Some Signal Processing, Information-Theoretic, and Coding-Theoretic Aspects of Multi-Antenna Communications

Babak Hassibi

Mathematics of Communications Research
Bell Laboratories, Lucent Technlogies

April 13, 2000

General Overview

Multiple antennas provide many exciting possibilities for *high data rate* wireless communications, since they can

- significantly boost channel capacity
- lower the probability of error

of a wireless communications link. (Key: *spatial diversity*)

Applications abound and include:

- wireless LAN, fixed wireless access, mobile wireless, wireless Internet, etc.

Why Multiple Antennas?

Traditionally, it was believed that there are two ways to increase channel capacity:

- increase transmit power: $C = \log(1 + \rho)$
- increase bandwidth: $C = \lim_{B \to \infty} B \log(1 + \rho/B) = \rho$

Neither of which is particularly exciting :(

But what about mutliple antennas? Well, pre-1995:

- fading is bad, scattering environment is bad
- line-of-sight is good
- beam-forming, angle-of-arrival estimation are the way to go
- capacity grows logarithmically in number of receive antennas

Things changed around 1995 (Foschini, Telatar)

Now we know better:

- Fading is good! Rich-scattering environment is good!
- Capacity increases *linearly* in the minimum of the number of receive and transmit antennas.

This is now an exciting solution :)

Research Challenges

- signal processing
- information theory
- coding theory (space-time codes)
- experimental — the propogation environment
  - Rayleigh vs. Rician fading, rich-scattering vs. line-of-sight
- RF circuits, antenna design
- system issues, network issues, multiple access, etc.

Outline

Will consider the first three different aspects

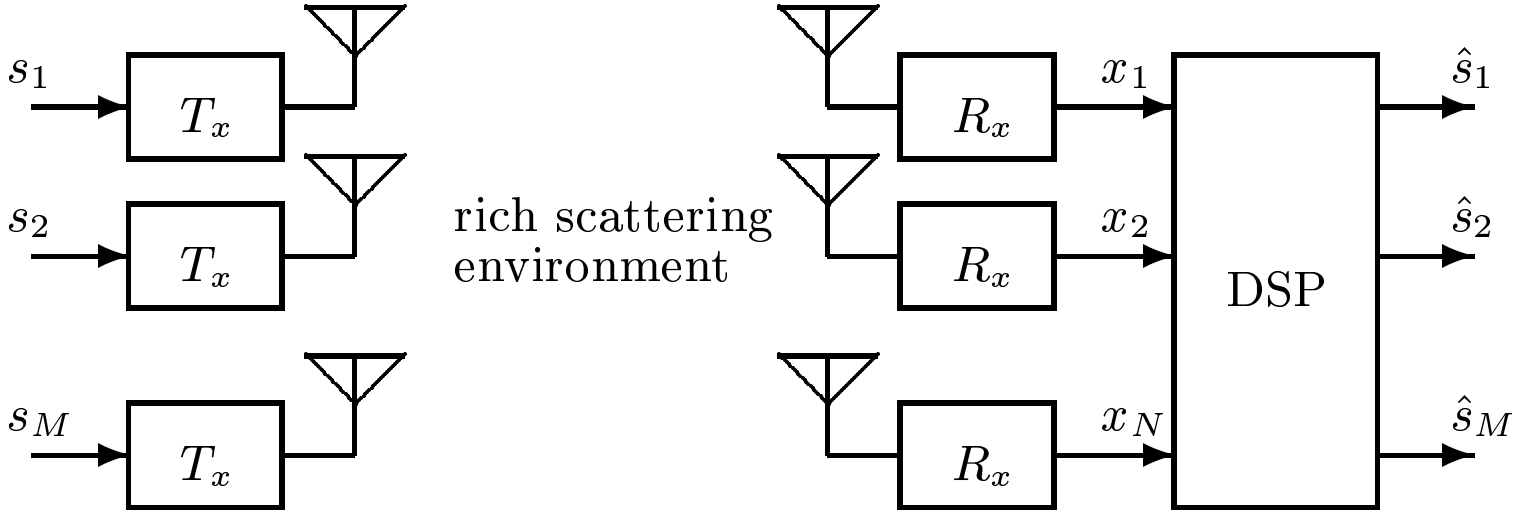
- **Signal Processing:**
  - an efficient square-root algorithm for Bell Labs Layered Space-Time (BLAST)
- **Information Theory:**
  - *autocapacity* – information transmission at Shannon capacity via coding over a *single* coherence interval
- **Coding Theory:**
  - multi-antenna signal constellation design via group representation theory

Signal Processing

- an efficient square-root algorithm for Bell Labs Layered Space-Time (BLAST)

Basic Model for BLAST

- Consider M signals impinging on an array of N ($N \geq M$) receivers via a *rich scattering* environment.



- The transmitted signals may come from an array of transmit antennas (as in BLAST), or from M separate transmit antennas (as in the uplink of a wireless LAN, etc.).

$$x = Hs + v.$$

Basic Idea of BLAST

Since maximum-likelihood detection out of the question, Foschini et al suggested the following three-step procedure:

- **Estimate the channel matrix via training sequence**
- **Find MMSE nulling vectors and optimal detection order**
  - determine "strongest" signal (the one with the smallest MMSE) and its corresponding nulling vector
  - consider deflated channel matrix and find next "strongest" signal and nulling vector
  - continue
- **Process the payload**
  - 1. MMSE nulling, 2. slicing (decoding), 3. symbol cancellation

Computational Complexity

For simplicity, let $M = N$

- Channel estimation: $2M^2 \log_2 L_T$
- Determining the nulling vectors and optimal ordering: $27M^4/4$
  - since we must compute M pseudo-inverses
- Processing the payload: $2M^2 L_P$

To see what these numbers mean for an actual systems, consider a target of 1 Mb/s data transmission over a 30 kHz wireless channel:

- $1/T = 24.3$ ksymbol/sec, 16-QAM
- $M = N = 14$, $L_T = 32$ and $L_P = 100$
- Required DSP integrated into a single chip solution

	Flops/burst	MegaFlops/s	%
channel estimation	7,840	1.44	0.65
nulling vectors and ordering	1,036,000	190.8	86.3
payload processing	156,800	28.9	13.1
TOTAL	1,200,000	221.2	100

The dominant portion of the computation involves determining the nulling vectors and optimal ordering. *Can this computation be reduced in a numerically stable way?*

Have developed an algorithm that

- is **cost-efficient:** requires only one (implicit) pseudoinverse computation and has complexity $29M^3/3$
- is **numerically stable:** is division-free and uses only unitary transformations
- is suitable for implemention in fixed-point, rather than floating-point architectures

	Flops/burst	MegaFlops/s	%
channel estimation	7,840	1.44	2.9
nulling vectors and ordering	106,100	19.5	39.1
payload processing	156,800	28.9	58.0
TOTAL	270,400	49.8	100

Information Theory

- *autocapacity* – information transmission at Shannon capacity via coding over a *single* coherence interval

Coding for Fading Channels

Consider the single-antenna additive Gaussian noise fading channel:

$$x = \sqrt{\rho} s h + w, \quad h \sim \mathcal{CN}(0,1),\ w \sim \mathcal{CN}(0,1),\ E|s|^2 = 1$$

Assume a *block-fading* model: h is fixed for a "coherence-interval" of T time samples, after which it changes to an independent value.

- If we code over block sizes of length QT (i.e., over Q coherence intervals), then Shannon theory asserts that for all rates $R < C$, where C is the *Shannon capacity*, we can achieve

$$P_e \to 0 \ \text{ as } \ Q \to \infty.$$

- If we code over only one coherence interval (even if $T \to \infty$) we cannot achieve $P_e \to 0$ for any rate $R > 0$. (Since there is always a nonzero probability that the channel is "bad").

- In conclusion, for a fading channel, to achieve Shannon capacity, we need *temporal diversity* and hence "channel-coding".

What about multi-antenna fading channels? Assume M transmit and N receive antennas, and a coherence interval of T time samples. Thus:

$$X = \sqrt{\frac{\rho}{M}} SH + W, \qquad E\text{trace}SS^* = TM$$

where H and W are $M \times N$ matrix and $T \times N$ matrices of independent $\mathcal{CN}(0,1)$ entries, respectively, and S is the $T \times M$ signal matrix.

- Once more, if we code over Q coherence intervals, we can achieve $P_e \to 0$, as $Q \to \infty$, for all rates less than the Shannon capacity.
- It thus appears that to achieve Shannon capacity we still need to resort to channel coding. This can be extremely computationally intensive in the multi-antenna case. (We already have seen how computationally intensive things can get even without channel coding — BLAST.)

Multi-Antenna Shannon Capacity

- When H is known to the receiver, the so-called *perfect knowledge* Shannon capacity is given by:

$$C = E \log \det \left(I_N + \rho \frac{H^* H}{M} \right)$$

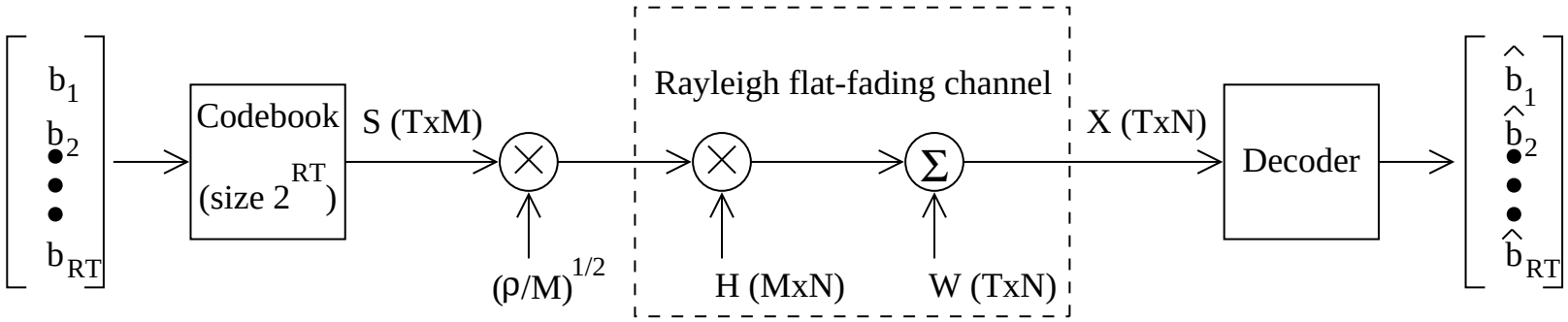
  In particular, as $M \to \infty$:

$$C = N \log(1 + \rho)!$$

- When $M \to \infty$, the assumption that the receiver knows the channel becomes less and less tenable, since we require longer and longer training sequences to identify the channel.

- Computing the Shannon capacity in the unknown channel case for an arbitrary M, N, and T is still an "open problem". As $T \to \infty$, however, the unknown channel capacity approaches the perfect knowledge capacity.

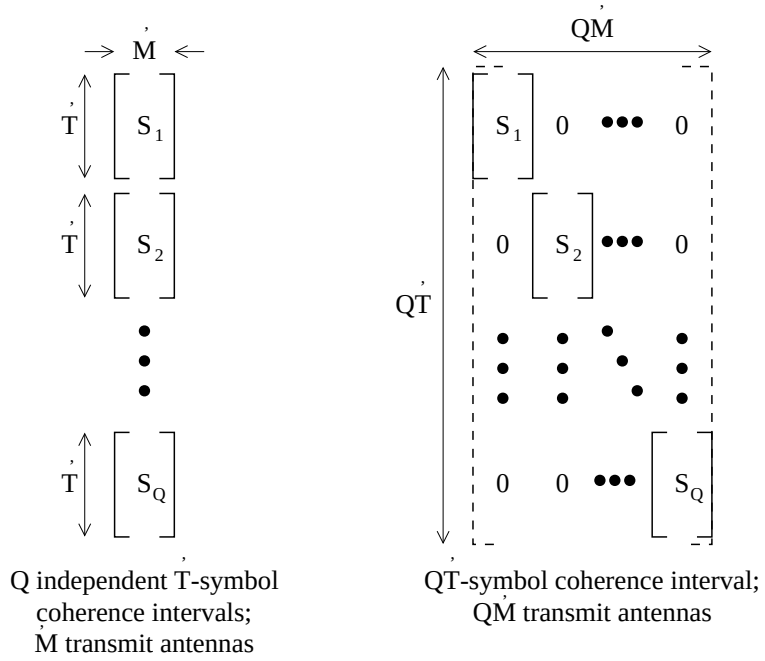
Coding over One Coherence Interval



- Hold constant:
  - total transmit power, ρ
  - number of receive antennas, N
  - $\beta = \frac{T}{M} = \dfrac{\text{coherence time}}{\text{number of transmit antennas}}$
- **Autocoding:** does there exists an *autocapacity*, C_a, such that for all $R < C_a$, we have $P_e \to 0$, as $(T, M) \to \infty$, but with $Q = 1$?
  - will not work if either T or M is fixed

Existence of Autocapacity

Is the autocapacity nonzero?



- $P_e \to 0$, as $Q \to \infty$, $\forall R < C(\rho, T, M, N)$.
- Since the block-diagonal signal structure is not necessarily optimal:

$$C_a(\rho, \beta, N) \geq \sup_{T, M : \frac{T}{M} = \beta} C(\rho, T, M, N).$$

Formula for Autocapacity

One can show that the lower bound is, in fact, tight:

$$C_a(\rho, \beta, N) = \sup_{T,M:\, \frac{T}{M}=\beta} C(\rho, T, M, N)$$

This allows us to explicitly compute the autocapacity as follows

$$C_a(\rho, \beta, N) \;\geq\; C(\rho, LT, LM, N) \;\geq\; C(\rho, LT, M, N)$$

Letting $L \to \infty$, the RHS converges to the perfect knowledge Shannon capacity. Thus

$$C_a(\rho, \beta, N) \geq E \log\det\left(I_N + \frac{\rho}{M} H^* H\right)$$

Further, letting $M \to \infty$

$$C_a(\rho, \beta, N) = N \log(1 + \rho)$$

How Many Antennas Do We Need?

Thus, irrespective of β, the value of the autocapacity is equal to the perfect knowledge Shannon capacity. *But we do not yet know* **how many** *antennas are required for the autocoding effect to kick in.*

- A *partial* answer to this question can be provided by studying the random coding exponent, or the so-called *cut-off rate*:

$$R_0(T, \beta, N, p(S)) = -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \int dX \sqrt{p(X \mid S_1) \cdot p(X \mid S_2)} \right\} \right],$$

  that allows us to explicitly bound the probability of error, P_e:

$$P_e \;\leq\; \exp\left\{-T \ln 2 \left[R_0(T, \beta, N, p(S)) - R\right]\right\}.$$

- **Remark:** This can be obtained using the *union bound* applied to the *Chernoff bound* on the pairwise probability of error.

Bounding the Cut-Off Rate

Optimizing the cut-off rate over the input density $p(S)$ appears to be intractable, so we compute it for the input densities:

- *Isotropically-distributed unitary matrix*: This is capacity-achieving when the channel is unknown in the high SNR regime.
- *Gaussian matrices with i.i.d. elements*: This is capacity-achieving when the channel is known.

In fact, for the above distributions, in addition to the cut-off rates, we have also analytically computed the random coding exponent, as well as the pairwise probability of error.

Random Matrices

The computation of these quantities requires the (asymptotic) eigenvalue distribution of various classes of random matrices — several of which were not known. The techniques involved in determining these distributions are quite interesting and have connections to

- orthogonal polynomials
- the saddlepoint method
- Wishart matrices
- Hankel operators
- hypergeometric functions

Hypergeometric Functions

"Hypergeometric functions are one of the paradises of nineteenth century mathematics that remain unknown to mathematicians of our day. Hypergeometric functions of several variables are an even better paradise: they will soon crop up in about everything."

-Gian Carlo Rota

Isotropically-Distributed Unitary Matrix

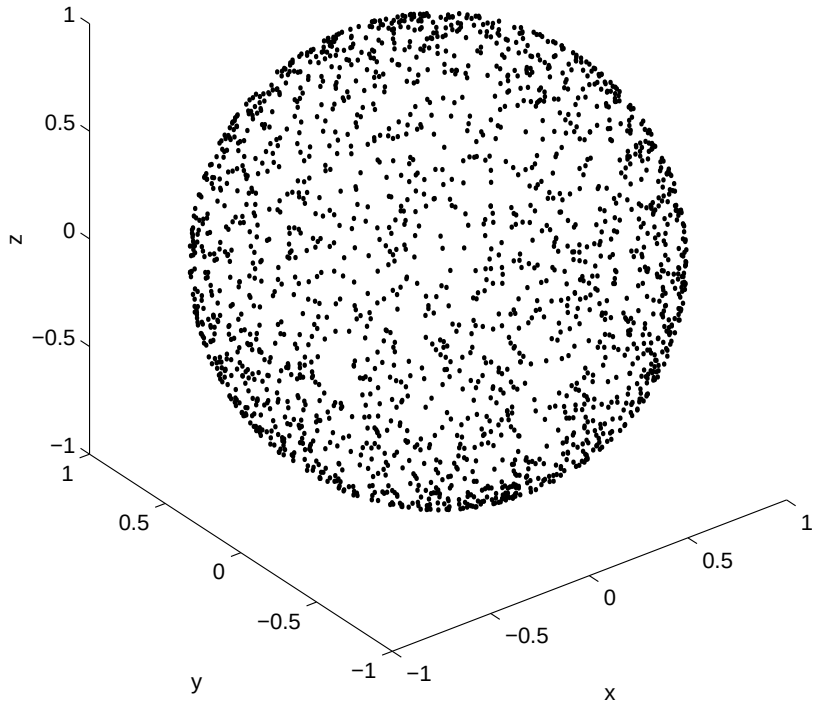
An isotropically-distributed $n \times n$ unitary matrix Ψ is one whose probability density function is invariant under pre- or post-multiplication by any fixed unitary matrix:

$$p(\Psi) = p(\Theta\Psi) = p(\Psi\Theta), \qquad \forall \Theta \text{ s.t. } \Theta\Theta^* = \Theta^*\Theta = I$$

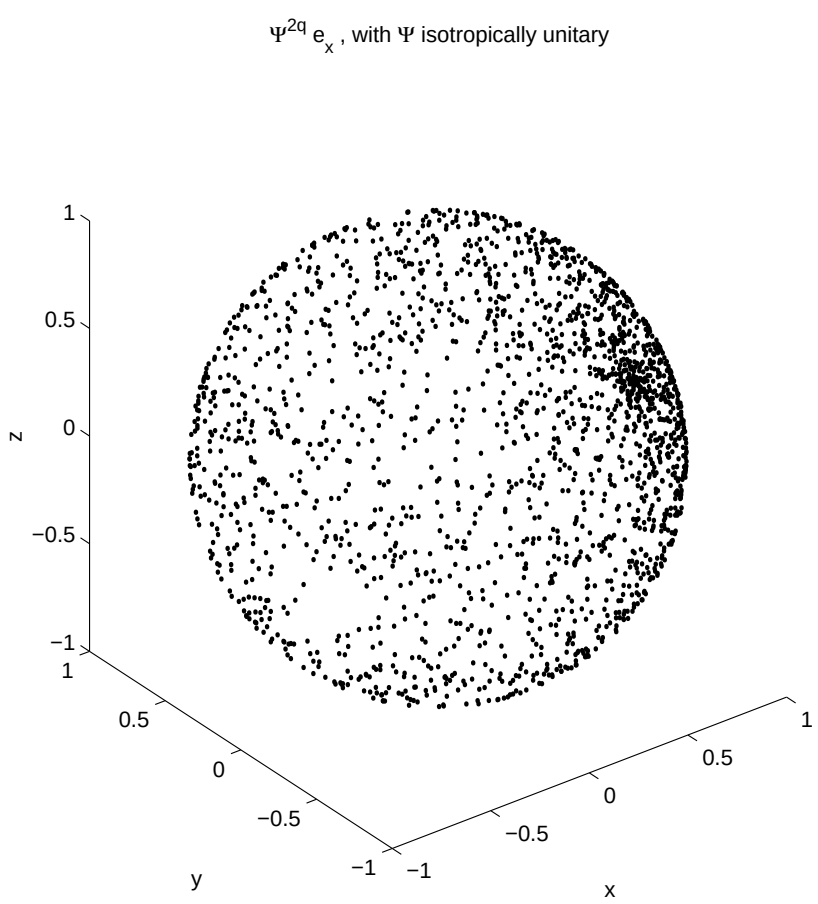
In particular, for any unit vector ξ, the vector $\Psi\xi$ is equally likely to point in any direction.

- Ψ^ℓ is **not** isotropically-distributed for $n \geq 3$ in the real case and $n \geq 2$ in the complex case
- In the complex case, Ψ^ℓ for $\ell \geq n$ has independent eigenvalues

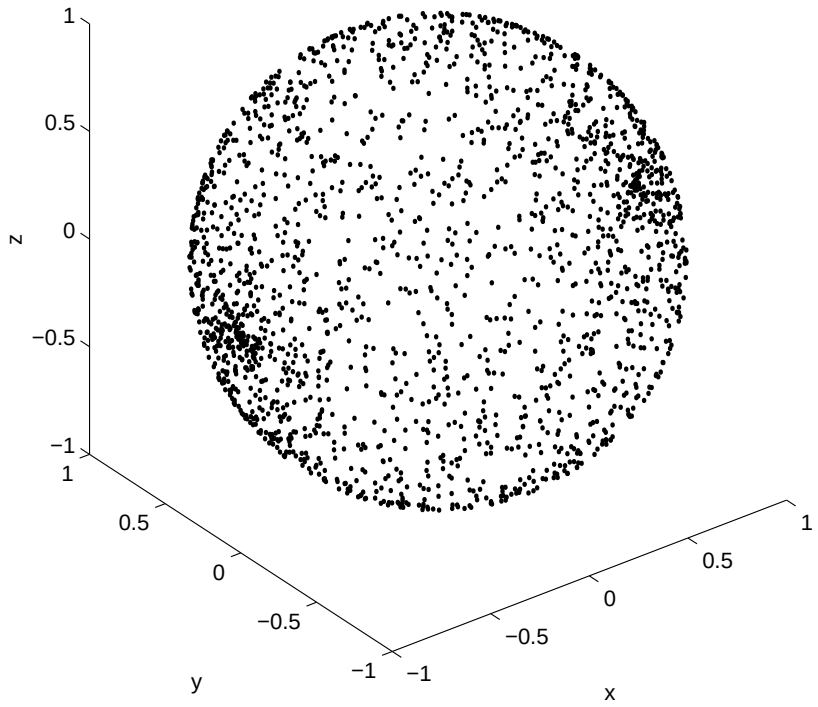
Ψe_x with Ψ Isotropically Unitary



$\Psi^2 e_x$ with Ψ Isotropically Unitary



$\Psi^3 e_x$ with Ψ Isotropically Unitary



When does Autocoding Kick In?

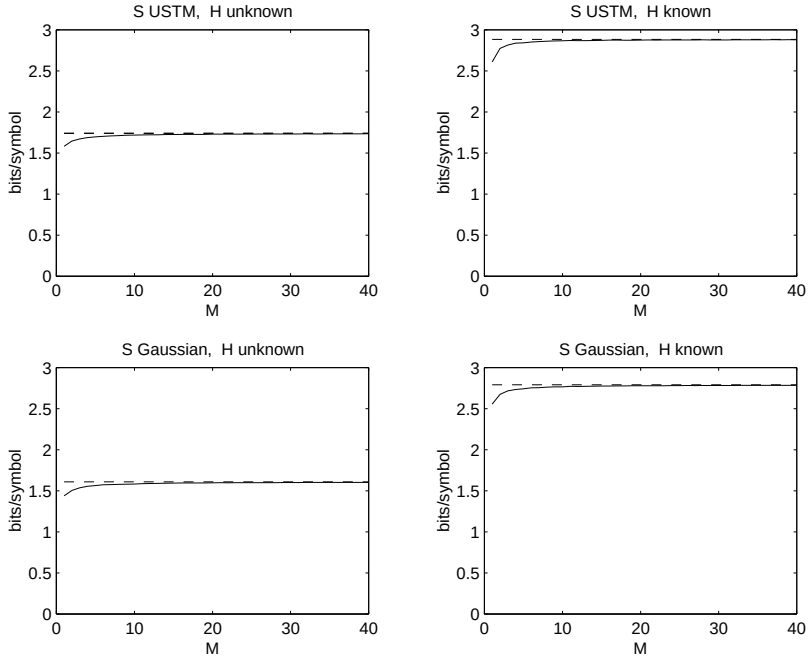


Figure 1: Cutoff rate bounds (bits/symbol) as a function of M for $\rho = 18$ dB, $\beta = 2$, and $N = 1$. The dashed lines are the asymptotic cutoff rate bounds ($M \to \infty$).

Even for small values of M the R_0 bounds are close to their asymptotic values. Thus autocoding takes effect for relatively small values of M.

Pairwise Probability of Error vs. M

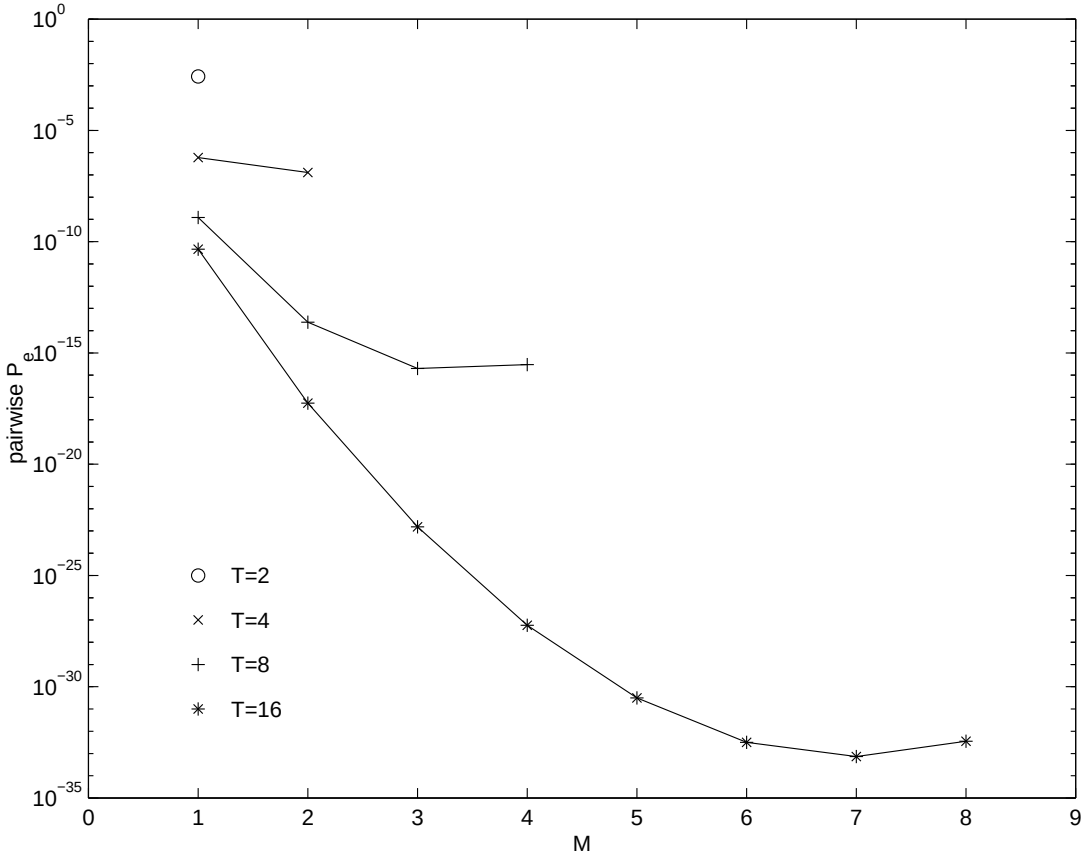


Figure 2: Pairwise probability of error for isotropicaly distributed signal for $N = 4$ receive antennas and $\rho = 18db$.

P_e vs. Transmission Rate

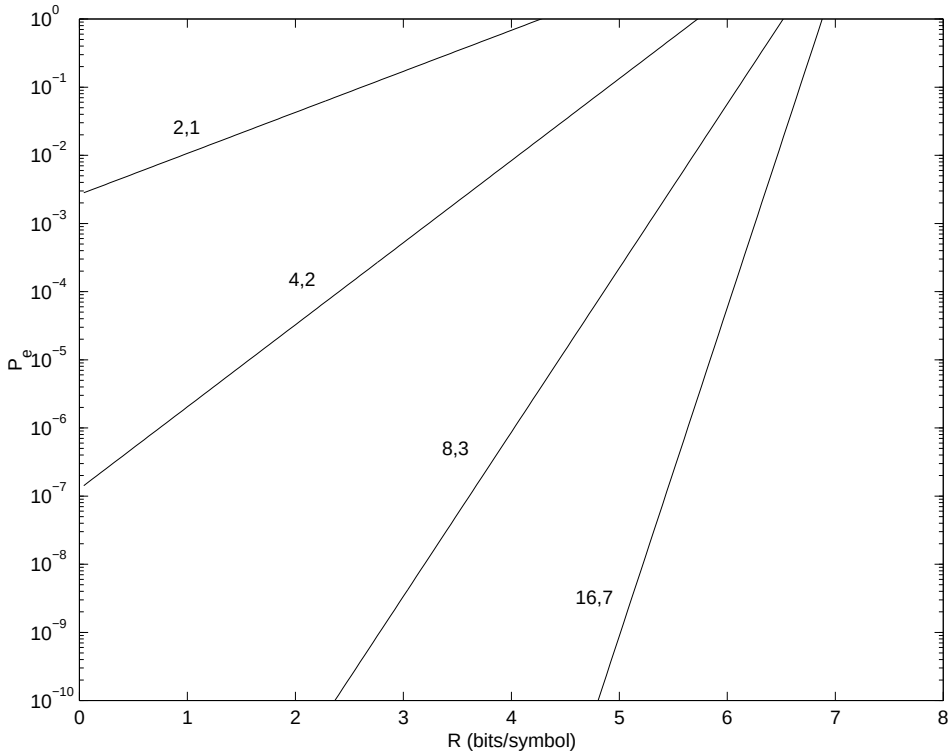


Figure 3: Upper bound on block probability of error versus transmission rate (bits/symbol) for random constellation of Unitary Space-Time signals, for $N = 4$, $\rho = 18$ dB, and $(T, M) = (2, 1), (4, 2), (8, 4), (16, 7)$.

Autocoding and Autocapacity – Summary

- Autocoding – coding within one block *without outage*. Perfect knowledge Shannon capacity can be achieved.
  - Kicks in with even relatively small (T, M).
  - Temporal diversity not needed – replaced by spatial diversity.
  - Shannon capacity for unknown channel not so important.
- Burden is shifted away from channel coding and onto decoding constellations of matrix-valued signals.
  - isotropically-distribued unitary signals are good.
- $T = 16$, $M = 7$, $N = 4$, $\rho = 18$ dB, $R = 5$, $P_e < 10^{-9}$. Constellation size is $2^{80} = 10^{24}$—yikes!
- Joint work with T. Marzetta and B. Hochwald.

Coding Theory

- multi-antenna signal constellation design via group representation theory

Constellation Design for Multiple Antennas

The preceding discussions suggest that a crucial issue in multi-antenna communications is the design of good signal constelations.

- One good candidate is a constellation of unitary signals (henceforth USTM - unitary space-time modulation)
  - In the known channel case, we shall take $T = M$ ($\beta = 1$), so that the constellation is composed of $M \times M$ unitary matrices:

$$\mathcal{V} = \{V_0, \ldots, V_{L-1}\}, \quad V_\ell V_\ell^* = V_\ell^* V_\ell = I_M$$

Multiple-Antenna Differential Modulation

In the unknown channel case, one can take $\beta = 2$ and employ a *differential-modulation* scheme as follows (Hughes 99, Hochwald and Sweldens 99):

$$S_i = V_i S_{i-1} = V_i V_{i-1} \dots V_0$$

with the same constellation, $\mathcal{V}$, as above. To see why, assume, momentarily, that there is no additive noise. Then, since the channel is constant over $T = 2M$ time samples:

$$X_i = S_i H = V_i S_{i-1} H = V_i X_{i-1}$$

so that we can decode the i-th signal V_i from X_i and X_{i-1}, without needing to know the channel matrix, H. When there is additive noise, the maximum-likelihood decoder is given by

$$\hat{V}_i = \arg \min_{\ell=0,\dots,\mathcal{L}-1} \|X_i - V_\ell X_{i-1}\|_F$$

The quality of a constellation $\mathcal{V}$ is determined by the probability of error of mistaking one symbol of $\mathcal{V}$ for another. It can be shown that, at high SNR, the probability of mistaking V_ℓ with $V_{\ell'}$ is dominantly dependant on the determinant of $V_\ell - V_{\ell'}$.

- In particular, we shall measure the quality of a constellation $\mathcal{V}$ by

$$\zeta_{\mathcal{V}} = \frac{1}{2} \min_{0 \le \ell < \ell' < \mathcal{L}} |\det(V_\ell - V_{\ell'})|^{1/M}$$

"Fully Diverse" Constellations

- Our design problem is thus reduced to the following:
  - *given M and R, find a set $\mathcal{V}$ of $L = 2^{MR}$, $M \times M$ unitary matrices, such that the minimum of the absolute value of the determinant of their pairwise differences is as large as possible.*
- We shall call any constellation $\mathcal{V}$ with the property that the determinants of the pairwise differences are all nonzero, *fully diverse.*
- The reason is the following: For any channel matrix H,

$$V_\ell H \neq V_{\ell'} H \quad \text{whenever} \quad \ell \neq \ell'$$

  In other words, there exists no channel H for which any two elements of $\mathcal{V}$ respond identically.

The Design Problem

The design problem is especially confounded for the following two reasons:

- The space of all $M \times M$ unitary matrices is very difficult to parametrize (it is the so-called complex *Steifel manifold*).
  - in particular, there are M^2 real free parameters in any $M \times M$ complex unitary matrix.
- The objective cost, i.e., the absolute value of a determinant, is *not* a norm.

Moreover, the size of the problem (we are seeking $L = 2^{MR}$ signals) makes an exact solution intractable.

Orthogonal Designs (Tarokh et al, 1998)

- Let x and y be complex numbers such that $|x|^2 + |y|^2 = 1$. Then an *orthogonal design* is the unitary matrix:

$$V = \begin{bmatrix} x & -y^* \\ y & x^* \end{bmatrix}$$

- Orthogonal designs have the property that any linear combination of them is a matrix with orthogonal columns

- In particular,

$$\det(V_1 - V_2) = \det \begin{bmatrix} x_1 - x_2 & -(y_1 - y_2)^* \\ y_1 - y_2 & (x_1 - x_2)^* \end{bmatrix} = |x_1 - x_2|^2 + |y_1 - y_2|^2$$

  Thus, the design problem reduces to the design of spherical codes on the 2-dimensional complex (or 4-dimensional real) sphere. This is a well-studied problem.

Diagonal Constellations

- The drawback of orthogonal designs is that they only exist for 2×2 matrices in the complex case, and 4×4 matrices in the real case. Thus one cannot use them to construct constellations for an arbitrary number of antennas.
- To alleviate this, diagonal constellations have been introduced:

$$V_\ell = \begin{bmatrix} e^{j2\pi u_1/L} & & & \\ & e^{j2\pi u_2/L} & & \\ & & \ddots & \\ & & & e^{j2\pi u_M/L} \end{bmatrix}^\ell,$$

  where $u_1, \ldots, u_M$ are integers.
- Optimization over $u_1, \ldots, u_M$ yields reasonably good constellations. However, much better constellations may be possible since the diagonal constraint appears to be quite restrictive.

Constellations from Groups

To break the logjam, let us investigate the case where $\mathcal{V}$ forms a group under matrix multiplication.

- This has useful practical implications (especially in the differential-modulation scheme) since matrix multiplication can be performed by table look-up.
- The design problem also simplifies somewhat under the group assumption, since

$$\begin{aligned}
\zeta_{\mathcal{V}} &= \frac{1}{2} \min_{0 \le \ell < \ell' < \mathcal{L}} \left| \det(V_\ell - V_{\ell'}) \right|^{1/M} \\
&= \frac{1}{2} \min_{0 \le \ell < \ell' < \mathcal{L}} \left| \det(V_\ell) \det\left(I - V_\ell^{-1} V_{\ell'}\right) \right|^{1/M} \\
&= \frac{1}{2} \min_{I \ne V \in \mathcal{V}} \left| \det(I - V) \right|^{\frac{1}{M}} .
\end{aligned}$$

Group Representations

Our constellation construction relies on the following crucial result:

- *any finite group, G, has a representation with unitary matrices.*

Take, for example, the L-th order cyclic group:

$$G_L = \langle \sigma | \sigma^L = 1 \rangle.$$

One representation can be obtained by representing σ as $e^{j2\pi/L}$. Other representations can be obtained by representing σ as:

$$\begin{bmatrix} e^{j2\pi u_1/L} & & & \\ & e^{j2\pi u_2/L} & & \\ & & \ddots & \\ & & & e^{j2\pi u_M/L} \end{bmatrix},$$

which yields the diagonal signal constellations introduced earlier.

Fixed-Point-Free Groups

- Although our aim is to maximize

$$\zeta_{\mathcal{V}} = \frac{1}{2} \min_{I \neq V \in \mathcal{V}} |\det(I - V)|^{\frac{1}{M}} ,$$

  it is at this point not even clear whether, or when, this quantity is zero for a given group G.

- Clearly, $\zeta_{\mathcal{V}}$ will be nonzero if and only if all the matrices $V \neq I$ have no eigenvalues at unity.

- Groups that have representations with this property are referred to as *fixed-point-free groups* — an eigenvalue at unity implies $Vx = x$, for some vector x, i.e., that V has a fixed-point.

- Therefore we need to search for fixed-point-free groups.

Zassenhaus' Characterization

Zassenhaus in 1936 gave an *almost* complete characterization of fixed-point-free groups. Here is his result for groups of odd order:

Theorem 1 (Zassenhauss) *Let G be a fixed point free group of odd order L. Then there exist integers m and r such that G is isomorphic to the group*

$$G_{m,r} = \langle \sigma, \tau \mid \sigma^m = 1, \tau^M = \sigma^t, \tau\sigma\tau^{-1} = \sigma^r \rangle,$$

where

(i) $L = mM$.

(ii) *M is the smallest integer such that* $r^M \equiv 1 \mathrm{mod}(m)$.

(iii) $\gcd(M, t) = 1$, *where* $t = \frac{m}{\gcd(r-1,m)}$.

All Odd-Order Fixed-Point-Free Groups

Using Zassenhaus' partial characterization, we have been able to show the following result.

Theorem 2 *A finite group G of odd order, L, is fixed point free if and only if it is isomorphic to a group*

$$G_{m,r} = \langle \sigma, \tau \mid \sigma^m = 1, \tau^M = \sigma^t, \tau\sigma\tau^{-1} = \sigma^r \rangle,$$

for some integers m and r such that:

(i) $L = mM$.

(ii) *M is the smallest integer such that* $r^M \equiv 1 \mathrm{mod}(m)$.

(iii) $\gcd(M, t) = 1$, *where* $t = \frac{m}{\gcd(r-1,m)}$.

(iv) *All prime divisors of M divide* $\gcd(r-1, m)$.

The Group Representation

The representation of the group takes the form:

$$\mathcal{V} = \left\{ \Delta(\sigma)^{\ell} \Delta(\tau)^{k} \, | 0 \leq \ell \leq m-1, 0 \leq k \leq M-1 \right\},$$

where

$$\Delta(\sigma) = \begin{pmatrix} \eta & 0 & \cdots & 0 \\ 0 & \eta^{r} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \eta^{r^{M-1}} \end{pmatrix}, \quad \Delta(\tau) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \eta^{t} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

and $\eta = e^{j2\pi/m}$.

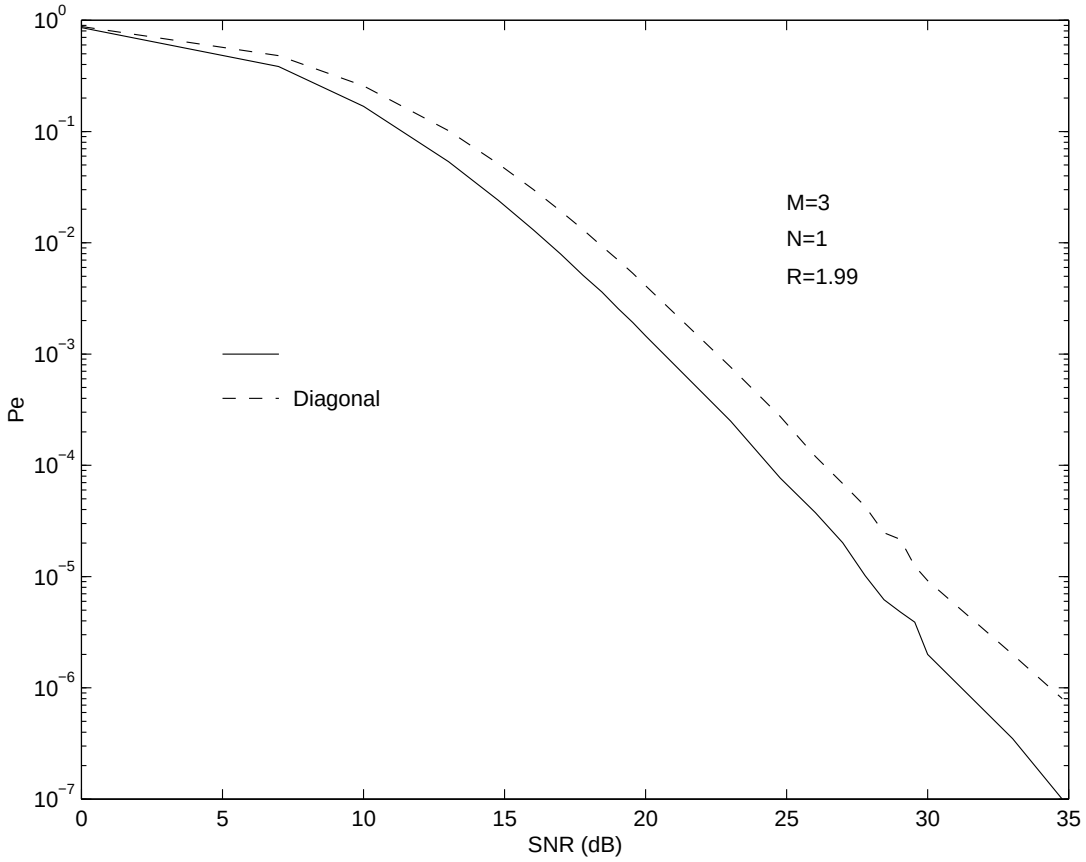
Example, $M = 3$

Let $M = 3$ and take $r = 4$ and $m = 21$. Then we have $t = 7$, and it can be verified that conditions (i)-(iv) of the theorem are all satisfied. Thus, define $\eta = e^{j2\pi/21}$, and set

$$A = \begin{pmatrix} \eta & 0 & 0 \\ 0 & \eta^4 & 0 \\ 0 & 0 & \eta^{16} \end{pmatrix} \quad , \quad B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \eta^7 & 0 & 0 \end{pmatrix} .$$

The 63 matrices $A^\ell B^k$, where $0 \leq \ell \leq 20$ and $0 \leq k \leq 2$ form a group under matrix multiplication and the corresponding ζ-value can be computed to be $\zeta = 0.3851$. This 3-antenna, 63-element, constellation is by one element shy of producing a constellation of rate 2.

$G_{21,4}$ vs. Diagonal Constellation



Example, $M = 9$

Let $M = 9$. With $r = 4$ and $m = 57$ all the conditions of the theorem are satisfied. Thus, if we define $\eta = e^{2\pi i/57}$, and set

$$A = \begin{pmatrix} \eta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \eta^4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \eta^{16} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta^7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \eta^{28} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \eta^{55} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta^{49} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta^{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \eta^{43} \end{pmatrix}, B = \begin{bmatrix} & & & & & & & & I_8 \\ \eta^{19} & & & & & & & & \end{bmatrix}.$$

we obtain a 9-antenna, 513-element, constellation ($R = 1$) with $\zeta = 0.36$.

Even-Order Fixed-Point-Free Groups

- The classification of even-order fixed-point-free groups is slightly more involved.
- In addition to $G_{m,r}$, there are five other group types.
- One interesting even-order fixed-point-free group is $SL_2(\mathcal{F}_5)$, group of 2×2 matrices over $\mathcal{F}_5$ with determinant unity. This group has 120 elements and can be expressed as

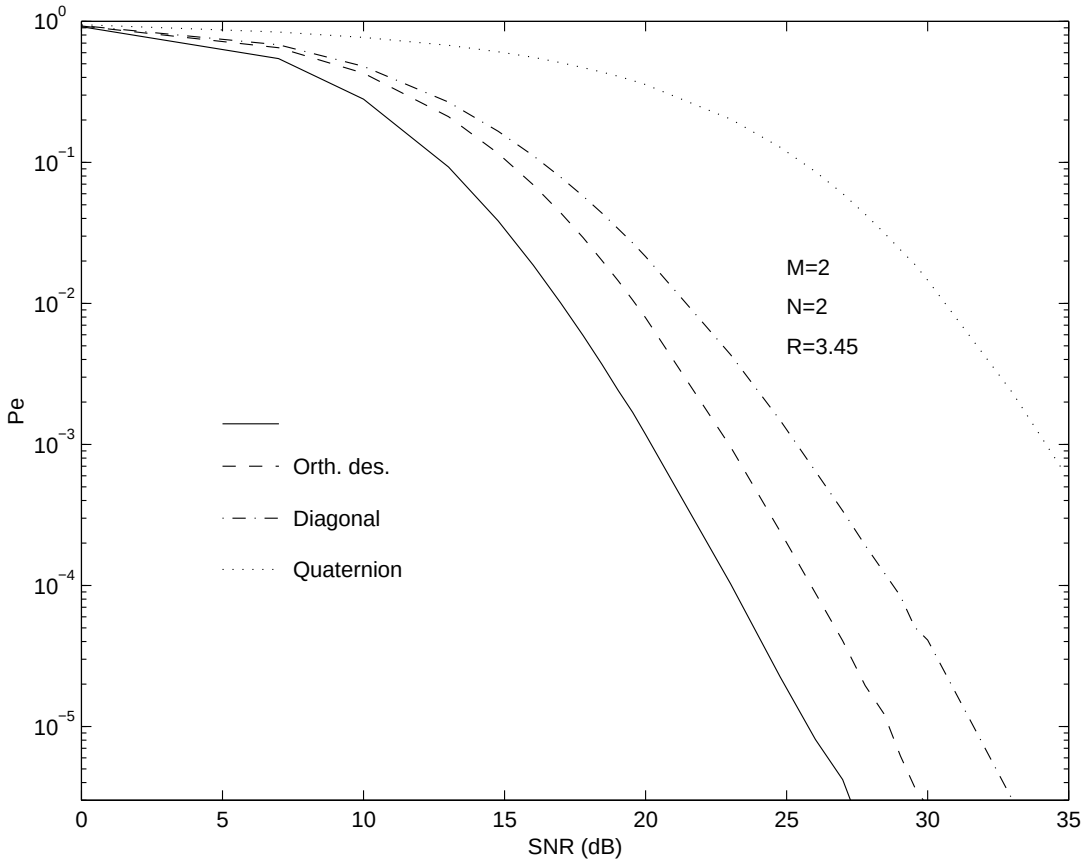
$$SL_2(\mathcal{F}_5) = \langle \mu, \gamma \mid \mu^2 = \gamma^3 = (\mu\gamma)^5, \mu^4 = 1 \rangle.$$

The representation of its generators is given by

$$\Delta(\mu) = \frac{1}{\sqrt{5}} \begin{bmatrix} \eta^2 - \eta^3 & \eta - \eta^4 \\ \eta - \eta^4 & \eta^3 - \eta^2 \end{bmatrix}, \quad \Delta(\gamma) = \frac{1}{\sqrt{5}} \begin{bmatrix} \eta - \eta^2 & \eta^2 - 1 \\ 1 - \eta^3 & \eta^4 - \eta^3 \end{bmatrix}$$

where $\eta = e^{2\pi i/5}$.

$M = 2,\ N = 2$ **and** $R = 3.45$



Beyond Groups

- Although we have been able to characterize all fixed-point-free groups, it turns out that such groups are few and far between.
- Thus, even though they yield constellations with a very acceptable ζ, such groups do not exist for any arbitrary M and any target rate, R.
- For example, it is not possible to use our theorem to construct constellations of "close to" rate 1 for matrix dimensions $M = 5$ and $M = 7$, since there exist no fixed-point free group representations for $M = 5$ and $M = 7$ matrix dimensions that have "close to" $L = 32$ and $L = 128$ elements, respectively.
- Thus, to construct constellations for arbitrary M and arbitrary R, it appears that we need to move beyond the group constructions considered so far.
- The group constructions, nonetheless, suggest structures that may be used to construct good non-group constellations.

Performance vs. Diagonal Constellations

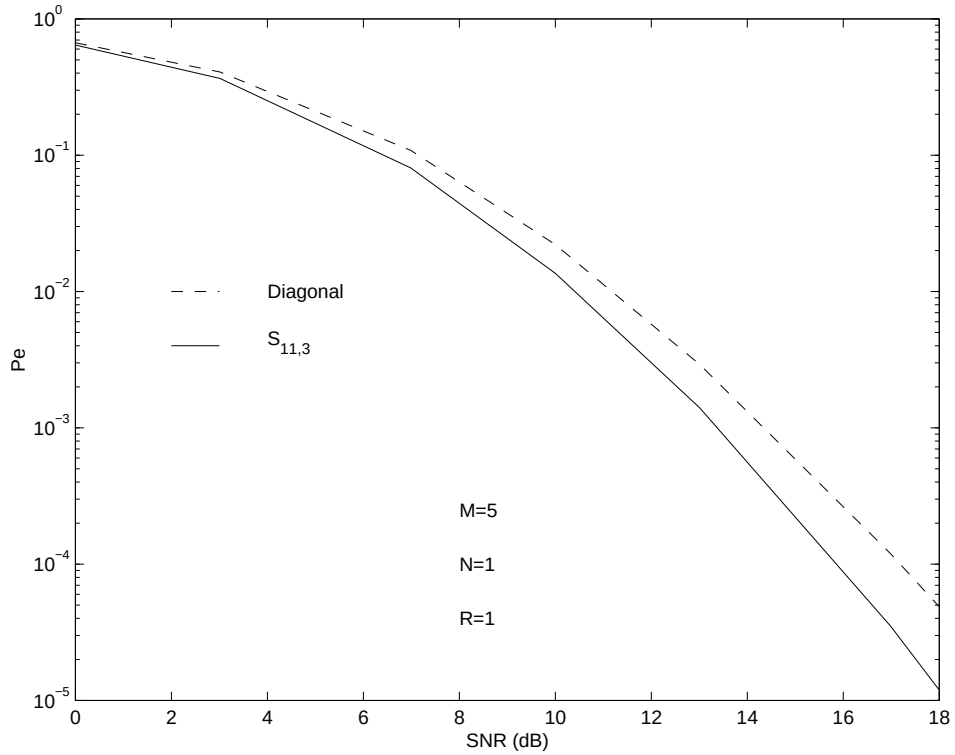
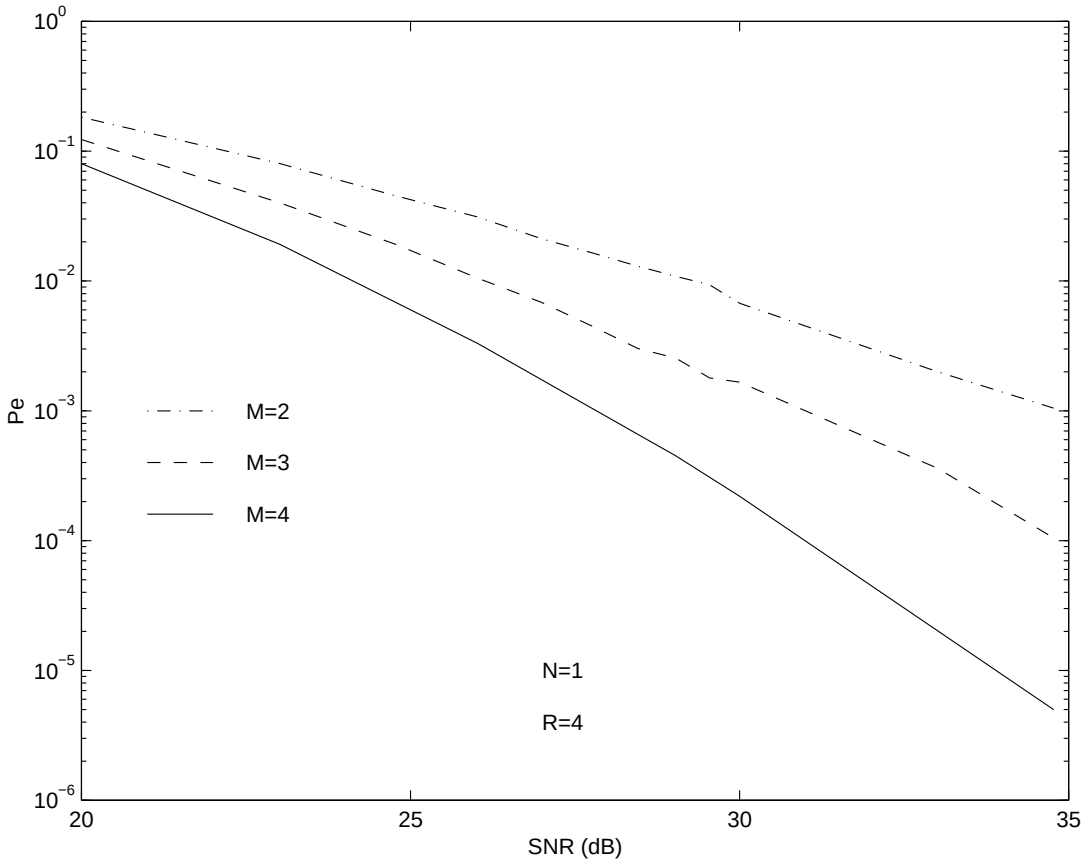


Figure 4: Block-error rate performance for $M = 5$ transmitter antennas and rate $R = 1$. The dashed line is the best diagonal construction. The solid line is our best $M = 5$ construction.

$R = 4$ Non-Group Constellations



Final Remarks

- We considered the construction of *fully diverse* USTM signals that are suitable for known channel, and unknown channel differential-modulation environments.
- We described a construction based on the representation theory of finite *fixed-point-free* groups.
- Unlike earlier *orthogonal designs*-based methods, our framework allows for any number of antennas.
- The special structure of the constellation allows for efficient decoding schemes, so that the exhaustive search is not necessary.
- For low to moderately high rates, the resulting constellations have excellent performance. We are currently investigating the construction of very high rate non-group constellations.
- Joint work with A. Shokrollahi, B. Hochwald, and W. Sweldens.

Summary

Multi-antenna communications has a great potential to increase the capacity and quality wireless communications. It also presents great challenges and many interesting problems.

To summarize, let us repeat the outline.

- **Signal Processing:**
  - an efficient square-root algorithm for Bell Labs Layered Space-Time (BLAST).
- **Information Theory:**
  - *autocapacity* – information transmission at Shannon capacity via coding over a *single* coherence interval.
- **Coding Theory:**
  - multi-antenna signal constellation design via group representation theory.

Other Work

- In the multiple-antenna area:
  - determining the optimal amount of training
- Other areas:
  - robust estimation and control
  - adaptive filtering
  - channel equalization (blind and robust)
  - neural networks

Additional Material - Not Covered in Talk

Description of Efficient Square-Root Algorithm

BLAST Algorithm

1. Find $H_\alpha^\dagger$ and $P = H_\alpha^\dagger (H_\alpha^\dagger)^*$.
2. Find the smallest diagonal entry of P and reorder the entries of s so that the smallest diagonal entry is the last (M-th) one.
3. Form the least-mean-squares estimate $\hat{s}_M = H_{\alpha,M}^\dagger x$.
4. Obtain s_M (via slicing) from $\hat{s}_M = H_{\alpha,M}^\dagger x$.
5. Cancel the effect of s_M and consider the *reduced-order* problem:

$$x - \underline{h}_M s_M = H^{(M-1)} s^{(M-1)} + v,$$

where

$$H^{(M-1)} = \left[\begin{array}{ccc} \underline{h}_1 & \dots & \underline{h}_{M-1} \end{array} \right] \text{ and } s^{(M-1)} = \left[\begin{array}{ccc} s_1 & \dots & s_{M-1} \end{array} \right]^T.$$

6. Continue to find $H_\alpha^{(M-1)\dagger}$ and $P^{(M-1)} = H_\alpha^{(M-1)\dagger} (H_\alpha^{(M-1)\dagger})^*$.

Computational Complexity

For simplicity, let $M = N$.

- Channel estimation: $2M^2 \log_2 L_T$.
- Determining the nulling vectors and optimal ordering: $27M^4/4$.
- Processing the payload: $2M^2 L_P$.

To see what these numbers mean for actual systems, consider the next target application for BLAST:

- 1 Mb/s data transmission over a 30 kHz wireless channel.
- $1/T = 24.3$ ksymbol/sec, 16-QAM.
- $M = N = 14$, $L_T = 32$ and $L_P = 100$.
- Required DSP integrated into a single chip solution.

	Flops/burst	MegaFlops/s	%
channel estimation	7,840	1.44	0.65
nulling vectors and ordering	1,036,000	190.8	86.3
payload processing	156,800	28.9	13.1
TOTAL	1,200,000	221.2	100

The dominant portion of the computation involves determining the nulling vectors and optimal ordering. *Can this computation be reduced in a numerically stable way?*

Objectives of the Algorithm

1. The algorithm must be **cost efficient**: *Is it possible to find* $H_\alpha^{(M-1)\dagger}$ *and* $P^{(M-1)}$ *from* $H_\alpha^\dagger$ *and* P*, without having to "re-solve" the reduced-order problem all over again?*

2. The algorithm must be **numerically stable and robust**.

   - Avoid "squaring" things (forming H^*H, for example, is undesirable). This increases the dynamic range of the quantities involved, the condition numbers, etc.
   - Avoid "inverting" things (inverting $(\alpha I + H^*H)$ to obtain P is undesirable).
   - Make as much use as possible of *unitary* transformations.

QR Decomposition

Let us begin with the QR decomposition:

$$\begin{bmatrix} H \\ \sqrt{\alpha} I_M \end{bmatrix} = QR = \begin{bmatrix} Q_\alpha \\ Q_2 \end{bmatrix} R,$$

where Q is an $(N+M) \times M$ matrix with orthonormal columns, and R is $M \times M$ and nonsingular. It can be shown that

$$P^{1/2} = R^{-1} \ , \ H_\alpha^\dagger = P^{1/2} Q_\alpha^* \quad \text{where} \quad P^{1/2} P^{*/2} = P.$$

Before addressing the question of how to compute $P^{1/2}$ and Q_α, let us focus on:

1. How to find the smallest diagonal entry of P?
2. How to find the square-root factor of $P^{(M-1)}$ from $P^{1/2}$?
3. How to find the nulling vectors?

Answers

1. Since $P^{1/2}P^{*/2} = P$, the diagonal entries of P are simply the squared lengths of the *rows* of $P^{1/2}$. Thus: *to find the minimum diagonal entry of P, we need only to find the minimum length row of $P^{1/2}$.*

2. Suppose now that we have reordered the entries of s so that the M-th diagonal entry of P is the smallest. Consider any unitary transformation Σ that rotates (or reflects) the M-th row of $P^{1/2}$ to lie along the direction of the M-th unit vector. In other words,

$$P^{1/2}\Sigma = \begin{bmatrix} P^{(M-1)/2} & P_M^{(M-1)/2} \\ 0 & p_M^{1/2} \end{bmatrix},$$

where $p_M^{1/2}$ is a scalar. *Then $P^{(M-1)/2}$ is a square-root of $P^{(M-1)}$.*

3. Suppose that we have repeated the above steps 1 and 2 until $P^{1/2}$ is transformed to an upper triangular matrix. Moreover, let $\underline{q}_{\alpha,i}$, $i = 1, \ldots, M$ denote the resulting columns of Q_α, *i.e.*,

$$Q_\alpha = \left[\begin{array}{ccc} \underline{q}_{\alpha,1} & \cdots & \underline{q}_{\alpha,M} \end{array} \right].$$

Then the nulling vectors for the signals s_1 to s_M are given by:

$$H^{\dagger}_{\alpha,i} = p_i^{1/2} \underline{q}^*_{\alpha,i},$$

where $p_i^{1/2}$ denotes the i-th diagonal entry of $P^{1/2}$.

Conclusion: Once $P^{1/2}$ and Q_α are computed, there is no need to recompute them for the deflated channel matrix $H^{(M-1)}$. All the information we need is already in $P^{1/2}$ and Q_α.

But what is the best way to compute $P^{1/2}$ and Q_α? (Inverting R to obtain $P^{1/2}$ is undesirable.)

Summary of Algorithm

1. Compute $P^{1/2}$ and Q.

   - Propagate the following square-root algorithm:

$$\begin{bmatrix} 1 & H_i P^{1/2}_{|i-1} \\ 0 & P^{1/2}_{|i-1} \\ -e_i & Q_{i-1} \end{bmatrix} \Theta_i = \begin{bmatrix} r^{1/2}_{e,i} & 0 \\ \bar{K}_{p,i} & P^{1/2}_{|i} \\ A_i & Q_i \end{bmatrix}, \quad P^{1/2}_{|0} = \frac{1}{\sqrt{\alpha}} I, \; Q_0 = 0$$

     where e_i is the i-th unit vector of dimension N, and Θ_i is any unitary transformation that block lower triangularizes the pre-array. After N steps, we have

$$P^{1/2} = P^{1/2}_{|N} \quad \text{and} \quad Q_\alpha = Q_N.$$

2. Find the minimum length row of $P^{1/2}$ and permute it to be the last (Mth) row. Permute s accordingly.

3. Find a unitary Σ such that $P^{1/2}\Sigma$ is block upper triangular:

$$P^{1/2}\Sigma = \begin{bmatrix} P^{(M-1)/2} & P_M^{(M-1)/2} \\ 0 & p_M^{1/2} \end{bmatrix}.$$

4. Update Q_α to $Q_\alpha\Sigma$.

5. The nulling vector for the M-th signal is given by $p_M^{1/2}\underline{q}^*_{\alpha,M}$, where $\underline{q}^*_{\alpha,M}$ is the M-th row of Q^*_α.

6. Go back to step 3, but now with $P^{(M-1)/2}$ and $Q_\alpha^{(M-1)}$ (the first $M-1$ columns of Q_α).

The complexity of the algorithm can be shown to be $29M^3/3$.

Remarks

- The above algorithm satisfies our objectives of cost-efficiency:
  - we have avoided computing the pseudo-inverse (or QR decomposition) for each deflated subchannel matrix. This reduced the computational complexity from $27M^4/4$ to $29M^3/3$, i.e., roughly by a factor of $0.7M$.

  and numerical stability and robustness:

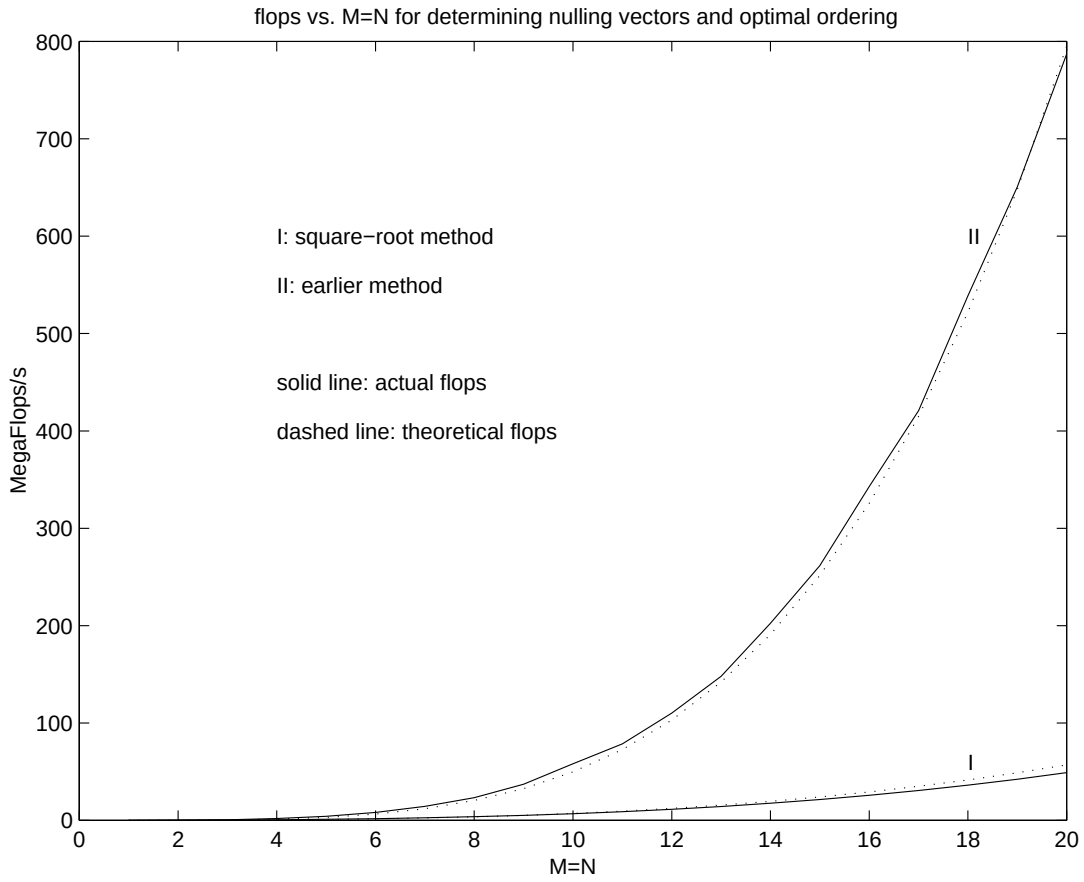
  - we have avoided squaring any of the quantities.
  - we have avoided computing inverses (and even scalar divisions) altogether.
  - we have used unitary transformations as much as possible.

But what does all this mean for our example (1 Mb/s, 30 kHz channel, 24.3 ksymbol/sec, 16-QAM, $M = N = 14$, $L_T = 32$, $L_P = 100$)? Well,...

	Flops/burst	MegaFlops/s	%
channel estimation	7,840	1.44	2.9
nulling vectors and ordering	106,100	19.5	39.1
payload processing	156,800	28.9	58.0
TOTAL	270,400	49.8	100

Thus the total computation has been reduced from 221 MegaFlops/s to 50 MegaFlops/s.

Flops ver. $M = N$


flops vs. M=N for determining nulling vectors and optimal ordering
MegaFlops/s
M=N
I: square−root method
II: earlier method
solid line: actual flops
dashed line: theoretical flops
II
I
0
100
200
300
400
500
600
700
800
0
2
4
6
8
10
12
14
16
18
20

Further Remarks

- The prominent component of the algorithm is the use of unitary transformations which introduce zeros in prescribed entries of given row vectors. These can be performed by either using a Householder reflection, or a sequence of Givens rotations.
- In hardware, the sequence of Givens rotations can be implemented using "division-free" methods, such as the CORDIC method. They can also be parallelized by means of a systolic-array-type architecture.
- The numerical stability of the algorithm makes it attractive for implementation in fixed-point, rather than floating-point, architectures.
- Finally, the savings in computational complexity make it possible to implement the algorithm using a single commercial DSP processor. Thus, there is no need to design a new chip.

Further Material Related to Autocapacity

Matrix Channel with Correlated Entries

When the elements of H are correlated the perfect knowledge Shannon capacity is still given by

$$C = E \log \det \left(I_N + \rho \frac{H^* H}{M} \right).$$

If we denote the eigenvalues of $\frac{H^* H}{M}$ by $\lambda_1, \ldots, \lambda_N$, then

$$C = E \log \prod_{i=1}^{N} (1 + \rho \lambda_i) = E \sum_{i=1}^{N} \log (1 + \rho \lambda_i) = N E \log (1 + \rho \lambda)$$

When the elements of H are uncorrelated $\lambda \to 1$ as $M \to \infty$ and we obtain $C = N \log(1 + \rho)$. When H is rank one (corresponding to a single planar wave line-of-site situation), we have

$$C = \log(1 + N\rho)$$

The general case depends on how "full-rank" H is.

How Many Antennas?

- Answer not given by the autocapacity formulas
- *Partial* answer can be provided by studying the random coding exponent

$$Pe \;\leq\; \exp\left\{-QT\ln 2\left[E_0(T,\beta,\rho,p(S)) - \mu R\right]\right\}$$

where

$$E_0(T,\beta,\rho,p(S)) = -\frac{1}{T}\log\left[\int_X \left(E_S\left\{p^{\frac{1}{1+\mu}}(X|S)\right\}\right)^{1+\mu} dX\right]$$

Here Q is the number of independent coherence intervals, and R is the rate of transmission

- We are interested in $Q = 1$, and wish to show that as $T \to \infty$, the exponent $[E_0(T,\beta,\rho,p(S)) - \mu R]$ is positive

The Cutoff Rate

- Joint optimization over $p(S)$ and μ is formidable. Therefore, we restrict our attention to $\mu = 1$, and "judicious" choices of the density $p(S)$.
- This leads to

$$P_e \;\leq\; \exp\left\{-T \ln 2 \left[R_0(T, \beta, N, p(S)) - R\right]\right\},$$

where

$$R_0(T, \beta, N, p(S)) = -\frac{1}{T} \log \left[\mathrm{E}_{S_1, S_2} \left\{ \int dX \sqrt{p(X \mid S_1) \cdot p(X \mid S_2)} \right\} \right]$$

is the so-called *cutoff rate*.

- **Remark:** This can also be obtained using the *union bound* applied to the *Chernoff bound* on the pairwise probability of error.

Asymptotic Cutoff Rate

Suppose H is unknown and that the input signals are chosen as isotropically-distributed unitary matrices. Then

$$\lim_{T\to\infty} R_0(T,\beta,N,p(S)) = \frac{N}{\beta}\left[\min(1,\beta-1)\log(1+\alpha)+ \beta\log\left(\frac{1+\sqrt{1-\frac{a\alpha}{1+\alpha}}}{2}\right)+ |\beta-2|\log\left(\frac{1+\sqrt{1-a}}{\sqrt{1-a}+\sqrt{1-\frac{a\alpha}{1+\alpha}}}\right)\right],$$

where $\alpha = \frac{(\rho\beta)^2}{4(1+\rho\beta)}$, and $a = 1-\left(\frac{\beta-2}{\beta}\right)^2$.

- Similar closed-form expressions can be found for other cases.

Additional Material on Constellation Design

The Fixed-Point-Free Group Types

1. $\boldsymbol{G_{m,r}}$:

$$G_{m,r} = \langle \sigma, \tau \mid \sigma^m = 1, \tau^n = \sigma^t, \sigma^\tau = \sigma^r \rangle$$

where (m, r) is admissible. The order of $G_{m,r}$ is $L = mn$.

2. $\boldsymbol{D_{m,r,\ell}}$:

$$D_{m,r,\ell} = \langle \sigma, \tau, \gamma \mid \sigma^m = 1, \tau^n = \sigma^t, \sigma^\tau = \sigma^r, \sigma^\gamma = \sigma^\ell, \tau^\gamma = \tau^\ell, \gamma^2 = \tau^{nr_0/2} \rangle,$$

where nr_0 is even, (m, r) is admissible, $\ell^2 \equiv 1 \bmod m$, $\ell \equiv 1 \bmod n$, and $\ell \equiv -1 \bmod s$, where s is the highest power of 2 dividing mn. The order of $D_{m,r,\ell}$ is $L = 2mn$.

3. $\boldsymbol{E_{m,r}}$:

$$\begin{aligned} E_{m,r} \quad = \quad & \langle \sigma, \tau, \mu, \gamma \mid \sigma^m = 1, \tau^n = \sigma^t, \sigma^\tau = \sigma^r, \mu^{\sigma^{m/t}} = \mu, \gamma^{\sigma^{m/t}} = \gamma, \\ & \mu^4 = 1, \mu^2 = \gamma^2, \mu^\gamma = \mu^{-1}, \mu^\tau = \gamma, \gamma^\tau = \mu\gamma \rangle, \end{aligned}$$

where (m, r) is admissible, mn is odd, and nr_0 is divisible by 3. The order of $E_{m,r}$ is $8mn$.

4. $\boldsymbol{F_{m,r,\ell}}$:

$$\begin{aligned} F_{m,r,\ell} \quad = \quad & \langle \sigma, \tau, \mu, \gamma, \nu \mid \sigma^m = 1, \tau^n = \sigma^t, \sigma^\tau = \sigma^r, \mu^{\sigma^{m/t}} = \mu, \\ & \gamma^{\sigma^{m/t}} = \gamma, \mu^\tau = \gamma, \gamma^\tau = \mu\gamma, \mu^4 = 1, \mu^2 = \gamma^2, \mu^\gamma = \mu^{-1}, \\ & \nu^2 = \mu^2, \sigma^\nu = \sigma^\ell, \tau^\nu = \tau^\ell, \mu^\nu = \gamma^{-1}, \gamma^\nu = \mu^{-1} \rangle, \end{aligned}$$

where (m, r) is admissible, mn is odd, r_0 is divisible by 3, n is not divisible by 3, $\ell^2 \equiv 1 \bmod m$, $\ell \equiv 1 \bmod n$, and $\ell \equiv -1 \bmod 3$. The order of $F_{m,r,\ell}$ is $16mn$.

5. $\boldsymbol{J_{m,r}}$:

$$J_{m,r} := SL_2(\mathcal{F}_5) \times G_{m,r},$$

where (m, r) is admissible, $\gcd(mn, 120) = 1$, and $SL_2(\mathcal{F}_5)$ is the group of 2×2-matrices over $\mathcal{F}_5$ with determinant 1. $SL_2(\mathcal{F}_5)$ has the generators and relations

$$SL_2(\mathcal{F}_5) = \langle \mu, \gamma \mid \mu^2 = \gamma^3 = (\mu\gamma)^5, \mu^4 = 1 \rangle.$$

The order of $J_{m,r}$ is $120mn$.

6. $\boldsymbol{K_{m,r,\ell}}$:

$$K_{m,r,\ell} = \langle J_{m,r}, \nu \rangle$$

with the relations

$$\nu^2 = \mu^2, \mu^\nu = (\mu\gamma)^7(\gamma\mu)^2\gamma(\gamma\mu)^2, \gamma^\nu = \gamma, \sigma^\nu = \sigma^\ell, \tau^\nu = \tau^\ell,$$

where μ and γ are as in $\boldsymbol{J_{m,r}}$, and where $\ell^2 \equiv 1 \bmod m$, $\ell \equiv 1 \bmod n$. The order of $K_{m,r,\ell}$ is $240mn$.