} \cdot p^{(t+1)^2}.
\]
General Stall Patterns

\[ (5,5)\text{-stall} \]

<table>
<thead>
<tr>
<th>( K )</th>
<th>( L )</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>( 3.55 \times 10^{-21} )</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>( 7.81 \times 10^{-28} )</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>( 2.54 \times 10^{-22} )</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>( 2.21 \times 10^{-28} )</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>( 1.40 \times 10^{-23} )</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>( 1.49 \times 10^{-29} )</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>( 8.53 \times 10^{-25} )</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>( 1.83 \times 10^{-32} )</td>
</tr>
</tbody>
</table>

Contribution of \((K, L)\)-stalls, \( p = 4.8 \times 10^{-3} \).
Multi-Edge-Type Representation

\[ \Pi_B \]

most reliable

\[ B_{i-1} \quad B_i \quad B_{i+1} \quad B_{i+2} \quad B_{i+3} \]

C

\[ \Pi \]

least reliable

C
Movie Time

\[ B_{i-1} \quad B_i \quad B_{i+1} \quad B_{i+2} \quad B_{i+3} \quad B_{i+4} \]
Braided Block Codes: Construction

Our experiments show:

- For high-rate (239/255), braided block codes yield approximately the same performance as the corresponding product code.
- “Multi-edge gain” diminished, since fewer bits (previous parity only) serve as “clean”.

But:

FPGA-based Simulation Results

Code Parameters

- \( m = 510, \ r = 32 \), triple-error-correcting BCH component code
Part II
Spectrally-Efficient Fiber-Optic Communications
The Kerr Effect

Kerr electro-optic effect (DC Kerr effect)

- An effect discovered by John Kerr in 1875
- It produces a change of refractive index in the direction parallel to the externally applied electric field
- The change of index is proportional to the square of the magnitude of the external field
The Kerr Effect in Optical Fibers

**Optical Kerr Effect (or AC Kerr effect)**

- No externally applied electric field is necessary
- The signal light itself produces the electric field that changes the index of refraction of the material (fused silica)
- The change in index in turn changes the signal field
- The change in index of refraction is proportional to the square of the field magnitude

A signal in a certain frequency band can distort the signal in a different frequency band *without spectral overlap*
Nonlinear Effects in Fibers

NL=nonlinear; SPM=self-phase modulation; MI=modulation instability; XPM=cross-phase modulation; IXPM=intrachannel XPM
FWM=four-wave mixing; IFWM=intrachannel FWM; WDM=wavelength division multiplexing
- coherent fiber-optic communication system
- standard-single-mode fiber
- ideal distributed Raman amplification

But, could also consider

- systems with inline dispersion-compensating fiber
- lumped amplification
Generalized Nonlinear Schrödinger Equation

- \( A(z, t) \) is the complex baseband representation of the signal (the full field, representing co-propagating DWDM signals
- Transmitter sends \( A(0, t) \)
- Receiver gets \( A(L, t) \), where \( L \) is the total system length

**Evolution of \( A(z, t) \)**

The generalized non-linear Schrödinger (GNLS) equation expresses the evolution of \( A(z, t) \):

\[
\frac{\partial A}{\partial z} + \frac{j\beta_2}{2} \frac{\partial^2 A}{\partial t^2} - j\gamma |A|^2 A = n(z, t).
\]

No loss term (since ideal distributed Raman amplification is assumed).

\( n(z, t) \) is a circularly symmetric complex Gaussian noise process with autocorrelation

\[
\mathcal{E} [n(z, t)n^*(z', t')] = \alpha h\nu_s K_T \delta(z - z', t - t'),
\]
### System Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Second-order dispersion $\beta_2$</td>
<td>$-21.668 \text{ ps}^2/\text{km}$</td>
</tr>
<tr>
<td>Loss $\alpha$</td>
<td>$4.605 \times 10^{-5} \text{ m}^{-1}$</td>
</tr>
<tr>
<td>Nonlinear coefficient $\gamma$</td>
<td>$1.27 \text{ W}^{-1}\text{km}^{-1}$</td>
</tr>
<tr>
<td>Center carrier frequency $\nu_s$</td>
<td>$193.41 \text{ THz}$</td>
</tr>
<tr>
<td>Phonon occupancy factor $K_T$</td>
<td>$1.13$</td>
</tr>
</tbody>
</table>
Solving the GNLS Equation

Throughout propagation over an optical fiber, stochastic effects (noise), linear effects (dispersion) and nonlinear effects (Kerr nonlinearity) interact.

Even in the absence of noise, solving the GNLS equation requires numerical techniques.
Split-Step Fourier Method

- divide fiber length into short segments
- consider each segment as the concatenation of (separable) nonlinear and linear transforms
- for distributed amplification, an additive noise is added after the linear step.

\[ A(z_0, t) \rightarrow A(z_0 + h, t) \quad \text{step size } h \]
In the absence of linear effects, the GNLS equation has the form

\[ \frac{\partial A}{\partial z} = j \gamma |A|^2 A, \]

with solution

\[ A(z_0 + h, t) = A(z_0, t) \exp(j \gamma |A(z_0, t)|^2 h). \]
Linear Step

We now use the previous solution as the input to the linear step, i.e., let

$$\tilde{A}(z_0, t) = A(z_0, t) \exp(j\gamma|A(z_0, t)|^2 h)$$

be the input to the linear step. The linear form of the GNLS equation is

$$\frac{\partial A}{\partial z} = -\frac{\alpha}{2} A - \frac{j\beta_2}{2} \frac{\partial^2 A}{\partial t^2},$$

which can be efficiently solved in the frequency domain. Defining

$$A(z, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \tilde{A}(z, \omega) \exp(j\omega t) d\omega,$$

it can be shown that

$$\tilde{A}(z_0 + h, \omega) = \tilde{A}(z_0, \omega) \exp \left( \left( j \frac{\beta_2}{2} \omega^2 - \frac{\alpha}{2} \right) h \right).$$
Putting this together, we have

\[ A(z_0 + h, t) = \mathcal{F}^{-1} \left\{ \mathcal{F} \left\{ \hat{A}(z_0, t) \right\} \exp \left( \left( j \frac{\beta^2}{2} \omega^2 - \frac{\alpha}{2} \right) h \right) \right\}, \]

where \( \mathcal{F} \) is the Fourier transform operator, and where

\[ \hat{A}(z_0, t) = A(z_0, t) \exp(j \gamma |A(z_0, t)|^2 h) \]
Digital backpropagation = split-step Fourier method, using a negative step-size $h$, performed at the receiver.

Full compensation (involving multiple WDM channels) generally impossible, even in absence of noise (due to wavelength routing).

Noise is neglected (cf. zero-forcing equalizer).

Single-channel backpropagation typically performed, after extraction of desired channel using a filter.
Channel \( l \) signal:

\[
X_l(t) = \sum_{k=-\infty}^{\infty} \frac{\phi_{k,l}}{\sqrt{T_s}} \text{sinc} \left( \frac{t - kT_s}{T_s} \right),
\]

where \( \text{sinc}(\theta) = \frac{\sin \pi \theta}{\pi \theta} \).

\( \phi_{k,l} \) are elements of a discrete-amplitude continuous-phase input constellation \( \mathcal{M} \), i.e., for \( N \) rings, \( \theta \in [0, 2\pi) \), and \( r \geq 0 \),

\[
\mathcal{M} = \{ m \cdot r \exp (j\theta) | m \in \{1, 2, \ldots, N\} \}.
\]

Each ring is assumed equiprobable, and for a given ring, the phase distribution is uniform.
Multi-channel systems

In the general case of a multi-channel system having \(2B + 1\) channels with a channel spacing \(1/T_s\) Hz, the input to the fiber has the form

\[
A(z = 0, t) = \sum_{k=-\infty}^{\infty} \sum_{l=-B}^{B} \phi_{k,l} \frac{\sin \left( \frac{t - kT_s}{T_s} \right)}{\sqrt{T_s}} e^{j2\pi lt/T_s}.
\]
Back-rotation:

\[ \tilde{\phi}_{k,l} = \hat{\phi}_{k,l} \exp \left( -j(\Phi_{XPM} + \angle \phi_{k,l}) \right), \]

where \( \Phi_{XPM} \) is a constant (input-independent) phase rotation contributed by cross-phase modulation (XPM).
Gaussian fitting

For each $i$ and a fixed $l$ (the channel of interest), we calculate the mean $\mu_i$ and covariance matrix $\Omega_i$ (of the real and imaginary components) of those $\tilde{\phi}_{k,l}$ corresponding to the $i$-th ring, and model the distribution of those $\tilde{\phi}_{k,l}$ by $\mathcal{N}(\mu_i, \Omega_i)$.

From the rotational invariance of the channel, the channel is modeled as

$$f \left( y \mid x = r \cdot i \exp(j\phi) \right) \sim \mathcal{N}(\mu_i \exp(j\phi), \Omega_i),$$

where the (constant) phase rotation due to $\Phi_{\text{XPM}}$ is ignored, since it can be canceled in the receiver.
The mutual information of the (assumed) memoryless channel is

$$I(X; Y) = \int \int f(x, y) \log_2 \frac{f(y|x)}{f(y)} \, dx \, dy,$$

where $f(x)$ represents the input distribution on $\mathcal{M}$ with equiprobable rings and a uniform phase distribution, which provides an estimate of the capacity of an optically-routed fiber-optic communication system.
## Signaling Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baud rate $1/T_s$</td>
<td>100 GHz</td>
</tr>
<tr>
<td>Channel bandwidth $W$</td>
<td>101 GHz</td>
</tr>
<tr>
<td>Number of rings $N$</td>
<td>64</td>
</tr>
<tr>
<td>Number of channels $2B + 1 = 5$</td>
<td></td>
</tr>
</tbody>
</table>
Achievable Rates from Memoryless Capacity Estimate

(BP adds 0.55 to 0.75 bits/s/Hz relative to EQ)
Pragmatic Coded Modulation via Staircase Codes

Approach: BICM + shaping, modulation $2^{K+2}$-QAM, hard-decisions

Syndrome-former matrix

$$H^T_U = [1 + D + D^2, 1 + D^2]^T.$$
## Achievable Rates

<table>
<thead>
<tr>
<th>Fiber System</th>
<th>$K$</th>
<th>$p_{\text{avg}}$</th>
<th>$P_{\text{in}}$</th>
<th>$I_P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L = 500 \text{ km, EQ}$</td>
<td>8</td>
<td>$1.61 \times 10^{-2}$</td>
<td>$-6$</td>
<td>8.05</td>
</tr>
<tr>
<td>$L = 500 \text{ km, BP}$</td>
<td>8</td>
<td>$3.52 \times 10^{-3}$</td>
<td>$-4$</td>
<td>8.73</td>
</tr>
<tr>
<td>$L = 1000 \text{ km, EQ}$</td>
<td>6</td>
<td>$3.88 \times 10^{-3}$</td>
<td>$-6$</td>
<td>6.78</td>
</tr>
<tr>
<td>$L = 1000 \text{ km, BP}$</td>
<td>8</td>
<td>$2.22 \times 10^{-2}$</td>
<td>$-4$</td>
<td>7.77</td>
</tr>
<tr>
<td>$L = 2000 \text{ km, EQ}$</td>
<td>6</td>
<td>$2.52 \times 10^{-2}$</td>
<td>$-6$</td>
<td>5.98</td>
</tr>
<tr>
<td>$L = 2000 \text{ km, BP}$</td>
<td>6</td>
<td>$5.16 \times 10^{-3}$</td>
<td>$-4$</td>
<td>6.72</td>
</tr>
</tbody>
</table>

(These achieve within 0.4 to 0.6 bits/s/Hz of estimated channel capacity.)
Staircase code design

First, design a collection of staircase codes of various (appropriate) rates:

![Graph showing BER performance with different rates](image-url)
Then, simulate their performance on the actual channel:

Performance to within 0.62 bits of estimated capacity is achieved!
(Much of the gap is due to quantization, i.e., hard-decisions.)
<table>
<thead>
<tr>
<th>Fiber System</th>
<th>$m$</th>
<th>$t$</th>
<th>$R$</th>
<th>Spec. Eff. (bits/s/Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L = 500$ km, EQ</td>
<td>190</td>
<td>4</td>
<td>77/95</td>
<td>7.48</td>
</tr>
<tr>
<td>$L = 500$ km, BP</td>
<td>255</td>
<td>3</td>
<td>239/255</td>
<td>8.50</td>
</tr>
<tr>
<td>$L = 1000$ km, EQ</td>
<td>255</td>
<td>3</td>
<td>239/255</td>
<td>6.62</td>
</tr>
<tr>
<td>$L = 1000$ km, BP</td>
<td>144</td>
<td>4</td>
<td>3/4</td>
<td>7.00</td>
</tr>
<tr>
<td>$L = 2000$ km, EQ</td>
<td>120</td>
<td>4</td>
<td>11/15</td>
<td>5.40</td>
</tr>
<tr>
<td>$L = 2000$ km, BP</td>
<td>628</td>
<td>4</td>
<td>146/157</td>
<td>6.58</td>
</tr>
</tbody>
</table>
Conclusions

- At rate $239/255$, staircase codes provide best-in-class 9.41 dB NCG.

- For high-spectral efficiency modulation, a pragmatic coding approach using staircase codes yields performance with 0.6 bits/s/Hz (at $L = 2000$ km) with practical decoding complexity.
Future Directions

- Can these capacity estimates be improved?
- Can soft-decisions be incorporated while retaining low complexity?
- Are there (well-principled) alternatives to digital backpropagation?
Future Directions

- Can these capacity estimates be improved?
- Can soft-decisions be incorporated while retaining low complexity?
- Are there (well-principled) alternatives to digital backpropagation?

Thank you for your attention!