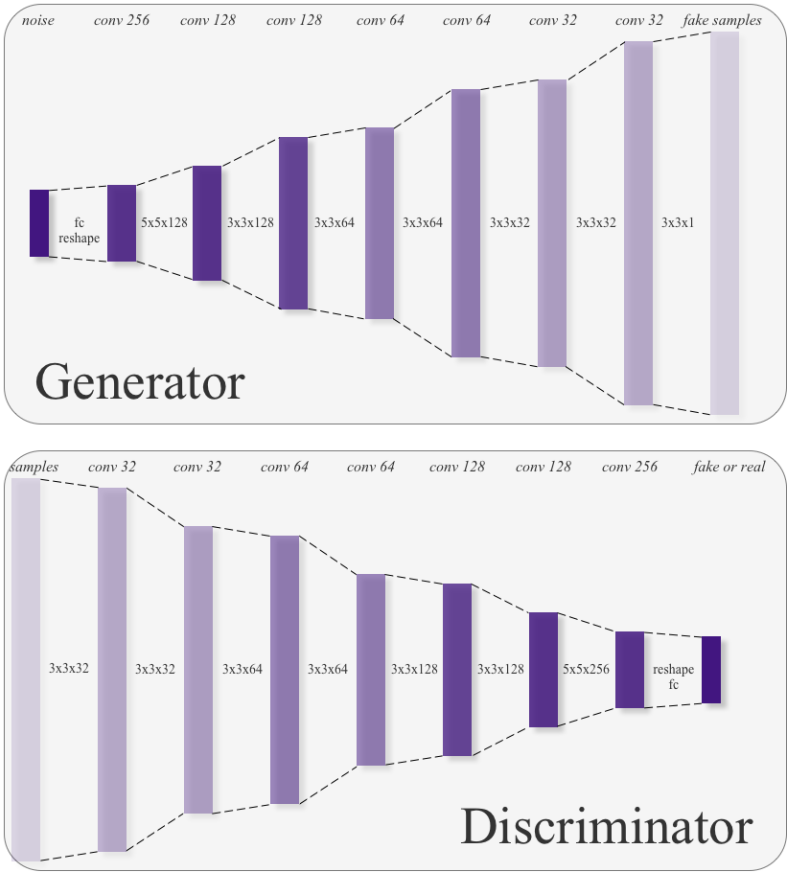
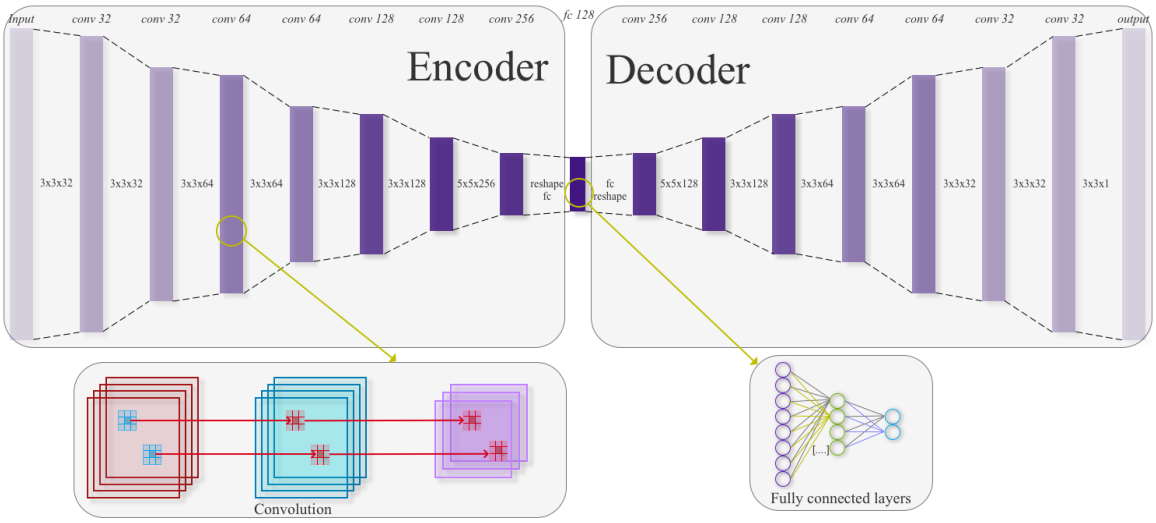
Supplementary Information

Generative adversarial networks (GAN) based efficient sampling of chemical composition space for inverse design of inorganic materials

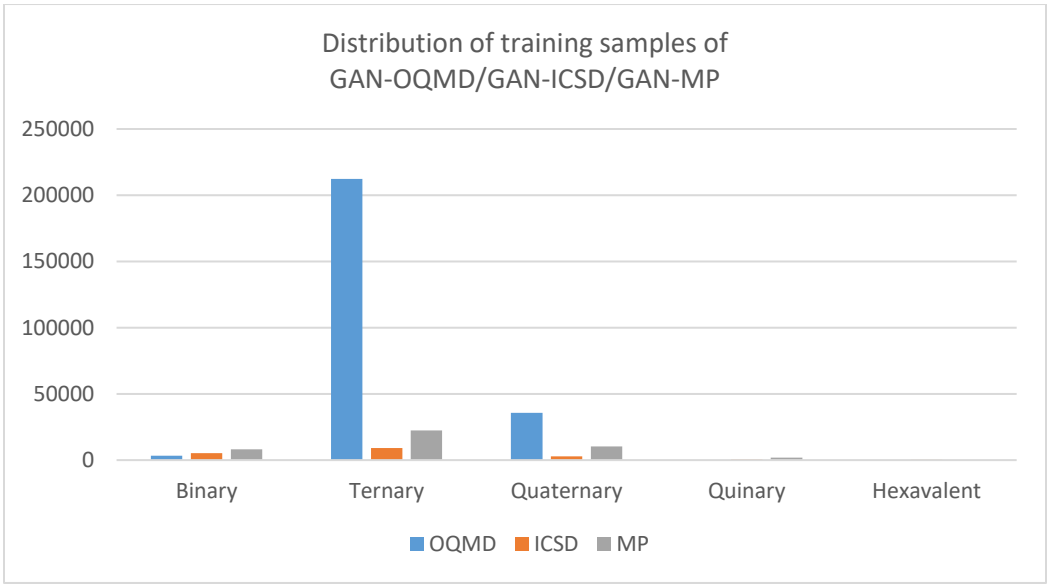
Supplementary Figures



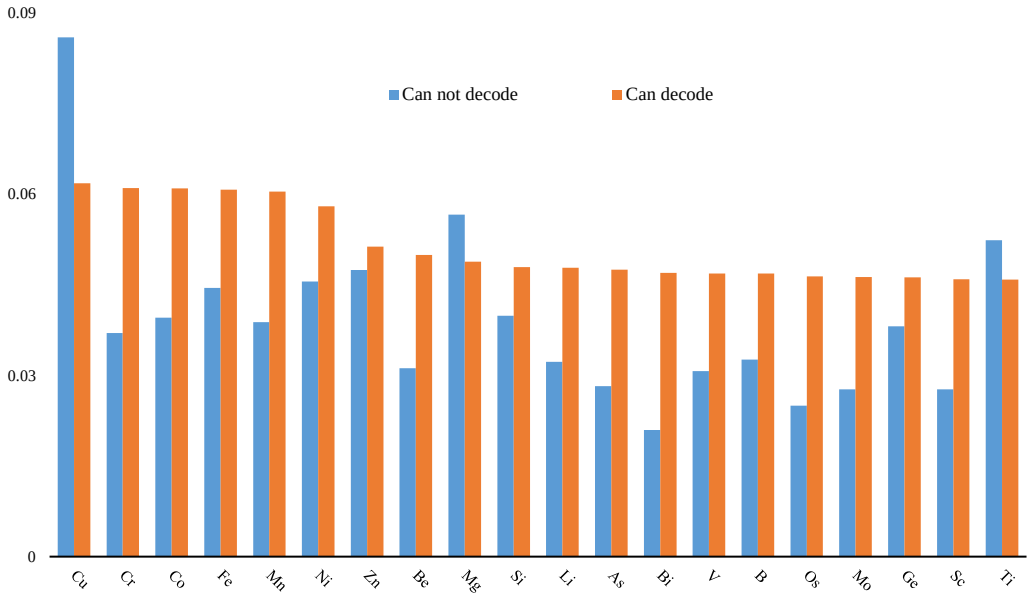
Supplementary Figure 1| Detailed architecture of the generator and discriminator in MatGAN model



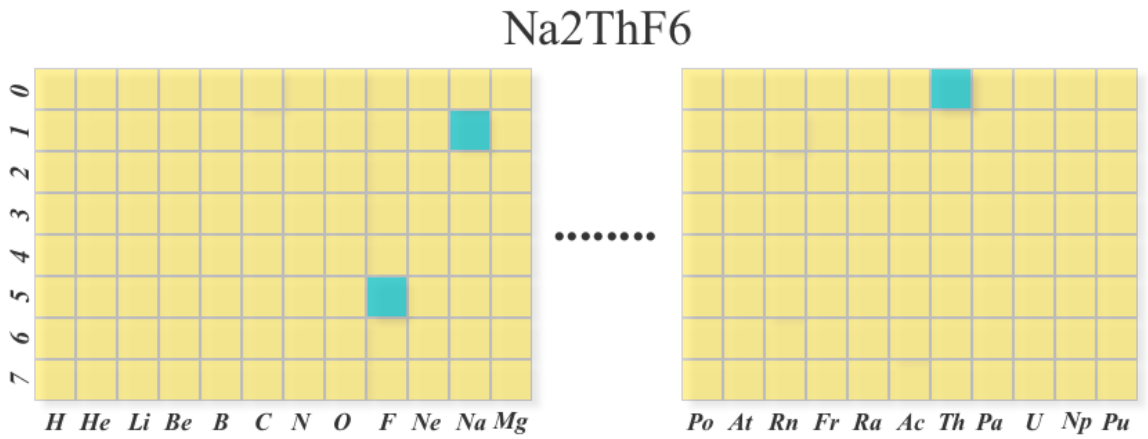
Supplementary Figure 2| Detailed architecture of the autoencoder model



Supplementary Figure 3| Comparison of the distributions of training samples of three GANs. The OQMD is strongly biased with a large percentage of ternary compounds.



Supplementary Figure 4| Comparison of enriched element distribution of decodable and non-decodable materials.



Supplementary Figure 5| One-hot encoding of materials for effective convolutional neural network learning

Supplementary Tables

Supplementary Table 1.| Parameters of GAN-OQMD generator

Model	Layer	Input Shape	Filter	Kernel Size	Stride
Generator	Fc1	[batch, 128]	-	-	-
	Reshape	[batch, 6×4×256]	-	-	-
	DeConv1	[batch, 6, 4, 256]	128	(5, 5, 256)	(2, 2)
	DeConv2	[batch, 11, 8, 128]	128	(3, 3, 128)	(2, 1)
	DeConv3	[batch, 22, 8, 128]	64	(3, 3, 128)	(1, 1)
	DeConv4	[batch, 22, 8, 64]	64	(3, 3, 64)	(2, 1)
	DeConv5	[batch, 43, 8, 64]	32	(3, 3, 64)	(1, 1)
	DeConv6	[batch, 43, 8, 32]	32	(3, 3, 32)	(2, 1)
	DeConv7	[batch, 85, 8, 32]	1	(3, 3, 32)	(1, 1)
Discriminator	Conv1	[batch, 85, 8, 1]	32	(3, 3, 1)	(1, 1)
	Conv2	[batch, 85, 8, 32]	32	(3, 3, 32)	(2, 1)
	Conv3	[batch, 43, 8, 32]	64	(3, 3, 32)	(1, 1)
	Conv4	[batch, 43, 8, 64]	64	(3, 3, 64)	(2, 1)
	Conv5	[batch, 22, 8, 64]	128	(3, 3, 64)	(1, 1)
	Conv6	[batch, 22, 8, 128]	128	(3, 3, 128)	(2, 1)
	Conv7	[batch, 11, 8, 128]	256	(5, 5, 128)	(2, 2)
	Reshape	[batch, 6, 4, 256]	-	-	-
	Fc1	[batch, 6×4×256]	-	-	-

Parameters of GAN-ICSD and GAN-MP generators

Model	Layer	Input Shape	Filter	Kernel Size	Stride
Generator	Fc1	[batch, 128]	-	-	-
	Reshape	[batch, 6×4×128]	-	-	-
	DeConv1	[batch, 6, 4, 128]	64	(5, 5, 128)	(2, 2)
	DeConv2	[batch, 11, 8, 64]	32	(3, 3, 64)	(2, 1)
	DeConv3	[batch, 22, 8, 32]	16	(3, 3, 32)	(2, 1)
	DeConv4	[batch, 43, 8, 16]	1	(3, 3, 16)	(2, 1)
Discriminator	Conv1	[batch, 85, 8, 1]	16	(3, 3, 1)	(2, 1)
	Conv2	[batch, 43, 8, 16]	32	(3, 3, 16)	(2, 1)
	Conv3	[batch, 22, 8, 32]	64	(3, 3, 32)	(2, 1)
	Conv4	[batch, 11, 8, 64]	128	(5, 5, 64)	(2, 2)
	Reshape	[batch, 6, 4, 128]	-	-	-
	Fc1	[batch, 6×4×128]	-	-	-

Supplementary Table 2| Parameters of Autoencoder

Model	Layer	Input Shape	Filter	Kernel Size	Stride
Encoder	Conv1	[batch, 85, 8, 1]	32	(3, 3, 1)	(1, 1)
	Conv2	[batch, 85, 8, 32]	32	(3, 3, 32)	(2, 1)
	Conv3	[batch, 43, 8, 32]	64	(3, 3, 32)	(1, 1)
	Conv4	[batch, 43, 8, 64]	64	(3, 3, 64)	(2, 1)
	Conv5	[batch, 22, 8, 64]	128	(3, 3, 64)	(1, 1)
	Conv6	[batch, 22, 8, 128]	128	(3, 3, 128)	(2, 1)
	Conv7	[batch, 11, 8, 128]	256	(5, 5, 128)	(2, 2)
	Reshape	[batch, 5, 4, 256]	-	-	-
	Fc1	[batch, 5×4×256]	-	-	-
Decoder	Fc1	[batch, 128]	-	-	-
	Reshape	[batch, 5×4×256]	-	-	-
	DeConv1	[batch, 5, 4, 256]	128	(5, 5, 256)	(2, 2)
	DeConv2	[batch, 11, 8, 128]	128	(3, 3, 128)	(2, 1)
	DeConv3	[batch, 22, 8, 128]	64	(3, 3, 128)	(1, 1)
	DeConv4	[batch, 22, 8, 64]	64	(3, 3, 64)	(2, 1)
	DeConv5	[batch, 43, 8, 64]	32	(3, 3, 64)	(1, 1)
	DeConv6	[batch, 43, 8, 32]	32	(3, 3, 32)	(2, 1)
	DeConv7	[batch, 85, 8, 32]	1	(3, 3, 32)	(1, 1)

Supplementary Table 3| Distribution of training/validation/generated samples and the recovery rates in terms of element number of three GANs: GAN-OQMD, GAN-ICSD and GAN-MP

GAN-OQMD	Training recovery %	Leave out recovery %	New rate %	Training sample number	Leave out sample number	Generated sample number
Binary	95.27%	95.73%	73.32%	3364	398	12012
Ternary	55.03%	54.90%	44.98%	212245	23606	212289
Quaternary	36.54%	36.69%	98.23%	35759	3925	741905
Quinary	0.00%	0.00%	100.00%	0	0	12012
Hexavalent	0.00%	0.00%	100.00%	0	0	212289
Seven	0.00%	0.00%	100.00%	0	0	741905
Octavalent	0.00%	0.00%	100.00%	0	0	664511

GAN-ICSD	Training recovery %	Leave out recovery %	New rate %	Training sample number	Leave out sample number	Generated sample number
Binary	78.12%	82.71%	83.15%	5145	561	24511
Ternary	30.41%	31.23%	98.68%	9092	1028	209520
Quaternary	3.26%	5.19%	99.98%	2849	308	512598
Quinary	0.00%	0.00%	100.00%	311	36	24511
Hexavalent	0.00%	0.00%	100.00%	24	2	209520
Seven	0.00%	0.00%	100.00%	1	0	512598
Octavalent	0.00%	0.00%	100.00%	0	0	554790

GAN-MP	Training recovery %	Leave out recovery %	New rate %	Training sample number	Leave out sample number	Generated sample number
Binary	84.42%	81.92%	72.22%	8060	874	24385
Ternary	38.94%	38.80%	95.86%	22403	2505	210745
Quaternary	14.97%	14.14%	99.68%	10218	1139	480498
Quinary	0.79%	0.50%	100.00%	1772	199	512603
Hexavalent	0.00%	0.00%	100.00%	0	0	394067
Seven	0.00%	0.00%	100.00%	0	0	253331
Octavalent	0.00%	0.00%	100.00%	0	0	141421

Supplementary Video 1: the demo video with the animation of the materials generation process of GAN-ICSD to show how it fills the composition design space of inorganic materials

https://youtu.be/psneoau1m-8